Machine Learning, 2021 Spring
Homework 2

Due on 23:59 MAR 28, 2021

Problem 1

Prove that $f : \mathbb{R}^n \to \mathbb{R}$ is affine if and only if f is both convex and concave. [2pts]

1. "$\Rightarrow$" $f: \mathbb{R}^n \to \mathbb{R}$ is affine, $f(x) = a^T x + b$

$\forall \lambda \in [0,1]$, $x, y \in \mathbb{R}^n$: $f(\lambda x + (1-\lambda) y) = a^T(\lambda x + (1-\lambda) y) + b$

$= \lambda(a^T x + b) + (1-\lambda)(a^T y + b)$

$= \lambda f(x) + (1-\lambda) f(y)$

$\because \begin{cases} f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y) \\ f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \end{cases}$, f is both convex and concave

"$\Leftarrow$" $f: \mathbb{R}^n \to \mathbb{R}$ is both convex and concave

$\forall \lambda \in [0,1]$, $x, y \in \mathbb{R}^n$: $f(\lambda x + (1-\lambda) y) = \lambda f(x) + (1-\lambda) f(y)$

Let $g(x) = f(x) - f(0)$, g is also convex and concave

$\forall \lambda \in [0,1]$, $x, y \in \mathbb{R}^n$: $g(\lambda x + (1-\lambda) y) = \lambda g(x) + (1-\lambda) g(y)$, $g(0) = 0$

Lemma1: $\forall x \in \mathbb{R}^n$, $\forall \lambda \in [0, +\infty)$: $g(\lambda x) = \lambda g(x)$

Pf: 1° Let $y=0$, $\forall x \in \mathbb{R}^n$, $\lambda \in [0,1]$: $g(\lambda x) = \lambda g(x)$

2° Let $y=0$, $\forall x \in \mathbb{R}^n$, $\lambda \in (1, +\infty)$, $g(\frac{1}{\lambda}(\lambda x)) = \frac{1}{\lambda} g(\lambda x)$

$g(\lambda x) = \lambda g(x)$

$\therefore$ 1° 2° $\Rightarrow \forall x \in \mathbb{R}^n, \forall \lambda \in [0, +\infty)$: $g(\lambda x) = \lambda g(x)$

Lemma2: $\forall x, y \in \mathbb{R}^n$, $g(x+y) = g(x) + g(y)$

Pf: $g(\frac{1}{2}(2x) + \frac{1}{2}(2y)) = \frac{1}{2} g(2x) + \frac{1}{2} g(2y)$, $\forall x, y \in \mathbb{R}^n$

Based on Lemma1, $\frac{1}{2} g(2x) = g(x)$, $\frac{1}{2} g(2y) = g(y)$

$\therefore \forall x, y \in \mathbb{R}^n$: $g(x+y) = g(x) + g(y)$

Based on Lemma1 and Lemma2:

$\forall x, y \in \mathbb{R}^n$: $g(x+y) = g(x) + g(y)$

$\therefore g(\lambda x + (-\lambda x)) = 0 = g(\lambda x) + g(-\lambda x)$, $\forall \lambda \in [0, +\infty)$, $x \in \mathbb{R}^n$

$g(-\lambda x) = -g(\lambda x) = -\lambda g(x)$

$\therefore g(\lambda x) = \lambda g(x)$, $\forall \lambda \in (-\infty, 0]$

$\therefore \begin{cases} \forall x, y \in \mathbb{R}^n, \lambda \in \mathbb{R}^n: g(\lambda x) = \lambda g(x) \\ \forall x, y \in \mathbb{R}^n: g(x+y) = g(x) + g(y) \end{cases}$

$\therefore g$ is linear function

$\therefore f(x) = g(x) + f(0)$ is affine

Problem 2

Suppose A and B are both convex sets, prove that $C = A \cap B$ is also convex. [1pts]

