Chapter 4: Unconstrained Optimization

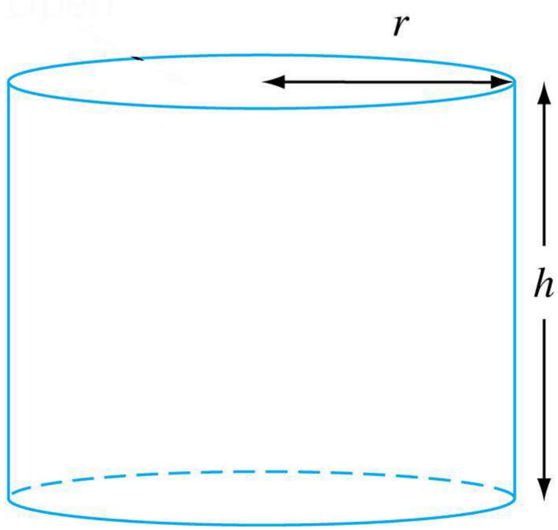
- Unconstrained optimization problem $\min_x F(x)$ or $\max_x F(x)$
- Constrained optimization problem

$$\begin{aligned} & \min_x F(x) \;\; \text{or} \;\; \max_x F(x) \\ \text{subject to} \;\; & g(x) = 0 \\ \text{and/or} \;\; & h(x) < 0 \text{ or } h(x) > 0 \end{aligned}$$

Example: minimize the outer area of a cylinder subject to a fixed volume. Objective function

$$F(x) = 2\pi r^2 + 2\pi r h, \;\; x = \begin{bmatrix} r \\ h \end{bmatrix}$$

Constraint: $2\pi r^2 h = V$



Outline:

- Part I: one-dimensional unconstrained optimization
  - Analytical method
  - Newton's method
  - Golden-section search method
- Part II: multidimensional unconstrained optimization
  - Analytical method
  - Gradient method — steepest ascent (descent) method
  - Newton's method

PART I: One-Dimensional Unconstrained Optimization Techniques

1 Analytical approach (1-D)

$\min_x F(x)$ or $\max_x F(x)$

- Let $F^{'}(x) = 0$ and find $x = x^*$.
- If $F^{''}(x^*) > 0$, $F(x^*) = \min_x F(x)$, x^* is a local minimum of $F(x)$;
- If $F^{''}(x^*) < 0$, $F(x^*) = \max_x F(x)$, x^* is a local maximum of $F(x)$;
- If $F^{''}(x^*) = 0$, x^* is a critical point of $F(x)$

Example 1: $F(x) = x^2$, $F^{'}(x) = 2x = 0$, $x^* = 0$. $F^{''}(x^*) = 2 > 0$. Therefore, $F(0) = \min_x F(x)$

Example 2: $F(x) = x^3$, $F^{'}(x) = 3x^2 = 0$, $x^* = 0$. $F^{''}(x^*) = 0$. x^* is not a local minimum nor a local maximum.

Example 3: $F(x) = x^4$, $F^{'}(x) = 4x^3 = 0$, $x^* = 0$. $F^{''}(x^*) = 0$.

In example 2, $F^{'}(x) > 0$ when $x < x^*$ and $F^{'}(x) > 0$ when $x > x^*$.

In example 3, x^* is a local minimum of $F(x)$. $F^{'}(x) < 0$ when $x < x^*$ and $F^{'}(x) > 0$ when $x > x^*$.

