Optimization in Continuous Time:

Gradient Flows

Let a C^1 function $f: \mathbb{R}^n \to \mathbb{R}$ be given.
A point $x^* \in \mathbb{R}^n$ is a global minimum of f if:

$$f(x^*) \leq f(x) \qquad \forall x \in \mathbb{R}^n$$

A necessary condition: $\nabla f(x^*) = 0$

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right)^T$$

Steepest descent (gradient descent):

$$\dot{x} = -\nabla f(x) \qquad\qquad x(0) = x_0$$

[cont. time limit of $x_{k+1} = x_k - h \nabla f(x_k)$, $k = 0, 1, \dots$ as $h \downarrow 0$]

Candidate LF: $V(x) = f(x)$ [assume w.l.o.g. $f(\cdot) \geq 0$]

$$\dot{V}(x) = -\nabla f(x)^T \dot{x} = -|\nabla f(x)|^2$$

If $x(t)$ remains bdd along $\dot{x} = -\nabla f(x)$, then

$$|\nabla f(x(t))| \to 0 \quad \text{as } t \to \infty.$$

[Use weak Lyapunov criterion with $W(x) = |\nabla f(x)|^2$]

How can we ensure $x(t)$ remains bdd?

E.g. f is radially unbounded:

$$f(x) \to \infty \quad \text{as} \quad |x| \to \infty$$

$\Rightarrow$ f has compact level sets

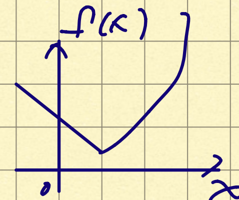
$$S_f(c) := \{x \in \mathbb{R}^n : f(x) \le c\}$$

$$\dot{V}(x(t)) \le 0 \quad \Rightarrow \quad f(x(t)) \le f(x(0)) \quad \forall t \ge 0$$

$$\Rightarrow \quad x(t) \in S_f(f(x(0)))$$

since $S_f(c)$ is compact for all $c \ge 0$, $|x(t)| \le R < \infty$ for some $R > 0$ determined by $x(0)$.

Convex functions $\quad f: \mathbb{R}^n \to \mathbb{R}$

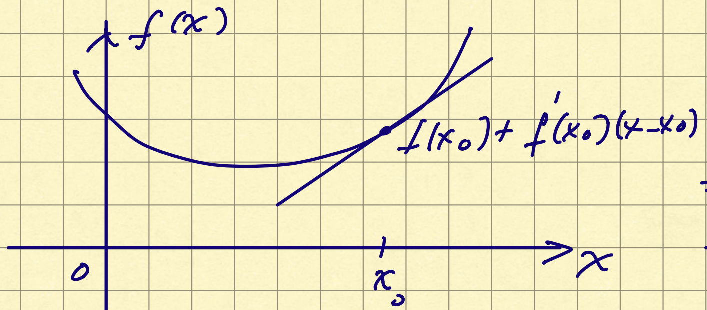


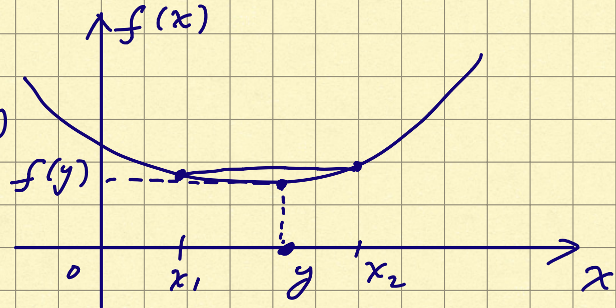
convex, not diff.

$n=1$: $f: \mathbb{R} \to \mathbb{R}$ is convex if

$$f(\lambda x + (1-\lambda)x') \le \lambda f(x) + (1-\lambda) f(x')$$

for all $x, x' \in \mathbb{R}$ and all $\lambda \in [0,1]$.





$$y = \lambda x_1 + (1-\lambda)x_2$$
$$0 \le \lambda \le 1$$

$n>1$: $f: \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f(\lambda x + (1-\lambda)x') \le \lambda f(x) + (1-\lambda) f(x')$$

for all $x, x' \in \mathbb{R}^n$ and all $\lambda \in [0,1]$.

