

Biomath 210 Homework 3

Huwenbo Shi (603-778-363) shihuwenbo@ucla.edu

November 7, 2015

Problem 4.12

Finding the spectral norm $\|\mathbf{M}\|$ of a matrix \mathbf{M} is equivalent to solving the problem

$$\begin{aligned} & \text{maximize} && \mathbf{v}^* \mathbf{M}^* \mathbf{M} \mathbf{v} \\ & \text{s.t.} && \mathbf{v}^* \mathbf{v} = 1. \end{aligned} \tag{1}$$

Finding the Lagrangian of the problem above, we get

$$\mathcal{L}(\mathbf{v}) = \mathbf{v}^* \mathbf{M}^* \mathbf{M} \mathbf{v} + \lambda \mathbf{v}^* \mathbf{v} - \lambda, \tag{2}$$

which has the Lagrangian

$$\nabla \mathcal{L}(\mathbf{v}) = 2\mathbf{M}^* \mathbf{M} \mathbf{v} + 2\lambda \mathbf{v}. \tag{3}$$

Setting the gradient to 0, we arrive at the following condition for a stationary point

$$\mathbf{M}^* \mathbf{M} \mathbf{v} = -\lambda \mathbf{v}. \tag{4}$$

When $-\lambda = \rho(\mathbf{M}^* \mathbf{M})$, the largest eigenvalue of $\mathbf{M}^* \mathbf{M}$, and \mathbf{v} , the corresponding normalized eigenvector, the above equality is satisfied. Because $\mathbf{x}^* \mathbf{M}^* \mathbf{M} \mathbf{x} \geq 0$ for all \mathbf{x} , $\mathbf{M}^* \mathbf{M}$ is positive semi-definite, and $\rho(\mathbf{M}^* \mathbf{M}) \geq 0$. Therefore,

$$\|\mathbf{M}\| = \sqrt{\mathbf{v}^* \mathbf{M}^* \mathbf{M} \mathbf{v}} = \sqrt{\mathbf{v}^* \rho(\mathbf{M}^* \mathbf{M}) \mathbf{v}} = \sqrt{\rho(\mathbf{M}^* \mathbf{M})}. \tag{5}$$

Similarly, we can show that

$$\|\mathbf{M}^* \mathbf{M}\| = \sqrt{\rho((\mathbf{M}^* \mathbf{M})^* \mathbf{M}^* \mathbf{M})} = \sqrt{\rho(\mathbf{M}^* \mathbf{M} \mathbf{M}^* \mathbf{M})}. \tag{6}$$

Let $\mathbf{M}^* \mathbf{M} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^*$ be the eigen-decomposition of $\mathbf{M}^* \mathbf{M}$, where \mathbf{Q} is a unitary matrix and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\mathbf{M}^* \mathbf{M}$. Then,

$$(\mathbf{M}^* \mathbf{M})(\mathbf{M}^* \mathbf{M}) = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^* \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^* = \mathbf{Q} \mathbf{\Lambda}^2 \mathbf{Q}^*, \tag{7}$$

from which we obtain

$$\|\mathbf{M}^* \mathbf{M}\| = \sqrt{\rho(\mathbf{M}^* \mathbf{M} \mathbf{M}^* \mathbf{M})} = \sqrt{\rho(\mathbf{M}^* \mathbf{M})^2} = \rho(\mathbf{M}^* \mathbf{M}) \tag{8}$$

From Equation (5) and (8), we conclude

$$\|\mathbf{M}\|^2 = \|\mathbf{M}^* \mathbf{M}\|. \quad (9)$$

To prove the inequality, let \mathbf{v} be the eigenvector corresponding to the largest eigenvalue, $\rho(\mathbf{M}^* \mathbf{M})$, of $\mathbf{M}^* \mathbf{M}$. First, it can be easily shown that for $\mathbf{M} \in \mathbb{R}^{m \times n}$,

$$\|\mathbf{M}\|_1 = \max_j \sum_{i=1}^m |x_{ij}|, \quad \|\mathbf{M}\|_\infty = \max_i \sum_{j=1}^n |x_{ij}|, \quad (10)$$

