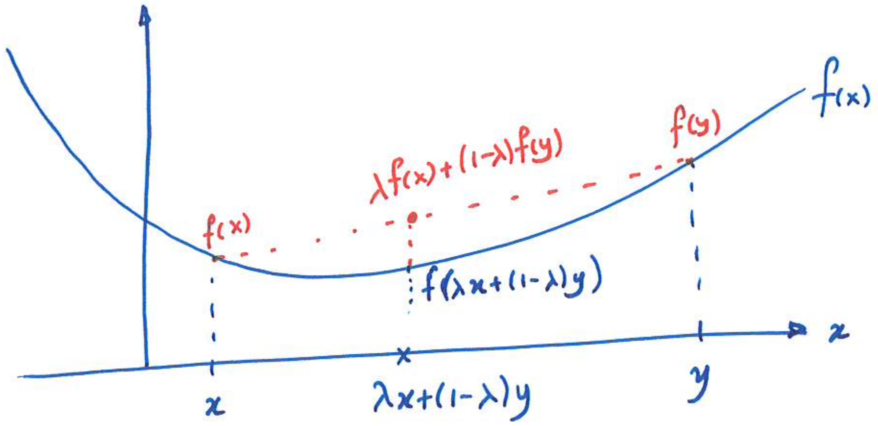
Convex Functions

Defn: A function $f: C \to \mathbb{R}$ defined on a convex set $C \subseteq \mathbb{R}^n$ is called convex if $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$ for any $x, y \in C$, $\lambda \in [0,1]$.



f is called strictly convex if $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$ for all $x, y \in C$ with $x \neq y$ and $\lambda \in (0,1)$.

e.g. Any norm is convex. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$; $x, y \in \mathbb{R}^n$, $\lambda \in [0,1]$.

$$\|\lambda x + (1-\lambda) y\| \le \|\lambda x\| + \|(1-\lambda)y\| \quad \text{(triangle ineq.)}$$
$$= \lambda \|x\| + (1-\lambda)\|y\| \quad \text{since } \lambda \geq 0,\ (1-\lambda) \geq 0.$$

Jensen's inequality: Let $f: C \to \mathbb{R}$ be a convex function over a convex set $C \subseteq \mathbb{R}^n$, $x_1, \ldots, x_k \in C$, $\lambda_1, \ldots, \lambda_k \geq 0$ with $\sum_{i=1}^{k} \lambda_i = 1$. Then,

$$f\left(\sum_{i=1}^{k} \lambda_i x_i\right) \le \sum_{i=1}^{k} \lambda_i f(x_i)$$

This can be proven by induction!

Extended Real Valued Functions

$\tilde{f}: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

The (effective) domain of $\tilde{f}$ is $\operatorname{dom} \tilde{f} = \{x \in \mathbb{R}^n \mid f(x) \in \mathbb{R}\}$

Note that if $f: \mathbb{R}^n \to \mathbb{R}$ is a function which is not well-defined on the whole space, then it's possible to assign value "$\pm\infty$" for those values. When we talk about convex functions we only consider $+\infty$.

The extended value extension of f is:

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in \operatorname{dom} f \\ +\infty & \text{otherwise.} \end{cases}$$

Recall that for convexity of f, we need to check two properties:

1.) $\operatorname{dom} f \subseteq \mathbb{R}^n$ is convex.

2.) $\forall x, y \in \mathbb{R}^n$, $\forall \theta \in [0,1]$: $f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$.

When we consider the extended value extension, 2.) implies 1.).

Why? If $x, y \in \operatorname{dom} f \Rightarrow f(x), f(y) \in \mathbb{R}$ and $\theta f(x) + (1-\theta) f(y) \in \mathbb{R}$.

2 implies that $f(\theta x + (1-\theta) y) \in \mathbb{R}$ as well. Then, $\theta x + (1-\theta) y \in \operatorname{dom} f$.

This proves that the domain is a convex set.

- An important example of an extended real valued function is the "indicator function" of a set $C \subseteq \mathbb{R}^n$: (<u>Notation</u>: $I_C(x)$ or $\delta_C(x)$)

$$I_C(x) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C. \end{cases}$$

Note that if the set C is convex, then I_C is a convex function. Clearly $\text{dom}\, I_C = C$ is convex. Moreover the function is constant on its domain.

A use of indicator function:

$$\left(\underset{x \in C}{\text{minimize}}\ f(x)\right) \text{ is equivalently } \left(\underset{x \in \mathbb{R}^n}{\text{minimize}}\ \big(f(x) + I_C(x)\big)\right).$$

- Another example is the <u>support function</u> of S for $S \subseteq \mathbb{R}^n$.

