1 Derivatives with scalars, vectors and matrices

When it comes to taking derivatives with scalars, vectors, and matrices we stick to the following conventions.

Scalars, matrices and vectors We will denote scalars, vectors and matrices as follows.

1. **Scalars** are denoted in the usual math font, e.g., $x \in \mathbb{R}$
2. **Vectors** are denoted with boldface lower case, e.g., $\mathbf{x} \in \mathbb{R}^n$. If we want to be explicit about the objects being column vectors we may write $\mathbf{x} \in \mathbb{R}^{n\times 1}$ (with n rows and 1 column), and for row vectors we can say $\mathbf{x} \in \mathbb{R}^{1\times n}$ (with 1 row and n columns). We rarely define a vector as a row vector but often write $\mathbf{x}^T$ to turn a column vector $\mathbf{x} \in \mathbb{R}^{n\times 1}$ into a row vector. The individual elements may be given by x_i with i the row index.
3. **Matrices** are denoted with boldface upper case, e.g., $\mathbf{X} \in \mathbb{R}^{m\times n}$. Such a matrix has m rows and n columns. The elements of the matrices may be denoted with X_{ij} with i the row index, and j the column index.

Derivatives The derivative of a function $f(x)$ with respect to input x will be denoted with either $\nabla_x f(x)$ or $\frac{df}{dx}(x)$. We then distinguish several cases for what kind of objects x and f are.

1. **Scalar input, scalar output:** $\nabla_x f(x) \in \mathbb{R}$ when $x \in \mathbb{R}$ and $f(x) \in \mathbb{R}$.
2. **Vector input, scalar output:** $\nabla_{\mathbf{x}} f(\mathbf{x}) \in \mathbb{R}^{1\times n}$ when $\mathbf{x} \in \mathbb{R}^n$ and $f(x) \in \mathbb{R}$. The result is thus a row vector.
3. **Vector input, vector output:** $\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}) \in \mathbb{R}^{m\times n}$ when $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{f}(x) \in \mathbb{R}^m$. The result is a matrix in which each row represents the derivative $\nabla_{\mathbf{x}} f_i(\mathbf{x})$ of the i-th output of $\mathbf{f}$. This matrix is known as the *Jacobian matrix*.
4. **Matrix input, scalar output:** $\nabla_{\mathbf{X}} f(\mathbf{X}) \in \mathbb{R}^{m\times n}$ when $\mathbf{X} \in \mathbb{R}^{m\times n}$ and $f(\mathbf{X}) \in \mathbb{R}$. The result is a matrix in which each element $[\nabla_{\mathbf{X}} f(\mathbf{X})]_{ij} = \frac{df}{dX_{ij}}(\mathbf{X})$ is the derivative of f to matrix element X_{ij}.

Other conventions In other books and papers you may encounter different conventions. E.g., you may encounter the derivative of a scalar function to a vector input to be a column vector (instead of what we define to be a row vector). This is all fine as long as the conventions are consistently applied. We encourage you to stick to our recommended convention, but if you decide to deviate from it please explicitly define your convention.

Always specify what kind of object you are working with! This allows you to easily verify if your computations are correct and it safes you a lot of mistakes. E.g., If you expect a column vector after your derivations and get a row vector, it indicates you probably got a transpose wrong somewhere.