and that $\|\mathbf{M}^*\|_1 = \|\mathbf{M}\|_\infty$. Then, it follows that

$$\rho(\mathbf{M}^* \mathbf{M}) \|\mathbf{v}\|_1 = \|\rho(\mathbf{M}^* \mathbf{M}) \mathbf{v}\|_1 = \|\mathbf{M}^* \mathbf{M} \mathbf{v}\|_1 \leq \|\mathbf{M}^*\|_1 \|\mathbf{M}\|_1 \|\mathbf{v}\|_1, \quad (11)$$

from which we conclude

$$\begin{aligned} \rho(\mathbf{M}^* \mathbf{M}) &\leq \|\mathbf{M}^*\|_\infty \|\mathbf{M}\|_1 \\ \|\mathbf{M}\| &\leq \sqrt{\|\mathbf{M}^*\|_\infty \|\mathbf{M}\|_1}. \end{aligned} \quad (12)$$

Problem 4.16

When the center \mathbf{x} of the sphere is fixed, then

$$\frac{\partial f(r, \mathbf{x})}{\partial r} = \sum_{i=1}^p 2(\|\mathbf{y}_i - \mathbf{x}\| - r) = -2pr + \sum_{i=1}^p 2\|\mathbf{y}_i - \mathbf{x}\|. \quad (13)$$

Setting the partial derivative to zero and solving for r gives

$$r = \frac{1}{p} \sum_{i=1}^p \|\mathbf{y}_i - \mathbf{x}\|. \quad (14)$$

For f fixed,

$$\begin{aligned} f(r, \mathbf{x}) &= \sum_{i=1}^p [\|\mathbf{y}_i - \mathbf{x}\|^2 - 2r\|\mathbf{y}_i - \mathbf{x}\| + r^2] \\ &= \sum_{i=1}^p [\|\mathbf{y}_i - \mathbf{x}\|^2 - 2r\|\mathbf{y}_i - \mathbf{x}\|] + pr^2 \end{aligned} \quad (15)$$

By Cauchy-Schwarz inequality, we have

$$(\mathbf{y}_i - \mathbf{x}_n)^* (\mathbf{y}_i - \mathbf{x}) \leq \|\mathbf{y}_i - \mathbf{x}_n\| \|\mathbf{y}_i - \mathbf{x}\|. \quad (16)$$

Thus,

$$\begin{aligned} \frac{1}{\|\mathbf{y}_i - \mathbf{x}_n\|} (\mathbf{y}_i - \mathbf{x}_n)^* (\mathbf{y}_i - \mathbf{x}) &\leq \|\mathbf{y}_i - \mathbf{x}\| \\ -2r \frac{1}{\|\mathbf{y}_i - \mathbf{x}_n\|} (\mathbf{y}_i - \mathbf{x}_n)^* (\mathbf{y}_i - \mathbf{x}) &\geq -2r \|\mathbf{y}_i - \mathbf{x}\|. \end{aligned} \quad (17)$$

And we can conclude that the function

$$g(\mathbf{x}|\mathbf{x}_n) = \sum_{i=1}^p \left[\|\mathbf{y}_i - \mathbf{x}\|^2 - 2r \frac{1}{\|\mathbf{y}_i - \mathbf{x}_n\|} (\mathbf{y}_i - \mathbf{x}_n)^* (\mathbf{y}_i - \mathbf{x}) \right] + pr^2 \quad (18)$$

majorizes $f(r, \mathbf{x})$ for r fixed.

Simplifying $g(\mathbf{x}|\mathbf{x}_n)$ gives

$$g(\mathbf{x}|\mathbf{x}_n) = \sum_{i=1}^p \left[-2\mathbf{x}^* \mathbf{y}_i + \mathbf{x}^* \mathbf{x} + 2r \frac{\mathbf{y}_i - \mathbf{x}_n}{\|\mathbf{y}_i - \mathbf{x}_n\|} \mathbf{x} \right] + c, \quad (19)$$

where c is an irrelevant constant. Taking the gradient of $g(\mathbf{x}|\mathbf{x}_n)$ with respect to \mathbf{x} , we get

$$\nabla g(\mathbf{x}|\mathbf{x}_n) = 2p\mathbf{x} + 2 \sum_{i=1}^p \left[-\mathbf{y}_i + r \frac{\mathbf{y}_i - \mathbf{x}_n}{\|\mathbf{y}_i - \mathbf{x}_n\|} \right]. \quad (20)$$

