Tasks

Task 1:

Answer: k = n

Task 2

Answer: dimension of M is (n, n)

Task 3

Derive second to fifth terms in Taylor series expansion for the natural logarithm with respect to point $a \geq 0$

Solution:

$$log(x) = log(a) + \frac{1}{a}(x-a) - \frac{1}{2a^2}(x-a)^2 + \\ \frac{2}{6a^3}(x-a)^3 - \frac{6}{24a^4}(x-a)^4 + o((x-a)^4)$$

Task 4

Find minimum (i.e. both point x^* and function value $f^* = f(x^*)$) of function with respect to parameters a, b, c:

$$f(x) = ax^2 + bx + c$$

Solution:

$$\frac{df}{dx}(x) = 2ax + b = 0 \implies x^* = -\frac{b}{2a}$$

$$\text{Substitute } x^* to f(x) \implies f(x^*) = a(-\frac{b}{2a})^2 - \frac{b}{2a}b + c = \frac{b}{4a^3} - \frac{b^2}{2a} + c$$

Task 5

What are dimensions of gradient of a function $h(x) = f(Ax)$, constructed of function $f : R^m \to R$ and matrix $A \in R^{m \times k}$.

Answer:

$$\nabla h(x) \in R^m$$

Task 6

Prove that for a strongly convex function with parameter μ holds:

$$\frac{\mu}{2}||x - x^*||_2^2 \leq f(x) - f^*$$

Proof

$$f^* = f(x^*)$$

$$f(x) - f(x^*) = \Bigg[f(x) = f(x^*) + f'(x^*)(x - x^*)$$

$$+ \frac{f''(x^*)}{2}(x - x^*)^2 + o(||x - x^*||_2^2)$$

$$\text{Derivative at } x^* \text{ is } 0 \Bigg] \implies$$

$$f(x) - f(x^*) = \frac{f''(x^*)}{2}(x - x^*)^2 + o(||x - x^*||_2^2)$$

Moreover, f strongly convex if and only if $f''(x) \geq m > 0$ for all x $\implies$
$\frac{f''(x^*)}{2}(x - x^*)^2 + o(||x - x^*||_2^2)|_{f''(x^*)=\mu} = \frac{\mu}{2}||x - x^*||_2^2 + o(||x - x^*||_2^2) \implies \frac{\mu}{2}||x - x^*||_2^2 \leq f(x)$ -

Task 7 ¶

Find conjugate $f^*(y)$ and it's domain for function

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

with **dom** $f = \{x : x > 0\}$

Solution

$$f^*(y) \doteq sup_{x \in \mathbf{dom}\, f} \left(y^T x - f(x)\right)$$

$$f^*(y) \doteq sup_{x \in \mathbf{dom}\, f} \left(y^T x - \frac{1}{x}\right)$$

Here, dimension of domain space is 1, thus

$$f^*(y) \doteq sup_{x \in \mathbf{dom}\, f} \left(yx - \frac{1}{x}\right) \implies \frac{d\left(yx - \frac{1}{x}\right)}{dx} = y + \frac{1}{x^2} = 0$$

$$y + \frac{1}{x^2} = 0 \implies -y = \frac{1}{x^2} \implies x = \frac{1}{\sqrt{-y}}$$

for all $y < 0$

$$f^*(y) = y\frac{1}{\sqrt{-y}} - \sqrt{-y},\ \mathbf{dom}\, f^* = \{y < 0\}$$

Task 8

Given conjugate function for $f^*(y)$ for $f(x)$, find conjugate function and its domain for

$$g(x) = f(x) + (c, x) + d$$
$$c \in R^n, d \in R$$

Solution

$$g^*(y) = sup_{x \in \mathbf{dom}\, f} \left(y^T x - g(x)\right) =$$
$$= sup_{x \in \mathbf{dom}\, f} \left(y^T x - f(x) - (c, x) - d\right)$$

On the other hand, we have

$$f^*(y) = sup_{x \in \mathbf{dom}\, f} \left(y^T x - f(x)\right) \implies$$
$$g^*(t)|_{t=y-c} = sup_{x \in \mathbf{dom}\, f} \left(t^T x - f(x)\right) - d = f^*(t) - d =$$
$$\implies g^*(y) = f^*(y - c) - d$$

Task 9

Derive gradient and hessian for $f(x) = (c, x)^2, x \in R^n$

Solution

$$f(x) = (c, x)^2$$
$$\nabla f(x) = 2 \left(c^\top \cdot x\right) c$$
$$\nabla^2 f(x) = 2 \left(c \cdot c^\top\right)$$

Task 10

Derive Hessian matrix for $f(x) = g(Ax + b)$, assuming diffenetiable $g : R^m \to R$, with dimensions $A \in R^{m \times n}, x \in R^n$

