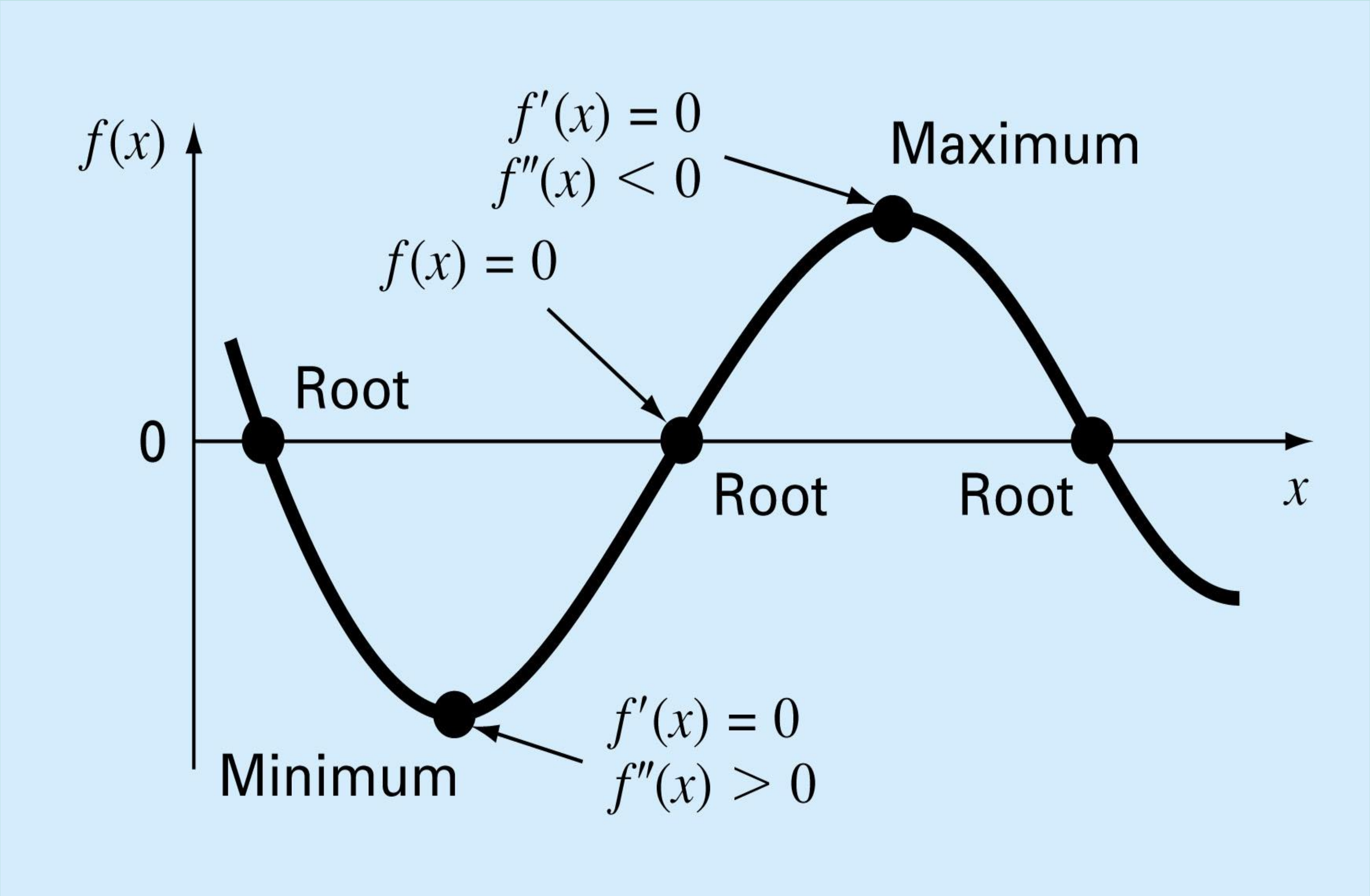
Optimization

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



Mathematical Background

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

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

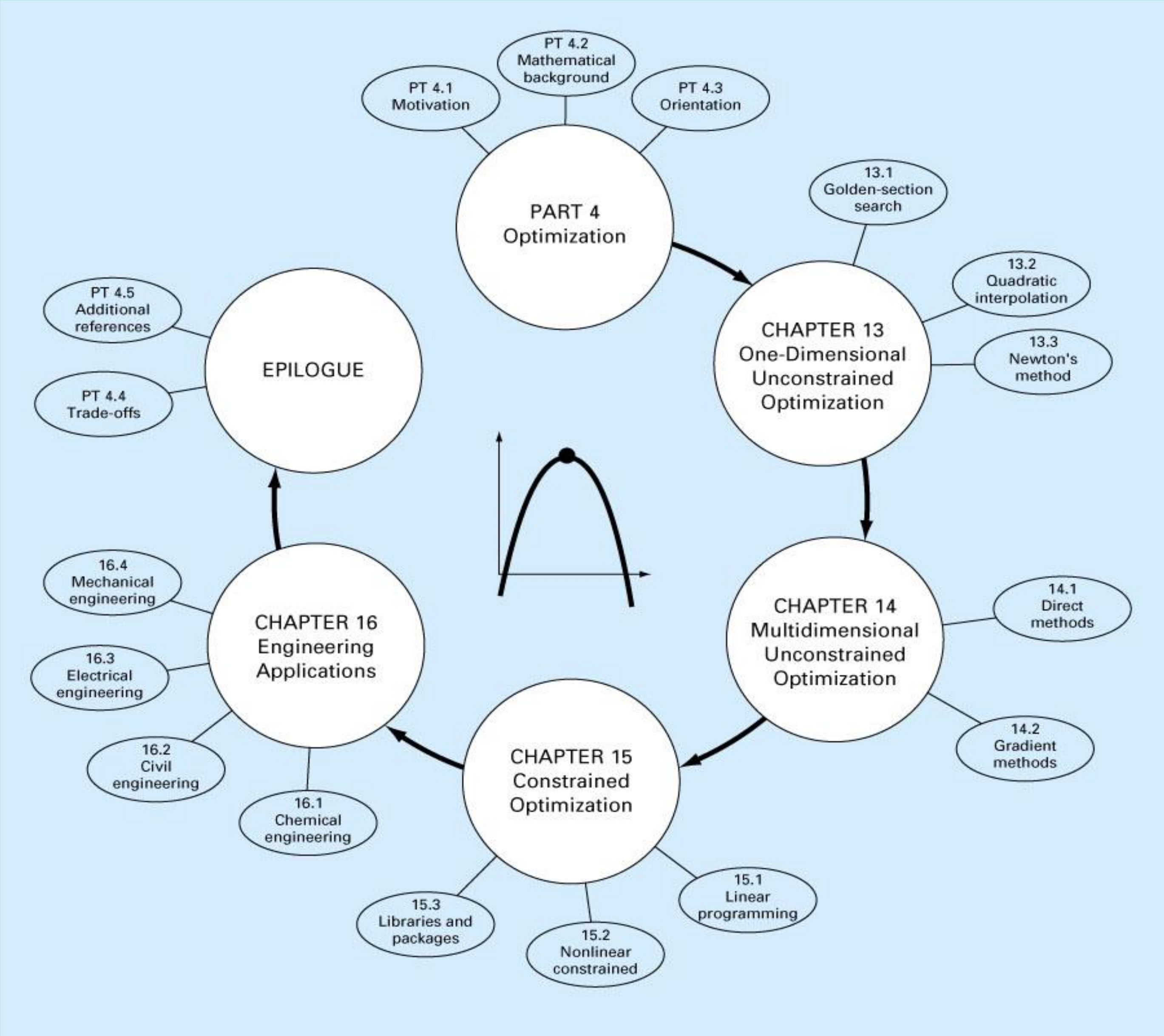
$$d_i(x) \leq a_i \quad i = 1, 2, \ldots, m^*$$

