



(c) Now consider the case when  $\mathbf{x} \in \mathbb{R}^n$ . We assume that

$$g(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top \mathbf{Q}(\mathbf{x} - \mathbf{x}^*)$$

where  $\mathbf{Q}$  is a symmetric, positive-definite matrix. The Hessian,  $\mathbf{H}$ , of  $g(\mathbf{x})$  is a matrix with components

$$H_{ij} = \frac{\partial^2 g(\mathbf{x})}{\partial x_i \partial x_j}.$$

By writing out  $g(\mathbf{x})$  as a double sum over the components compute the Hessian  
[2 marks]


2

(d) Gradient descent in  $\mathbb{R}^n$  is given by

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - r \nabla g(\mathbf{x}).$$

Using the definition of  $g(\mathbf{x})$  write down a recursion relation between  $\mathbf{x}^{(t+1)}$  and  $\mathbf{x}^{(t)}$ .  
[2 marks]


2

(e) Defining  $\Delta^{(t)} = (\mathbf{x}^{(t)} - \mathbf{x}^*)$  obtain a recursion relation between  $\Delta^{(t+1)}$  and  $\Delta^{(t)}$ .  
(This is easy if you subtract  $\mathbf{x}^*$  from both sides of the recursion equation for  $\mathbf{x}^{(t+1)}$ .)  
[2 marks]


2

- (f) Using the eigenvalue decomposition  $\mathbf{Q} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$  and defining  $z^{(t)} = \mathbf{V}^T \Delta^t$  write out a recursion relation between  $z^{(t+1)}$  and  $z^{(t)}$ . (This is helped by multiplying the recursion relation on the left by  $\mathbf{V}^T$  and using the fact that  $\mathbf{V}$  is an orthogonal matrix.) [3 marks]

---

---

---

---

---

3

- (g) Solve the recursion relation to obtain a formula for  $x^{(t)}$  in terms of the initial state  $x^{(0)}$ . Express this formula for the  $i^{th}$  component of  $x^{(t)}$  and hence find a condition on the learning rate  $r$  to ensure convergence. [4 marks]

---

---

---

---

---

---

---

---

---

---

4

End of question 1

(a) $\frac{1}{2}$ (b) $\frac{1}{5}$ (c) $\frac{1}{2}$ (d) $\frac{1}{2}$ (e) $\frac{1}{2}$ (f) $\frac{1}{3}$ (g) $\frac{1}{4}$ Total $\frac{1}{20}$
--

$$f(x) = \frac{c}{2}(x - x^*)^2 + \frac{d}{6}(x - x^*)^3$$
$$x^{(t+1)} = x^{(t)} - \frac{f'(x^{(t)})}{f''(x^{(t)})}.$$

(To do this we need to expand a term with the structure

$$\frac{r \vdash s \in}{u \vdash v \in}.$$

$$\frac{1}{u+v\epsilon} = \frac{1}{u} \frac{1}{1+\frac{v}{u}\epsilon} = \frac{1}{u} \left( 1 - \frac{v}{u}\epsilon + \left(\frac{v}{u}\epsilon\right)^2 - \dots \right)$$

[5 marks]

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and extend across the width of the page. There is a vertical margin line on the left side, creating a narrow left margin. The paper appears to be from a notebook or a standard ruled document.

(b) Consider the function

$$h(x) = -x \log(x)$$

defined for  $0 < x \leq 1$ . By computing  $h'(x)$  and setting  $h'(x) = 0$  compute the value of  $x^*$  that maximises  $h(x)$ . [2 marks]


2
---

(c) Compute  $h''(x)$  and thus compute the Newton update function

$$n(x) = x - \frac{h'(x)}{h''(x)}$$

(the answer is rather surprising and in no way general).