2. $\forall x, y \in C = A \cap B$

$x, y \in A$, $x, y \in B$

$\therefore \forall \theta \in [0,1]$: $\theta x + (1-\theta) y \in A$

$\theta x + (1-\theta) y \in B$

$\therefore \theta x + (1-\theta) y \in A \cap B$

$\therefore C = A \cap B$ is convex set

Problem 3

Suppose your algorithm for solving the problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \; f(\boldsymbol{x}) \tag{1}$$

takes iteration:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha_k \boldsymbol{p}^k \tag{2}$$

where $\boldsymbol{p}^k = \boldsymbol{H}^k \nabla f(\boldsymbol{x}^k)$. What kind of $\boldsymbol{H}^k$ can guarantee that $\boldsymbol{p}^k$ is a descent direction ? [2pts]

3. the definition of descent direction: $d \in \mathbb{R}^n$ such that $\langle \nabla f(x), d \rangle < 0$

let $H^k \prec 0$, $\langle \nabla f(x^k), p^k \rangle = \nabla f(x^k)^T H^k \nabla f(x^k) < 0 \quad \forall x^k \in \mathbb{R}^n$

$\therefore H^k$ be negative definite can guarantee p^k be a descent direction

Problem 4

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. For a given $\boldsymbol{x} \in \mathbb{R}^n$, show that moving along $-\nabla f(\boldsymbol{x}) \neq 0$ with sufficiently small stepsize causes decrease on f, that is,

$$f(\boldsymbol{x} - \alpha \nabla f(\boldsymbol{x})) < f(\boldsymbol{x}) \tag{3}$$

for sufficiently small $\alpha > 0$. [2pts]

4. fix x, $f(x-\alpha\nabla f(x)) = f(x) - \alpha\|\nabla f(x)\|^2 + o(\alpha)$ $\alpha > 0$

$$\frac{f(x-\alpha\nabla f(x)) - f(x)}{\alpha} = -\|\nabla f(x)\|^2 + \frac{o(\alpha)}{\alpha}$$

$\because \lim_{\alpha\to 0} \frac{o(\alpha)}{\alpha} = 0$

$\therefore \exists \hat{\alpha} > 0, \forall \alpha \in (0, \hat{\alpha})$: $\frac{o(\alpha)}{\alpha} < \frac{1}{2}\|\nabla f(x)\|^2$

$$\frac{f(x-\alpha\nabla f(x)) - f(x)}{\alpha} < -\frac{1}{2}\|\nabla f(x)\|^2 \leq 0$$

$\therefore f(x-\alpha\nabla f(x)) < f(x)$ for sufficient small α, $\alpha > 0$

Problem 5

Use gradient descent to solve the underdetermined linear system:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \tag{4}$$

with stepsize chosen as exact line search, initial point $\boldsymbol{x}^0 = 0$ and maximum iteration 1000. Plot :

1. The objective value against the iteration.(Use log scale for y-axis)
2. The ℓ_2 norm of gradient against the iteration.(Use log scale for y-axis)
3. The stepsize against the iteration.

The data $\boldsymbol{A} \in \mathbb{R}^{500\times 1000}, \boldsymbol{b} \in \mathbb{R}^{500\times 1}$ is attached in data/A.csv and data/b.csv with comma-separated (delimiter=','). [Hint: what is the solution to the exact line search for quadratic function?][3pts]