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

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.

Figure PT4.5



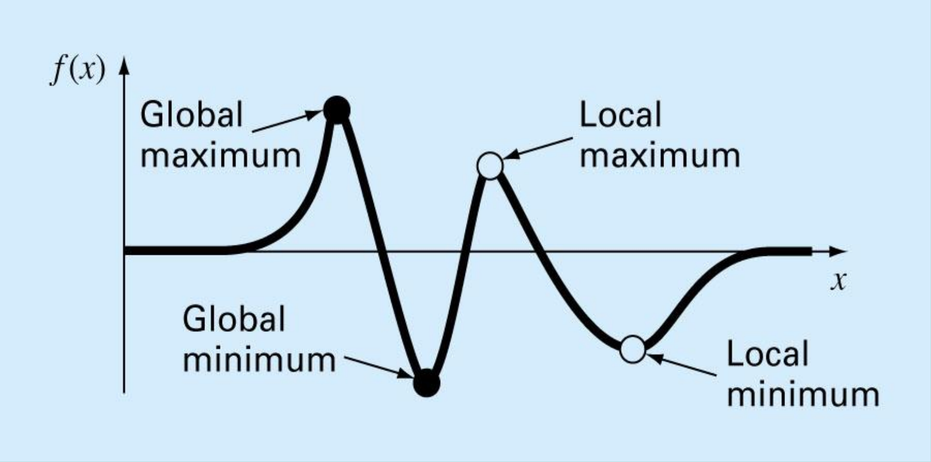
PT 4.2
Mathematical
background
PT 4.1
Motivation
PT 4.3
Orientation
PART 4
Optimization
13.1
Golden-section
search
13.2
Quadratic
interpolation
CHAPTER 13
One-Dimensional
Unconstrained
Optimization
13.3
Newton's
method
PT 4.5
Additional
references
EPILOGUE
PT 4.4
Trade-offs
14.1
Direct
methods
CHAPTER 14
Multidimensional
Unconstrained
Optimization
14.2
Gradient
methods
16.4
Mechanical
engineering
CHAPTER 16
Engineering
Applications
16.3
Electrical
engineering
16.2
Civil
engineering
16.1
Chemical
engineering
CHAPTER 15
Constrained
Optimization
15.1
Linear
programming
15.3
Libraries and
packages
15.2
Nonlinear
constrained

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



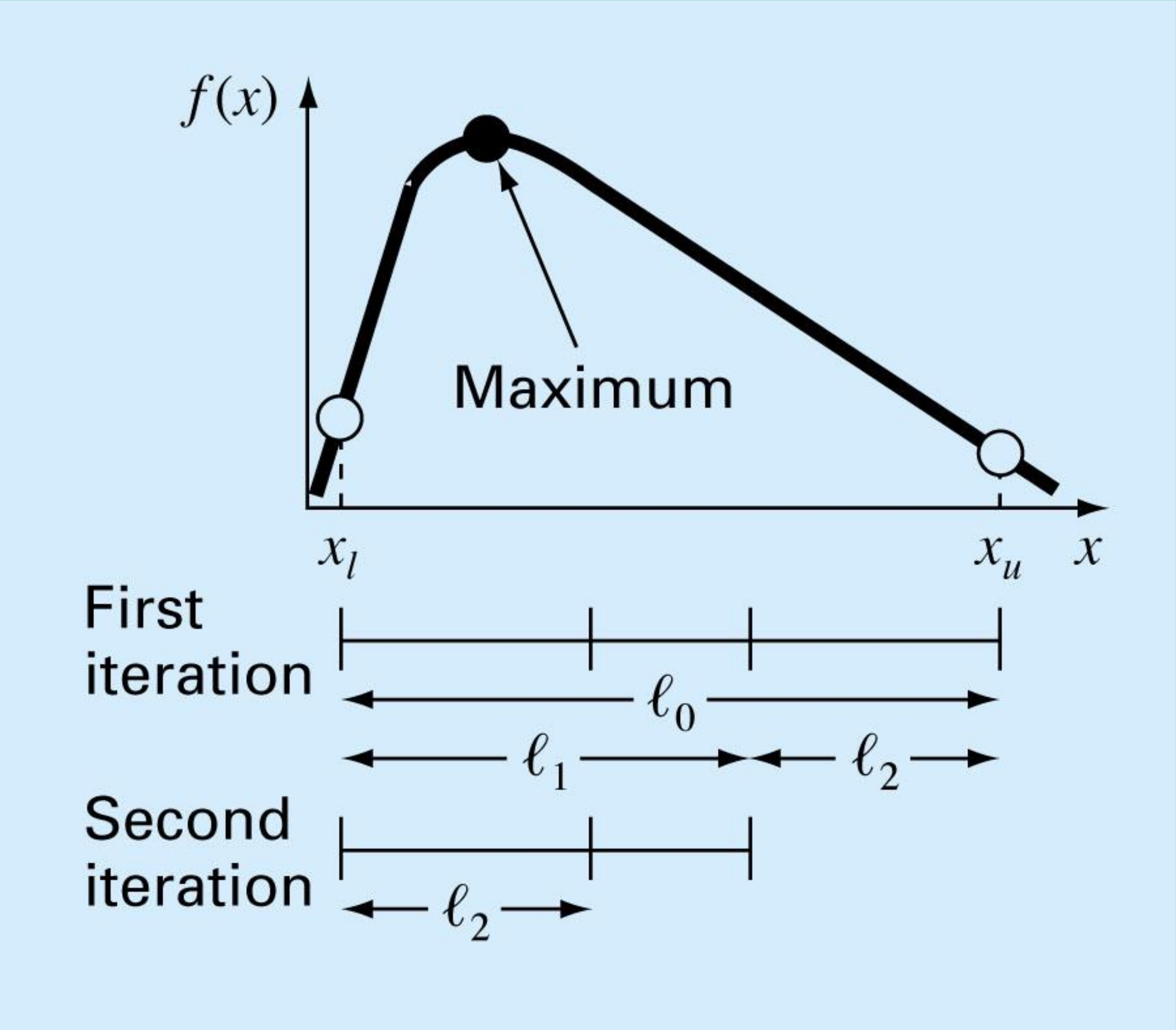
How do we distinguish global optimum from local one?