$\circ$ - Hamadard product (elementwise)

Solution

$$\nabla_x f(x) = \nabla_x g(Ax + b)$$

Let k_i - i-th component of $g(.)$, thus

$$\nabla_x f_i = \sum_{j=1}^{m} \frac{\partial g}{\partial k_j} \frac{\partial k_j}{\partial x_i} = \sum_{j=1}^{m} \frac{\partial g}{\partial k_j} \left(\frac{\partial}{\partial x_i} \sum_{k=1}^{n} a_{j,k} x_k + b_j \right) = \sum_{j=1}^{m} \frac{\partial g}{\partial k_j} a_{j,i}$$

or in matrix form

$$\nabla_x f(x) = div_{y=Ax+b}\, g(y) \circ A^\top \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

The next subtask, $\nabla_x^2 f(x)$:

$$\nabla_x^2 f(x)_{i,j} = \sum_{l=1}^{m} \sum_{p=1}^{m} \frac{\partial^2 g}{\partial k_l \partial k_p} \frac{\partial k_l}{\partial x_i} \frac{\partial k_p}{\partial x_j} = \sum_{l=1}^{m} \sum_{p=1}^{m} \left[\frac{\partial^2 g}{\partial k_l \partial k_p} \left(\frac{\partial}{\partial x_i} \sum_{t=1}^{n} a_{l,t} x_t + b_k \right) \left(\frac{\partial}{\partial x_j} \sum_{t=1}^{n} a_{p,t} x_t + b_k \right) \right] =$$

$$\nabla_x^2 f(x) = (\nabla_{y=Ax+b} \cdot \nabla_{y=Ax+b}^\top) g(y) \circ \left(A^\top \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \left(A^\top \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)^\top \right)$$

Task 11

Prove sufficient first order optimality condition for (everywhere) differentiable convex function $f(x)$: If $\nabla f(x^*) = 0$, then x^* is a global minimum of f".

Solution

From "Extra task 1(3)" we have:

$$f(y) \geq f(x) + (\nabla f(x), y - x) + \mu ||x - y||_2^2$$

All we need it let $x = x^*$:

$$f(y) \geq f(x^*) + (\nabla f(x^*), y - x) + \mu ||x - y||_2^2 \implies f(y) \geq f(x^*) + \mu ||x - y||_2^2 \implies f(y) > f(x^*)$$

For all $y \in \ \mathbf{dom} f$ and $x^* \neq y$

Task 12

Solve optimal step-size problem for the quadratic function, with symmetric positive definite matrix $A > 0, A \in R^{n \times n}$, and $x, b, d \in R^n$. Your goal is to find optimal γ^* for given A, b, d, x. The resulting expression must be written in terms of inner products (..., ...)

$$f(\gamma) = (A(x + \gamma d), x + \gamma d) + (b, x + \gamma d) \to min_{\gamma \in R}$$

Solution

$$f(\gamma) = (A(x + \gamma d), x + \gamma d) + (b, x + \gamma d) \to min_{\gamma \in R}$$

$$\frac{df(\gamma)}{d\gamma} = \frac{d}{d\gamma}\left(x^T Ax + \gamma x^T d + \gamma d^T Ax + \gamma^2 d^T Ad + b^T x + \gamma b^T d\right) =$$

$$= x^T d + d^T Ax + 2\gamma d^T Ad + b^T d = 0$$

$$\gamma = -\frac{b^T d + x^T d + d^T Ax}{2d^T Ad}$$

$$\gamma = -\frac{(b + x + Ax, d)}{2(Ad, d)}$$

Task 13

Derive subgradient (subdifferential) for the function $f(x) = \left[x^2 - 1\right]_+, x \in R$ subgradient method).

Solution

By definition

$$\partial f(x) = \{g | g^T (y - x) \leq f(y) - f(x), y \in \mathbf{dom}\, f\}$$

$$f(x) = \left[x^2 - 1\right]_+ = \begin{cases} 0, -1 < x < 1, \\ x^2 - 1, \{x \geq 1\} \bigcup \{x \leq -1\} \end{cases}$$

$f(x)$ not differentiable at $x = -1, x = 1$, thus:

$$\implies \partial f(x) = \begin{cases} 0, -1 < x < 1, \\ \partial(x^2 - 1), \{x > 1\} \bigcup \{x < -1\} \end{cases} = \begin{cases} 0, -1 < x < 1, \\ 2x, \{x > 1\} \bigcup \{x < -1\}, \\ [0, 2], x = 1, \\ [-2, 0], x = -1 \end{cases}$$

Extra tasks

Task 1

$$1. f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \mu\frac{\alpha(1-\alpha)}{2}||x-y||_2$$

$$3. f(y) \geq f(x) + (\nabla f(x), y-x) + \mu||x-y||_2^2$$

$$4. (\nabla f(x) - \nabla f(y), x-y) \geq \mu||x-y||_2$$

- a) Derive 4th definition of strong convexity from the 3rd one (function is differentiable).
- b) Derive 3rd definition of strong convexity from the 1st one (function is differentiable). Hint: you may need limits.

