# Derivative as a linear map

**Tangent space:** Let $x \in \mathbf{R}^n$ and consider displacement vectors from $x$. These displacements, usually denoted $\Delta x$, form a vector space called the tangent space. The tangent space is just another copy of $\mathbf{R}^n$ but with the origin shifted to $x$. The components of a displacement vector $\Delta x$ with respect to the standard basis are denoted $\Delta x_j$.

**Slope:** Given a function $f\colon \mathbf{R} \to \mathbf{R}$, suppose we can draw a tangent line to the graph of $f$ at a point $(x, f(x))$. This line does not necessarily go through the origin, but if we shift the origin to $(x, f(x))$ and think of the line as the graph of a function on the tangent space, it does. We get a covector — a linear map of $\Delta x$ and denote it $df$. In fact, $df(\Delta x) = m\,\Delta x$, where $m$ is the *slope* of the line.

**Osculatory approximation:** The graph of the tangent line is very close to the graph of $f$ near $x$. More precisely we can say that the difference between these graphs: $\Delta f - df(\Delta x)$, where $\Delta f = f(x+\Delta x) - f(x)$, becomes very small as $\Delta x \to 0$, faster than $\Delta x$ itself. Mathematically this means that as $\Delta x \to 0$

$$\frac{\Delta f - df(\Delta x)}{\Delta x} = \frac{f(x+\Delta x) - f(x) - m\,\Delta x}{\Delta x} \to 0.$$

This means that $m = \lim_{\Delta x \to 0} \dfrac{f(x+\Delta x) - f(x)}{\Delta x}$. Thus, the slope $m$ is the derivative of $f$ at $x$ and is denoted $f'(x)$.
Taking the special case of the indentity function $f(x) = x$ we obtain $df(\Delta x) = \Delta x$. In this case $df$ is denoted $dx$, which is none other than the indentity map of $\Delta x$. We may now rewrite $df(\Delta x) = m\,\Delta x = f'(x)\,\Delta x = f'(x)\,dx(\Delta x)$, so dropping the variable $\Delta x$, we obtain $df = f'(x)\,dx$.

The formula $df = f'(x)\,dx$ is the source of the alternate notation for the derivative $f'(x) = \dfrac{df}{dx}$.

**Linear map $df$ for vector variables:** If $f\colon \mathbf{R}^n \to \mathbf{R}^m$, we define $df$ to be the linear map of $\Delta x$ such that as $\Delta x \to 0$.

$$\frac{\Delta f - df(\Delta x)}{|\Delta x|} \to 0.$$

Note that this is a vector formula with the numerator in $\mathbf{R}^m$.

**Partial derivatives, the derivative matrix:** Let us take a special case $\Delta x = he_j$. Then $|\Delta x| = |he_j| = |h|$ and

$$\frac{\Delta f - df(he_j)}{|h|} \to 0, \quad \text{so} \quad \frac{f(x+he_j) - f(x) - h\,df(e_j)}{h} \to 0.$$

Therefore $df(e_j) = \lim_{h \to 0} \dfrac{f(x+he_j) - f(x)}{h}$, which the partial derivative of $f$ with respect to $x_j$ and is denoted $\dfrac{\partial f}{\partial x_j}$.
We see that $df$ is represented by an $m \times n$ matrix, called the derivative matrix, whose columns are partial derivatives of $f$.
If $\Delta x = \sum_{j=1}^{n} h_j e_j$, then $dx_i(\Delta x) = dx_i\left(\sum_{j=1}^{n} h_j e_j\right) = \sum_{j=1}^{n} h_j dx_i(e_j) = h_i$, so $\Delta x = \sum_{j=1}^{n} dx_j(\Delta x) e_j$. Therefore, $df(\Delta x) = df\left(\sum_{j=1}^{n} h_j e_j\right) = \sum_{j=1}^{n} h_j df(e_j) = \sum_{j=1}^{n} df(e_j)\,dx_j(\Delta x) = \sum_{j=1}^{n} \dfrac{\partial f}{\partial x_j}\,dx_j(\Delta x)$. Dropping the variable $\Delta x$ we get $df = \sum_{j=1}^{n} \dfrac{\partial f}{\partial x_j}\,dx_j$.
The derivative matrix is also known as the Jacobian matrix. In the special case $m = 1$, the $1 \times n$ derivative matrix may be thought of as a row vector of partial derivatives, known as the *gradient* and denoted $\operatorname{grad} f$ or $\nabla f$.

**Rules of differentiation:** The important rules are

(a) Constant: $d(c) = 0$
(b) Linearity: $d(f+g) = df + dg$
(c) Product: $d(f \cdot g) = df \cdot g + f \cdot dg$
(d) Chain: $d(f(g)) = f'(g)\,dg$