Fundamentals of Optimization

Mathematical foundation

Pham Quang Dung

dungpq@soict.hust.edu.vn

Department of Computer Science

Optimization problems

- Maximize or minimize some function relative to some set (range of choices)
- The function represents the quality of the choice, indicating which is the "best"
- Example
  - A shipper need to find the shortest route to deliver packages to customers 1, 2, …, N

Notations

- $x \in R^n$: vector of decision variables $x_{j,}$ $j = 1, 2, \ldots, n$
- $f : R^n \rightarrow R$ is the objective function (**dom** $f = R^n$)
- $g_i : R^n \rightarrow R$ is the constraint function defining restriction on x, $i = 1, 2, \ldots, m$

minimize $f(x)$ over $x = (x_1, x_2, \ldots, x_n) \in X \subset R^n$
satisfying a property P:

$$g_i(\mathrm{x}) \leq b_i,\ i = 1, 2, \ldots, s$$
$$g_i(\mathrm{x}) = d_i,\ i = s + 1, 2, \ldots, m$$

Convex sets

- S is called a convex set if: $\forall\ u_1, u_2, \ldots, u_k$ in S, $\forall$ non-negative numbers $\lambda_1, \lambda_2, \ldots, \lambda_k$ such that $\sum_{i=1}^{k} \lambda_i = 1$, then $\sum_{i=1}^{k} u_i \lambda_i$ is in S

Convex set

Non Convex set

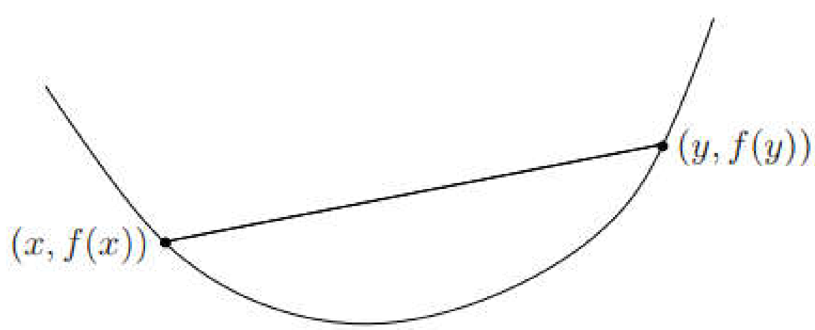
Convex functions

- Linear function: $f(x) = Ax$
- Affine function: $f(x) = Ax + b$
- Convex function
  - f is called convex if $\forall x_1, x_2$ and $\forall \lambda \in (0,1)$:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

  - f is called strictly convex if $\forall x_1 \neq x_2$ and $\forall \lambda \in (0,1)$:

$$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2)$$



Convex functions

- Example: $f(x) = 2x + 3$
  - $f(\lambda x_1 + (1- \lambda)x_2) = 2(\lambda x_1 + (1- \lambda)x_2) + 3 = (2\lambda x_1 + 3\lambda) + (2(1- \lambda)x_2) + (1- \lambda)3) = \lambda f(x_1) + (1- \lambda)f(x_2)$

Convex functions

- Examples
  - $f(x) = x^2$
  - $f(x) = e^{ax}$, a is a constant
  - $f(x) = x\ln x$

