DSA3102: Lecture 13. Problems 1, 2, 3, 4(a)

function need not be convex

Recap: Gradient descent. $f: \mathbb{R}^n \to \mathbb{R}$

$\Delta x = -\nabla f(x)$

$x^{(k+1)} = x^{(k)} - t_k \nabla f(x^{(k)}) = x^{(k)} + t_k (\Delta x^{(k)})$. update estimate of optimal solution

Selection of step size

Exact line search: $t^{(k)} = \underset{s \geq 0}{\operatorname{argmin}} f(x^{(k)} - s \nabla f(x^{(k)}))$

minimisation problem

f: convex $\qquad S = \{x \in \mathbb{R}^n : f(x) \leq f(x^{(0)})\}$ function value less than equal to starting value

Def: m-strongly convex $\qquad \nabla^2 f(x) \succeq mI$ for all $x \in S$.

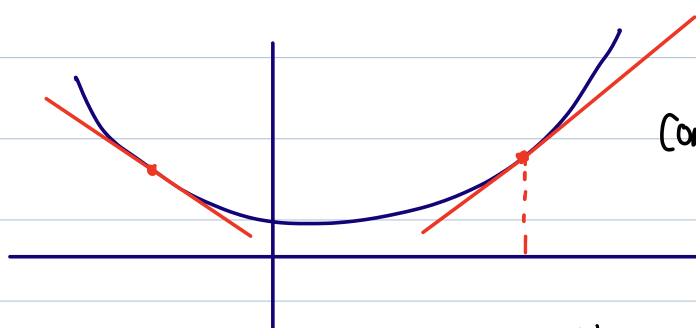
stronger than convex $\qquad$ hessian

Eg: $f(x) = ax+b \qquad f: \mathbb{R} \to \mathbb{R}$ $\qquad$ cannot find $M>0$ s.t $f''(x)$ is +ve

$f''(x) = 0$ not m-s.c. for any $m > 0$

$f(x) = ax^2 + bx + c \quad (a>0)$

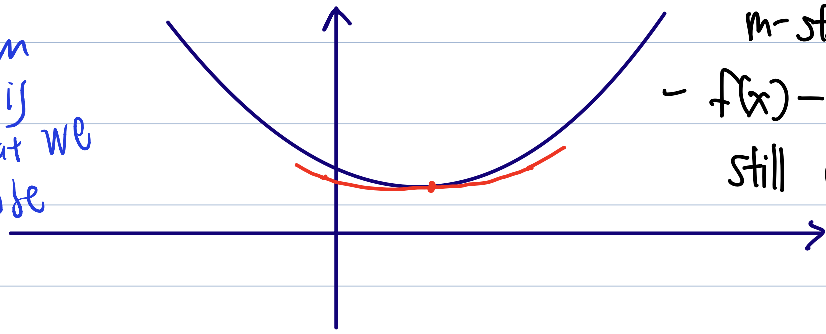
$f''(x) = 2a \qquad$ 2a-strongly convex



Convex function.

the tangent lies below curve for every point can put quadratic under the curve

maximum curvature is the m that we can choose



m-strongly convex

- $f(x) - \frac{m}{2}\|x\|^2$ is still convex

Def: M-smoothness $\quad \forall x \in S \quad \nabla^2 f(x) \leq MI.$

Thm: If f is m-s.c. & M-smooth,

$$f(x^{(k+1)}) - p^* \leq c\,(f(x^{(k)}) - p^*)$$

$p^* = \min_x f(x)$ $\quad$ go to c^k

$$c = 1 - \frac{m}{M}.$$

Problem 1: $\quad f(x) = \frac{1}{2}x^TQx - b^Tx, \quad Q \in S^n_{++}.$

$x \in \mathbb{R}^n.$

a) $\lambda_m(Q)$: m^{th} largest e-value of Q.

$$\lambda_1(Q) \geq \lambda_2(Q) \geq \cdots \geq \lambda_n(Q) > 0.$$

all eigenvalues +ve since matrix is PD.

choose L to be one of the numbers above, find which one?