Setting the gradient to 0 and solving for \mathbf{x} gives the update

$$\mathbf{x}_{n+1} = \frac{1}{p} \sum_{i=1}^p \left[\mathbf{y}_i - r \frac{\mathbf{y}_i - \mathbf{x}_n}{\|\mathbf{y}_i - \mathbf{x}_n\|} \right]. \quad (21)$$

Problem 4.17

By the definition of subdifferential,

$$\begin{aligned} \partial f(\mathbf{x}) &= \{\mathbf{v} : f(\mathbf{z}) \geq f(\mathbf{x}) + \mathbf{v}^* (\mathbf{z} - \mathbf{x})\} \\ \partial g(\mathbf{x}|\mathbf{x}) &= \{\mathbf{v} : g(\mathbf{z}|\mathbf{x}) \geq g(\mathbf{x}|\mathbf{x}) + \mathbf{v}^* (\mathbf{z} - \mathbf{x})\}. \end{aligned} \quad (22)$$

Let $\mathbf{u} \in \partial f(\mathbf{x})$, then

$$f(\mathbf{z}) \geq f(\mathbf{x}) + \mathbf{u}^* (\mathbf{z} - \mathbf{x}). \quad (23)$$

Because $g(\mathbf{y}|\mathbf{x})$ majorizes $f(\mathbf{y})$ with anchor point \mathbf{x} , we have $g(\mathbf{z}|\mathbf{x}) \geq f(\mathbf{z})$, and $g(\mathbf{x}|\mathbf{x}) = f(\mathbf{x})$. Therefore, from inequality (23), we have

$$g(\mathbf{z}|\mathbf{x}) \geq g(\mathbf{x}|\mathbf{x}) + \mathbf{u}^* (\mathbf{z} - \mathbf{x}). \quad (24)$$

Thus, we can conclude $\mathbf{u} \in \partial g(\mathbf{x}|\mathbf{x})$ and $\partial f(\mathbf{x}) \subset \partial g(\mathbf{x}|\mathbf{x})$.

Let $f(x) = |x|$, and $g(y|x) = |2y|$ be the majorization of $f(x)$ at anchor point $x = 0$. Then $\partial f(0) = [-1, 1]$ and $\partial g(0|0) = [-2, 2]$. In this case, we have $\partial f(0) \subset \partial g(0|0)$, but $\partial f(0) \neq \partial g(0|0)$.

Problem 4.19

Since $[y_i - \mu_i(\boldsymbol{\theta})]^2 = [y_i - \mu_i(\boldsymbol{\theta}_m) + \mu_i(\boldsymbol{\theta}_m) - \mu_i(\boldsymbol{\theta})]^2$, for $w_i \in [0, 1]$ we have

$$\begin{aligned}
& w_i [y_i - \mu_i(\boldsymbol{\theta})]^2 \\
&= w_i [y_i - \mu_i(\boldsymbol{\theta}_m) + \mu_i(\boldsymbol{\theta}_m) - \mu_i(\boldsymbol{\theta})]^2 \\
&= w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]^2 + w_i [\mu_i(\boldsymbol{\theta}_m) - \mu_i(\boldsymbol{\theta})]^2 + 2w_i [y_i - \mu_i(\boldsymbol{\theta}_m)][\mu_i(\boldsymbol{\theta}_m) - \mu_i(\boldsymbol{\theta})] \\
&\leq w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]^2 + [\mu_i(\boldsymbol{\theta}_m) - \mu_i(\boldsymbol{\theta})]^2 + 2w_i [y_i - \mu_i(\boldsymbol{\theta}_m)][\mu_i(\boldsymbol{\theta}_m) - \mu_i(\boldsymbol{\theta})] \\
&= w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]^2 + \mu_i(\boldsymbol{\theta}_m)^2 - 2\mu_i(\boldsymbol{\theta}_m)\mu_i(\boldsymbol{\theta}) + \mu_i(\boldsymbol{\theta})^2 \\
&\quad + 2w_i [y_i - \mu_i(\boldsymbol{\theta}_m)][\mu_i(\boldsymbol{\theta}_m) - \mu_i(\boldsymbol{\theta})] \\
&= \mu_i(\boldsymbol{\theta})^2 - 2\mu_i(\boldsymbol{\theta}_m)\mu_i(\boldsymbol{\theta}) - 2w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]\mu_i(\boldsymbol{\theta}) \quad , \quad (25) \\
&\quad + w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]^2 + 2w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]\mu_i(\boldsymbol{\theta}_m) + \mu_i(\boldsymbol{\theta}_m)^2 \\
&= \mu_i(\boldsymbol{\theta})^2 - 2[w_i y_i + (1 - w_i)\mu_i(\boldsymbol{\theta}_m)]\mu_i(\boldsymbol{\theta}) + \mu_i(\boldsymbol{\theta}_m)^2 \\
&\quad + w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]^2 + 2w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]\mu_i(\boldsymbol{\theta}_m) \\
&= [w_i y_i + (1 - w_i)\mu_i(\boldsymbol{\theta}_m) - \mu_i(\boldsymbol{\theta})]^2 - [w_i y_i + (1 - w_i)\mu_i(\boldsymbol{\theta}_m)]^2 + \mu_i(\boldsymbol{\theta}_m)^2 \\
&\quad + w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]^2 + 2w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]\mu_i(\boldsymbol{\theta}_m) \\
&= [w_i y_i + (1 - w_i)\mu_i(\boldsymbol{\theta}_m) - \mu_i(\boldsymbol{\theta})]^2 + c_{i,m}
\end{aligned}$$

