ĐẠI HỌC BÁCH KHOA

25 YEARS ANNIVERSARY SOICT

HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

FUNDAMENTALS OF OPTIMIZATION

Unconstrained convex optimization

CONTENT

- Unconstrained optimization problems
- Descent method
- Gradient descent method
- Newton method
- Subgradient method

Unconstrained convex optimization

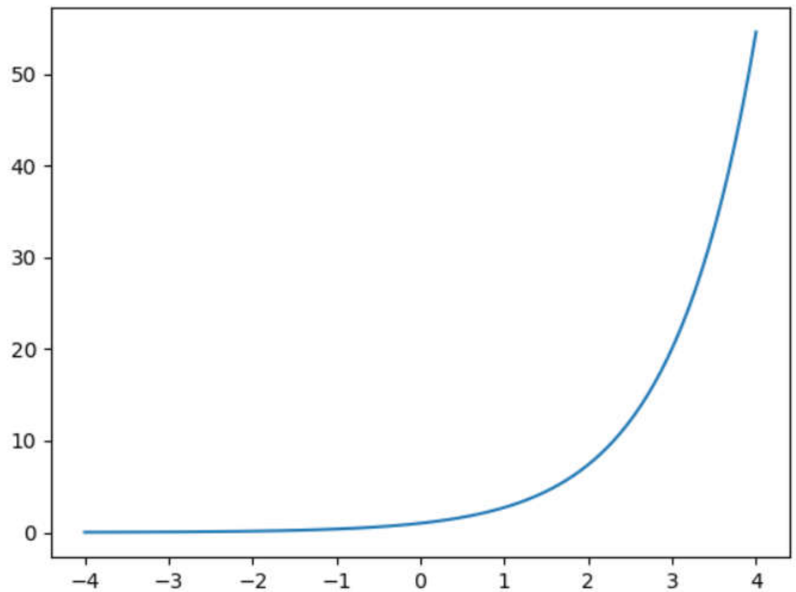
- Unconstrained, smooth convex optimization problem:

$$\min f(x)$$

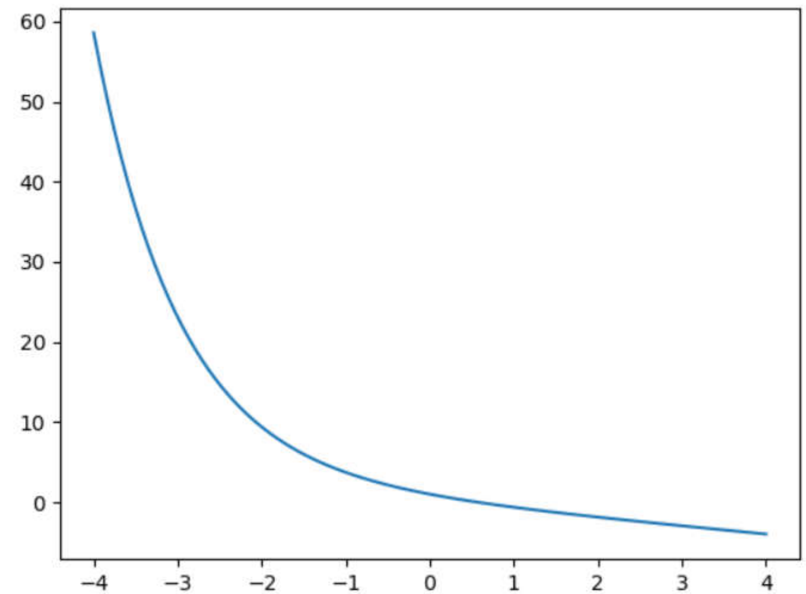
  - $f: R^n \rightarrow R$ is convex and twice differentiable
  - ***dom*** $f = R$: no constraint
  - Assumption: the problem is solvable with $f^* = \min_x f(x)$ and $x^* = \text{argMin}_x f(x)$
- To find x, solve equation $\nabla f(x^*) = 0$: not easy to solve analytically
- Iterative scheme is preferred: compute minimizing sequence $x^{(0)}, x^{(1)}, \ldots$ s.t. $f(x^{(k)}) \rightarrow f(x^*)$ as $k \rightarrow \infty$
- The algorithm stops at some point $x^{(k)}$ when the error is within acceptable tolerance: $f(x^{(k)}) - f^* \leq \varepsilon$

Local minimizer

- x^* is a local minimizer for f: $R^n \rightarrow R$ if $f(x^*) \leq f(x)$ for $||x^*-x|| \leq \varepsilon$ ($\varepsilon > 0$ is a constant)
- x^* is a global minimizer for f: $R^n \rightarrow R$ if $f(x^*) \leq f(x)$ for all $x \in R^n$