Basis

- $f(x_1, x_2, \ldots, x_n)$ is a multivariable function

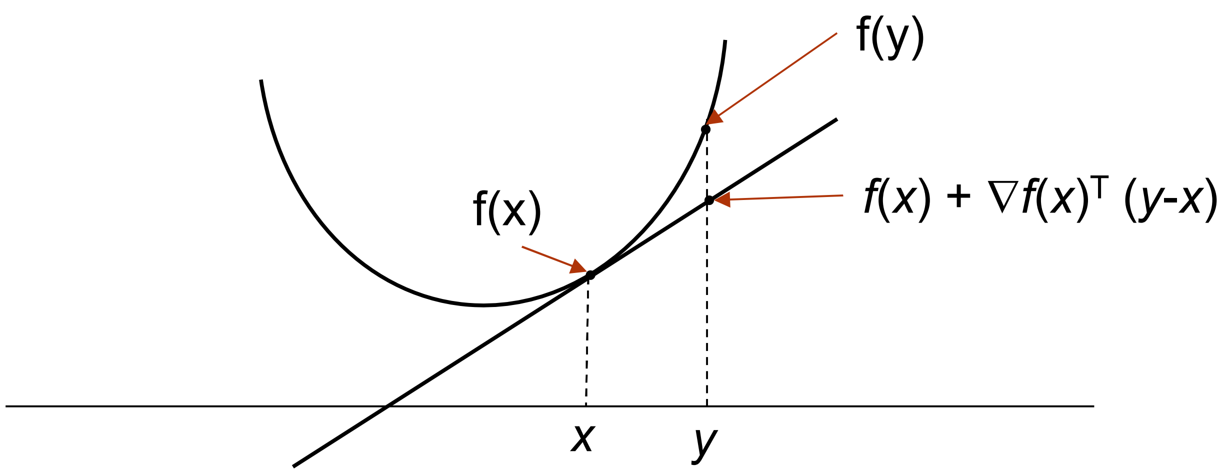
$$\nabla f(x) \quad (\text{or } f'(x)) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \ldots \\ \frac{\partial f}{\partial xn}(x) \end{pmatrix}$$

$$\nabla^2 f(x) \quad (\text{or } f''(x)) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right)$$ called Hessian matrix

Basis

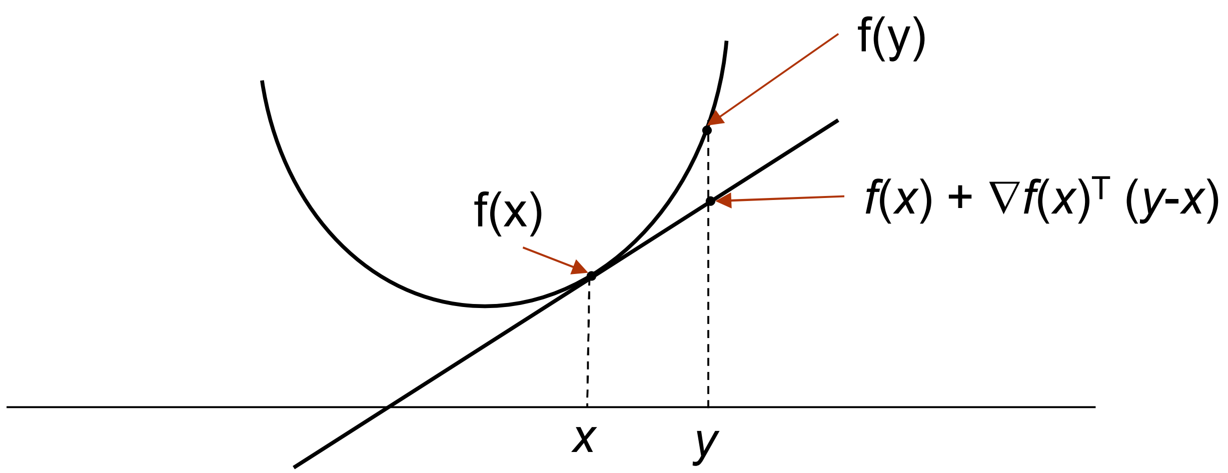
- If f is convex and differentiable, then

$$f(x) + \nabla f(x)^T (y-x) \le f(y),\ \forall x,y \in \textbf{dom}\ f$$



Convex functions

- First-order condition
  - Suppose f is differentiable (i.e., its gradient exists at all points in **dom** f, which is open). f is convex if and only if **dom** f is convex and $f(x) + \nabla f(x)^T (y-x) \le f(y)$, $\forall x, y \in$ **dom** f



Convex functions

- Proof (First-order condition, case n = 1)
  - Suppose *f* is differentiable and is convex
    - $\forall\ 0 < t \le 1$, $x + t(y-x) \in$ **dom** *f*
    - By convexity of *f*, $\forall x,y \in$ **dom** *f*, we have

$$f(x + t(y-x)) \le (1-t)\, f(x) + t\, f(y)$$

$\rightarrow f(y) \ge f(x) + \frac{f(x+t(y-x))-f(x)}{t}$. By taking the limit as t $\rightarrow$ 0 yields $f(x) + \nabla f(x)^T (y-x) \le f(y)$,

