problem 1

$\{(x_1, x_2) : x_1^2 + x_2^2 = 1\}$

$\{(x_1, x_2) : |x_1| + |x_2| \leq 1\}$

1. For $\{(x_1, x_2) : x_1^2 + x_2^2 = 1\}$, suppose there are two points $\mathbf{x}$ and $\mathbf{y}$. $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2)$

and $x_1^2 + x_2^2 = 1$, $y_1^2 + y_2^2 = 1$

$\forall \mathbf{x}, \mathbf{y} \in S$, $\forall \lambda \in (0, 1)$

$$\lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in S$$

$$(\lambda x_1 + (1-\lambda) y_1 , \lambda x_2 + (1-\lambda) y_2) \in S$$

$$[\lambda x_1 + (1-\lambda) y_1]^2 + [\lambda x_2 + (1-\lambda) y_2]^2$$

$$= \lambda^2 x_1^2 + (1-\lambda)^2 y_1^2 + 2\lambda(1-\lambda) x_1 y_1 + \lambda^2 x_2^2 + (1-\lambda)^2 y_2^2 + 2\lambda(1-\lambda) x_2 y_2$$

$$= \lambda^2 (x_1^2 + x_2^2) + (1-\lambda)^2 (y_1^2 + y_2^2) + 2\lambda(1-\lambda)(x_1 y_1 + x_2 y_2)$$

$$= \lambda^2 \times 1 + (1-\lambda)^2 \times 1 + 2\lambda(1-\lambda)(x_1 y_1 + x_2 y_2)$$

$$= \lambda^2 + 1 + \lambda^2 - 2\lambda + 2\lambda(1-\lambda)(x_1y_1 + x_2y_2)$$

$$= 1 + 2\lambda^2 - 2\lambda + 2\lambda(1-\lambda)(x_1y_1 + x_2y_2)$$

when $x_1y_1 + x_2y_2 = 1$, $1 + 2\lambda^2 - 2\lambda + 2\lambda(1-\lambda) \times 1$

$$= 1 + 2\lambda^2 - 2\lambda + 2\lambda - 2\lambda^2$$

$$= 1$$

However, we don't know whether $x_1y_1 + x_2y_2 = 1$ or not.

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

2. For $\{(x_1, x_2) : |x_1| + |x_2| \leq 1\}$, suppose there are two points x and y. $x = (x_1, x_2)$, $y = (y_1, y_2)$

and $|x_1| + |x_2| \leq 1$, $|y_1| + |y_2| \leq 1$

$\forall x, y \in S$, $\forall \lambda \in (0, 1)$

$\lambda x + (1-\lambda) y \in S$

$x, y \in S$

$\lambda x + (1-\lambda) y \in S$

$$(\lambda x_1 + (1-\lambda) y_1 \ , \ \lambda x_2 + (1-\lambda) y_2) \in S$$

Prove : $|\lambda x_1 + (1-\lambda) y_1| + |\lambda x_2 + (1-\lambda) y_2| \leq 1$

$$|\lambda x_1 + (1-\lambda) y_1| + |\lambda x_2 + (1-\lambda) y_2|$$

$$\leq |\lambda x_1| + |(1-\lambda) y_1| + |\lambda x_2| + |(1-\lambda) y_2|$$

$$= \lambda(|x_1| + |x_2|) + (1-\lambda)(|y_1| + |y_2|)$$

$$\leq \lambda + (1-\lambda)$$

$$= 1$$

Thus, $\{(x_1, x_2) : |x_1| + |x_2| \leq 1\}$ is convex.

Problem 2

a) View x_1 as a variable and x_2 as a constant.

$$f(x_1, x_2) = x_1^2 + x_2^2 - 4x_1x_2$$

$$\nabla f(x_1) = 2x_1 - 4x_2$$

$$\nabla^2 f(x_1) = 2 \geq 0$$

Thus, f is convex in x_1.

b) View x_2 as a variable and x_1 as a constant.

$$f(x_1, x_2) = x_1^2 + x_2^2 - 4x_1x_2$$

$$\nabla f(x_2) = 2x_2 - 4x_1$$

$$\nabla^2 f(x_2) = 2 \geq 0$$

Thus, f is convex in x_2.

(c) $f(x_1, x_2) = x_1^2 + x_2^2 - 4x_1x_2$

$\nabla_{x_1} f(x_1) = 2x_1 - 4x_2$

$\nabla_{x_2} f(x_2) = 2x_2 - 4x_1$

$$H(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \end{pmatrix}_{2\times2} = \begin{pmatrix} 2 & -4 \\ -4 & 2 \end{pmatrix}_{2\times2}$$

$$H(x) - \lambda \cdot I = \left| \begin{pmatrix} 2 & -4 \\ -4 & 2 \end{pmatrix}_{2\times2} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right|$$

$$= \left| \begin{pmatrix} 2-\lambda & -4 \\ -4 & 2-\lambda \end{pmatrix}_{2\times2} \right| = 0$$

$(2-\lambda)^2 - 16 = 0$

$4 - 4\lambda + \lambda^2 - 16 = 0$

$\lambda^2 - 4\lambda - 12 = 0$

$(\lambda - 6)(\lambda + 2) = 0$

$\lambda_1 = 6 \quad \lambda_2 = -2$

-2 is smaller than 0, then the function is not convex.