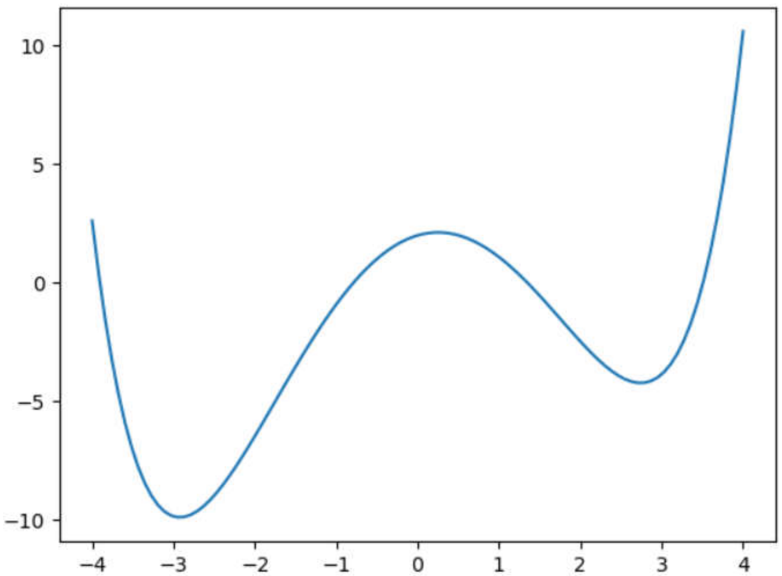
$f(x) = e^x$ has no minimizer



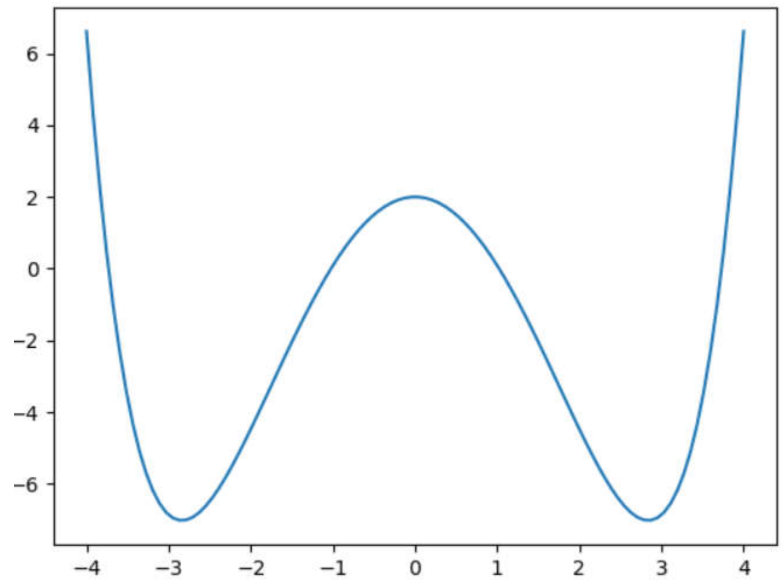
$f(x) = -x + e^{-x}$ has no minimizer

Local minimizer

- x^* is a local minimizer for f: $R^n \rightarrow R$ if $f(x^*) \le f(x)$ for $||x^*-x|| \le \varepsilon$ ($\varepsilon > 0$ is a constant)
- x^* is a global minimizer for f: $R^n \rightarrow R$ if $f(x^*) \le f(x)$ for all $x \in R^n$



$f(x) = e^x + e^{-x} - 3x^2 + x$ has one local minimizer and one global minimizer



$f(x) = e^x + e^{-x} - 3x^2$ has two global minimizers

Local minimizer

- **Theorem** (Necessary condition for local minimum) If x^* is a local minimizer for f: $R^n \rightarrow R$, then $\nabla f(x^*) = 0$ (x^* is also called *stationary point* for f)

Local minimizer

Example

- $f(x,y) = x^2 + y^2 - 2xy + x$
- $\nabla f(x) = \begin{pmatrix} 2x - 2y + 1 \\ 2y - 2x \end{pmatrix} = 0$ has no solution

→ there is no minimizer of $f(x,y)$

Local minimizer

- **Theorem** (Sufficient condition for a local minimum) Assume x^* is a stationary point and that $\nabla^2 f(x^*)$ is positive definite, then x^* is a local minimizer

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial xn} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial xn} \\ \cdots & & & \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial xn \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial xn \partial xn} \end{pmatrix}$$

Local minimizer

- Matrix A_{nxn} is called positive definite if

$$A^i = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,i} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,i} \\ & \cdots\cdots & & \\ a_{i,1} \;..\; & a_{i,2} & \cdots & a_{i,i} \end{pmatrix}, \quad \det(A^i) > 0,\ i = 1,\ldots,n$$

Local minimizer

- **Example** $f(x,y) = e^{x^2+y^2}$

$\nabla f(x,y) = \begin{pmatrix} 2xe^{x^2+y^2} \\ 2ye^{x^2+y^2} \end{pmatrix}$ = 0 has unique solution $x^* = (0,0)$

$\nabla^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ > 0 → (0,0) is a minimizer of f

Local minimizer

- **Example** Find a minimizer of $f(x,y) = x^2 + y^2 - 2xy - x$?

