Chapter 5 Solutions

Question 1

Firstly, let's rewrite $f(x)$ as $f(x) = 4\log(x)\sin(x^3)$.

Then, $f'(x) = \frac{4}{x}\sin(x^3) + 12x^2\log(x)\cos(x^3)$.

Question 2

If we rewrite our function as $f(x) = (1+\exp(-x))^{-1}$, then we have $f'(x) = \frac{\exp(-x)}{(1+\exp(-x))^2}$.

Question 3

We have $f'(x) = \frac{\mu - x}{\sigma^2} f(x)$.

Question 4

We compute the first five derivatives of our function at 0. We have $f(0) = f'(0) = 1$, $f^{(2)}(0) = f^{(3)}(0) = -1$, and $f^{(4)}(0) = f^{(5)}(0) = 1$.

The Taylor polynomial $T_5(x) = 1 + x - \frac{1}{2}x^2 - \frac{1}{6}x^3 + \frac{1}{24}x^4 + \frac{1}{120}x^5$. The lower-order Taylor polynomials can be found by truncating this expression appropriately.

Question 5

Part a

We can see below that $\frac{\partial f_1}{\partial x}$ has dimension 1×2; $\frac{\partial f_2}{\partial x}$ has dimension $1 \times n$; and $\frac{\partial f_3}{\partial x}$ has dimension $n^2 \times n$.

Part b

We have $\frac{\partial f_1}{\partial x} = \begin{bmatrix} \cos(x_1)\cos(x_2) & -\sin(x_1)\sin(x_2) \end{bmatrix}$; $\frac{\partial f_2}{\partial x} = y^\top$. (Remember, y is a column vector!)

Part c

Note that $xx^\top$ is the matrix $\begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2^2 & \cdots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_n^2 \end{bmatrix}$. Thus its derivative will be a higher-order tensor.
However, if we consider the matrix to be an n^2-dimensional object in its own right, we can compute the Jacobian. Its first row consists of $\begin{bmatrix} 2x_1 & x_2 & \cdots & x_n \mid x_2 & 0 & \cdots & 0 \mid \cdots \mid x_n & 0 & \cdots & 0 \end{bmatrix}$, where I have inserted a vertical bar every n columns, to aid readability.

Question 6

We have $\frac{df}{dt} = \cos(\log(t^\top t)) \cdot \frac{1}{t^\top t} \cdot \begin{bmatrix} 2t_1 & 2t_2 & \cdots & 2t_D \end{bmatrix} = \cos(\log(t^\top t)) \cdot \frac{2t^\top}{t^\top t}$.

For g, if we explicitly compute AXB and find its trace, we have that $g(X) = \sum_{k=1}^{D} \sum_{j=1}^{F} \sum_{i=1}^{E} a_{ki} x_{ij} b_{jk}$.
Thus we have, $\frac{\partial g}{\partial x_{ij}} = \sum_{k=1}^{D} b_{jk} a_{ki}$, and this is the (i, j)-th entry of the required derivative. Hence $\frac{dg}{dX} = B^\top A^\top$.

Question 7

Part a

The chain rule tells us that $\frac{df}{dx} = \frac{df}{dz}\frac{dz}{dx}$, where $\frac{df}{dz}$ has dimension 1×1, and $\frac{dz}{dx}$ has dimension $1 \times D$. We know $\frac{dz}{dx} = 2x^\top$ from f in Question 6. Also, $\frac{df}{dz} = \frac{1}{1+z}$.

Therefore, $\frac{df}{dx} = \frac{2x^\top}{1+x^\top x}$.

Part b

Here we have $\frac{df}{dz}$ is an $E \times E$ matrix, namely $\begin{bmatrix} \cos z_1 & 0 & \cdots & 0 \\ 0 & \cos z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cos z_E \end{bmatrix}$.

Also, $\frac{dz}{dx}$ is an $E \times D$-dimensional matrix, namely A itself.

The overall derivative is obtained by multiplying these two matrices together, which will again give us an $E \times D$-dimensional matrix.

Question 8

Part a

We have $\frac{df}{dz}$ has dimension 1×1, and is simply $-\frac{1}{2}\exp(-\frac{1}{2}z)$.

Now, $\frac{dz}{dy}$ has dimension $1 \times D$, and is given by $y^\top(S^{-1} + (S^{-1})^\top)$.

Finally, $\frac{dy}{dx}$ has dimension $D \times D$, and is just the identity matrix.

Again, we multiply these all together to get our final derivative.

Part b

If we explicitly write out $xx^\top + \sigma^2 I$, and compute its trace, we find that $f(x) = x_1^2 + \cdots + x_n^2 + n\sigma^2$.

Hence, $\frac{df}{dx} = 2x^\top$.

Part c

Here, $\frac{df}{dz} = \begin{bmatrix} \frac{1}{\cosh^2 z_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\cosh^2 z_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\cosh^2 z_M} \end{bmatrix}$, while $\frac{dz}{dx} = A$, as in Question 7b.

Finally, $\frac{df}{dx}$ is given by the product of these two matrices.

Question 9

Piecing this together, replacing z with $t(\epsilon, \nu)$ throughout, we have that $g(\nu) = \log(p(x, t(\epsilon, \nu))) - \log(q(t(\epsilon, \nu), \nu))$.

Therefore, $\frac{dg}{d\nu} = \frac{p'(x,t(\epsilon,\nu)) \cdot t'(\epsilon,\nu)}{p(x,t(\epsilon,\nu))} - \frac{q'(t(\epsilon,\nu),\nu) \cdot t'(\epsilon,\nu)}{q(t(\epsilon,\nu),\nu)}$.