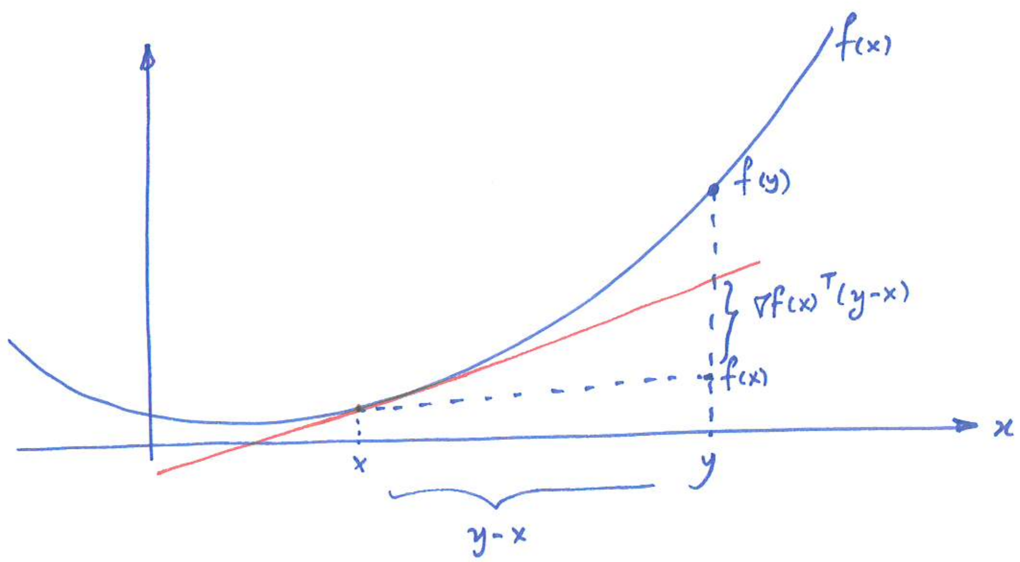
$$\sigma_S(x) = \sup_{y \in S} x^T y \quad \left(= \max_{y \in S} x^T y\right)$$

We will see later that this is a convex extended real valued function.

First order condition for convexity

<u>Theorem:</u> Let $f : C \to \mathbb{R}$ be a continuously differentiable function defined on a convex set $C \subseteq \mathbb{R}^n$. f is convex if and only if

$$\forall\, x, y \in C : \quad f(y) \geq f(x) + \nabla f(x)^T (y - x). \qquad (\star)$$



Proof: $\Rightarrow$ Assume f is convex. Let's prove (*). Let $x, y \in C$ be arbitrary.

If $x = y$, then (*) holds. Assume $x \neq y$. Let $\lambda \in (0,1]$. Then,

$$f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x) \quad \text{holds.}$$

$$\Rightarrow \quad f(x + \lambda(y-x)) - f(x) \leq \lambda f(y) - \lambda f(x)$$

$$\Rightarrow \quad \frac{f(x+\lambda(y-x))}{\lambda} \leq f(y) - f(x) \qquad (\text{as } \lambda > 0)$$

Taking the limit as $\lambda \downarrow 0$, we obtain:

$$\underbrace{f'(x; y-x)}_{\nabla f(x)^T (y-x)} \leq f(y) - f(x). \implies (*) \text{ holds.}$$

as f is conts. diff.

$\Leftarrow$ Assume (*) holds. Let $x, y \in C$, $\lambda \in (0,1)$. We want to show

that $f(\underbrace{\lambda x + (1-\lambda) y}_{u}) \leq \lambda f(x) + (1-\lambda) f(y)$

Note that as C is convex, $u \in C$, where $u := \lambda x + (1-\lambda) y$.

Let's apply (*) to x & u and y & u separately.

(1) $f(x) \geq f(u) + \nabla f(u)^T \underbrace{(x-u)}_{\substack{= x - \lambda x - (1-\lambda) y \\ = (1-\lambda)(x-y)}}$ (2) $f(y) \geq f(u) + \nabla f(u)^T \underbrace{(y-u)}_{\substack{= y - \lambda x - (1-\lambda) y \\ = \lambda (y-x)}}$

So, we have

$$① \quad f(x) \geq f(u) + (1-\lambda) \nabla f(u)^T (x-y) \qquad / \text{Multiply with } \frac{\lambda}{1-\lambda}$$

$$② \quad f(y) \geq f(u) + (-\lambda) \nabla f(u)^T (x-y) \qquad / \text{add}$$

$$\Rightarrow \quad \frac{\lambda}{1-\lambda} f(x) + f(y) \geq \frac{\lambda}{1-\lambda} f(u) + f(u) + \lambda \nabla f(u)^T (x-y) - \lambda \nabla f(u)^T (x-y)$$

$$\Rightarrow \quad \frac{\lambda f(x) + (1-\lambda) f(y)}{1-\lambda} \geq \frac{\lambda f(u) + (1-\lambda) f(u)}{1-\lambda}$$

$$\Rightarrow \quad \lambda f(x) + (1-\lambda) f(y) \geq \lambda f(u) + (1-\lambda) f(u) = f(u) = f(\lambda x + (1-\lambda) y).$$

$\blacksquare$

Second Order Conditions

Theorem: Let $f: C \to \mathbb{R}$ be twice continuously differentiable over C, $C \subseteq \mathbb{R}^n$ is an open convex set. f is convex if and only if $\nabla^2 f(x) \succcurlyeq 0$ for any $x \in C$.

