Homework 2

Frederick J. Boehm

9/21/2020

Exercise 1

i $\iff$ iii

we want to show that i $\iff$ iii.

We follow the hint.

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(\eta)(x - \eta)^2$$

We then see that $f''(\eta) \geq 0$ means that the last term in the right-hand side is non-negative. Thus, $f(x) \geq f(x_0) + f'(x_0)(x - x_0)$. For the reverse implication, reverse the above steps.

We thus see that i $\iff$ iii.

ii $\implies$ iii

We follow the hint.

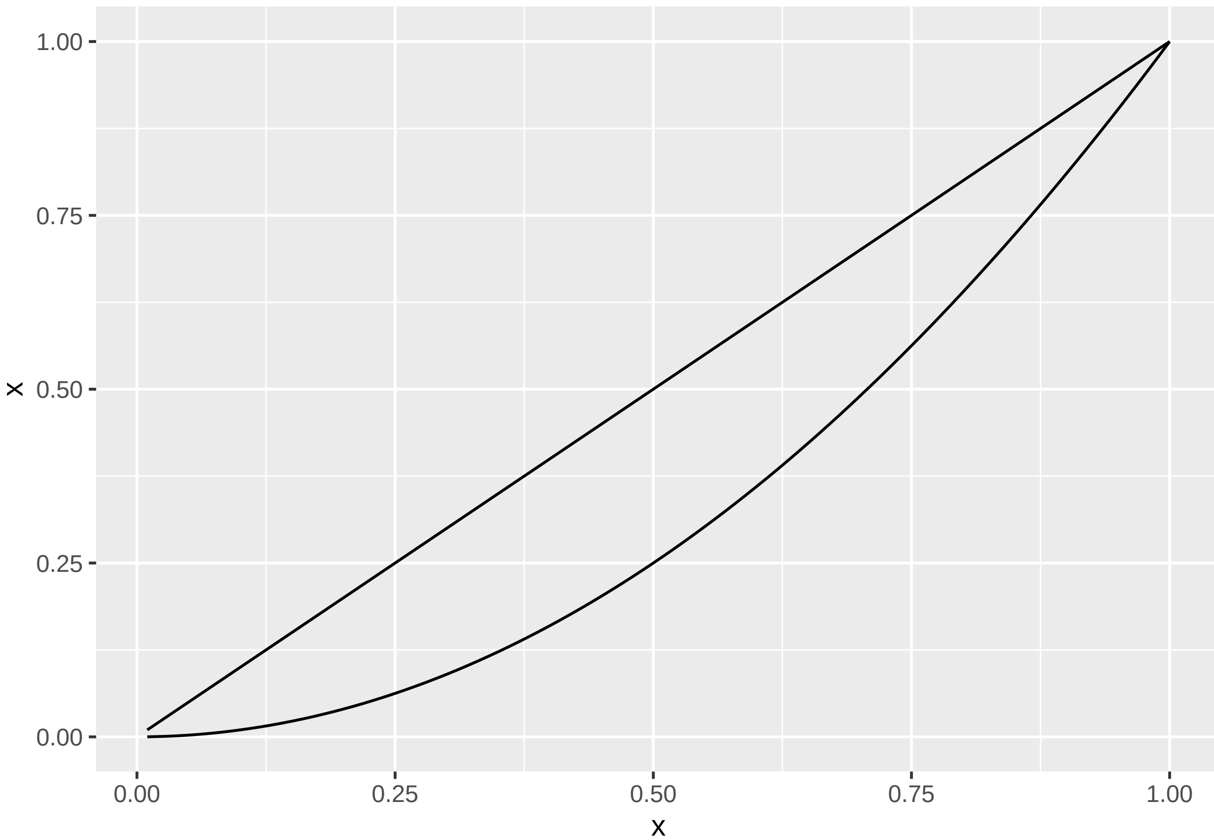
$$\begin{aligned}
f(x + h(x - x_0)) &= f((1 + h)x - hx_0) \\
&= f((1 - (-h))x + (-h)x_0) \\
&\leq -hf(x_0) + (1 + h)f(x) \\
&= f(x) + h[f(x) - f(x_0)] \\
&\implies f(x + h(x - x_0)) \leq h[f(x) - f(x_0)] \\
&\implies \frac{f(x + h(x - x_0))}{h} \leq f(x) - f(x_0) \\
&\implies f'(x_0)(x - x_0) \leq f(x) - f(x_0)
\end{aligned}$$

iii $\implies$ ii

Plot for ii

We let $f(x) = x^2$ with domain $[0, 1]$. We vary λ from zero to one below.

```
library(magrittr)
library(ggplot2)
tibble::tibble(x = 1:100 / 100, xsq = x ^ 2)  %>%
  ggplot() +
  geom_line(aes(y = x, x = x)) +
  geom_line(aes(y = xsq, x = x))
```