If f is C^1, equivalent condition is: for all x, x',

$$f(x') \ge f(x) + \nabla f(x)^T (x' - x)$$

Claim: this is equivalent to the original def'n.

proof:

1) Assume f is convex

$\forall\, x, x' \in \mathbb{R}^n, \quad \forall\, \lambda \in (0,1]$

$$\lambda f(x) + (1-\lambda) f(x') \geq f(\lambda x + (1-\lambda) x')$$

$$f(x) \geq \frac{f(\lambda x + (1-\lambda)x') - f(x')}{\lambda} + f(x')$$

take $\lambda \downarrow 0$: $\quad f(x) \geq f(x') + \nabla f(x')^T (x - x')$

[Note: $\lambda \mapsto f(\lambda x + (1-\lambda) x')$ is C^1].

2) Assume $f(x) \geq \underbrace{f(x') + \nabla f(x')^T (x - x')}_{\text{1st-order apx to } f \text{ around } x'}$

Take $x, x' \in \mathbb{R}^n$, $\lambda \in [0,1]$

$$\lambda f(x) + (1-\lambda) f(x') \geq f(\lambda x + (1-\lambda) x') \quad \leftarrow \text{goal}$$

$$f(x) \geq f(\overbrace{\lambda x + (1-\lambda) x'}^{:= y}) + \nabla f(y)^T (x - y)$$

$$f(x') \geq f(y) + \nabla f(y)^T (x' - y)$$

$\Rightarrow \lambda f(x) + (1-\lambda) f'(x)$

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

$$+ (1-\lambda) f(y) + (1-\lambda) \nabla f(y)^T (x' - y)$$

$$= f(y) + \nabla f(y)^T [\underbrace{\lambda x + (1-\lambda) x'}_{= y} - y]$$

$\Rightarrow \lambda f(x) + (1-\lambda) f(x') \geq f(\lambda x + (1-\lambda) x')$. $\blacksquare$

If x^* is a global min. of f and f is C^1 and convex, then

$$f(x) \geq f(x^*) + \nabla f(x^*)^T (x - x^*)$$

$$f(x) - f(x^*) \geq \nabla f(x^*)^T (x - x^*)$$

$\Rightarrow \quad \nabla f(x^*)^T (x - x^*) \leq 0 \qquad \forall x \in \mathbb{R}^n$

$\Leftrightarrow \quad \nabla f(x^*)^T v \leq 0 \qquad \forall v \in \mathbb{R}^n$

$\Leftrightarrow \quad \nabla f(x^*) = 0$

If x^* is a global min, then $\nabla f(x^*) = 0$.

Conversely, suppose $\nabla f(x^*) = 0$. Then

$$f(x) \geq f(x^*) + \underbrace{\nabla f(x^*)^T}_{=0}(x - x^*), \quad \forall x \in \mathbb{R}^n$$

$\Rightarrow f(x^*) \leq f(x), \ \forall x \in \mathbb{R}^n$ (x^* is a global min.)

A point $x^* \in \mathbb{R}^n$ is a global min of a convex C^1 fctn $f : \mathbb{R}^n \to \mathbb{R}$ iff $\nabla f(x^*) = 0$.

Let's get back to $\dot{x} = -\nabla f(x)$ [f convex]

Candidate LF: $V(x) = \frac{1}{2}|x - x^*|^2$ $\quad x^*$ – global min.

$\nabla V(x) = x - x^*$

$\dot{V}(x) = -\nabla V(x)^T \nabla f(x)$

$\quad = \nabla f(x)^T (x^* - x)$

$$f(x^*) \geq f(x) + \underbrace{\nabla f(x)^T (x^* - x)}_{\dot{V}(x)}$$ [by 1st-order cond.]