where

$$\begin{aligned}
c_{i,m} &= -[w_i y_i + (1 - w_i)\mu_i(\boldsymbol{\theta}_m)]^2 + w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]^2 \\
&\quad + 2w_i [y_i - \mu_i(\boldsymbol{\theta}_m)]\mu_i(\boldsymbol{\theta}_m) + \mu_i(\boldsymbol{\theta}_m)^2 \quad (26) \\
&= (w_i - w_i^2)y_i^2 + (w_i - w_i^2)\mu_i(\boldsymbol{\theta}_m)^2 - 2(w_i - w_i^2)y_i\mu_i(\boldsymbol{\theta}_m) \\
&= w_i(1 - w_i)[y_i - \mu_i(\boldsymbol{\theta}_m)]^2
\end{aligned}$$

The inequality becomes equality when $\boldsymbol{\theta} = \boldsymbol{\theta}_m$. Therefore,

$$\begin{aligned}
g(\boldsymbol{\theta}|\boldsymbol{\theta}_m) &= \sum_i [w_i y_i + (1 - w_i)\mu_i(\boldsymbol{\theta}_m) - \mu_i(\boldsymbol{\theta})]^2 + c_m \\
c_m &= \sum_i w_i(1 - w_i)[y_i - \mu_i(\boldsymbol{\theta}_m)]^2 \quad (27)
\end{aligned}$$

is indeed a majorization of $f(\boldsymbol{\theta})$.

Problem 4.20

Let $\theta = \sigma^2$, and $f(\theta, \mu) = -\frac{m}{2} \ln \theta - \frac{1}{2\theta} \sum_{i=1}^m (y_i - \mu_n)^2 - \rho|\mu_n|$. For μ fixed, the derivative of $f(\theta, \mu)$ with respect to θ is

$$\frac{df}{d\theta} = -\frac{m}{2} \frac{1}{\theta} + \frac{1}{2} \theta^{-2} \sum_{i=1}^m (y_i - \mu_n)^2. \quad (28)$$

Setting the derivative to 0 gives the update

$$\theta_{n+1} = \frac{1}{m} \sum_{i=1}^m (y_i - \mu_n)^2. \quad (29)$$

For θ fixed, the subgradient of $f(\theta, \mu)$ with respect to μ is

$$\frac{df}{d\mu} = \frac{1}{\theta_n} \sum_{i=1}^m y_i - \frac{1}{\theta_n} m\mu - \rho \begin{cases} 1 & \text{if } \mu > 0 \\ [-1, 1] & \text{if } \mu = 0 \\ -1 & \text{if } \mu < 0 \end{cases}. \quad (30)$$

Setting the derivative to 0, we get

$$\mu = \frac{1}{m} \sum_{i=1}^m y_i - \begin{cases} \frac{\rho\theta_n}{m} & \text{if } \mu > 0 \\ [-\frac{\rho\theta_n}{m}, \frac{\rho\theta_n}{m}] & \text{if } \mu = 0 \\ -\frac{\rho\theta_n}{m} & \text{if } \mu < 0 \end{cases}. \quad (31)$$

For $\mu > 0$, we have $\mu = \frac{1}{m} \sum_{i=1}^m y_i - \frac{\rho\theta_n}{m}$ when $\frac{1}{m} \sum_{i=1}^m y_i \in (\frac{\rho\theta_n}{m}, +\infty)$.