- 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



f(x)
Maximum
x_l
x_u
x
First iteration
ℓ_0
ℓ_1
ℓ_2
Second iteration
ℓ_2

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

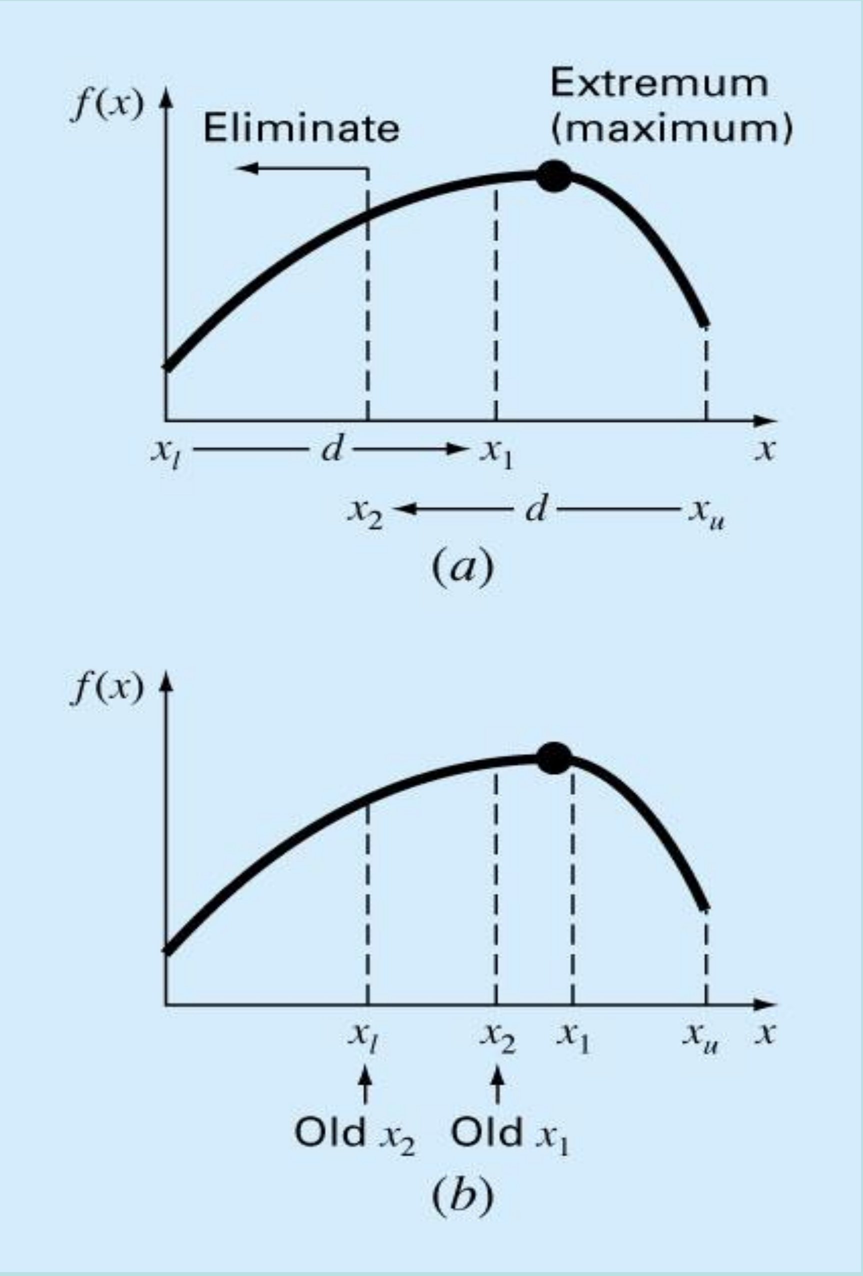
•The first condition specifies that the sum of the two sub lengths l_1 and l_2 must equal the original interval length.

•The second say that the ratio of the length must be equal

$$\frac{l_1}{l_1 + l_2} = \frac{l_2}{l_1} \qquad R = \frac{l_2}{l_1}$$
$$1 + R = \frac{1}{R} \qquad R^2 + R - 1 = 0$$
$$R = \frac{-1 + \sqrt{1 - 4(-1)}}{2} = \frac{\sqrt{5} - 1}{2} = 0.61803$$

Golden Ratio

Figure 13.4



f(x)
Eliminate
Extremum
(maximum)
x_l
d
x_1
x
x_2
d
x_u
(a)
f(x)
x_l
x_2
x_1
x_u
x
Old x_2
Old x_1
(b)

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.

