Review of Simple Matrix Derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $y = f(\mathbf{x}) = f(x_1, \ldots, x_n)$.

Definition: Gradient

The gradient vector, or simply the gradient, denoted ∇f, is a column vector containing the first-order partial derivatives of f:

$$\nabla f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{pmatrix}$$

Definition: Hessian

The Hessian matrix, or simply the Hessian, denoted $\mathbf{H}$, is an $n \times n$ matrix containing the second derivatives of f:

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 y}{\partial x_1^2} & \cdots & \frac{\partial^2 y}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 y}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 y}{\partial x_n^2} \end{pmatrix} = \nabla^2 f(\mathbf{x}) = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T}$$

Application: Differentiating Linear Form

$$\mathbf{x}^T\mathbf{b} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = b_1x_1 + \cdots + b_nx_n$$

Then,

$$\begin{aligned} \frac{\partial \mathbf{x}^T\mathbf{b}}{\partial \mathbf{x}} &= \begin{bmatrix} \frac{\partial \mathbf{x}^T\mathbf{b}}{\partial x_1} \\ \vdots \\ \frac{\partial \mathbf{x}^T\mathbf{b}}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial x_1}(b_1x_1 + \cdots + b_nx_n) \\ \vdots \\ \frac{\partial}{\partial x_n}(b_1x_1 + \cdots + b_nx_n) \end{bmatrix} \\ &= \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \\ &= \mathbf{b} \end{aligned}$$

Application: Differentiating Quadratic Form

$$
\begin{aligned}
\mathbf{x}^T \mathbf{A} \mathbf{x} &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} (a_{11}x_1 + \cdots + a_{n1}x_n) & \cdots & (a_{1n}x_1 + \cdots + a_{nn}x_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{n} a_{i1}x_i & \cdots & \sum_{i=1}^{n} a_{in}x_i \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= x_1 \sum_{i=1}^{n} a_{i1}x_i + \cdots + x_n \sum_{i=1}^{n} a_{in}x_i \\
&= \sum_{j=1}^{n} x_j \sum_{i=1}^{n} a_{ij}x_i \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij}x_i x_j
\end{aligned}
$$

Application: Differentiating Quadratic Form

$$\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial x_1} \\ \vdots \\ \frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial x_n} \end{bmatrix}$$

Consider the k^{th} row in the above vector:

$$\begin{aligned}
\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial x_k} &= \frac{\partial}{\partial x_k} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j \right) \\
&= \frac{\partial}{\partial x_k} \left(x_1 \sum_{i=1}^{n} a_{i1} x_i + \cdots + x_k \sum_{i=1}^{n} a_{ik} x_i + \cdots + x_n \sum_{i=1}^{n} a_{in} x_i \right) \\
&= x_1 a_{k1} + \cdots + \left(\sum_{i=1}^{n} a_{ik} x_i + x_k a_{kk} \right) + \cdots + x_n a_{kn} \\
&= \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i \\
&= (k^{\text{th}} \text{ row of } \mathbf{A}) \mathbf{x} + (\text{transpose of } k^{\text{th}} \text{ column of } \mathbf{A}) \mathbf{x} \\
&= \Big[(k^{\text{th}} \text{ row of } \mathbf{A}) + (\text{transpose of } k^{\text{th}} \text{ column of } \mathbf{A}) \Big] \mathbf{x}
\end{aligned}$$

Application: Differentiating Quadratic Form

Therefore,

$$
\begin{aligned}
\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= \begin{bmatrix} \left[(1^{\text{st}} \text{ row of } \mathbf{A}) + (\text{transpose of } 1^{\text{st}} \text{ column of } \mathbf{A})\right]\mathbf{x} \\ \vdots \\ \left[(n^{\text{th}} \text{ row of } \mathbf{A}) + (\text{transpose of } n^{\text{th}} \text{ column of } \mathbf{A})\right]\mathbf{x} \end{bmatrix} \\
&= \begin{bmatrix} \left[(1^{\text{st}} \text{ row of } \mathbf{A}) + (\text{transpose of } 1^{\text{st}} \text{ column of } \mathbf{A})\right] \\ \vdots \\ \left[(n^{\text{th}} \text{ row of } \mathbf{A}) + (\text{transpose of } n^{\text{th}} \text{ column of } \mathbf{A})\right] \end{bmatrix}\mathbf{x} \\
&= \left(\begin{bmatrix} (1^{\text{st}} \text{ row of } \mathbf{A}) \\ \vdots \\ (n^{\text{th}} \text{ row of } \mathbf{A}) \end{bmatrix} + \begin{bmatrix} (\text{transpose of } 1^{\text{st}} \text{ column of } \mathbf{A}) \\ \vdots \\ (\text{transpose of } n^{\text{th}} \text{ column of } \mathbf{A}) \end{bmatrix} \right)\mathbf{x} \\
&= \left(\mathbf{A} + \mathbf{A}^T\right)\mathbf{x}
\end{aligned}
$$