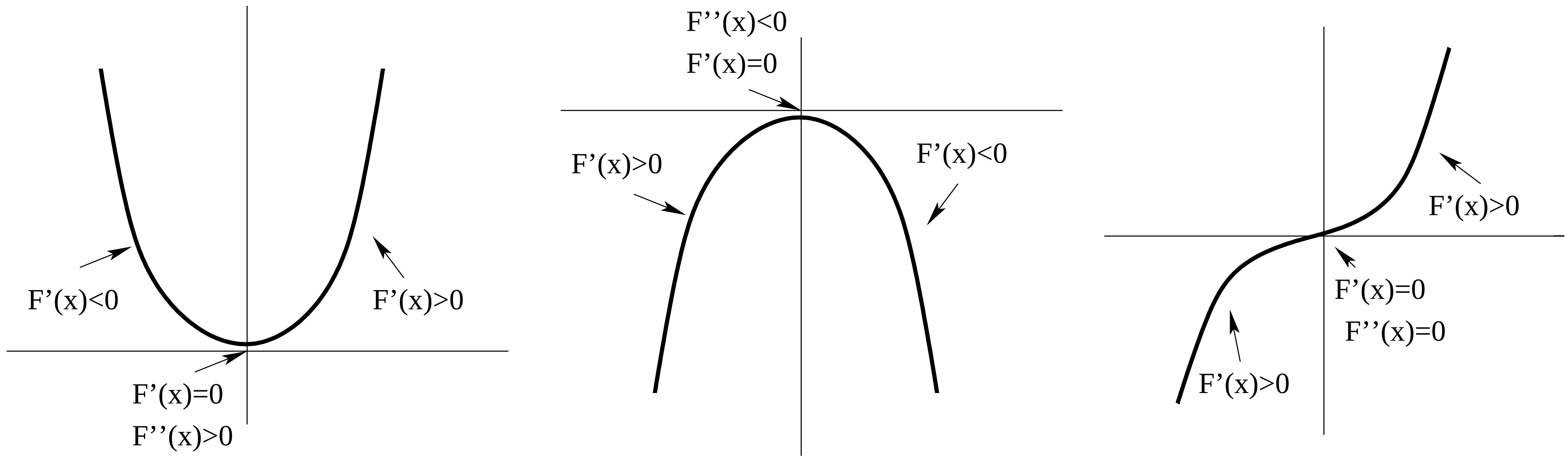


Figure 1: Example of constrained optimization problem

2 Newton's Method

$\min_x F(x)$ or $\max_x F(x)$

Use x_k to denote the current solution.

$$
\begin{aligned}
F(x_k + p) &= F(x_k) + pF'(x_k) + \frac{p^2}{2}F''(x_k) + \ldots \\
&\approx F(x_k) + pF'(x_k) + \frac{p^2}{2}F''(x_k)
\end{aligned}
$$

$$\begin{aligned} F(x^*) &= \min_x F(x) \approx \min_p F(x_k + p) \\ &\approx \min_p \left[F(x_k) + pF^{'}(x_k) + \frac{p^2}{2} F^{''}(x_k) \right] \end{aligned}$$

Let

$$\frac{\partial F(x)}{\partial p} = F^{'}(x_k) + pF^{''}(x_k) = 0$$

we have

$$p = -\frac{F^{'}(x_k)}{F^{''}(x_k)}$$

Newton's iteration

$$x_{k+1} = x_k + p = x_k - \frac{F^{'}(x_k)}{F^{''}(x_k)}$$

Example: find the maximum value of $f(x) = 2\sin x - \frac{x^2}{10}$ with an initial guess of $x_0 = 2.5$.

Solution:

$$f^{'}(x) = 2\cos x - \frac{2x}{10} = 2\cos x - \frac{x}{5}$$

$$f''(x) = -2\sin x - \frac{1}{5}$$

$$x_{i+1} = x_i - \frac{2\cos x_i - \frac{x_i}{5}}{-2\sin x_i - \frac{1}{5}}$$

$x_0 = 2.5$, $x_1 = 0.995$, $x_2 = 1.469$.

Comments:

- Same as N.-R. method for solving $F'(x) = 0$.
- Quadratic convergence, $|x_{k+1} - x^*| \leq \beta|x_k - x^*|^2$
- May diverge
- Requires both first and second derivatives
- Solution can be either local minimum or maximum

3 Golden-section search for optimization in 1-D

$\max_x F(x)$ ($\min_x F(x)$ is equivalent to $\max_x -F(x)$)
Assume: only 1 peak value (x^*) in (x_l, x_u)
Steps:

1. Select $x_l < x_u$
2. Select 2 intermediate values, x_1 and x_2 so that $x_1 = x_l + d$, $x_2 = x_u - d$, and $x_1 > x_2$.
3. Evaluate $F(x_1)$ and $F(x_2)$ and update the search range
   - If $F(x_1) < F(x_2)$, then $x^* < x_1$. Update $x_l = x_l$ and $x_u = x_1$.
   - If $F(x_1) > F(x_2)$, then $x^* > x_2$. Update $x_l = x_2$ and $x_u = x_u$.
   - If $F(x_1) = F(x_2)$, then $x_2 < x^* < x_1$. Update $x_l = x_2$ and $x_u = x_1$.
4. Estimate
   $x^* = x_1$ if $F(x_1) > F(x_2)$, and
   $x^* = x_2$ if $F(x_1) < F(x_2)$