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

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

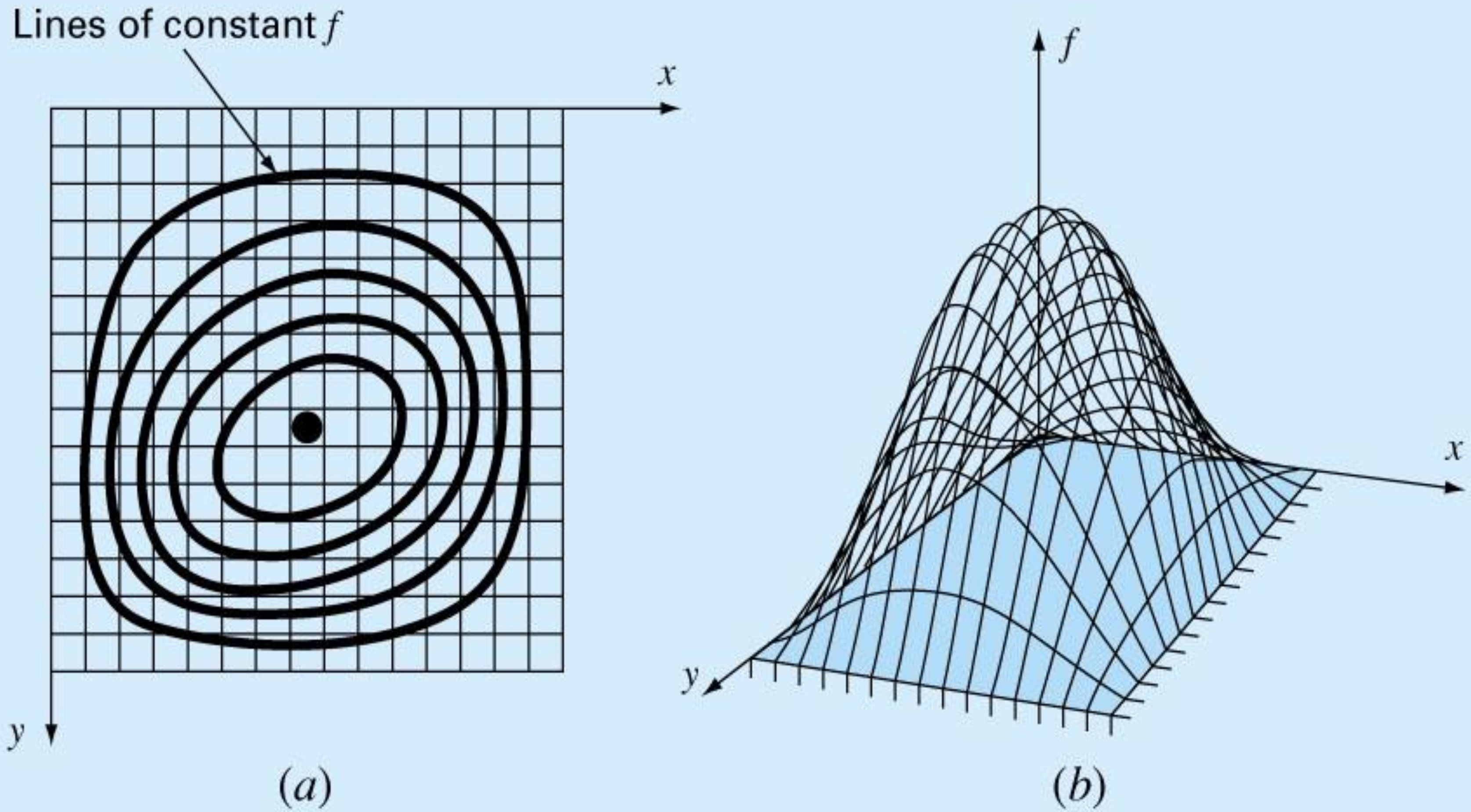
  We can use the following as a technique to the extremum of *f(x)*.