5. $\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}\|Ax-b\|_2^2$

$\nabla f(x) = A^T(Ax-b)$, $x^{k+1} = x^k - \alpha\nabla f(x)$

$f(x - s\nabla f(x))$

$= \frac{1}{2}[A(x-s\nabla f(x))-b]^T[A(x-s\nabla f(x))-b]$

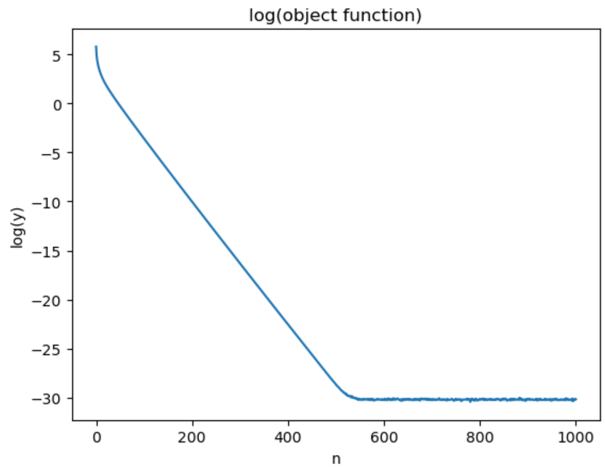
$= \frac{1}{2}(x-s\nabla f(x))^T A^T A(x-s\nabla f(x)) - b^T A(x-s\nabla f(x)) + \frac{1}{2}b^T b$

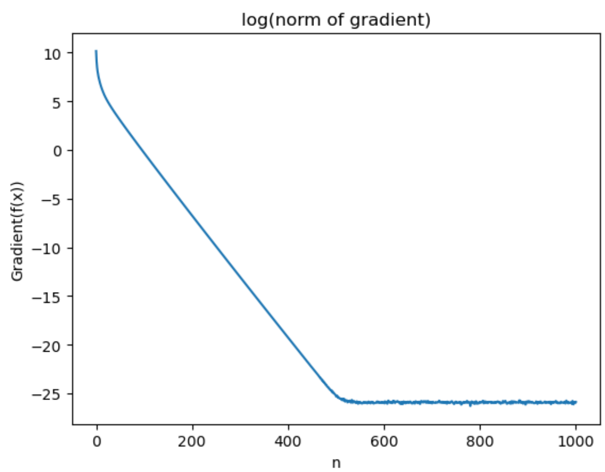
$= \frac{1}{2}x^T A^T A x - b^T A x + \frac{1}{2}b^T b + \frac{1}{2}\nabla f(x)^T A^T A\nabla f(x) s^2 - \nabla f(x)^T A^T A x s + b^T A \nabla f(x) s$

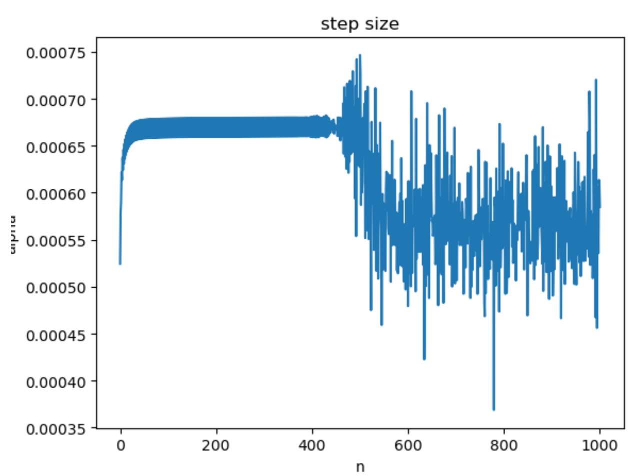
$= \frac{1}{2}\|A\nabla f(x)\|^2 s^2 + (b^T A\nabla f(x) - \nabla f(x)^T A^T A x) s + \frac{1}{2}x^T A^T A x - b^T A x + \frac{1}{2}b^T b$

$= \frac{1}{2}\|A\nabla f(x)\|^2 \left(s + \frac{b^T A\nabla f(x) - \nabla f(x)^T A^T A x}{\|\nabla f(x)^T A\|^2}\right)^2 + \square$

$\alpha = \arg\min_{s\geq 0} f(x - s\nabla f(x)) = \frac{\nabla f(x)^T A^T A x - b^T A\nabla f(x)}{\|A\nabla f(x)\|^2}$, α is the stepsize chosen by exact line search


log(object function)
log(y)
n


log(norm of gradient)
Gradient(f(x))
n


step size
n