Lipschitz: $\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\| \quad \forall x,y \in \mathbb{R}^n.$

$\|v\|^2 = v^Tv$

Sol: $\nabla f(x) = Qx - b$

$$LHS^2 = \|\nabla f(x) - \nabla f(y)\|^2 = \|\underbrace{Q(x-y)}_{v}\|^2$$

$$= (Q(x-y))^T Q(x-y)$$

$$= (x-y)^T Q^T Q (x-y).$$

$M \leq \lambda_{max}(M) I.$

upper bound M in terms of its maximum eigenvalue

$$\leq (x-y)^T \lambda_{max}(Q^T Q) I (x-y).$$

$$= \lambda_{max}(Q^T Q) \|x-y\|^2$$

$$\lambda_{max}(Q^T Q) = \lambda_{max}(Q)^2 = \lambda_1(Q)^2$$

Take root on both sides

$$LHS \leq \underbrace{\lambda_1(Q)}_{L} \|x-y\|$$

$M = 1.$

(b) $\min_x f(x)$ $\quad f(x) = \frac{1}{2} x^T Q x - b^T x$

Differentiate and set to 0 $\quad \nabla f(x) = Qx - b = 0, \quad x^* = Q^{-1} b.$

(c) modified gradient method $x^{(k+1)} = x^{(k)} - s D \nabla f(x^{(k)})$, $D \in S^n_{++}$. positive definite

matrix

Find $0 < s < \frac{2}{M}$ s.t. the above alg. converges to x^*.

Sol: $\quad x^{(k+1)} = x^{(k)} - sD(Qx^{(k)} - b).$ gradient from prev part

$$x^{(k+1)} - x^* = x^{(k)} - sD\boxed{(Qx^{(k)} - b)} - x^*$$

$$= x^{(k)} - sD\boxed{Q(x^{(k)} - \underbrace{Q^{-1}b}_{x^*})} - x^*$$

$$= (x^{(k)} - x^*) - sDQ(x^{(k)} - x^*)$$

e.g. $a_{k+1} = 0.8 a_k$

does not matter what a_0 is, will always converge to 0

analyse this

$$= \underbrace{(I - sDQ)}_{} (x^{(k)} - x^*)$$

$D = D^{1/2} D^{1/2}$

$y^{(k)} = D^{-1/2} x^{(k)}$, $y^* = D^{-1/2} x^*$

$$\rightarrow \quad D^{1/2}(y^{(k+1)} - y^*) = (I - sDQ) D^{1/2} (y^{(k)} - y^*)$$

$$y^{(k+1)} - y^* = (I - sD^{1/2} Q D^{1/2})(y^{(k)} - y^*)$$

$$y^{(k+1)} - y^* = D^{-1/2}(I - sDQ) D^{1/2} (y^{(k)} - y^*)$$

$$= (I - sD^{1/2} Q D^{1/2})(y^{(k)} - y^*).$$

$\rightarrow$ symmetric

The system converges $y^{(k)} \to y^*$ (equiv to $x^{(k)} \to x^*$)

iff $|\lambda_{max}(A)| < 1$

largest eigenvalue of $I - B$

$$\lambda_{max}(I - B) = 1 - \lambda_{max}(B).$$

$$|\lambda_{max}(I - \underbrace{s D^{1/2} Q D^{1/2}}) | < 1$$

$$0 < \lambda_{max}(s D^{1/2} Q D^{1/2}) < 2$$

$$0 < s \, \lambda_{max}(D^{1/2} Q D^{1/2}) < 2$$

$$0 < s < 2/M \quad , \quad M = \lambda_{max}(D^{1/2} Q D^{1/2})$$

Recap: $\min_{x \in \mathbb{R}^n} f_0(x)$ s.t. $f_i(x) \le 0 \quad \forall i = 1, \dots, m$

$Ax = b.$

f_i : convex $\Rightarrow$ convex opt. problem.

When the opt. problem is convex, then the KKT conditions if can find this that satisfies KKT for convex opt problem, then they are primal dual optimum

are necessary & sufficient for determining that (x^*, λ^*, v^*) are primal-dual optimal.

KKT conditions: (x, λ, v) satisfy KKT if

1) Stationarity: $\nabla_x L(x, \lambda, v) = 0$ (n equations)

2) Primal feasibility: $f_i(x) \le 0 \quad \forall i = 1, \dots, m$ $\qquad Ax = b$

3) Dual feasibility: $\lambda \ge 0$. component wise

4) Complementary slackness: $\lambda_i f_i(x) = 0 \quad \forall i = 1, \dots, m$.