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

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



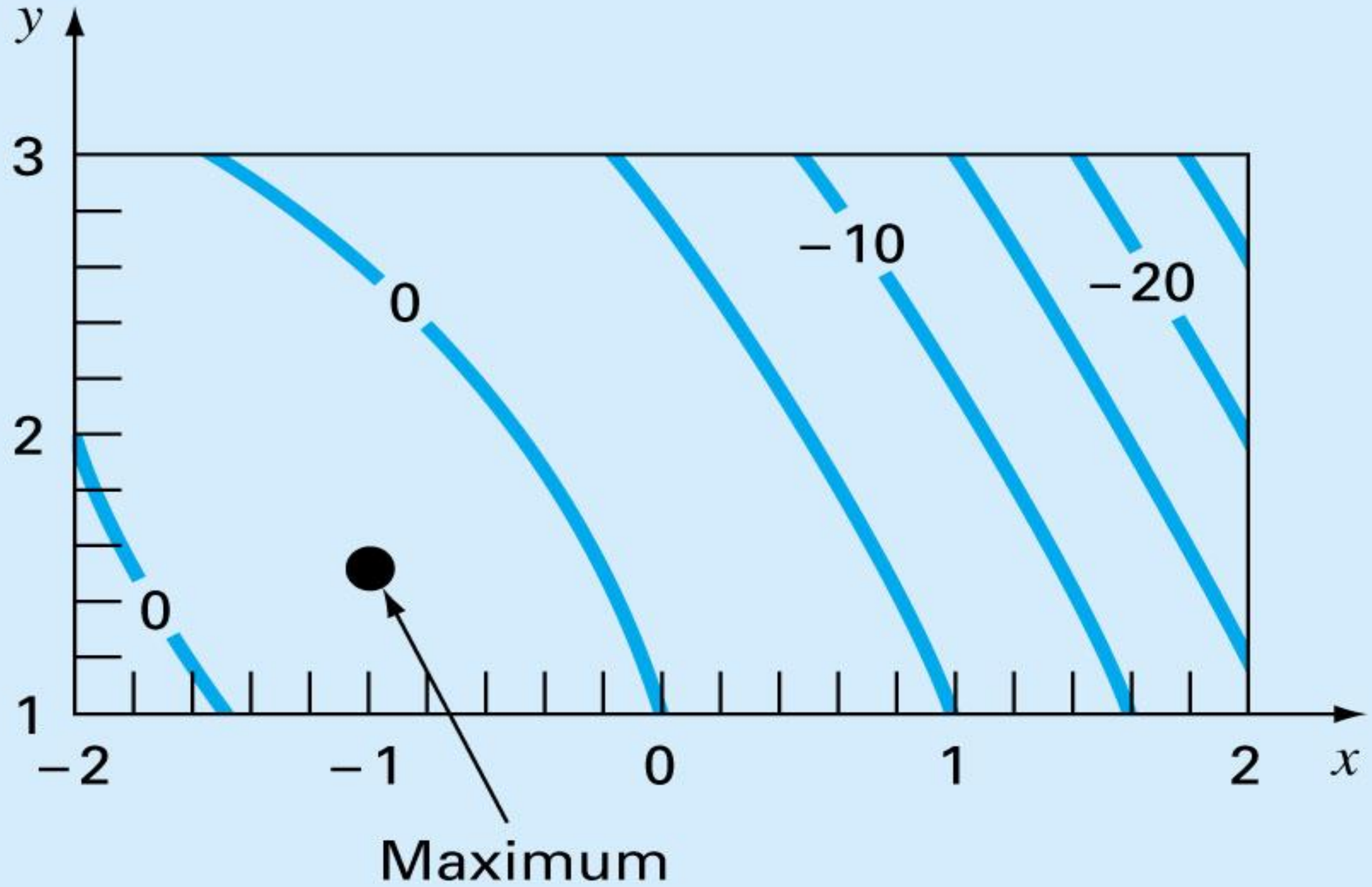
DIRECT METHODS
Random Search

- 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



Advantages/

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

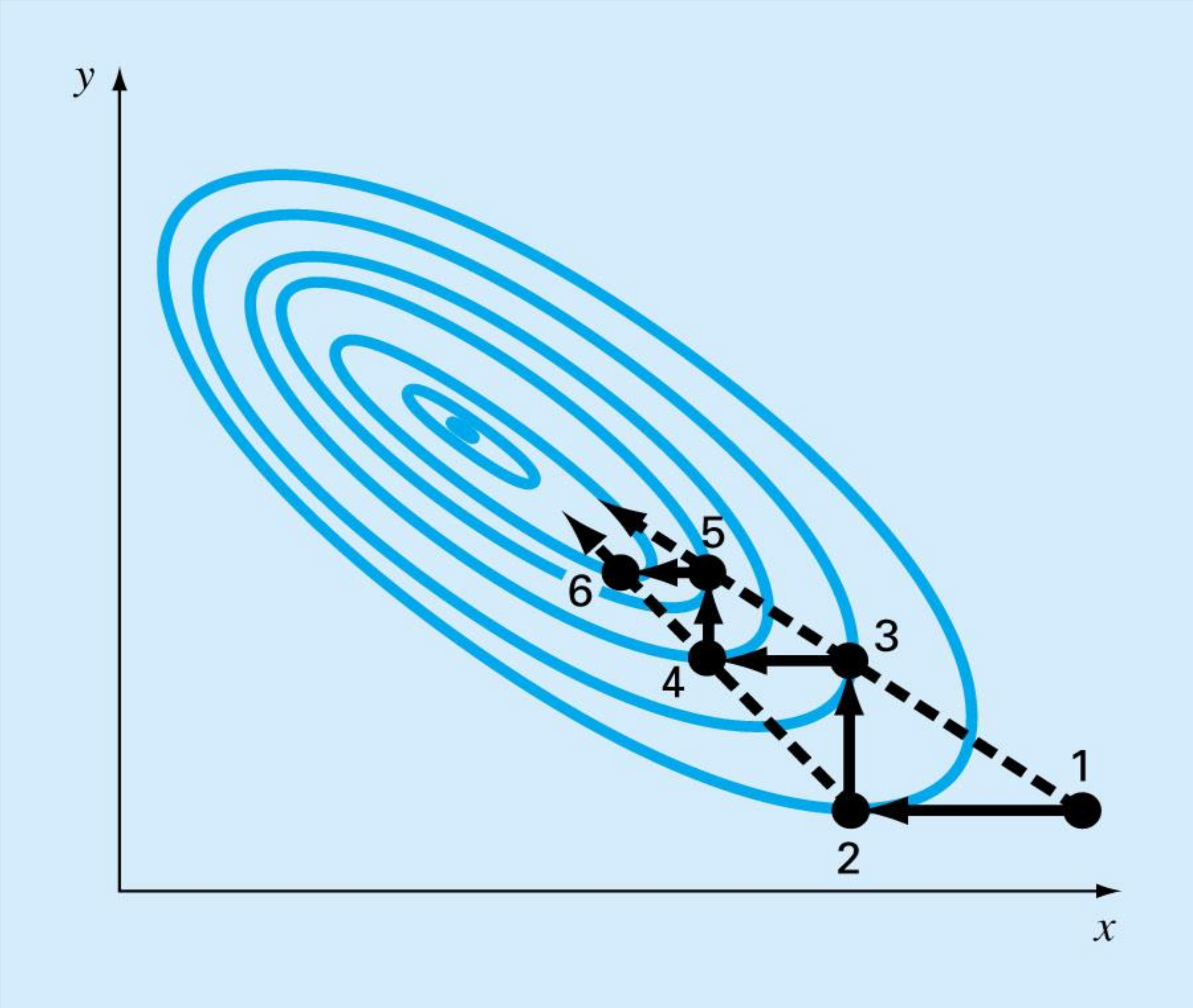