Proof: $\Leftarrow$: Assume that $\nabla^2 f(x) \succcurlyeq 0$ for any $x \in C$.

Let $x, y \in C$ be arbitrary. By the linear approximation theorem, $\exists z \in [x,y]$ s.t.

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2} (y-x)^T \nabla^2 f(z) (y-x).$$

Note that $z \in C$, hence $\nabla^2 f(z) \succcurlyeq 0$ holds, which implies

$\frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x) \geq 0$ for all $x, y \in C$. Then,

$f(y) \geq f(x) + \nabla f(x)^T (y-x)$ holds for all $x, y \in C$.

This is the first order condition for convexity. Hence, f is convex.

$\Longrightarrow$: Assume that f is convex over C. Let $x \in C$, be arbitrary, $y \in \mathbb{R}^n$.

Since C is an open set, $\exists\, \varepsilon > 0$ s.t. $x + \lambda y \in C$ for all $\lambda \leq \varepsilon$.

By the first order condition, we have

$$f(x+\lambda y) \geq f(x) + \nabla f(x)^T [x + \lambda y - x]$$

$$\Rightarrow \quad f(x+\lambda y) \geq f(x) + \lambda \nabla f(x)^T y. \quad (*)$$

By quadratic approximation theorem we have

$$f(x+\lambda y) = f(x) + \nabla f(x)^T (x+\lambda y - x) + \frac{1}{2}(x+\lambda y - x)^T \nabla^2 f(x)(x+\lambda y - x) + o(\|x+\lambda y - x\|^2)$$

$$\Rightarrow f(x+\lambda y) = f(x) + \lambda \nabla f(x)^T y + \frac{\lambda^2}{2} y^T \nabla^2 f(x) y + \underbrace{o(\lambda^2 \|y\|^2)}_{\text{A function } h \text{ satisfying } \lim_{\lambda \downarrow 0} \frac{h(\lambda)}{\lambda^2} = 0.}$$

$$\text{by } (*) \quad \geq f(x) + \lambda \nabla f(x)^T y.$$

Hence, $\frac{\lambda^2}{2} y^T \nabla^2 f(x) y + o(\lambda^2 \|y\|^2) \geq 0$ holds for any $0 < \lambda \leq \varepsilon$.

Let's divide everything by $\lambda^2 > 0$: $\quad \frac{1}{2} y^T \nabla^2 f(x) y + \frac{o(\lambda^2 \|y\|^2)}{\lambda^2} \geq 0.$

Taking the limit as $\lambda \downarrow 0$ yields that

$$\frac{1}{2} y^T \nabla^2 f(x) y \geq 0.$$

Recall that $x \in C, y \in \mathbb{R}^n$ are arbitrary. Hence, the inequality holds for all such x, y. This implies that $\nabla^2 f(x) \succeq 0$ for all $x \in C$.

$\blacksquare$

Theorem: let $f : C \to \mathbb{R}^n$ be twice diff. continuous over $C \subseteq \mathbb{R}$. C: open, convex. If $\nabla^2 f(x) \succ 0$ for any $x \in C$, then f is strictly convex over C.

- Note that the reverse of this statement is not necessarily correct.

Consider $f(x) = x^4$ over $\mathbb{R}$. f is strictly convex over $\mathbb{R}$.

$f'(x) = 4x^3$, $f''(x) = 12x^2$.

Note that $f''(0) = 0 \not\succ 0$, even though f is strictly convex.

Example: let $f : \mathbb{R} \times \mathbb{R}_{++} \to \mathbb{R}$, $\mathbb{R} \times \mathbb{R}_{++} = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 > 0\}$

$f(x) = \dfrac{x_1^2}{x_2}$. let's show that f is convex by checking its hessian.

$$\nabla f(x) = \begin{bmatrix} \dfrac{2x_1}{x_2} \\ -\dfrac{x_1^2}{x_2^2} \end{bmatrix}, \quad \nabla^2 f(x) = \begin{bmatrix} \dfrac{2}{x_2} & \dfrac{-2x_1}{x_2^2} \\ \dfrac{-2x_1}{x_2^2} & \dfrac{2x_1^2}{x_2^3} \end{bmatrix} = 2 \begin{bmatrix} \dfrac{1}{x_2} & \dfrac{-x_1}{x_2^2} \\ \dfrac{-x_1}{x_2^2} & \dfrac{x_1^2}{x_2^3} \end{bmatrix}$$

We want to check if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbb{R} \times \mathbb{R}_{++}$.