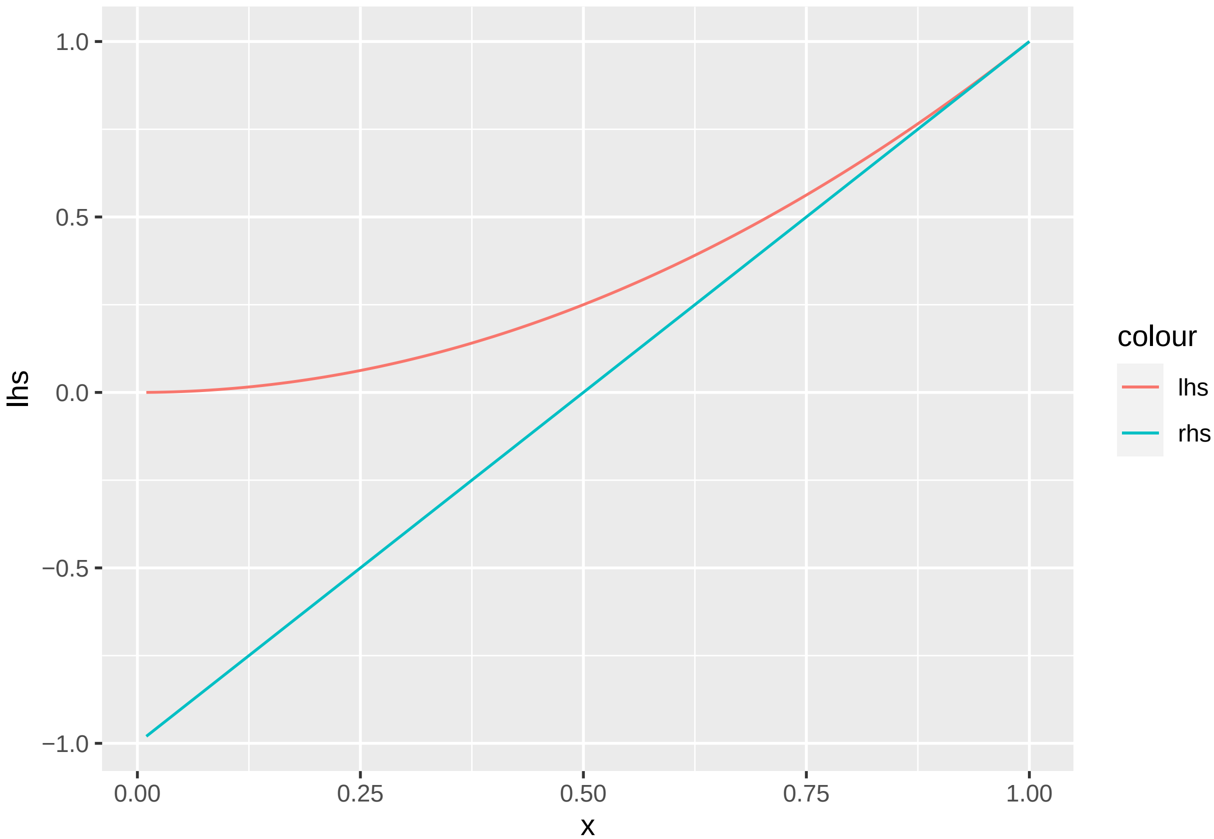
For x = 0, y = 1, we see that

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

Plot for iii

we set $x_0 = 1$ and vary x.

```
tibble::tibble(x = 1:100 / 100, lhs = x ^ 2, rhs = 1 ^ 2 + 2 * (x - 1) )  %>%
  ggplot() + geom_line(aes(y = lhs, x = x, colour = "lhs")) + geom_line(aes(y = rhs, x = x, colour = "r
```