Disadvantages/

- 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



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

Figure 14.4

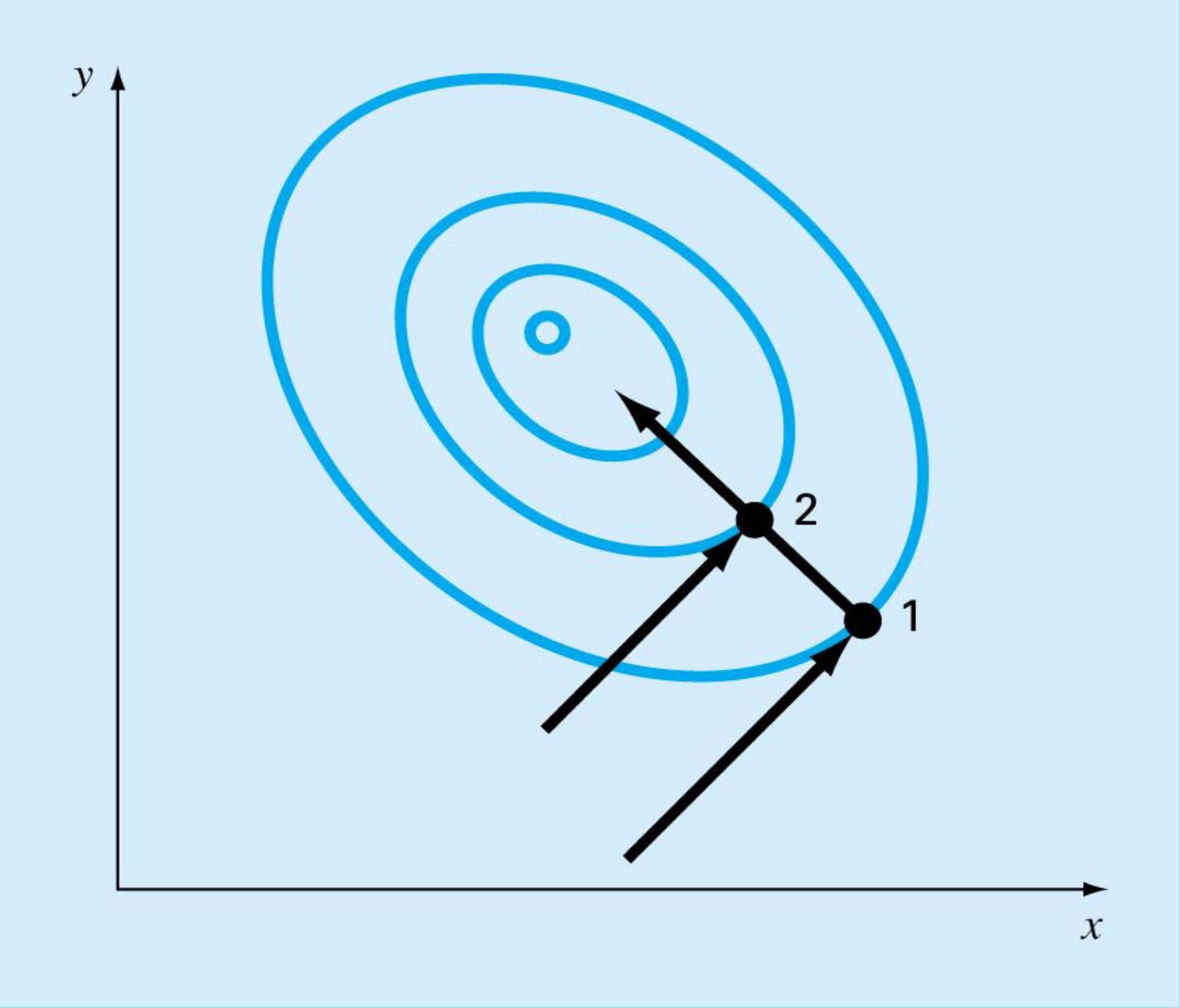
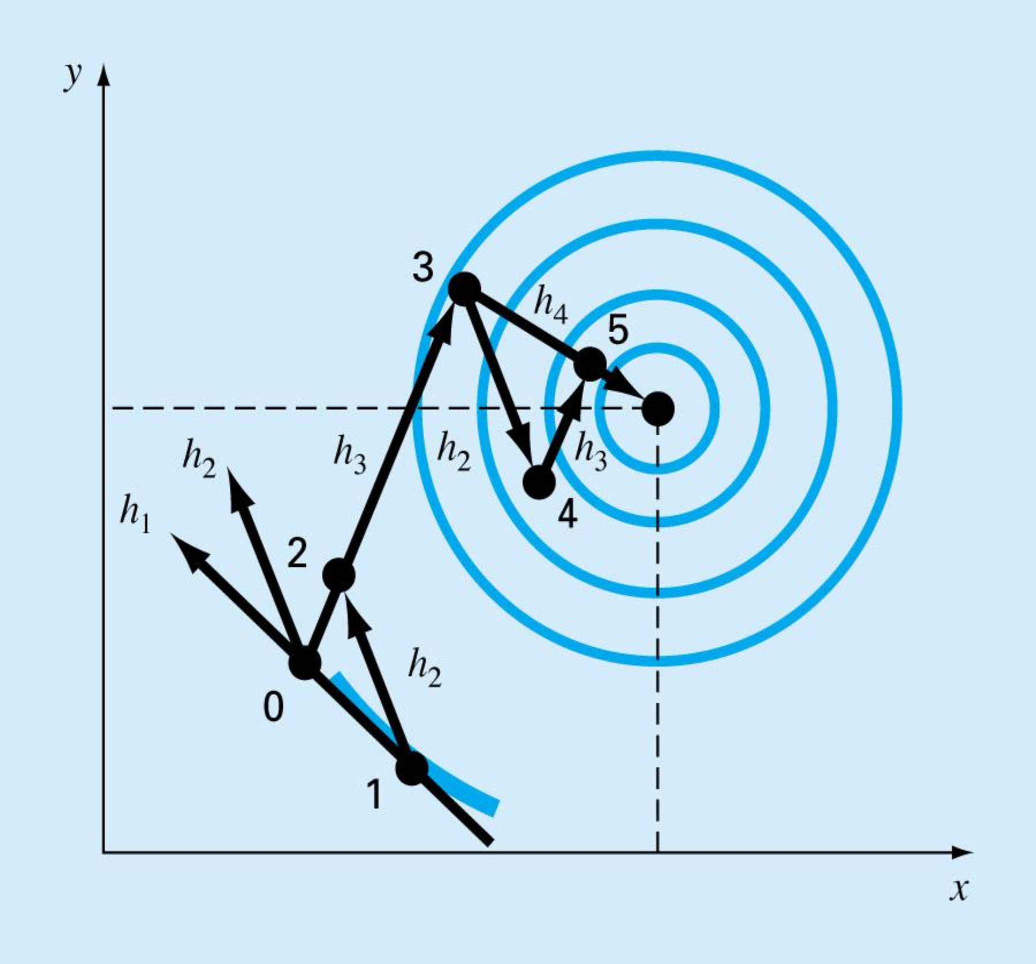


Figure 14.5

- 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*.



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?