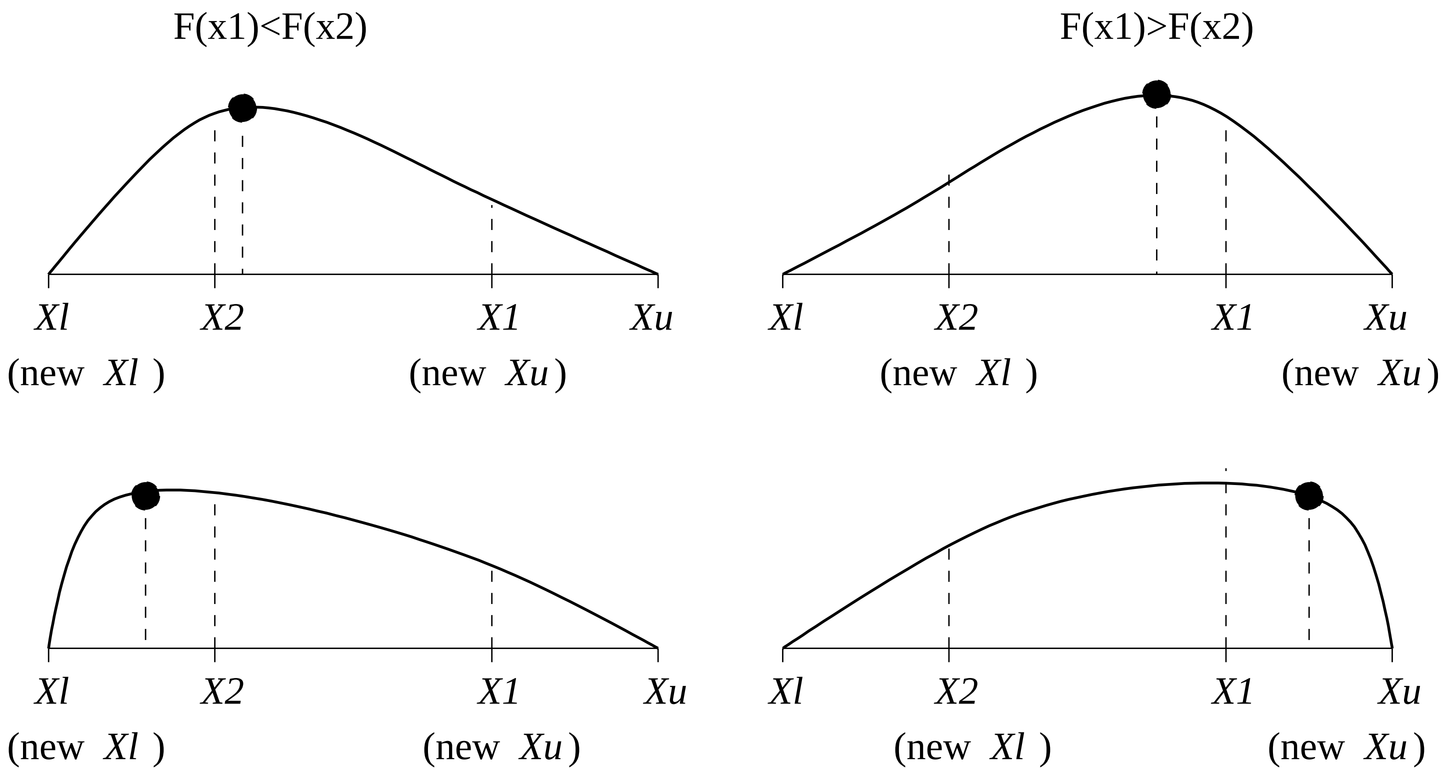


Figure 2: Golden search: updating search range

- Calculate ϵ_a. If $\epsilon_a < \epsilon_{threshold}$, end.

$$\epsilon_a = \left| \frac{x_{\text{new}} - x_{\text{old}}}{x_{\text{new}}} \right| \times 100\%$$

The choice of d

- Any values can be used as long as $x_1 > x_2$.
- If d is selected appropriately, the number of function evaluations can be minimized.

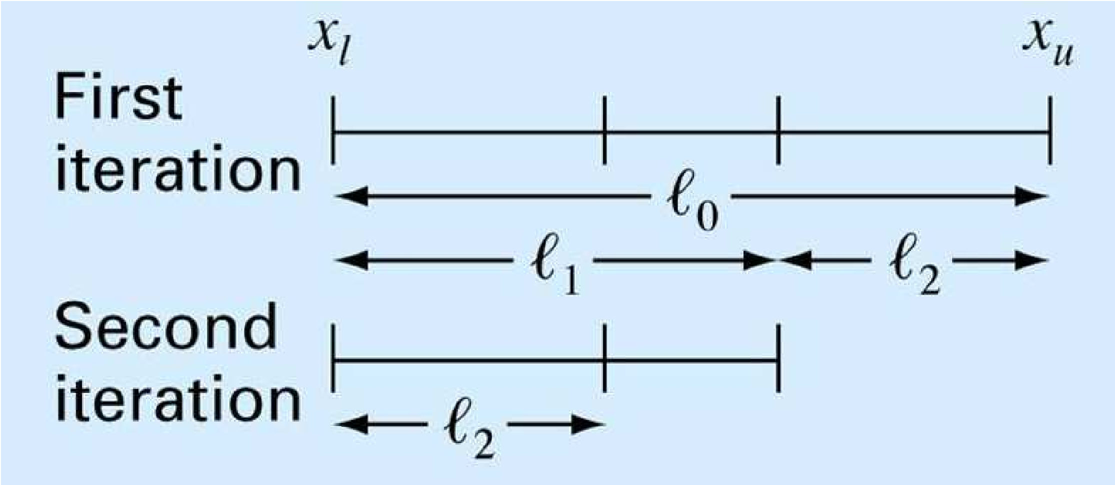


Figure 3: Golden search: the choice of d

$d_0 = l_1$, $d_1 = l_2 = l_0 - d_0 = l_0 - l_1$. Therefore, $l_0 = l_1 + l_2$.
$\frac{l_0}{d_0} = \frac{l_1}{d_1}$. Then $\frac{l_0}{l_1} = \frac{l_1}{l_2}$.
$l_1^2 = l_0 l_2 = (l_1 + l_2) l_2$. Then $1 = \left(\frac{l_2}{l_1}\right)^2 + \frac{l_2}{l_1}$.