We plot along the horizontal axis the values of x and along the vertical axis the values of $f(x_0)+f'(x_0)(x-x_0)$ and those of $f(x)$. "lhs" denotes $f(x)$, while "rhs" denotes $f(x_0)+f'(x_0)(x-x_0)$

Exercise 2

Discrete random variables

A discrete random variable X taking countably many values x_i has expectation $\mathbb{E}X = \sum_{i=1}^{\infty} p_i x_i$.

$$f(\mathbb{E}X) = f(\sum_i p_i x_i) \leq \sum_i p_i f(x_i) = \mathbb{E}\left(f(X)\right)$$

where we use induction on the natural numbers to argue that the inequality ii from question 1 extends to countably many summands.

General random variables

We follow the hint.

$$\begin{aligned}
f(X) &\geq f(\mathbb{E}X) + f'(\mathbb{E}X)\left([X - \mathbb{E}X]\right) \\
&\implies \mathbb{E}\left(f(X)\right) \geq f(\mathbb{E}X) + f'(\mathbb{E}X)\left(\mathbb{E}\left[X - \mathbb{E}X\right]\right) \\
&\implies \mathbb{E}\left(f(X)\right) \geq f(\mathbb{E}X)
\end{aligned}$$

Note that the last inequality follows from line 2 because $\mathbb{E}\left[X - \mathbb{E}X\right] = \mathbb{E}X - \mathbb{E}X = 0$.

Exercise 3

Q3, Part i

We recall the definition of f continuous: f continuous if for any $\epsilon > 0$ there is a $\delta > 0$ such that $d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \epsilon$. To see that f is continuous, let $\epsilon > 0$. Then, observe that $d_X(x, y) < \frac{\epsilon}{L} \implies d_Y(f(x), f(y)) < \epsilon$, so $\frac{\epsilon}{L}$ is the desired δ.

Q3, Part ii

I had to depart from the hint here. I started with the statement of the mean value theorem.

We note that f is continuous, since it is differentiable, so the mean value theorem applies. For some $c \in (x, y)$ we have

$$f'(c) = \frac{f(x) - f(y)}{||x - y||} \implies f(x) - f(y) = f'(c)[||x - y||]$$
$$\leq L||x - y||$$

Q3, Part iii

f is not differentiable because $\lim_{x \downarrow 0} f'(x) = 1$ while $\lim_{x \uparrow 0} f'(x) = -1$.

To show that f is 1-lipchitz on $\mathbb{R}$, consider the arbitrary interval (a, b). If $a \geq 0$ then $f(b) - f(a) = b - a$. If $b \leq 0$, then $f(b) - f(a) = -(b - a)$. If $a < 0$ and $b > 0$ then $f(b) - f(a) = |b - |a|| \leq |b - a|$.

Q3, Part iv

We note that the derivative is unbounded on the interval. This implies that there's a neighborhood near zero where $\frac{f(x)-f(y)}{x-y} > L$ for any finite L.

Q3, Part v

No. On $\mathbb{R}$, the slope, $f'(x) = 2x$ goes to infinity as $x \to \infty$.

To be precise, suppose that there is a constant C such that f is C-lipschitz. f C-lipschitz on $\mathbb{R}$ means that for any two points x and y, $|f(x) - f(y)| \leq C|x - y|$.

Choose $x = C$ and $y = C + 1$. Then $f(x) = C^2$ and $f(y) = C^2 + 2C + 1$, so $|f(x) - f(y)| = 2C + 1$, which is a contradiction. Thus, it must be that f is not L-lipschitz on $\mathbb{R}$.

Exercise 4

Q4, i

First, recall the definition of ϵ-net. A set $A_\epsilon \subset B$ is an ϵ-net for B if, for any point $x \in B$, $d(x, A_\epsilon) \leq \epsilon$.

We want to show that the algorithm in question creates an ϵ-net.

Suppose that we enumerate the elements of $T = \{x_1, \cdots, x_n\}$.

I claim that the set T, generated by the algorithm, is an ϵ-net for B.

Suppose T is not an ϵ-net for B. Then there is a point x_0 in B such that $d(x_0, B) \geq \epsilon$. This implies that $d(x_0, x_j) \geq \epsilon$, but that, in turn, implies that $x_0 \in B$, a contradiction. Thus, T is an ϵ-net for B.