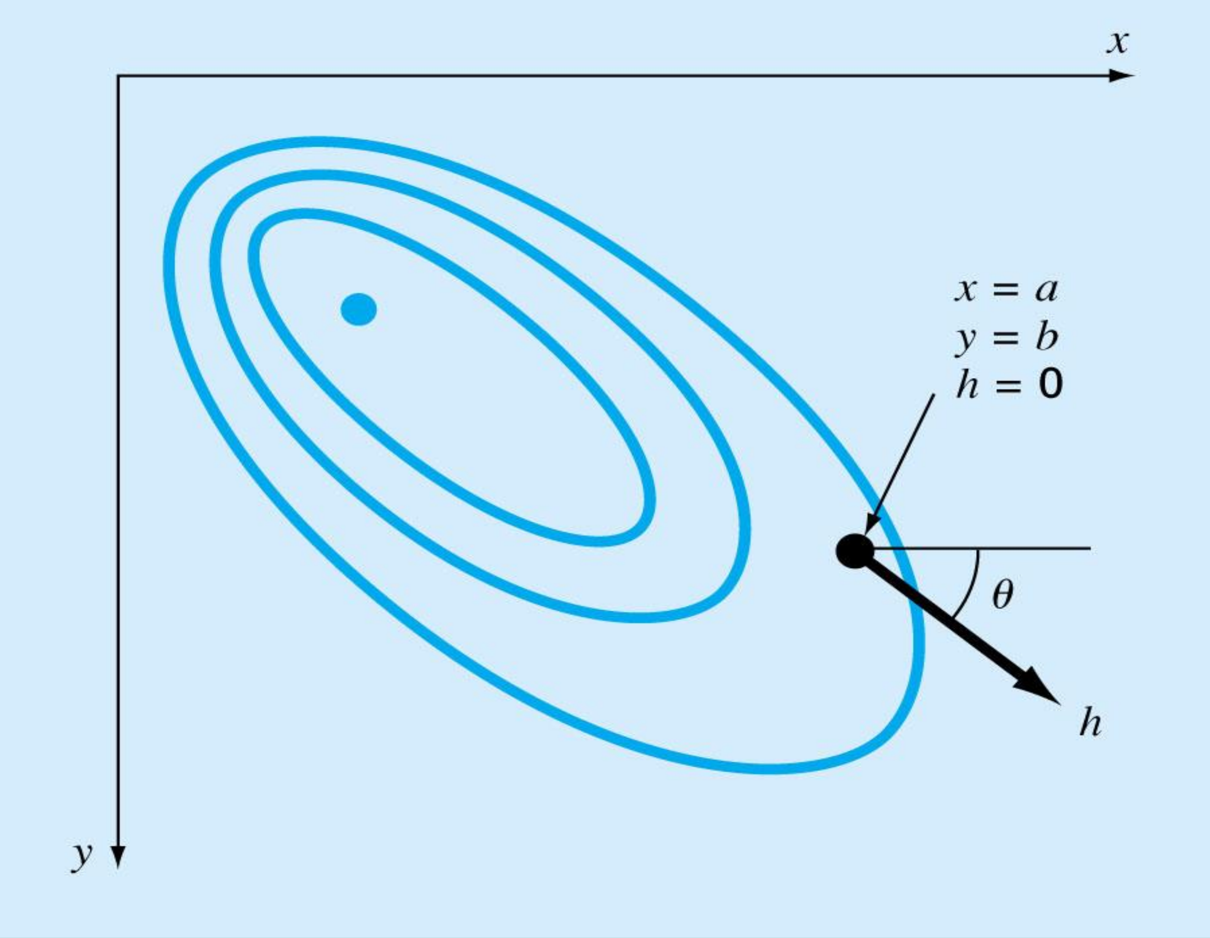
$$\nabla f = \frac{\partial f}{\partial x}\mathrm{i} + \frac{\partial f}{\partial y}\mathrm{j} \qquad \text{or } del\ f$$