Solution (b)

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \mu\frac{\alpha(1-\alpha)}{2}||x-y||_2 =$$

$$= [f(y) = f(x) + \nabla f(x)(y-x) + \ldots] = \alpha f(x) + (1-\alpha)\left(f(x) + \nabla f(x)(y-x) + \frac{\nabla^2 f(x)}{2}(y-x)^2 + o(||y$$

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)\left(f(x) + \nabla f(x)(y-x) + \frac{\nabla^2 f(x)}{2}(y-x)^2 + o(||y-x||_2^2)\right)$$

$$f(\alpha x + (1-\alpha)y) \leq f(x) + (1-\alpha)\left(\nabla f(x)(y-x) + o(||y-x||_2^2)\right) + (1-\alpha)\frac{\nabla^2 f(x)}{2}(y-x)^2$$

Let $\alpha = 0$, thus:

$$f(y) \leq f(x) + \nabla f(x)(y-x) + \frac{\nabla^2 f(x)}{2}(y-x)^2 \leftrightarrow f(y) \leq f(x) + \nabla f(x)(y-x) + \mu(y-x)^2$$

Task 2(a)

Solve Least Squares problem (find x^* and f^*) $min_x||Ax-b||_2$ b) $A \in R^{n\times n}, b \in R^m, m > n, detA \neq 0$

Solution

$$f(x) = ||Ax-b||_2 = (Ax-b)^\top(Ax-b) = (x^\top A^\top - b^\top)(Ax-b) = x^\top A^\top Ax - x^\top A^\top b - b^\top A$$

$$min_x f(x) = min_x\left[x^\top A^\top Ax - x^\top A^\top b - b^\top Ax + b^\top b\right] = min_x\left[x^\top A^\top Ax - x^\top A^\top b - b^\top A\right.$$

$$\frac{df(x)}{dx} = \frac{d}{dx}\left(x^\top A^\top Ax - x^\top A^\top b - b^\top Ax + b^\top b\right) = x^\top AA^\top + x^\top AA - b^\top A^\top - b^\top A = \vec{0} \leftrightarrow x^\top(AA^\top + $$

$$x^\top = b^\top A^{-1} \leftrightarrow x = (A^{-1})^\top b$$

Optimal point x^* to $f(x) \implies f^*$:

$$f(x) = ||Ax-b||_2 = f(x) = ||A(A^{-1})^\top b - b||_2 = 0$$

Task 2(b)

Solve Least Squares problem (find x^* and f^*) $min_x ||Ax - b||_2$ b) $A \in R^{m \times n}, b \in R^m, m > n$, assuming A has full column rank, i.e. $A^\top A$ is non-singular.

Hint: if $f(x) \geq 0$, then $(min_x f(x))^2 = min_x (f(x)^2)$.

Solution

$f(x) = ||Ax - b||_2 = (Ax - b)^\top (Ax - b) = (x^\top A^\top - b^\top)(Ax - b) = x^\top A^\top Ax - x^\top A^\top b - b^\top Ax + b^\top b$

$$\implies min_x f(x) = min_x \left[x^\top A^\top Ax - x^\top A^\top b - b^\top Ax + b^\top b\right] = min_x \left[x^\top A^\top Ax - x^\top A^\top b - b^\top Ax\right]$$

We have a hint, that: $f(x) \geq 0$, then $(min_x f(x))^2 = min_x (f(x)^2)$, obviously, $f(x) \geq 0 \implies$

$$\frac{df(x)}{dx} = \frac{d}{dx}\left(x^\top A^\top Ax - x^\top A^\top b - b^\top Ax + b^\top b\right)^2 = 2\left(x^\top A^\top Ax - x^\top A^\top b - b^\top Ax + b^\top b\right)\left(2(Ax^\top)^\top A\right.$$

$$\begin{cases} x^\top A^\top Ax - x^\top A^\top b - b^\top Ax + b^\top b = 0, \text{ hasn't solution} \\ 2(Ax^\top)^\top A - (A^T b) - (A^T b) = \vec{0} \end{cases} \implies 2(Ax^\top)^\top A - (A^T b) - (A^T b) = \vec{0}$$

$$x = ((A^\top A)^{-1})A^\top b$$

As it was in the hint, we have optimal x^* and we can substitute it to $f(x)$

$f(x) = (Ax - b)^\top (Ax - b) = (A(((A^\top A)^{-1})A^\top b - b)^\top (A((A^\top A)^{-1})A^\top b - b) = ||A((A^\top A)^{-1})A^\top b - b||_2$