Define $r = \frac{d_0}{l_0} = \frac{d_1}{l_1} = \frac{l_2}{l_1}$. Then $r^2 + r - 1 = 0$, and $r = \frac{\sqrt{5}-1}{2} \approx 0.618$
$d = r(x_u - x_l) \approx 0.618(x_u - x_l)$ is referred to as the golden value.

<u>Relative error</u>

$$\epsilon_a = \left| \frac{x_{\text{new}} - x_{\text{old}}}{x_{\text{new}}} \right| \times 100\%$$

Consider $F(x_2) < F(x_1)$. That is, $x_l = x_2$, and $x_u = x_u$.

For case (a), $x^* > x_2$ and x^* closer to x_2.

$$\begin{aligned}
\Delta x &\leq x_1 - x_2 = (x_l + d) - (x_u - d) \\
&= (x_l - x_u) + 2d = (x_l - x_u) + 2r(x_u - x_l) \\
&= (2r - 1)(x_u - x_l) \approx 0.236(x_u - x_l)
\end{aligned}$$

For case (b), $x^* > x_2$ and x^* closer to x_u.

$$\begin{aligned}
\Delta x &\leq x_u - x_1 \\
&= x_u - (x_l + d) = x_u - x_l - d \\
&= (x_u - x_l) - r(x_u - x_l) = (1 - r)(x_u - x_l) \\
&\approx 0.382(x_u - x_l)
\end{aligned}$$

Therefore, the maximum absolute error is $(1 - r)(x_u - x_l) \approx 0.382(x_u - x_l)$.

$$\begin{aligned}\epsilon_a &\le \left|\frac{\Delta x}{x^*}\right| \times 100\% \\ &\le \frac{(1-r)(x_u - x_l)}{|x^*|} \times 100\% \\ &= \frac{0.382(x_u - x_l)}{|x^*|} \times 100\%\end{aligned}$$

Example: Find the maximum of $f(x) = 2\sin x - \frac{x^2}{10}$ with $x_l = 0$ and $x_u = 4$ as the starting search range.

Solution:

Iteration 1: $x_l = 0$, $x_u = 4$, $d = \frac{\sqrt{5}-1}{2}(x_u - x_l) = 2.472$, $x_1 = x_l + d = 2.472$, $x_2 = x_u - d = 1.528$. $f(x_1) = 0.63$, $f(x_2) = 1.765$.
Since $f(x_2) > f(x_1)$, $x^* = x_2 = 1.528$, $x_l = x_l = 0$ and $x_u = x_1 = 2.472$.

Iteration 2: $x_l = 0$, $x_u = 2.472$, $d = \frac{\sqrt{5}-1}{2}(x_u - x_l) = 1.528$, $x_1 = x_l + d = 1.528$, $x_2 = x_u - d = 0.944$. $f(x_1) = 1.765$, $f(x_2) = 1.531$.
Since $f(x_1) > f(x_2)$, $x^* = x_1 = 1.528$, $x_l = x_2 = 0.944$ and $x_u = x_u = 2.472$.

4 Analytical Method

- Definitions:
  - If $f(x, y) < f(a, b)$ for all (x, y) near (a, b), $f(a, b)$ is a local maximum;
  - If $f(x, y) > f(a, b)$ for all (x, y) near (a, b), $f(a, b)$ is a local minimum.
- If $f(x, y)$ has a local maximum or minimum at (a, b), and the first order partial derivatives of $f(x, y)$ exist at (a, b), then

$$\frac{\partial f}{\partial x}|_{(a,b)} = 0, \text{ and } \frac{\partial f}{\partial y}|_{(a,b)} = 0$$

- If

$$\frac{\partial f}{\partial x}|_{(a,b)} = 0 \text{ and } \frac{\partial f}{\partial y}|_{(a,b)} = 0,$$

then (a, b) is a critical point or stationary point of $f(x, y)$.

- If

$$\frac{\partial f}{\partial x}|_{(a,b)} = 0 \text{ and } \frac{\partial f}{\partial y}|_{(a,b)} = 0$$

and the second order partial derivatives of $f(x, y)$ are continuous, then

- When $|H| > 0$ and $\frac{\partial^2 f}{\partial x^2}|_{(a,b)} < 0$, $f(a, b)$ is a local maximum of $f(x, y)$.
- When $|H| > 0$ and $\frac{\partial^2 f}{\partial x^2}|_{(a,b)} > 0$, $f(a, b)$ is a local minimum of $f(x, y)$.
- When $|H| < 0$, $f(a, b)$ is a saddle point.

Hessian of $f(x, y)$:

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

- $|H| = \frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} - \frac{\partial^2 f}{\partial x \partial y} \cdot \frac{\partial^2 f}{\partial y \partial x}$
- When $\frac{\partial^2 f}{\partial x \partial y}$ is continuous, $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$.
- When $|H| > 0$, $\frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} > 0$.

Example (saddle point): $f(x, y) = x^2 - y^2$.
$\frac{\partial f}{\partial x} = 2x$, $\frac{\partial f}{\partial y} = -2y$.
Let $\frac{\partial f}{\partial x} = 0$, then $x^* = 0$. Let $\frac{\partial f}{\partial y} = 0$, then $y^* = 0$.