Descent method

Determine starting point $x^{(0)} \in R^n$;

$k \leftarrow 0$;

while(stop condition not reach){

  Determine a search direction $p_k \in R^n$;

  Determine a step size $\alpha_k > 0$ s.t. $f(x^{(k)} + \alpha_k p_k) < f(x^{(k)})$;

  $x^{(k+1)} \leftarrow x^{(k)} + \alpha_k p_k$;

  $k \leftarrow k+1$;

}

Stop condition may be

- $|| \nabla f(x^k)|| \le \varepsilon$
- $||x^{k+1} - x^k|| \le \varepsilon$
- $k > K$ (maximum number of iterations)

Gradient descent method

- Gradient descent schema

$$x^{(k)} = x^{(k-1)} - \alpha_k \nabla f(x^{(k-1)})$$

init $x^{(0)}$;

$k = 1$;

while stop condition not reach{

  specify constant α_k;

  $x^{(k)} = x^{(k-1)} - \alpha_k \nabla f(x^{(k-1)})$;

  $k = k + 1$;

}

- α_k might be specified in such a way that $f(x^{(k-1)} - \alpha_k \nabla f(x^{(k-1)}))$ is minimized: $\frac{\partial f}{\partial \alpha_k} = 0$

Gradient descent method

Example $f(x_1,x_2,x_3) = x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_2x_3 + x_1 + x_3 \rightarrow \min$?

Newton method

- Second-order Taylor approximation g of f at x is

$$f(x+h) \approx g(x+h) = f(x) + h\nabla f(x) + \frac{1}{2}h^2\ \nabla^2 f(x)$$

- Which is a convex quadratic function of h
- g(x+h) is minimized when $\frac{\partial g}{\partial h} = 0 \rightarrow h = -\nabla^2 f(x)^{-1}\nabla f(x)$

Newton method

Generate $x^{(0)}$; // starting point

$k = 0$;

while stop condition not reach{

  $x^{(k+1)} \leftarrow x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$;

  $k = k + 1$;

}

Newton method

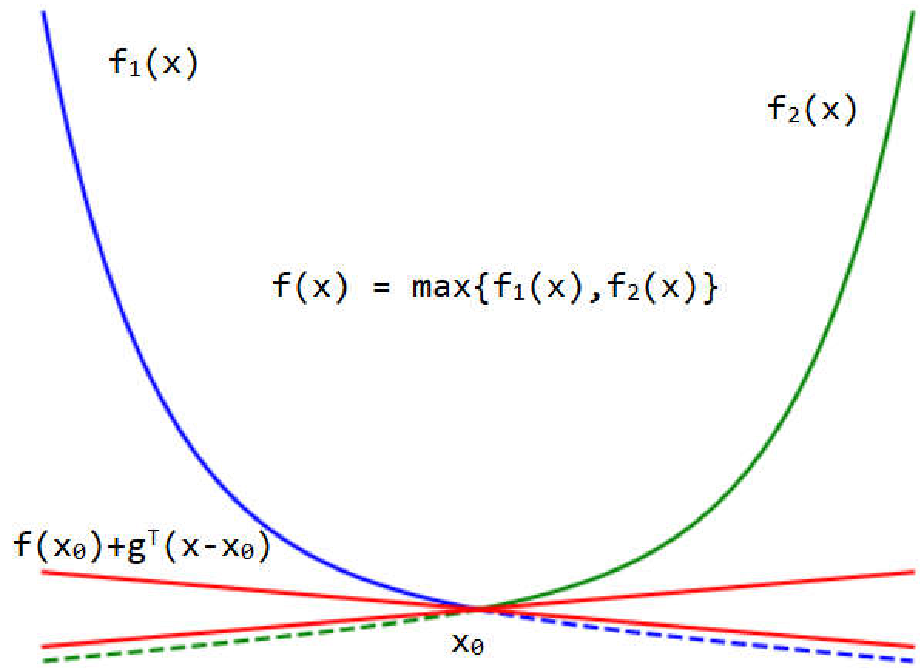
Step	x	y	f
Initialization	[0,0,0]	[1, 1, 1]	0
Step 1	[-1., -1., -1.]	[-2.46519033e-32 1.11022302e-16 2.22044605e-16]	-1.0000000000000004
Step 2	[-1., -1., -1.]	[0., 0., 0.]	-1

