Problem 1.

Solution

- $S_1 = \{(x_1, x_2) : x_1^2 + x_2^2 = 1\}$ not convex.

  Counterexample : $(x_1, x_2) = (1, 0)$ $\quad (y_1, y_2) = (0, 1)$ $\quad \lambda = \frac{1}{2}$

  $$\frac{1}{2}(x_1, x_2) + \frac{1}{2}(y_1, y_2) = \left(\frac{1}{2}, \frac{1}{2}\right) \notin S.$$

- $S_2 = \{(x_1, x_2) : |x_1| + |x_2| \leq 1\}$ convex

  Proof: For every $x = (x_1, x_2)$, $y = (y_1, y_2)$ and every $\lambda \in (0,1)$

  $$\lambda x + (1-\lambda) y = (\lambda x_1 + (1-\lambda) y_1, \ \lambda x_2 + (1-\lambda) y_2)$$

  Then we compute

  $$\begin{aligned}
  & |\lambda x_1 + (1-\lambda) y_1| + |\lambda x_2 + (1-\lambda) y_2| \\
  \leq \ & |\lambda x_1| + |(1-\lambda) y_1| + |\lambda x_2| + |(1-\lambda) y_2| && \text{[Triangular inequality]} \\
  = \ & \lambda |x_1| + (1-\lambda)|y_1| + \lambda |x_2| + (1-\lambda)|y_2| && [\lambda, 1-\lambda \text{ are positive}] \\
  = \ & \lambda (|x_1| + |x_2|) + (1-\lambda)(|y_1| + |y_2|) && \text{[rearrange terms]} \\
  \leq \ & \lambda + 1 - \lambda && [x, y \in S] \\
  = \ & 1
  \end{aligned}$$

  #

Problem 2.

a) View x_2 as a constant

$$\frac{df}{dx_1} = 2x_1 - 4x_2, \quad \frac{d^2 f}{dx_1^2} = 2 \geq 0 \quad \text{for every } x_1$$

Thus, f is convex in x_1.

b) View x_1 as a constant

$$\frac{df}{dx_2} = 2x_2 - 4x_1, \quad \frac{d^2 f}{dx_2} = 2 \geq 0 \quad \text{for every } x_2$$

Thus, f is convex in x_2

c) Consider some point, for example $(0, 0)$

$$H(0,0) = \begin{bmatrix} 2 & -4 \\ -4 & 2 \end{bmatrix}$$

By the definition of eigenvalues, $Hx = \lambda x$ for every $x \neq 0$

$$(H - \lambda I) x = 0$$

i.e., $\begin{bmatrix} 2-\lambda & -4 \\ -4 & 2-\lambda \end{bmatrix} = 0$

This gives us two eigenvalues 6 and -2.

Since one of them is negative, f is not convex in (x_1, x_2)

Problem 3.

Proof

Consider any two points $x, y \in \text{dom} f$

For every $\lambda \in (0,1)$

Since f is convex,

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

$$f(y) \geq f(x) + \frac{f(x+\lambda(y-x)) - f(x)}{\lambda}$$ [rearrange terms]

Take the limit of λ to 0^+ (meaning that $\lambda > 0$ could be arbitrarily small)

$$f(y) \geq f(x) + \lim_{\lambda \to 0^+} \frac{f(x+\lambda(y-x)) - f(x)}{\lambda}$$

$$= f(x) + \lim_{\lambda \to 0^+} \frac{[\nabla_x f(x)]^T \cdot \lambda (y-x)}{\lambda}$$

$$= f(x) + [\nabla_x f(x)]^T (y-x)$$

#

Problem 4

Proof

$$\|z - x\| = \|x - \lambda x + \lambda y - x\|$$ [Put z in]

$$= \left\| \frac{\varepsilon}{2\|y-x\|} \cdot (y-x) \right\|$$ [Put ε in]

$$= \frac{\varepsilon}{2} < \varepsilon$$

#

Problem 5

1) Proof

For every $y \in \text{dom} f$

$$f(y) \geq f(x) + [\nabla f(x)]^T (y-x)$$ [1^{st}-order condition]

$$= f(x)$$ [$\nabla f(x) = 0$]

Note: Proof in Problem 3 only used the differentiation at x.

#

2). Not unique. Example: $f(x) = 0$ for $x \in \mathbb{R}$. Every point is a global optimum.

Problem 6

a) $J = \|\mathbf{X}\mathbf{w} - \mathbf{t}\|^2 + \mathbf{w}^T\mathbf{w}$

b) $\nabla_{\mathbf{w}} J = 2\mathbf{X}^T\mathbf{X}\mathbf{w} - 2\mathbf{X}^T\mathbf{t} + 2\mathbf{w}$

$\nabla^2_{\mathbf{w}} J = 2\mathbf{X}^T\mathbf{X} + 2\mathbf{I}$ [$\mathbf{I}$ is an identity matrix]

$2\mathbf{X}\ \mathbf{X} + 2\mathbf{I}$ is positive definite because for every non-zero vector $\mathbf{u} \in \mathbb{R}^{d+1}$,

$$\mathbf{u}^T(\mathbf{X}^T\mathbf{X} + \mathbf{I})\,\mathbf{u} = \mathbf{u}^T\mathbf{X}^T\mathbf{X}\,\mathbf{u} + \mathbf{u}^T\mathbf{u} \geqslant 0$$

Therefore, J is convex

Problem 7. $J(w - \alpha u) = J(w) + [\nabla J(w)]^T(-\alpha u) + O(\alpha^2)$

When $\alpha \to 0^+$, $J(w-\alpha u) - J(w) = -\alpha\,[\nabla J(w)]^T \cdot u + O(\alpha^2)$

$\leqslant 0$ [because $[\nabla J(w)]^T u \geqslant 0$]

Problem 8.

Initialize w randomly

Loop until convergence:

Compute gradient $g = 2X^TXw^{(old)} - 2X^Tt + 2w^{(old)}$ [problem 6]

Update $w^{(new)} = w^{(old)} - \alpha \cdot g$

Return w.