2a) $$\min_{(x_0, x_1, \dots, x_n) \in \mathbb{R}^{n+1}} \underbrace{-\log(1+x_0) - \sum_{i=1}^{n} \log x_i}_{f_0(x_0, \dots, x_n)}$$

$-\log$ is convex

all x_i is symmetric (x_1, x_2 is symmetric)

$x_0 + x_i - 1 \le 0 \Leftarrow$ $\quad \lambda_i \quad x_0 + x_i \le 1, \quad \forall i = 1, \dots, n$

domain is +ve values

$\lambda_0 \quad x_0, \underbrace{x_1, \dots, x_n}_{\text{inactive}} \ge 0. \Rightarrow x_0 \ge 0.$

$-x_0 \le 0$

$$\text{dom } f_0 = \{(x_0, x_1, \dots, x_n) \in \mathbb{R}^{n+1} : \begin{array}{l} x_0 > -1 \\ x_i > 0, \; i = 1, \dots, n \end{array}\}.$$

$x = (x_0, x_1, \dots, x_n)$

$$L(x, \lambda) = -\log(1+x_0) - \sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} \lambda_i (x_0 + x_i - 1)$$
$$+ \lambda_0 (-x_0)$$

Stationarity:

$\frac{\partial L}{\partial x_0} = 0 \Rightarrow -\frac{1}{1+x_0} + \sum_{i=1}^{n} \lambda_i - \lambda_0 = 0.$ —(S1)

$\frac{\partial L}{\partial x_i} = 0 \Rightarrow -\frac{1}{x_i} + \lambda_i = 0 \quad \forall\, i = 1, \dots, n.$ —(S2)

P.F: $x_0 + x_i \leq 1 \quad \forall\, i = 1, \dots, n$

$x_0 \geq 0$

D.F. $\lambda_0 \geq 0, \quad \lambda_i \geq 0, \forall\, i = 1, \dots, n$

C.S. $\lambda_i(x_0 + x_i - 1) = 0 \quad \forall\, i = 1, \dots, n$ — (CS1)

$\lambda_0 x_0 = 0$ — (CS2)

First, note from (S2) that $\lambda_i = \frac{1}{x_i}$ for all $i = 1, \dots, n$.

But $x_i > 0 \quad \forall\, i = 1, \dots, n \Rightarrow \lambda_i > 0 \quad \forall\, i = 1, \dots, n.$ strictly positive so $x_0 + x_i - 1 = 0$

By (CS1), $x_0 + x_i = 1 \quad \forall\, i = 1, \dots, n$

$x_i = 1 - x_0 \quad \forall\, i = 1, \dots, n.$

Case (A): $\underline{\lambda_0 = 0}$. KKT conditions simplify. $x_0 > 0$

$-\frac{1}{1+x_0} + \sum_{i=1}^{n} \lambda_i = 0$ (stationarity condition 1) stationarity condition 2

$$\frac{1}{x_i} = \lambda_i \quad \forall i = 1, \dots, n.$$

$$x_i = 1 - x_0. \quad \forall i = 1, \dots, n$$

These equations imply that

$$0 = -\frac{1}{1+x_0} + \sum_{i=1}^{n} \frac{1}{x_i} = -\frac{1}{1+x_0} + \sum_{i=1}^{n} \frac{1}{1-x_0}$$

$$0 = -\frac{1}{1+x_0} + \frac{n}{1-x_0} \Rightarrow \frac{n}{1-x_0} = \frac{1}{1+x_0}.$$

$x_0 \in (0, 1]$

However, $\frac{n}{1-x_0} \geq n \geq 1 > \frac{1}{1+x_0}$ $\quad \Rightarrow\Leftarrow$.

Case (B): $\lambda_0 > 0$

(S1) $\quad -\frac{1}{1+x_0} + \sum_{i=1}^{n} \lambda_i - \lambda_0 = 0.$

(S2) $\quad -\frac{1}{x_i} + \lambda_i = 0 \quad \forall i = 1, \dots, n.$

$$\left.\begin{array}{ll} x_0 + x_i = 1 & \forall i = 1, \dots, n. \\ x_0 = 0 & \text{from (CS)} \end{array}\right\} \quad x_i = 1 \quad \forall i = 1, \dots, n.$$

By subst. $x_i = 1$ into (S2), $\lambda_i = \underset{\geq 0}{1} \quad \forall i = 1, \dots, n$

From (S1): $\lambda_0 = -\frac{1}{1+x_0} + \sum_{i=1}^{n} \lambda_i = -1 + n = n - 1 \geq 0$

C.S: $\lambda_0^* x_0^* = 0$ $\quad (x_0^* + x_i^* - 1)\lambda_i^* = 0 \quad \forall i = 1, \ldots, n$

$$x^* = (0, 1, \ldots, 1), \quad \lambda^* = (n-1, 1, \ldots, 1).$$

2b) $\min_{x \in \mathbb{R}^n} c^T x \quad Ax = b, \ x \geq 0.$

feasible (these conditions hold)