Therefore, $(0, 0)$ is a critical point.
$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(2x) = 2$, $\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}(-2y) = -2$
$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}(-2y) = 0$, $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}(2x) = 0$
$|H| = \frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} - \frac{\partial^2 f}{\partial x \partial y} \cdot \frac{\partial^2 f}{\partial y \partial x} = -4 < 0$
Therefore, $(x^*, y^*) = (0, 0)$ is a saddle maximum.

Example: $f(x, y) = 2xy + 2x - x^2 - 2y^2$, find the optimum of $f(x, y)$.

Solution:
$\frac{\partial f}{\partial x} = 2y + 2 - 2x$, $\frac{\partial f}{\partial y} = 2x - 4y$.
Let $\frac{\partial f}{\partial x} = 0$, $-2x + 2y = -2$.
Let $\frac{\partial f}{\partial y} = 0$, $2x - 4y = 0$.
Then $x^* = 2$ and $y^* = 1$, i.e., $(2, 1)$ is a critical point.
$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(2y + 2 - 2x) = -2$
$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}(2x - 4y) = -4$
$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}(2x - 4y) = 2$, or

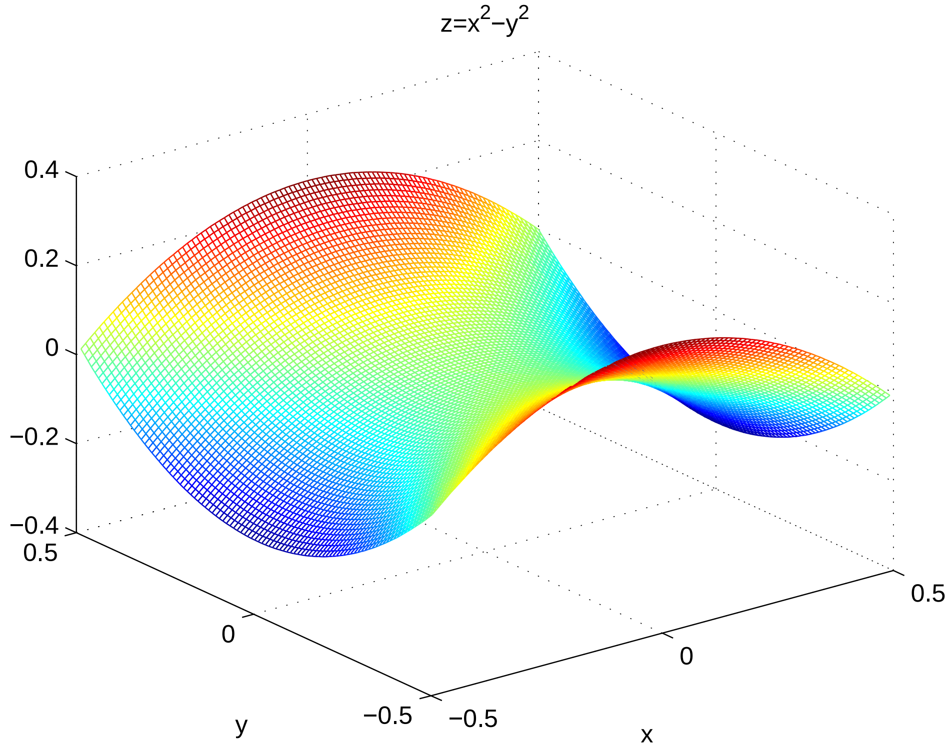


Figure 4: Saddle point

$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}(2y + 2 - 2x) = 2$
$|H| = \frac{\partial^2 f}{\partial x^2} \cdot \frac{\partial^2 f}{\partial y^2} - \frac{\partial^2 f}{\partial x \partial y} \cdot \frac{\partial^2 f}{\partial y \partial x} = (-2) \times (-4) - 2^2 = 4 > 0$
$\frac{\partial^2 f}{\partial x^2} < 0$. $(x^*, y^*) = (2, 1)$ is a local maximum.

5 Steepest Ascent (Descent) Method

Idea: starting from an initial point, find the function maximum (minimum) along the steepest direction so that shortest searching time is required.

Steepest direction: directional derivative is maximum in that direction — gradient direction.

Directional derivative

$$D_h f(x, y) = \frac{\partial f}{\partial x} \cdot \cos\theta + \frac{\partial f}{\partial y} \cdot \sin\theta = \langle [\frac{\partial f}{\partial x} \;\; \frac{\partial f}{\partial y}]' \cdot [\cos\theta \;\; \sin\theta]' \rangle$$

$\langle \cdot \rangle$: inner product

Gradient

When $[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y}]'$ is in the same direction as $[\cos\theta \quad \sin\theta]'$, the directional derivative is maximized. This direction is called gradient of $f(x, y)$.

The gradient of a 2-D function is represented as $\nabla f(x, y) = \frac{\partial f}{\partial x}\vec{i} + \frac{\partial f}{\partial y}\vec{j}$, or $[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y}]'$.