Application: Differentiating Quadratic Form

The following can be easily verified:

- If $\mathbf{A}$ is symmetric, then

$$\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}.$$

- Differentiating $\mathbf{x}^T\mathbf{A}\mathbf{x}$ w.r.t to x_k is equal to

$$\frac{\partial}{\partial x_k}\left(\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}x_ix_j\right) = \frac{\partial}{\partial x_k}\left(\sum_{i=1}^{n} a_{ik}x_ix_k + \sum_{j=1}^{n} a_{kj}x_kx_j\right)$$

- Differentiating a summation:

$$\begin{aligned}\frac{\partial}{\partial x_k}\left(\sum_{i=1}^{n} a_{ik}x_ix_k\right) &= \sum_{i=1}^{n}\frac{\partial}{\partial x_k}(a_{ik}x_ix_k)\\ &= \sum_{i=1}^{n} a_{ik}x_i\end{aligned}$$

Application: Taylor Expansion

k^{th} Order Taylor Expansion in $\mathbb{R}$

Suppose $f : \mathbb{R} \to \mathbb{R}$ is k times differentiable on $\mathbb{R}$. Then for any $x, \bar{x} \in \mathbb{R}$, there exists a $\hat{x}$ between x and $\bar{x}$ such that

$$
\begin{aligned}
f(x) = f(\bar{x}) &+ \frac{1}{1!}\frac{\partial f(\bar{x})}{\partial x}(x-\bar{x}) + \frac{1}{2!}\frac{\partial^2 f(\bar{x})}{\partial x^2}(x-\bar{x})^2 + \cdots + \frac{1}{(k-1)!}\frac{\partial^{k-1} f(\bar{x})}{\partial x^{k-1}}(x-\bar{x})^{k-1} \\
&+ \frac{1}{k!}\frac{\partial^k f(\hat{x})}{\partial x^k}(x-\bar{x})^k \\
= &\sum_{i=0}^{k-1} \frac{1}{i!}\frac{\partial^i f(\bar{x})}{\partial x^i}(x-\bar{x})^i + \frac{1}{k!}\frac{\partial^k f(\hat{x})}{\partial x^k}(x-\bar{x})^k.
\end{aligned}
$$

Things to note:

- $0! = 1$ and $\frac{\partial^0 f(\bar{x})}{\partial x^0} = f(\bar{x})$. Thus the first term in the summation (when $i = 0$) is $f(\bar{x})$.
- The first $(k-1)^{\text{th}}$ order derivative is evaluated at $\bar{x}$; whereas the k^{th} order derivative is evaluated at $\hat{x}$.

Application: Taylor Expansion

Second Order Taylor Expansion in $\mathbb{R}^n$

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable on $\mathbb{R}^n$. Then for any $\mathbf{x}, \bar{\mathbf{x}} \in \mathbb{R}^n$, there exists a $\hat{\mathbf{x}}$ between $\mathbf{x}$ and $\bar{\mathbf{x}}$,

$$
\begin{aligned}
f(\mathbf{x}) &= f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{H}(\hat{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) \\
&= f(\bar{\mathbf{x}}) + \sum_{i=1}^{n} \frac{\partial f(\bar{\mathbf{x}})}{\partial x_i}(x_i - \bar{x}_i) + \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f(\hat{\mathbf{x}})}{\partial x_i \partial x_j}(x_i - \bar{x}_i)(x_j - \bar{x}_j),
\end{aligned}
$$

where $\mathbf{H}(\hat{\mathbf{x}}) = \nabla^2 f(\hat{\mathbf{x}})$ is the Hessian evaluated at $\hat{\mathbf{x}}$.