Problem 3.

First-order condition is:

$$\forall x, y \in \text{dom} f, \quad f(y) \geqslant f(x) + [\nabla f(x)]^T \cdot (y-x)$$

Suppose f is a differentiable convex function, by definition

$$\forall x, y \in \text{dom} f, \forall \lambda \in [0,1],$$

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

$$f(x + \lambda(y-x)) \leqslant f(x) + \lambda(f(y) - f(x))$$

$$\Rightarrow f(y) - f(x) \geqslant \frac{f(x+\lambda(y-x)) - f(x)}{\lambda}, \forall \lambda \in (0,1]$$

As $\lambda \downarrow 0$, we get

$$f(y) - f(x) \geqslant \nabla f^T(x)(y-x)$$

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

Thus, f satisfies the first-order condition.

Problem 4.

$$\lambda = \frac{\varepsilon}{2\|y-x\|}$$

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

$$\begin{aligned}\|z-x\| &= \|(1-\lambda)x + \lambda y - x\| \\ &= \|\lambda y - \lambda x\| \\ &= \lambda \|y-x\| \\ &= \frac{\varepsilon}{2\|y-x\|}\|y-x\| \\ &= \frac{\varepsilon}{2} < \varepsilon\end{aligned}$$

Thus, z is indeed in the ε-neighbor of x.

Problem 5.

1) Suppose f is a convex function and the gradient $\nabla f(x) = 0$ at the point x.

By first order condition,

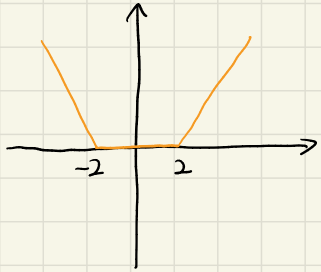
$$\forall x, y \in \mathrm{dom} f,\ f(y) \geq f(x) + [\nabla f(x)]^T \cdot (y - x)$$

$$\nabla f(x) = 0 \implies f(y) \geq f(x)$$

Thus, x is a global optimum of function f.

2) No, global optimum is not unique.

Counter example:



Suppose

$$f(x) = \begin{cases} 0 & \text{if } -2 < x < 2 \\ |x| - 2, & \text{otherwise} \end{cases}$$

$f(x)$ is a convex function, $\lambda f(x)+(1-\lambda)f(y) \geqslant f(\lambda x+(1-\lambda)y)$,
for $\forall\, x, y \in \mathrm{dom} f$, $\forall\, \lambda \in (0,1)$

And obviously, $f(x)$ has so many global optimum,
from -2 to 2.

Problem 6

a) $$J = \sum_{m=1}^{m} \left(\sum_{i=0}^{d} w_i x_i^{(m)} - t^{(m)} \right)^2 + \sum_{i=0}^{d} w_i^2$$

$$= \sum_{m=1}^{M} \left(\sum_{i=0}^{d} w_i x_i^{(m)} - t^{(m)} \right)^2 + W^T W$$

$$= \sum_{m=1}^{m} \left(X^{(m)} W - t^{(m)} \right)^2 + W^T W$$

$$= \|XW - t\|_2^2 + W^T W$$

$$= (XW - t)^T (XW - t) + W^T W$$

b) $$J = (XW - t)^T (XW - t) + W^T W$$

$$= (W^T X^T - t^T)(XW - t) + W^T W$$

$$= W^T X^T X W - W^T X^T t - t^T X W + t^T t + W^T W$$

$$= W^T X^T X W - 2 t^T X W + t^T t + W^T W$$

$$\nabla J(w) = 2X^T X W - 2X^T t + 2W$$

$$\nabla\nabla J(w) = 2X^T X + 2 \geq 0$$

Thus, J is convex in W.

Problem 7

Let u be a d-dimensional vector such that $[\nabla J(w)]^T u \geq 0$

prove that $J(w - \partial u) \leq J(w)$ for a small enough positive ∂.

Move a small step towards the opposite direction of

gradient u, and we get:

By Taylor approx,

As $\partial \to 0$, $J(w-\partial u) = J(w) + [\nabla_w J(w)]^T [-\partial u] + o(\partial^2)$

$$= J(w) - \partial \|u\|^2 + o(\partial^2)$$

$$\leq J(w)$$

small but positive

Thus, $J(w-\partial u) \leq J(w)$ for a small enough positive ∂.

Problem 8.

Initialize $w = w^{(0)}$ randomly

Loop over epochs $t = 0, 1, 2, \ldots$:

Compute gradient $\nabla J(w)\big|_{w=w^{(t)}}$

$$= 2X^T X w^{(t)} - 2X^T t + 2w^{(t)}$$

Update parameters

$$w^{(t+1)} := w^{(t)} - \lambda \cdot \nabla J(w)\big|_{w=w^{(t)}}$$

Until stopping criterion satisfies.