The gradient of an n-D function is represented as $\nabla f(\vec{X}) = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \ldots \quad \frac{\partial f}{\partial x_n}\right]'$, where $\vec{X} = [x_1 \quad x_2 \quad \ldots \quad x_n]'$

Example: $f(x, y) = xy^2$. Use the gradient to evaluate the path of steepest ascent at (2,2).

Solution:

$\frac{\partial f}{\partial x} = y^2$, $\frac{\partial f}{\partial y} = 2xy$.

$\frac{\partial f}{\partial x}|_{(2,2)} = 2^2 = 4$, $\frac{\partial f}{\partial y}|_{(2,2)} = 2 \times 2 \times 2 = 8$

Gradient: $\nabla f(x, y) = \frac{\partial f}{\partial x}\vec{i} + \frac{\partial f}{\partial y}\vec{j} = 4\vec{i} + 8\vec{j}$

$\theta = \tan^{-1}\frac{8}{4} = 1.107$, or 63.4^o.

$\cos\theta = \frac{4}{\sqrt{4^2+8^2}}$, $\sin\theta = \frac{8}{\sqrt{4^2+8^2}}$.

Directional derivative at (2,2): $\frac{\partial f}{\partial x} \cdot \cos\theta + \frac{\partial f}{\partial y} \cdot \sin\theta = 4\cos\theta + 8\sin\theta = 8.944$

If $\theta^{'} \neq \theta$, for example, $\theta^{'} = 0.5325$, then

$$D_{h'}f|_{(2,2)} = \frac{\partial f}{\partial x} \cdot \cos\theta^{'} + \frac{\partial f}{\partial y} \cdot \sin\theta^{'} = 4\cos\theta^{'} + 8\sin\theta^{'} = 7.608 < 8.944$$

Steepest ascent method

Ideally:

- Start from (x_0, y_0). Evaluate gradient at (x_0, y_0).
- Walk for a tiny distance along the gradient direction till (x_1, y_1).
- Reevaluate gradient at (x_1, y_1) and repeat the process.

Pros: always keep steepest direction and walk shortest distance
Cons: not practical due to continuous reevaluation of the gradient.

Practically:

- Start from (x_0, y_0).
- Evaluate gradient (h) at (x_0, y_0).

- Evaluate $f(x, y)$ in direction h.
- Find the maximum function value in this direction at (x_1, y_1).
- Repeat the process until (x_{i+1}, y_{i+1}) is close enough to (x_i, y_i).

<u>Find $\vec{X}_{i+1}$ from $\vec{X}_i$</u>

For a 2-D function, evaluate $f(x, y)$ in direction h:

$$g(\alpha) = f(x_i + \frac{\partial f}{\partial x}|_{(x_i, y_i)} \cdot \alpha, y_i + \frac{\partial f}{\partial y}|_{(x_i, y_i)} \cdot \alpha)$$

where α is the coordinate in h-axis.

For an n-D function $f(\vec{X})$,

$$g(\alpha) = f(\vec{X} + \nabla f|_{(\vec{X}_i)} \cdot \alpha)$$

Let $g^{'}(\alpha) = 0$ and find the solution $\alpha = \alpha^*$.

Update $x_{i+1} = x_i + \frac{\partial f}{\partial x}|_{(x_i, y_i)} \cdot \alpha^*$, $y_{i+1} = y_i + \frac{\partial f}{\partial y}|_{(x_i, y_i)} \cdot \alpha^*$.