$\dot{V}(x) \leq f(x^*) - f(x) \leq 0$

Let's fix some $T > 0$

$$V(x(T)) - V(x(0)) = \int_0^T \dot{V}(x(t))\, dt$$

$$= \int_0^T \nabla f(x(t))^T (x^* - x(t))\, dt$$

$$f(x^*) \geq f(x(t)) + \underbrace{\nabla f(x(t))^T (x^* - x(t))}_{\dot{V}(x(t))}$$

$$\leq \int_0^T [f(x^*) - f(x(t))]\, dt$$

$$\int_0^T f(x(t))\, dt \leq T f(x^*) + V(x(0)) - V(x(T))$$
$$\leq T f(x^*) + \frac{1}{2} |x(0) - x^*|^2$$

$$\Rightarrow \quad \boxed{\frac{1}{T} \int_0^T f(x(t))\, dt \leq f(x^*) + \frac{|x(0) - x^*|^2}{2T}}$$

Consequences:

1) $$\min_{0 \leq t \leq T} f(x(t)) - \min_{x \in \mathbb{R}^n} f(x) \leq \frac{|x(0) - x^*|^2}{2T}$$

$$\left[\int_0^T f(x(t))\, dt \geq T \min_{0 \leq t \leq T} f(x(t)) \right]$$

2) Let $\bar{x}(T) := \frac{1}{T} \int_0^T x(t)\, dt$

$$f(\bar{x}(T)) - \min_{x \in \mathbb{R}^n} f(x) \leq \frac{|x(0) - x^*|^2}{2T}$$

Follows from Jensen's inequality:

$$f\left(\frac{1}{T} \int_0^T x(t)\, dt \right) \leq \frac{1}{T} \int_0^T f(x(t))\, dt \quad .$$

Strong convexity

$f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex $(\mu \geq 0)$ if

$f_\mu(x) := f(x) - \frac{\mu}{2} |x|^2$ is convex.

[$\mu = 0$: just convexity]

$$f(x) = \frac{\mu}{2}|x|^2 \qquad \nabla f(x) = \mu x$$

$$\nabla^2 f(x) = \mu I_n \qquad (\mu > 0)$$

$$= \begin{pmatrix} \mu & & & \\ & \mu & & \\ & & \ddots & \\ & & & \mu \end{pmatrix}$$

Claim: Let a C^1 $f: \mathbb{R}^n \to \mathbb{R}$ be given. Then f is μ-strongly convex iff

$$f(x') \geq f(x) + \nabla f(x)^T (x' - x) + \frac{\mu}{2}|x' - x|^2$$

for all $x, x' \in \mathbb{R}^n$.

Proof (sketch)

Apply the first-order condition to $f_\mu(x) = f(x) - \frac{\mu}{2}|x|^2$.

Back to gradient flow:

$$\dot{x} = -\nabla f(x)$$

Assume f is C^1 and μ-SC.

$$V(x) = \frac{1}{2}|x - x^*|^2$$

$$\dot{V}(x) = \nabla f(x)^T (x^* - x)$$

By μ-SC,

$$f(x^*) \geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2}|x - x^*|^2$$

$$= f(x) + \dot{V}(x) + \mu V(x)$$

$$\Rightarrow \dot{V}(x) \leq f(x^*) - f(x) - \mu V(x)$$

$$\leq -\mu V(x)$$

$$\frac{d}{dt} V(x(t)) \leq -\mu V(x(t)) \qquad [\mu > 0]$$

Recall: diff. inequality $\frac{d}{dt} g(t) \leq C g(t)$

$$\frac{\frac{d}{dt} g(t)}{g(t)} \leq C$$

$$g(t) \leq g(0) e^{Ct}$$

$\Rightarrow \; V(x(t)) \leq V(x(0)) e^{-\mu t} \qquad (t \geq 0)$

$$\boxed{|x(t) - x^*|^2 \leq |x(0) - x^*|^2 e^{-\mu t}}$$

Ex if f is C^2 and μ-SC, then $\nabla^2 f(x) \succeq \mu I_n$.

$v^T \nabla^2 f(x) v \geq \mu |v|^2 \qquad \forall v, x$

$f(x) = x^T P x \qquad P \succ 0$

$\dot{x} = -2Px$