[3 marks]

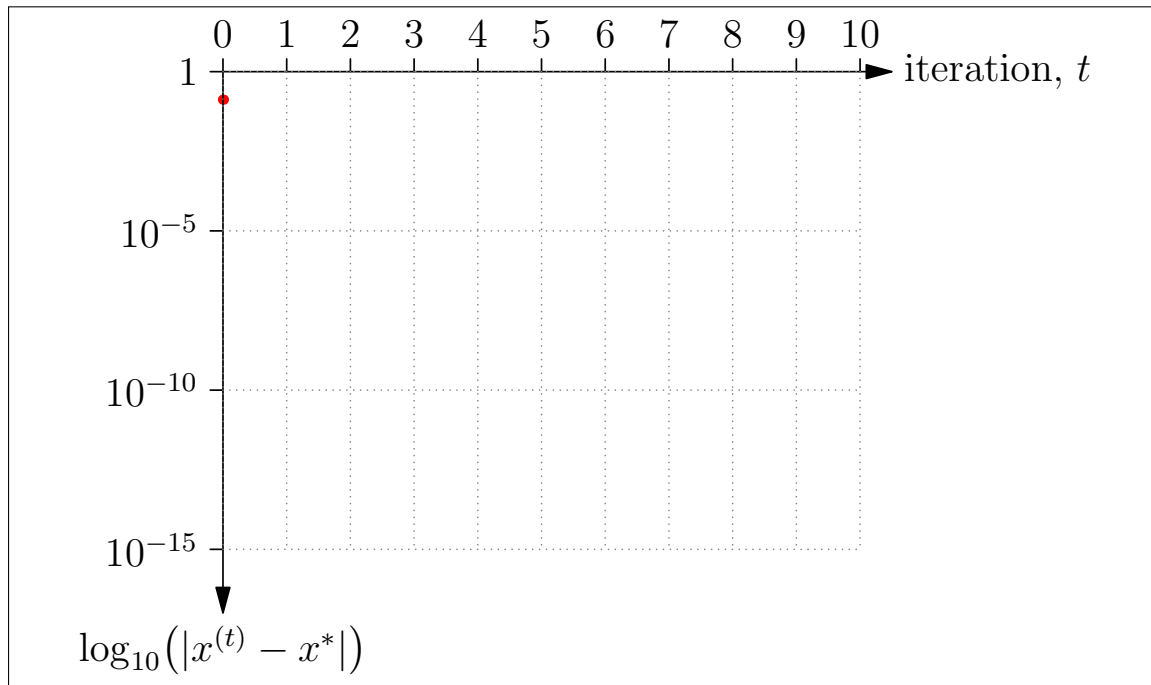

3
---

- (d) On the axes given below plot  $x^{(t)}$  for  $t = 0, 1, 2, \dots, 10$ , starting from  $x^{(0)} = 0.5$  where we use the gradient ascent updates

$$x^{(t+1)} = x^{(t)} + r h'(x^{(t)})$$

for  $r = \{0.3, 0.4, 0.5\}$  (that is you should plot three curves).

Also plot  $x^{(t)}$  where  $x^{(t+1)} = n(x^{(t)})$  (that is using Newton's update formula) for  $t = 0, 1, 2, 3$  and 4—note that to machine precision  $x^{(5)} = x^*$ . [10 marks]



10
----

End of question 2

(a) $\frac{5}{5}$	(b) $\frac{2}{2}$	(c) $\frac{3}{3}$	(d) $\frac{10}{10}$	Total $\frac{20}{20}$
-------------------	-------------------	-------------------	---------------------	-----------------------

**3**(a) Show that for  $p_i > 0$  the function

$$h(\mathbf{p}) = -\sum_i p_i \log(p_i)$$

is strongly convex-down. Hint: show that the Hessian matrix is negative-definite.  
[3 marks]

---

---

---

---

---

---

---

---

---

---

3

(b) Write down the Lagrangian,  $\mathcal{L}$ , for the problem of maximising  $h(\mathbf{p})$  subject to the constraints

$$\sum_i p_i = 1 \qquad \sum_i p_i E_i = U.$$

Then explain why there is a unique solution to this constrained optimisation problem.

[2 marks]

---

---

---

---

---

---

---

---

---

---

2

- (c) By setting  $\partial \mathcal{L} / \partial p_i = 0$  find the value of  $p_i$  that maximises  $\mathcal{L}$  in terms of  $E_i$  and the Lagrange multipliers. Use the constraint  $\sum_i p_i = 1$  to eliminate the Lagrange multiplier that enforces this constraint. [5 marks]

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and extend across the width of the page. There is a vertical margin line on the left side, creating a narrow left margin. The paper appears to be from a notebook or a standard ruled document.

15

End of question 3

(a)  $\frac{\quad}{3}$  (b)  $\frac{\quad}{2}$  (c)  $\frac{\quad}{5}$  Total  $\frac{\quad}{10}$