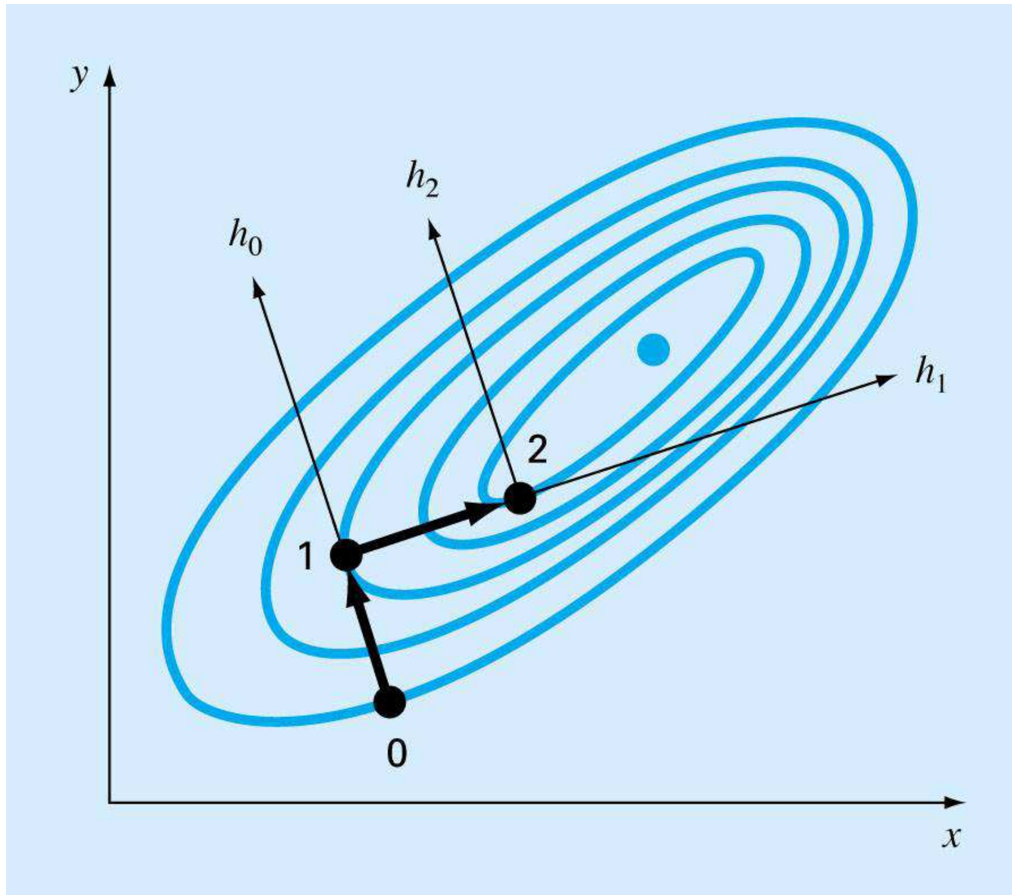


Figure 5: Illustration of steepest ascent

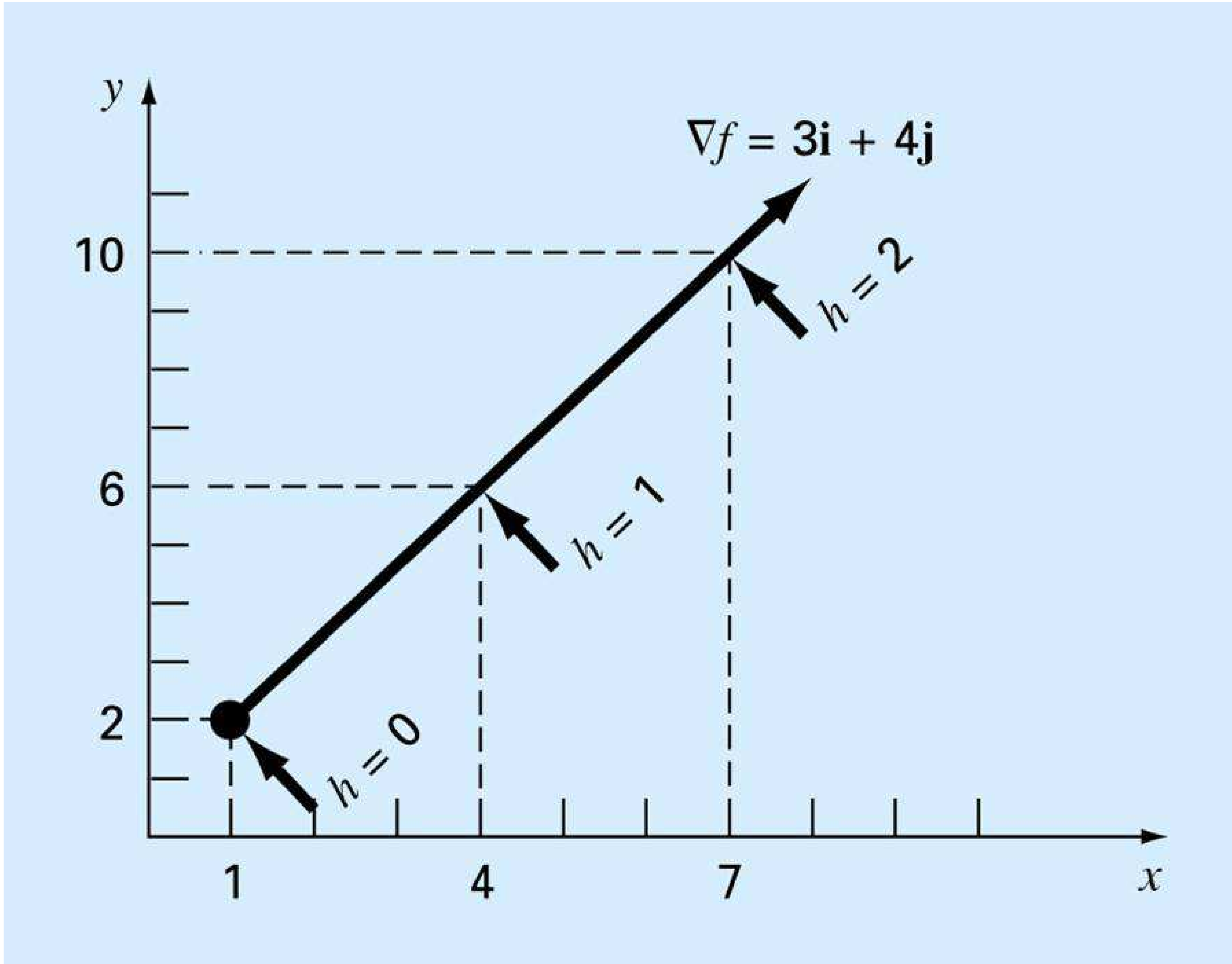


Figure 6: Relationship between an arbitrary direction h and x and y coordinates

Example: $f(x,y) = 2xy + 2x - x^2 - 2y^2$, $(x_0, y_0) = (-1, 1)$.