  - Suppose $f(x) + \nabla f(x)^T (y-x) \le f(y)$, $\forall x,y \in$ **dom** *f*
    - Let $z = \lambda x + (1-\lambda)y$
    - $f(x) \ge f(z) + f'(z)(x-z)$, $f(y) \ge f(z) + f'(z)(y-z) \rightarrow \lambda f(x) + (1-\lambda) f(y) \ge f(z) = f(\lambda x + (1-\lambda)y) \rightarrow$ *f* is convex

Convex functions

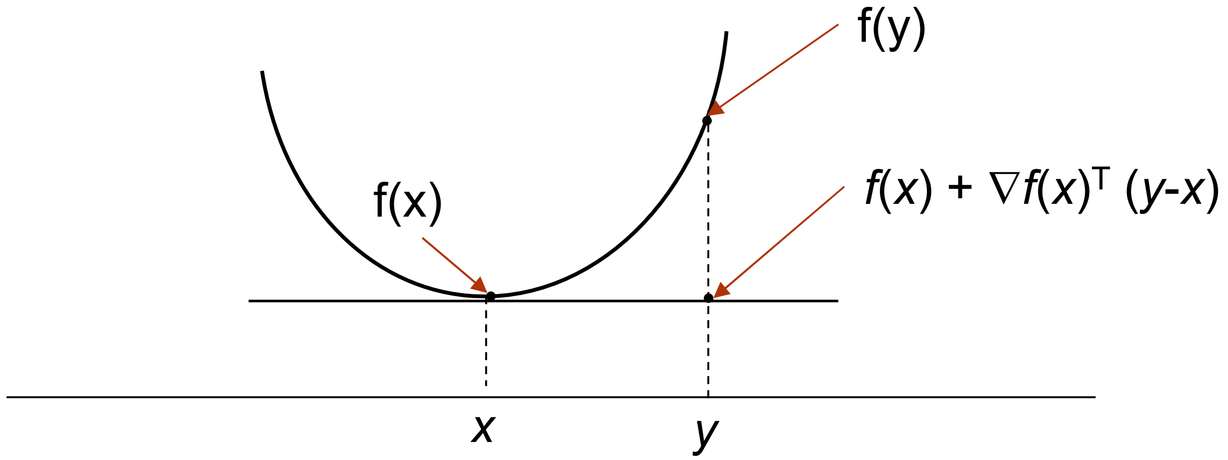
- Proof (First-order condition, general case)
  - Given $x,y \in R^n$, denote $g(t) = f(ty + (1-t)x)$, $g'(t) = \nabla f(ty + (1-t)x)(y-x)$
  - (1) Assume f is convex → g is convex. Hence $g(1) \geq g(0) + g'(0)$ (see case n = 1 above). This implies $f(y) \geq f(x) + \nabla f(x)^T(y-x)$
  - (2) Assume $f(y) \geq f(x) + \nabla f(x)^T(y-x)$, $\forall x,y \in$ **dom** f
    - If $py + (1-p)x \in$ **dom** f and $qy + (1-q)x \in$ **dom** f, we have

$$f(py + (1-p)x) \geq f(qy + (1-q)x) + \nabla f(qy + (1-q)x)^T(y-x)(p-q)$$

    i.e., $g(p) \geq g(q) + g'(q)(p-q)$ → g is convex (this implies f is convex)

Basis

- If $\nabla f(x) = 0$, then $f(y) \geq f(x)$, $\forall y \in$ **dom** f → x is global minimizer of the function f



Norms

- Norm: A real-valued function $f(x)$ on R^n is called a norm, if
  - $f(x) \geq 0$
  - $\lambda f(x) = f(\lambda x)$
  - $f(x + y) \leq f(x) + f(y)$ (triangle inequality)
- Examples
  - $||x||_p = (|x_1|^p + |x_2|^p + . . . + |x_n|^p)^{1/p}$
  - $||x||_1 = (|x_1| + |x_2| + . . . + |x_n|)$
  - $||x||_\infty = \max\{|x_1|, |x_2|, . . . , |x_n|\}$

Taylor approximation

- Single variable Taylor series

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots$$

- First-order Taylor approximation

$$f(x + h) \approx f(x) + h^{\mathrm{T}} \nabla(x)$$

- Second-order Taylor approximation

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