For $\mu = 0$, we have $0 \in \sum_{i=1}^m y_i - [-\frac{\rho\theta_n}{m}, \frac{\rho\theta_n}{m}]$ when $\frac{1}{m} \sum_{i=1}^m y_i \in [-\frac{\rho\theta_n}{m}, \frac{\rho\theta_n}{m}]$.

For $\mu < 0$, we have $\mu = \frac{1}{m} \sum_{i=1}^m y_i + \frac{\rho\theta_n}{m}$ when $\frac{1}{m} \sum_{i=1}^m y_i \in (-\infty, -\frac{\rho\theta_n}{m})$.

To summarize, the update for μ is

$$\mu_{n+1} = \begin{cases} 0 & \text{if } \sum_{i=1}^m y_i \in [-\frac{\rho\theta_n}{m}, \frac{\rho\theta_n}{m}] \\ \frac{1}{m} \sum_{i=1}^m y_i - \frac{\rho\theta_n}{m} & \text{if } \frac{1}{m} \sum_{i=1}^m y_i \in (\frac{\rho\theta_n}{m}, +\infty) \\ \frac{1}{m} \sum_{i=1}^m y_i + \frac{\rho\theta_n}{m} & \text{if } \frac{1}{m} \sum_{i=1}^m y_i \in (-\infty, -\frac{\rho\theta_n}{m}) \end{cases}. \quad (32)$$

Problem 4.21

Python script implementing the coordinate ascent algorithm is attached. 10,000 data points were drawn from the Gaussian distribution with mean 1.0, and variance 100.0. The simulated data has sample mean 0.878, and sample variance 99.538. μ and σ^2 were initialized to be 0. 10,000 iterations were applied in the coordinate ascent algorithm to ensure convergence. 6 different ρ 's ranging from 0 to 5 were tested. Table 1 shows the estimated parameters for different ρ . The magnitude of $\hat{\mu}$ shrinks as ρ increases. This is because the term $\rho|\mu|$ penalizes large $|\mu|$ in the objective.

```
import math
import numpy as np

# create simulated data
mu = 1.0
sigma2 = 100.0
m = 10000
data = np.random.normal(mu, math.sqrt(sigma2), m)
```

```

# compute sample mean, sample variance
sample_mean = np.mean(data)
sample_var = np.var(data)
print sample_mean, sample_var

# iterate through different rho
all_rho = [0.0, 1.0, 2.0, 3.0, 4.0, 5.0]
for rho in all_rho:

    # estimate mu and sigma2
    mu_est = 0.0
    sigma2_est = 0.0

    for i in range(0,10000):
        sigma2_est = np.sum(np.square(data-mu_est))/m
        tmp = rho*sigma2_est/m
        if(sample_mean >= -tmp and sample_mean <= tmp):
            mu_est = 0.0
        elif(sample_mean > tmp):
            mu_est = sample_mean - tmp
        else:
            mu_est = sample_mean + tmp

# print out result
print rho, mu_est, sigma2_est

```

ρ	$\hat{\mu}$	$\hat{\sigma}^2$
0.0	0.878	99.538
1.0	0.868	99.538
2.0	0.858	99.539
3.0	0.848	99.539
4.0	0.838	99.540
5.0	0.828	99.540

Table 1: Estimated μ and σ^2 for different ρ

Problem 4.25

Let

$$h(\mathbf{y}) = (f \square g)(\mathbf{y}) = \inf_{\mathbf{x}+\mathbf{z}=\mathbf{y}} [f(\mathbf{x}) + g(\mathbf{z})]. \quad (33)$$