First iteration:

$x_0 = -1$, $y_0 = 1$.

$\frac{\partial f}{\partial x}|_{(-1,1)} = 2y + 2 - 2x|_{(-1,1)} = 6$, $\frac{\partial f}{\partial y}|_{(-1,1)} = 2x - 4y|_{(-1,1)} = -6$

$\nabla f = 6\vec{i} - 6\vec{j}$

$$\begin{aligned} g(\alpha) &= f(x_0 + \frac{\partial f}{\partial x}|_{(x_0,y_0)} \cdot \alpha, y_0 + \frac{\partial f}{\partial y}|_{(x_0,y_0)} \cdot \alpha) \\ &= f(-1 + 6\alpha, 1 - 6\alpha) \\ &= 2 \times (-1 + 6\alpha) \cdot (1 - 6\alpha) + 2(-1 + 6\alpha) - (-1 + 6\alpha)^2 - 2(1 - 6\alpha)^2 \\ &= -180\alpha^2 + 72\alpha - 7 \end{aligned}$$

$g^{'}(\alpha) = -360\alpha + 72 = 0$, $\alpha^* = 0.2$.

Second iteration:

$x_1 = x_0 + \frac{\partial f}{\partial x}|_{(x_0,y_0)} \cdot \alpha^* = -1 + 6 \times 0.2 = 0.2$, $y_1 = y_0 + \frac{\partial f}{\partial y}|_{(x_0,y_0)} \cdot \alpha^* = 1 - 6 \times 0.2 = -0.2$

$\frac{\partial f}{\partial x}|_{(0.2,-0.2)} = 2y + 2 - 2x|_{(0.2,-0.2)} = 2 \times (-0.2) + 2 - 2 \times 0.2 = 1.2$,

$\frac{\partial f}{\partial y}|_{(0.2,-0.2)} = 2x - 4y|_{(0.2,-0.2)} = 2 \times 0.2 - 4 \times (-0.2) = 1.2$

$\nabla f = 1.2\vec{i} + 1.2\vec{j}$

$$\begin{aligned} g(\alpha) &= f(x_1 + \frac{\partial f}{\partial x}|_{(x_1,y_1)} \cdot \alpha, y_1 + \frac{\partial f}{\partial y}|_{(x_1,y_1)} \cdot \alpha) \\ &= f(0.2 + 1.2\alpha, -0.2 + 1.2\alpha) \\ &= 2 \times (0.2 + 1.2\alpha) \cdot (-0.2 + 1.2\alpha) + 2(0.2 + 1.2\alpha) \\ &\quad -(0.2 + 1.2\alpha)^2 - 2(-0.2 + 1.2\alpha)^2 \\ &= -1.44\alpha^2 + 2.88\alpha + 0.2 \end{aligned}$$

$g^{'}(\alpha) = -2.88\alpha + 2.88 = 0$, $\alpha^* = 1$.

Third iteration:

$x_2 = x_1 + \frac{\partial f}{\partial x}|_{(x_1,y_1)} \cdot \alpha^* = 0.2 + 1.2 \times 1 = 1.4$, $y_2 = y_1 + \frac{\partial f}{\partial y}|_{(x_1,y_1)} \cdot \alpha^* = -0.2 + 1.2 \times 1 = 1$

$\ldots$

$(x^*, y^*) = (2, 1)$

6 Newton's Method

Extend the Newton's method for 1-D case to multidimensional case.
Given $f(\vec{X})$, approximate $f(\vec{X})$ by a second order Taylor series at $\vec{X} = \vec{X}_i$:

$$f(\vec{X}) \approx f(\vec{X}_i) + \nabla f^{'}(\vec{X}_i)(\vec{X} - \vec{X}_i) + \frac{1}{2}(\vec{X} - \vec{X}_i)^{'} H_i(\vec{X} - \vec{X}_i)$$

where H_i is the Hessian matrix

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ & \cdots & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

At the maximum (or minimum) point, $\frac{\partial f(\vec{X})}{\partial x_j} = 0$ for all $j = 1, 2, \ldots, n$, or $\nabla f = \vec{0}$. Then

$$\nabla f(\vec{X}_i) + H_i(\vec{X} - \vec{X}_i) = 0$$

If H_i is non-singular,

$$\vec{X} = \vec{X}_i - H_i^{-1} \nabla f(\vec{X}_i)$$

Iteration: $\vec{X}_{i+1} = \vec{X}_i - H_i^{-1}\nabla f(\vec{X}_i)$

Example: $f(\vec{X}) = 0.5x_1^2 + 2.5x_2^2$

$\nabla f(\vec{X}) = \begin{bmatrix} x_1 \\ 5x_2 \end{bmatrix}$

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

$\vec{X}_0 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \vec{X}_1 = \vec{X}_0 - H^{-1}\nabla f(\vec{X}_0) = \begin{bmatrix} 5 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 5 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Comments: Newton's method

- Converges quadratically near the optimum
- Sensitive to initial point
- Requires matrix inversion
- Requires first and second order derivatives