Derivatives with respect to vectors

Let $x \in R^n$ (a column vector) and let $f : R^n \to R$. The derivative of f with respect to x is the row vector:

$$\frac{\partial f}{\partial x} = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})$$

$\frac{\partial f}{\partial x}$ is called the gradient of f.

The Hessian matrix is the square matrix of second partial derivatives of a scalar valued function f:

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1^2} \end{bmatrix} \tag{1}$$

The mixed derivatives of f are the entries off the main diagonal in the Hessian. Assuming that they are continuous, the order of differentiation does not matter.
If the gradient of f is zero at some point x, then f has a critical point at x. The determinant of the Hessian at x is then called the discriminant. If this determinant is zero then x is called a degenerate critical point of f. Otherwise it is non-degenerate.
For a non-degenerate critical point x, if the Hessian is positive definite at x, then f attains a local minimum at x. If the Hessian is negative definite at x, then f attains a local maximum at x. If the Hessian has both positive and negative eigenvalues then x is a saddle point for f (this is true even if x is degenerate). Otherwise the test is inconclusive.

Let $x \in R^n$ (a column vector) and let $f : R^n \to R^m$. The derivative of f with respect to x is the $m \times n$ matrix:

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f(x)_1}{\partial x_1} & \cdots & \frac{\partial f(x)_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial f(x)_m}{\partial x_1} & \cdots & \frac{\partial f(x)_m}{\partial x_n} \end{bmatrix} \tag{2}$$

$\frac{\partial f}{\partial x}$ is called the Jacobian matrix of f.

Examples:
Let $u, x \in R^n$ (column vectors).

1. The derivative of $u^T x = \sum_{i=1}^n u_i x_i$ with respect to x:

$$\frac{\partial \sum_{i=1}^n u_i x_i}{\partial x_i} = u_i \Rightarrow \frac{\partial u^T x}{\partial x} = (u_1, \ldots, u_n) = u^T \tag{3}$$

2. The derivative of $x^T x = \sum_{i=1}^n x_i$ with respect to x:

$$\frac{\partial \sum_{i=1}^n x_i^2}{\partial x_i} = 2x_i \Rightarrow \frac{\partial x^T x}{\partial x} = (2x_1, \ldots, 2x_n) = 2x^T \tag{4}$$

We will compute this derivative once again using the product rule: first holding x constant and then holding x^T constant.

$$\frac{\partial (f(x)^T g(x))}{\partial x} = \frac{\partial (f(x)^T g(\bar{x}))}{\partial x} + \frac{\partial (f(\bar{x})^T g(x))}{\partial x} \tag{5}$$

Placing a bar over a vector to indicate that it is being treated as constant, we have:

$$\frac{\partial x^T x}{\partial x} = \frac{\partial x^T \bar{x}}{\partial x} + \frac{\partial \bar{x}^T x}{\partial x} = x^T + x^T = 2x^T \tag{6}$$

3. Let $A \in R^{m\times n}$ and $x \in R^n$. Let $a_1^T, \ldots, a_n^T$ be the rows of A.

$$Ax = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix} \tag{7}$$

$$\frac{\partial Ax}{\partial x} = \begin{bmatrix} \frac{\partial a_1^T x}{\partial x} \\ \frac{\partial a_2^T x}{\partial x} \\ \vdots \\ \frac{\partial a_m^T x}{\partial x} \end{bmatrix} = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{bmatrix} \tag{8}$$

$$\Rightarrow \frac{\partial Ax}{\partial x} = A \tag{9}$$

4. Example: Let $A \in R^{m\times n}$ and $x \in R^n$.

$$\frac{\partial x^T A x}{\partial x} = \frac{\partial x^T A \bar{x}}{\partial x} + \frac{\partial \bar{x}^T A x}{\partial x} \tag{10}$$

To compute these derivatives we will use $\frac{\partial x^T u}{\partial x} = \frac{\partial u^T x}{\partial x} = u^T$ by substituting $u_1 = A\bar{x}$ and $u_2^T = \bar{x}^T A$.

$$\begin{aligned} \frac{\partial x^T A x}{\partial x} &= \frac{\partial x^T A \bar{x}}{\partial x} + \frac{\partial \bar{x}^T A x}{\partial x} = \\ \frac{\partial x^T u_1}{\partial x} + \frac{\partial u_2^T x}{\partial x} &= u_1^T + u_2^T = x^T A^T + x^T A = x^T(A + A^T) \end{aligned} \tag{11}$$

If A is symmetric then $A = A^T$ and $\frac{\partial x^T A x}{\partial x} = 2x^T A$. Taking the second derivative, we have:

$$\frac{\partial^2 x^T A x}{\partial x^2} = A + A^T \tag{12}$$

The chain rule: Let $U \subseteq R^n, V \subseteq R^n$ be open sets. Let $f : U \to V$ be differentiable in $a \in U$ and let $g : V \to R^p$ be differentiable in $b = f(a)$ then $g(f(x)) : U \to R^p$ is differentiable in a and $\frac{\partial g(f(a))}{\partial a} = \frac{\partial g(b)}{\partial b}\frac{\partial f(a)}{\partial a}$.
Example: We will compute again example 4, this time we will define $u = \bar{x}$ and use the chain rule:

$$\begin{aligned} \frac{\partial x^T A x}{\partial x} &= \frac{\partial x^T A \bar{x}}{\partial x} + \frac{\partial \bar{x}^T A x}{\partial x} = \\ &\frac{\partial (x^T A) u}{\partial x} + \frac{\partial u^T (Ax)}{\partial x} \end{aligned} \tag{13}$$

Now define $b_1 = A^T x$ and $b_2 = Ax$,

$$\frac{\partial (x^T A)u}{\partial x} + \frac{\partial u^T (Ax)}{\partial x} = \frac{\partial b_1^T u}{\partial b} \frac{\partial b_1}{\partial x} + \frac{\partial u^T b_2}{\partial b_2} \frac{\partial b_2}{\partial x} = \tag{14}$$
$$u^T A^T + u^T A = x^T (A + A^T)$$