Lagrangian of original LP

$$L(x, \lambda, \nu) = c^T x + \lambda^T(-x) + \nu^T(Ax - b).$$

Lagrange dual f^n $g(\lambda, \nu) = \inf_x L(x, \lambda, \nu).$

$$= \inf_x \left\{ (c - \lambda + A^T \nu)^T x - \nu^T b \right\}$$

linear: unbounded below

$\lambda = c + A^T \nu$

$$g(\lambda, \nu) = \begin{cases} -b^T \nu & c - \lambda + A^T \nu = 0 \\ -\infty & \text{o.w.} \end{cases}$$

Lagrange dual problem: $\max_{\nu, \lambda \geq 0} -b^T \nu$ s.t.

$$\max_{\nu} -b^T \nu \quad \text{s.t.} \quad A^T \nu + c \geq 0.$$

$$\min_{\nu} b^T \nu \quad \text{s.t.} \quad A^T \nu + c \geq 0.$$

the new primal

$$\tilde{L}(v, x) = b^T v + x^T(-c - A^T v)$$

Lagrange dual function:

$$\tilde{g}(x) = \inf_v \{ (b - Ax)^T v - c^T x \}$$

$$= \begin{cases} -c^T x & \text{s.t. } b = Ax \\ -\infty & \text{else.} \end{cases}$$

Lagrange dual problem:

$$\max_{x \geq 0} \; -c^T x \qquad b = Ax.$$

$$\min_x \; c^T x \quad , \quad b = Ax, \; x \geq 0. \quad \text{(original)}$$

4a) Recap: Newton's method.

$$x^{(k+1)} = x^{(k)} - t \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$$

$$x^{(k+1)} = x^{(k)} - t \frac{f'(x^{(k)})}{f''(x^{(k)})}$$

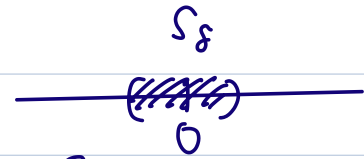
Univariate 1 dimensional case
where x is a scalar

Under suitable condition

quadratic convergence

$$\|x^{(k+1)} - x^*\| \leq \frac{LM}{2} \|x^{(k)} - x^*\|^2$$

$$\sup_{x \in S_\delta} \left\| \left(\nabla^2 f(x)\right)^{-1} \right\| \leq M.$$

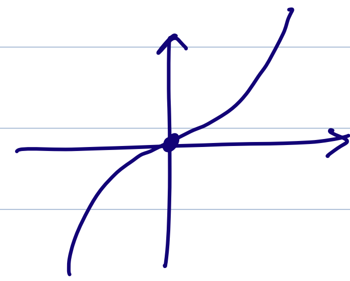


$S_\delta = \{x \in \mathbb{R} : |x - 0| < \delta\}$

$$\sup_{x \in S_\delta} \left| \left(f''(x)\right)^{-1} \right| \leq M$$

4a) $f(x) = x^3$

$f'(x) = 3x^2, \qquad f''(x) = 6x.$



step size is 1 so $t=1$

$$x^{(k+1)} = x^{(k)} - t \frac{f'(x^{(k)})}{f''(x^{(k)})}$$

$$x^{(k+1)} = x^{(k)} - \frac{3(x^{(k)})^2}{6x^{(k)}} = \frac{1}{2} x^{(k)}.$$

linear convergence

no quadratic convergence

$x^{(0)} = 8 \to x^{(1)} = 4, \quad x^{(2)} = 2, \ \ldots$

$$f''(x) = 6x. \Rightarrow \sup_{x \in S_\delta} \left\{ \frac{1}{f''(x)} = \frac{1}{6x} \right\} = \infty.$$

(cannot find M)

no bound on the inverse hessian

$x^* = 0, \qquad S_\delta = (-\delta, \delta)$

---

3. Subgradient of f at x is $\delta_x(f) \in \mathbb{R}^n$

convex but not necessarily differentiable function

$$f(x) + (y-x)^T \delta_x(f) \leq f(y) \quad \forall x, y \in \operatorname{dom} f.$$

f is differentiable, $\quad \delta_x(f) = \nabla f(x)$

f is diff^{ble} & convex

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

first order condition

a) $\partial f(x) = \{ \delta_x(f) : \delta_x(f) \text{ is a subgradient of } f \text{ at } x \}$.