Then

$$\begin{aligned}
(f \square g)^*(\mathbf{y}) &= h^*(\mathbf{y}) \\
&= \sup_{\mathbf{w}} [\mathbf{w}^* \mathbf{y} - h(\mathbf{w})] \\
&= \sup_{\mathbf{w}} \left\{ \mathbf{w}^* \mathbf{y} - \inf_{\mathbf{x} + \mathbf{z} = \mathbf{w}} [f(\mathbf{x}) + g(\mathbf{z})] \right\} \\
&= \sup_{\mathbf{w}} \left\{ \mathbf{w}^* \mathbf{y} + \sup_{\mathbf{x} + \mathbf{z} = \mathbf{w}} [-f(\mathbf{x}) - g(\mathbf{z})] \right\} \\
&= \sup_{\mathbf{w}} \left\{ \sup_{\mathbf{x} + \mathbf{z} = \mathbf{w}} [\mathbf{w}^* \mathbf{y} - f(\mathbf{x}) - g(\mathbf{z})] \right\} \\
&= \sup_{\mathbf{x}} [\mathbf{x}^* \mathbf{y} - f(\mathbf{x})] + \sup_{\mathbf{z}} [\mathbf{z}^* \mathbf{y} - g(\mathbf{z})] \\
&= f^*(\mathbf{y}) + g^*(\mathbf{y}).
\end{aligned} \tag{34}$$

The 6-th equality distributes the constraint $\mathbf{x} + \mathbf{z} = \mathbf{w}$ into finding the supremum of $\mathbf{x}^* \mathbf{y} - f(\mathbf{x})$ and $\mathbf{z}^* \mathbf{y} - g(\mathbf{z})$ separately. The sum of \mathbf{x}' and \mathbf{y}' at which the supremums are attained forms \mathbf{w}' , at which the supremum of $\mathbf{w}^* \mathbf{y} - h(\mathbf{w})$ is attained.

Problem 4.32

$$\partial f_1(x) = \begin{cases} \{0\} & \text{if } x \in [-1, 1] \\ \{1\} & \text{if } x \in (1, 2] \\ \{-1\} & \text{if } x \in [-2, -1) \\ [0, 1] & \text{if } x = 1 \\ [-1, 0] & \text{if } x = -1 \end{cases} \tag{35}$$

The function $y = \sqrt{1 - x^2}$ defines the top half of a circle with radius 1, as it can be written as $y^2 = 1 - x^2$ where $y \geq 0$. Therefore, $f_2(x)$ defines the bottom half of a circle with a radius 1 shifted upward by 1. At $x = 1$ and $x = -1$, the derivative of $f_2(x)$ with respect to x is ∞ and $-\infty$, respectively. The only supporting hyperplanes at these points are the vertical tangent lines. Therefore, $f_2(x)$ is not subdifferentiable at $x = 1$ and $x = -1$. For $x \in (-1, 1)$,

$$\frac{df_2(x)}{dx} = x(1 - x^2)^{-\frac{1}{2}} \tag{36}$$

Problem 4.33

Let $f(x) = |x|$.

When $x > 0$, $f(x) = x$, then $\partial f(x) = \{1\}$.

When $x < 0$, $f(x) = -x$, then $\partial f(x) = \{-1\}$.

When $x = -x = 0$,

$$\partial f(x) = \{\alpha(-1) + (1 - \alpha)(1) : \alpha \in [0, 1]\} = [-1, 1]. \tag{37}$$

Problem 4.34

Let

$$f(\mathbf{x}) = \|\mathbf{x}\|_1 = \max\{-x_1 - x_2, -x_1 + x_2, x_1 - x_2, x_1 + x_2\}. \quad (38)$$

If $x_1 > 0$ and $x_2 > 0$, $f(\mathbf{x}) = x_1 + x_2$, and $\partial f(\mathbf{x}) = \{(1, 1)\}$.

If $x_1 < 0$ and $x_2 < 0$, $f(\mathbf{x}) = -x_1 - x_2$, and $\partial f(\mathbf{x}) = \{(-1, -1)\}$.

If $x_1 < 0$ and $x_2 > 0$, $f(\mathbf{x}) = -x_1 + x_2$, and $\partial f(\mathbf{x}) = \{(-1, 1)\}$.

If $x_1 > 0$ and $x_2 < 0$, $f(\mathbf{x}) = x_1 - x_2$, and $\partial f(\mathbf{x}) = \{(1, -1)\}$.

These define the four corner points of the square on \mathbb{R}^2 .

If $x_1 = 0$ and $x_2 > 0$, for $\alpha \in [0, 1]$,

$$\partial f(\mathbf{x}) = \{\alpha(-1, 1) + (1 - \alpha)(1, 1) : \alpha \in [0, 1]\}. \quad (39)$$

Similarly, if $x_1 = 0$ and $x_2 < 0$, $\partial f(\mathbf{x}) = \{\alpha(-1, -1) + (1 - \alpha)(1, -1) : \alpha \in [0, 1]\}$.

