Practice Gradients/Derivatives - Solutions

Matrix and Vector Calculus Proofs

Given: vectors $x, y \in \mathbb{R}^d$, matrices $M \in \mathbb{R}^{k\times d}$ and $A \in \mathbb{R}^{d\times d}$

1. Proof: $\nabla_x(y^T x) = \nabla_x(x^T y) = y^T$

Let $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$ and $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix}$

Then:

$$y^T x = x^T y = \sum_{i=1}^{d} y_i x_i$$

The gradient is:

$$\nabla_x(y^T x) = \begin{bmatrix} \frac{\partial}{\partial x_1}(y^T x) & \frac{\partial}{\partial x_2}(y^T x) & \cdots & \frac{\partial}{\partial x_d}(y^T x) \end{bmatrix}$$

Computing each partial derivative:

$$\frac{\partial}{\partial x_j}\left(\sum_{i=1}^{d} y_i x_i\right) = y_j \quad \text{for } j = 1, \ldots, d$$

Therefore:

$$\nabla_x(y^T x) = \begin{bmatrix} y_1 & y_2 & \cdots & y_d \end{bmatrix} = y^T$$

Since $y^T x = x^T y$, we also have:

$$\nabla_x(x^T y) = y^T$$

2. Proof: $\nabla_x(Mx) = M$

Let $M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1d} \\ m_{21} & m_{22} & \cdots & m_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k1} & m_{k2} & \cdots & m_{kd} \end{bmatrix}$

Then:

$$Mx = \begin{bmatrix} \sum_{j=1}^{d} m_{1j}x_j \\ \sum_{j=1}^{d} m_{2j}x_j \\ \vdots \\ \sum_{j=1}^{d} m_{kj}x_j \end{bmatrix}$$

The gradient $\nabla_x(Mx)$ is a $k \times d$ matrix where the (i,j)-th element is:

$$\frac{\partial}{\partial x_j}\left(\sum_{l=1}^{d} m_{il}x_l\right) = m_{ij}$$

Therefore:

$$\nabla_x(Mx) = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1d} \\ m_{21} & m_{22} & \cdots & m_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k1} & m_{k2} & \cdots & m_{kd} \end{bmatrix} = M$$

3. Proof: $\nabla_x(x^T Ax) = x^T(A^T + A)$

Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dd} \end{bmatrix}$

Then:

$$x^T Ax = \sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}x_ix_j$$

We compute the partial derivative with respect to x_k:

$$\frac{\partial}{\partial x_k}(x^T Ax) = \frac{\partial}{\partial x_k}\left(\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}x_ix_j\right)$$

Using the product rule and noting that terms are zero unless $i = k$ or $j = k$:

$$\begin{aligned} \frac{\partial}{\partial x_k}(x^T Ax) &= \sum_{j=1}^{d} a_{kj}x_j + \sum_{i=1}^{d} a_{ik}x_i \\ &= \sum_{j=1}^{d} a_{kj}x_j + \sum_{i=1}^{d} a_{ik}x_i \end{aligned}$$

The first sum is the k-th element of Ax, and the second sum is the k-th element of $A^T x$. Therefore:

$$\nabla_x(x^T Ax) = \begin{bmatrix} \frac{\partial}{\partial x_1}(x^T Ax) & \frac{\partial}{\partial x_2}(x^T Ax) & \cdots & \frac{\partial}{\partial x_d}(x^T Ax) \end{bmatrix} = (Ax)^T + (A^T x)^T$$

Since $(Ax)^T = x^T A^T$ and $(A^T x)^T = x^T A$, we have:

$$\nabla_x(x^T Ax) = x^T A^T + x^T A = x^T(A^T + A)$$

4. Proof: For symmetric A, $\nabla_x(x^T Ax) = 2(Ax)^T$

If A is symmetric, then $A^T = A$.

From the previous result:

$$\nabla_x(x^T Ax) = x^T(A^T + A) = x^T(A + A) = 2x^T A$$

Since A is symmetric:

$$2x^T A = 2(A^T x)^T = 2(Ax)^T$$

Therefore:

$$\nabla_x(x^T Ax) = 2(Ax)^T$$

Summary

We have proven all four derivative formulas:

1. $\nabla_x(y^T x) = \nabla_x(x^T y) = y^T$
2. $\nabla_x(Mx) = M$
3. $\nabla_x(x^T Ax) = x^T(A^T + A)$
4. For symmetric A, $\nabla_x(x^T Ax) = 2(Ax)^T$