Subgradient method

- For minimize nondifferentiable convex function
- Subgradient method is not a descent method: the function value can increase

Subgradient method

- Subgradient of *f* at *x*
  - Any vector *g* such that $f(x') \geq f(x) + g^T(x'-x)$
  - If f is differentiable, only possible choice is $g^{(k)} = \nabla f(x^{(k)})$, → the subgradient method reduces to the gradient method



Basic subgradient method

$$x^{(k+1)} = x^{(k)} - \alpha_k g^{(k)}$$

- $x^{(k)}$: is at the k^{th} iteration
- $g^{(k)}$: any subgradient of f at $x^{(k)}$
- $\alpha_k > 0$ is the k^{th} step size
- Note: subgradient is not a descent method, thus $f_{best}^{(k)} = \min\{f(x^{(1)}), f(x^{(2)}), \ldots, f(x^{(k)})\}$

Convergence proof

- Notations: x^* is a minimizer of f
- Assumptions
  - Norm of the subgradients is bounded (with a constant G): $||g^{(k)}||_2 \leq G$ (this is the case if, for example, f satisfies the Lipschitz condition $|f(u) - f(v)| \leq G||u\text{-}v||_2$)
  - $||x^{(1)} - x^*||_2^2 \leq R$ (with a known constant R)
- We have $||x^{(k+1)} - x^*||_2^2 = ||x^{(k)} - \alpha_k g^{(k)} - x^*||_2^2$

$= ||x^{(k)} - x^*||_2^2 - 2\alpha_k g^{(k)T}(x^{(k)} - x^*) + \alpha_k^2||g^{(k)}||_2^2$

$\leq ||x^{(k)} - x^*||_2^2 - 2\alpha_k(f(x^{(k)}) - f(x^*)) + \alpha_k^2||g^{(k)}||_2^2$ (due to the fact that $f(x^*) \geq f(x^{(k)}) + g^{(k)T}(x^* - x^{(k)})$) (1)

Convergence proof

- Apply the inequality (1) recursively, we have

$||x^{(k+1)} - x^*||_2^2 \leq ||x^{(1)} - x^*||_2^2 - 2\sum_{i=1}^{k} \alpha_i(f(x^{(i)}) - f^*) + \sum_{i=1}^{k} \alpha_i^2 ||g^{(i)}||_2^2$ (where $f^* = f(x^*)$)

$\rightarrow 2\sum_{i=1}^{k} \alpha_i(f(x^{(i)}) - f^*) \leq R^2 + \sum_{i=1}^{k} \alpha_i^2 ||g^{(i)}||_2^2$

$\rightarrow R^2 + \sum_{i=1}^{k} \alpha_i^2 ||g^{(i)}||_2^2 \geq 2\sum_{i=1}^{k} \alpha_i(f(x^{(i)}) - f^*) \geq 2(\sum_{i=1}^{k} \alpha_i) \min_{i=1,\dots,k}(f(x^{(i)}) - f^*) = 2\sum_{i=1}^{k} \alpha_i(f_{best}^{(k)} - f^*)$

$$\rightarrow f_{best}^{(k)} - f^* \leq \frac{R^2 + \sum_{i=1}^{k} \alpha_i^2 ||g^{(i)}||_2^2}{2\sum_{i=1}^{k} \alpha_i} \qquad (2)$$

Convergence proof

- Different cases
  - Constant step size $\alpha_k = \alpha$

→ $f_{best}^{(k)} - f^* \leq \dfrac{R^2 + G^2\alpha^2 k}{2\alpha k}$

→ $f_{best}^{(k)} - f^*$ converges to $G^2\alpha/2$ when $k \to \infty$

  - Constant step length $\alpha_k = \gamma / ||g^{(k)}||_2$

→ $f_{best}^{(k)} - f^* \leq \dfrac{R^2 + \gamma^2 k}{2\gamma k/G}$

→ $f_{best}^{(k)} - f^*$ converges to $G\gamma/2$ when $k \to \infty$

Convergence proof

- Different cases
  - Square summable but not summable

$||\alpha||_2^2 = \sum_{i=1}^{\infty} \alpha_i^2 < \infty$ and $\sum_{i=1}^{\infty} \alpha_i = \infty$

$$\rightarrow f_{best}^{(k)} - f^* \leq \frac{R^2 + G^2||\alpha||_2^2}{2\sum_{i=1}^{k} \alpha_i}$$

$\rightarrow f_{best}^{(k)} - f^*$ converges to 0 as $k \rightarrow \infty$

Example

$$\text{minimize } f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i)$$

- Finding subgradient: given x, the index j for which

$$a_j^T x + b_j = \max_{i=1,\ldots,m}(a_i^T x + b_i)$$

→ subgradient at x is $g = a_j$

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Thank you for your attentions!