If $x_1 > 0$ and $x_2 = 0$, $\partial f(\mathbf{x}) = \{\alpha(1, -1) + (1 - \alpha)(1, 1) : \alpha \in [0, 1]\}$.

If $x_1 < 0$ and $x_2 = 0$, $\partial f(\mathbf{x}) = \{\alpha(-1, -1) + (1 - \alpha)(-1, 1) : \alpha \in [0, 1]\}$.

These define the four edges of the square.

For $x_1 = x_2 = 0$,

$$\partial f(\mathbf{x}) = \left\{ \alpha_1(-1, -1) + \alpha_2(-1, 1) + \alpha_3(1, -1) + \alpha_4(1, 1) : \alpha_i \geq 0, \sum_i \alpha_i = 1 \right\}. \quad (40)$$

This defines the entire square.

At $\mathbf{x} = (0, 0)$, $\partial f(\mathbf{x}) = [-1, 1] \times [-1, 1]$, which is a square on \mathbb{R}^2 . At $\mathbf{x} = (1, 0)$, $\partial f(\mathbf{x}) = \{1\} \times [-1, 1]$, which is a line segment from $(1, -1)$ to $(1, 1)$. At $\mathbf{x} = (1, 1)$, $\partial f(\mathbf{x}) = \{1\} \times \{1\}$, which is the point $(1, 1)$.

Let

$$f(\mathbf{x}) = \|\mathbf{x}\|_\infty = \max\{-x_1 + 0x_2, x_1 + 0x_2, 0x_1 - x_2, 0x_1 + x_2\}. \quad (41)$$

Following the same reasoning in the previous problem, it can be shown that

$$\partial f(\mathbf{x}) = \begin{cases} \{(-1, 0)\} & \text{if } |x_1| > |x_2|, x_1 < 0, x_2 \neq 0 \\ \{(1, 0)\} & \text{if } |x_1| > |x_2|, x_1 > 0, x_2 \neq 0 \\ \{(0, -1)\} & \text{if } |x_1| < |x_2|, x_2 < 0, x_1 \neq 0 \\ \{(0, 1)\} & \text{if } |x_1| < |x_2|, x_2 > 0, x_1 \neq 0 \\ \{\alpha(-1, 0) + (1 - \alpha)(0, 1) : \alpha \in [0, 1]\} & \text{if } |x_1| = |x_2| \neq 0, x_1 < 0, x_2 > 0 \\ \{\alpha(-1, 0) + (1 - \alpha)(0, -1) : \alpha \in [0, 1]\} & \text{if } |x_1| = |x_2| \neq 0, x_1 < 0, x_2 < 0 \\ \{\alpha(1, 0) + (1 - \alpha)(0, 1) : \alpha \in [0, 1]\} & \text{if } |x_1| = |x_2| \neq 0, x_1 > 0, x_2 > 0 \\ \{\alpha(1, 0) + (1 - \alpha)(0, -1) : \alpha \in [0, 1]\} & \text{if } |x_1| = |x_2| \neq 0, x_1 > 0, x_2 < 0 \\ \{\alpha_1(-1, 0) + \alpha_2(1, 0) + \alpha_3(0, -1) + \alpha_4(0, 1) : \alpha_i \geq 0, \sum_i \alpha_i = 1\} & \text{if } x_1 = x_2 = 0 \end{cases}. \quad (42)$$

In summary, the first four points define the corner points of the diamond on \mathbb{R}^2 . The next four set define the four edges of the diamond. The last set is the convex hull of the

four corner points, and defines the entire diamond. At $\mathbf{x} = (0, 0)$,

$$\partial f(\mathbf{x}) = \left\{ \alpha_1(-1, 0) + \alpha_2(1, 0) + \alpha_3(0, -1) + \alpha_4(0, 1) : \alpha_i \geq 0, \sum_i \alpha_i = 1 \right\} \quad (43)$$

which is a diamond on \mathbb{R}^2 . At $\mathbf{x} = (1, 0)$, $\partial f(\mathbf{x}) = \{(1, 0)\}$, which is a corner point of the diamond. At $\mathbf{x} = (1, 1)$,

$$\partial f(\mathbf{x}) = \{\alpha(1, 0) + (1 - \alpha)(0, 1) : \alpha \in [0, 1]\}, \quad (44)$$

which is an edge of the diamond.