Q4, ii

We want to show that

$$\cup_{x_i \in T} B_{\frac{\epsilon}{2}}(x_i) \subset B_{1+\frac{\epsilon}{2}}(0)$$

This is true because the elements of T are in $B = B_1(0)$, and thus, the $\frac{\epsilon}{2}$ balls centered at the elements of T are contained in the $1 + \frac{\epsilon}{2}$ ball centered at zero.

More precisely, choose a $y \in \cup_{x_i \in T} B_{\frac{\epsilon}{2}}(x_i)$. For some x_k, $y \in B_{\frac{\epsilon}{2}}(x_k)$. Note that the distance from 0 to y can be no more than $1 + \frac{\epsilon}{2}$, so $y \in B_{1+\frac{\epsilon}{2}}(0)$.

Now, we want to clarify why this statement implies that $N(B, \epsilon) \leq |T|$ and $|T| \leq \left(\frac{3}{\epsilon}\right)^d$.

We begin by verifying the first inequality. This follows directly from the definition of the algorithm. T is an ϵ-net, so, by definition, the covering number is no bigger than the cardinality of T.

Consider the case $d = 1$. Here, B is the interval $(-1, 1)$ on the real line. $\frac{3}{\epsilon}$ is the number of ϵ-balls needed when we permit overlap of balls, but prohibit any ball center from being in more than one ball.

Q4, iii

We first consider the case $d = 1$, where we see that we need at $\frac{1}{\epsilon}$ balls to ensure that every point in the 1-ball is within ϵ of an ϵ-ball. This is because the diameter of the ϵ-ball is 2ϵ, and the interval is length 2.

$$\frac{2}{2\epsilon} = \frac{1}{\epsilon}$$

We note that decreasing the number of ϵ-balls would not result in a ϵ-net.

Now, consider the case $d > 1$. We need to show that the $N \geq \left(\frac{1}{\epsilon}\right)^d$

We follow the approach given in Bartlett's notes (https://www.stat.berkeley.edu/~bartlett/courses/2013spring-stat210b/notes/12notes.pdf).

Namely, we observe that, for an ϵ-net of size N, we have

$$B \subset \cup_{i=1}^{N} (x_i + \epsilon B)$$

Consider, then, the volume of the ball.

$$\text{Volume}(B) \leq N \text{Volume}(\epsilon B) = N \epsilon^d \text{Volume}(B)$$

from which we see that

$$N \geq \frac{1}{\epsilon^d}$$

Q4, iv

To answer the question - I'm not sure because I can't recall if all norms on $\mathbb{R}^d$ arise as $||x||_p = (x_1^p + \ldots + x_d^p)^{\frac{1}{p}}$. It seems that for any natural number $p \in \mathbb{N}$, $||x||_p$ yields the results above. If it's true that all norms on $\mathbb{R}^d$ look like a p-norm for some p, then the answer to the question is "no" because all norms give the same result.

Exercise 5

Q5, i

We try to follow the hint. We have $\frac{1}{\epsilon}$ "starting points". Each "starting point" has two choices for slope, either positive or negative at every increment of ϵ, yielding $2^{\frac{1}{\epsilon}}$ "paths" for each starting point.

Adding the number of "paths", we get an upper bound of $\frac{1}{\epsilon}2^{\frac{1}{\epsilon}}$.

Q5, ii

Q5, iii

I drew on Bartlett's explanation (12notes.pdf, see above url) when attempting the solution below.

It's slightly trickier, but we can approach this by again using a grid. Instead of a square grid with equal increment sizes, we now have a grid with ϵ increments on the vertical axis and $\frac{\epsilon}{L}$ increments on the horizontal axis. We again have $\frac{1}{\epsilon}$ starting points, but, now, we have $2^{\frac{L}{\epsilon}}$ paths.

Exercise 6

Q6, i

Q6, ii

We first use Dudley's result.

$$\mathbb{E}\left[sup_{\theta\in\Theta}|Y_n^{(f)}|\right] \leq 12K\int_0^K \sqrt{\frac{\log N}{n}}d\epsilon \leq 12K\int_0^K \sqrt{\frac{d\log\frac{3}{\epsilon}}{n}}$$

Q6, iii

d and n balance each other. Linear increases in d can be offset by linear increases in n.