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

Figure 14.6

- For n dimensions

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



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

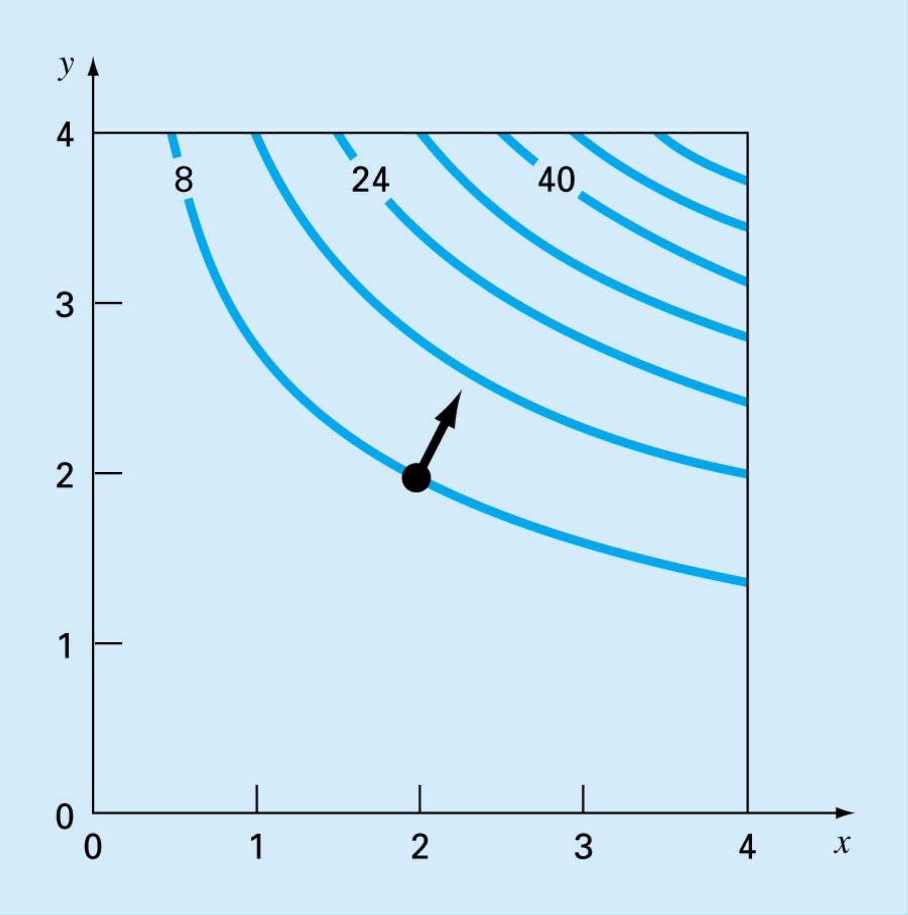
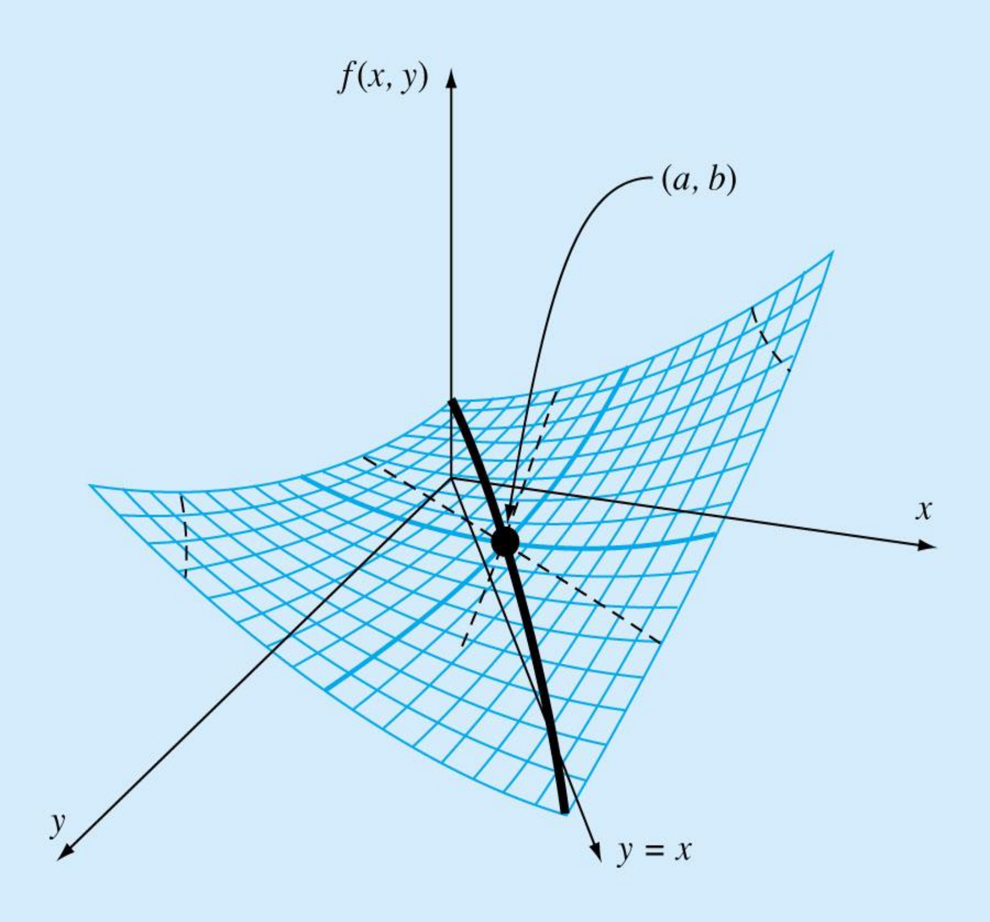


Figure 14.8



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

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

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

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

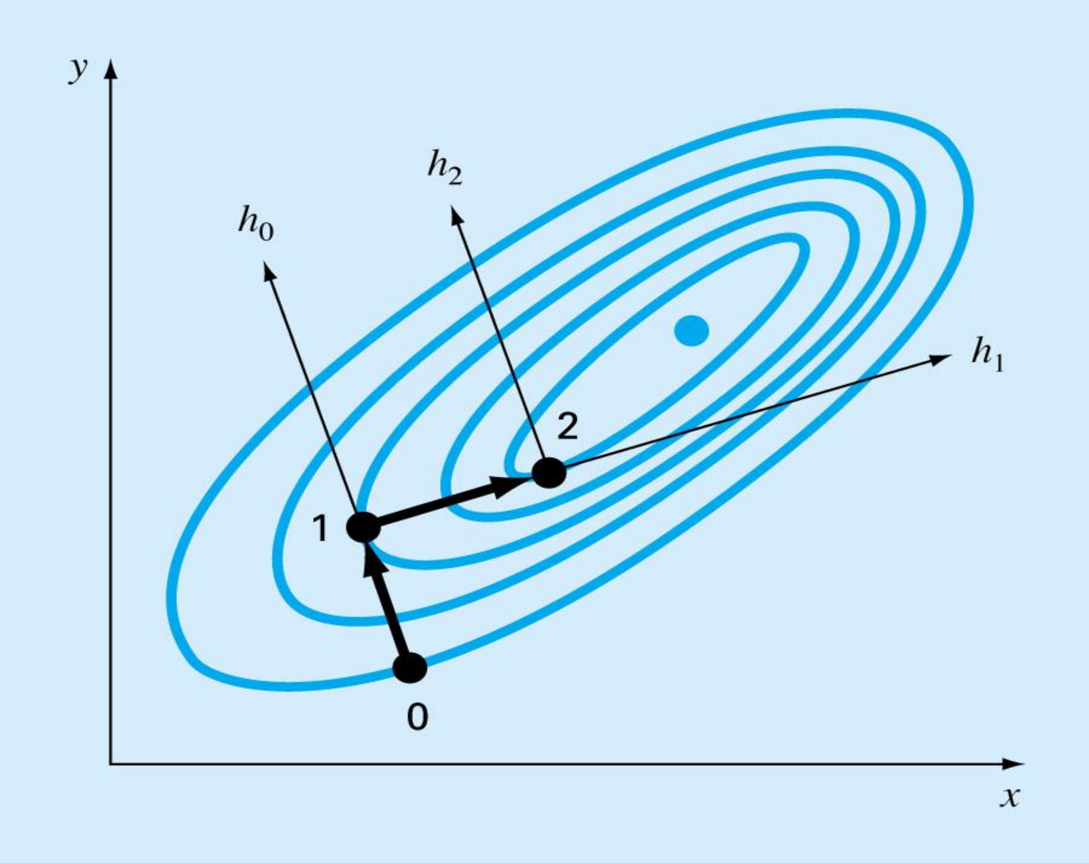
$$\text{If } \left|\text{H}\right| \prec 0\text{, then f(x, y) has a saddle point}$$

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



- 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

Figure 14.10

- If $x_o=1$ and $y_o=2$

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

$$x = 1 + 3h$$

$$y = 2 + 4h$$