Claim: $\partial f(x)$ is a convex set.

collect all the subgradient and form sub differential

every subgradient must satisfy this

$$\text{Pf: } \partial f(x) = \{ v \in \mathbb{R}^n : f(x) + v^T(y-x) \leq f(y) \quad \forall x, y \in \text{dom} f \}$$

$$= \bigcap_{x,y \in \text{dom} f} \underbrace{\{ v \in \mathbb{R}^n : f(x) + v^T(y-x) \leq f(y) \}}_{\text{II}}$$

$$= \bigcap_{x,y \in \text{dom} f} S_{x,y} \quad : \text{convex}$$

show that each one is convex then whole thing is convex

$$S_{x,y} := \{ v \in \mathbb{R}^n : \underbrace{f(x)}_{\text{fixed}} + v^T \underbrace{(y-x)}_{\text{fixed}} \leq \underbrace{f(y)}_{\text{fixed}} \}$$

$S_{x,y}$ is a halfspace, hence convex.

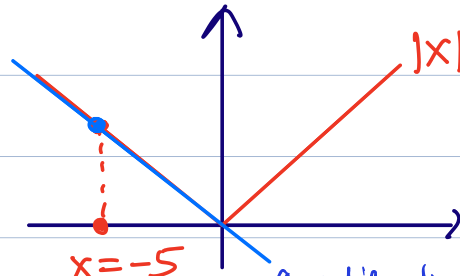
Intersection of an arbitrary family of convex sets is convex.

$\Rightarrow \partial f(x)$ is convex.

not differentiable at 0

b) $f: \mathbb{R} \to \mathbb{R}$ $\quad f(x) = |x|$

i) $\delta_x(f)$ when $x < 0$



$f(y) \geq f(x) + \delta_x(f)(y-x)$

$f(y) = 5 + \delta_x(f)(y+5)$ $\quad y = -6.$

$6 = 5 + \delta_x(f)(-6+5)$ $\quad y = +7$

$\delta_x(f) = -1$

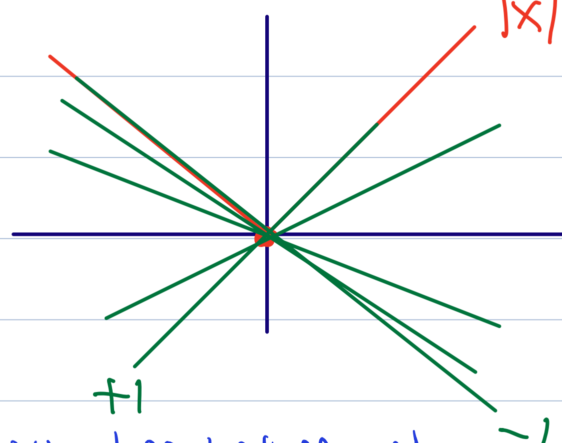
gradient here is exactly -1 (find supporting hyperplane)

$\partial f(x) = \{-1\}$

$x<0$ $\quad \delta_x(f) = -1$

ii) $\delta_x(f) = 1$ $\quad x > 0$

$\partial f(x) = \{1\}$



iii) $x = 0$

$\partial_x(f) = [-1, 1]$.

function is not differentiable here

(every line with slope between +1 and -1 supports the epigraph of $f(x)$)

---

c) $f_0, f_1 : \mathbb{R} \to \mathbb{R}$ convex

$$\min_{x \in \mathbb{R}} f_0(x), \quad f_1(x) \leq 0$$

$g(\lambda)$: Lagrange dual function.

this is concave

applied to concave functions

Show that a supergradient of $g(\lambda)$ at λ is given by $f_1(x_\lambda)$ where

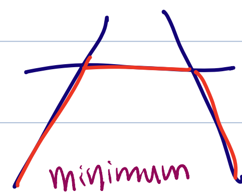
$$x_\lambda = \operatorname{argmin}\{f_0(x) + \lambda f_1(x)\}.$$

Sol: Lagrangian: $L(x,\lambda) = f_0(x) + \lambda f_1(x)$

Lagrange dual function:

$$g(\lambda) = \inf_x \{f_0(x) + \lambda f_1(x)\}$$

$$g(\lambda) = f_0(x_\lambda) + \lambda f_1(x_\lambda)$$



NTP: $$g(\lambda) + (\mu - \lambda) f_1(x_\lambda) \geq g(\mu)$$

By direct calculation: $\inf_x h(x) \leq h(x_{favorite})$

$$\begin{aligned} g(\mu) &= \inf\{f_0(x) + \mu f_1(x)\} \\ &\leq f_0(x_\lambda) + \mu f_1(x_\lambda) \\ &= f_0(x_\lambda) + (\mu - \lambda) f_1(x_\lambda) + \lambda f_1(x_\lambda) \\ &= g(\lambda) + (\mu - \lambda) f_1(x_\lambda). \end{aligned}$$

(qed).