Let \( p(x_1, x_2, ..., x_d) \) be a point in \( \mathbb{R}^d \). We will often view it as a \( d \)-dimensional vector \([x_1, x_2, ..., x_d]\). As a convention, if it has been clear from the context that \( p \) is a point, then \( p \) represents this corresponding vector.

1 Curves and Tangent Vectors

**Curves.** Imagine that you move a point around in \( \mathbb{R}^d \). The locus of the point forms a curve. Intuitively, a curve is a 1d geometric object. Indeed, we can represent a curve using a vector function \( r(t) \):

\[
r(t) = [x_1(t), x_2(t), ..., x_d(t)]
\]

where \( t \) is a real value in a certain range, and each function \( x_i(t) \) (with \( i \in [1, d] \)) returns a real value. For each \( t \), \( (x_1(t), ..., x_d(t)) \) defines a point, and \( r(t) \) gives the corresponding vector.

For example, \( r(t) = [\cos t, \sin t] \) for \( t \in [0, 2\pi] \) defines a circle in \( \mathbb{R}^2 \), whereas \( r(t) = [\cos t, \sin t, t] \) for \( t \in [0, 2\pi] \) defines a circular helix in \( \mathbb{R}^3 \) as shown below:

As yet another example, given constant \( d \)-dimensional vectors \( p \) and \( q \) with \( q \neq 0 \), function \( r(t) = p + tq \) for \( t \in (-\infty, \infty) \) gives a line in \( \mathbb{R}^d \).

**Tangent Vectors.** We are ready to introduce:

**Definition 1.** Let \( r(t) \) be a curve, \( t_0 \) be a value of \( t \), and \( p \) be the point corresponding to \( r(t_0) \). If \( r(t) \) is differentiable at \( t_0 \), then the vector \( r'(t_0) \) is the **tangent vector** of the curve at \( p \).
The tangent vector has an intuitive geometric interpretation. Let $q$ be the point that corresponds to $f(t_0 + \Delta t)$; see the figure below. Let us focus on the direction of the directed segment $\overrightarrow{p, q}$. Now, imagine $q$ moving along the curve towards $p$ (namely, $\Delta t$ tends to 0). The direction of the directed segment gradually converges to the direction of the tangent vector at $p$.

![Tangent Vector Diagram]

We will refer to $u(t_0) = \frac{r'(t)}{|r'(t)|}$ as the unit tangent vector of the curve at $p$. Note that $|u(t_0)| = 1$.

As an example, consider the helix mentioned earlier: $r(t) = [\cos t, \sin t, t]$ for $t \in [0, 2\pi)$. Let $p$ be the point corresponding to $r(1)$. Then, the tangent vector of the curve at $p$ is $r'(1) = [-\sin(1), \cos(1), 1]$. The unit tangent vector at $p$ is therefore $[-\frac{\sin(1)}{\sqrt{2}}, \frac{\cos(1)}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$.

### 2 Gradient

Let $f(x_1, x_2, ..., x_d)$ be a scalar function of real-valued parameters $x_1, ..., x_d$. In other words, for each point $p(x_1, ..., x_d)$ of $\mathbb{R}^d$, $f(x_1, x_2, ..., x_d)$ returns a real value, if it is defined at $p$. For simplicity, sometimes we may write $f(x_1, x_2, ..., x_d)$ simply as $f(p)$. Next, we introduce a concept called gradient for such functions:

**Definition 2.** Let $f(x_1, ..., x_d)$ be a function defined as above. Consider a point $(t_1, t_2, ..., t_d)$ at which the partial derivative $\frac{\partial f}{\partial x_i}(t_1, ..., t_d)$ exists for all $i \in [1, d]$. Then, the **gradient** of $f(x_1, ..., x_d)$ at $(t_1, t_2, ..., t_d)$ is the vector:

$$\nabla f(t_1, ..., t_d) = \left[ \frac{\partial f}{\partial x_1}(t_1, ..., t_d), \frac{\partial f}{\partial x_2}(t_1, ..., t_d), ..., \frac{\partial f}{\partial x_d}(t_1, ..., t_d) \right].$$

For example, suppose that $f(x, y, z) = x^3 + 2xy + 3xz^2$. We know that $\frac{\partial f}{\partial x} = 3x^2 + 2y + 3z^2$, $\frac{\partial f}{\partial y} = 2x$, and $\frac{\partial f}{\partial z} = 6x$. Therefore,

$$\nabla f(t_1, t_2, t_3) = [3t_1^2 + 2t_2 + 3t_3^2, 2t_1, 6t_1].$$

We can as well just write the gradient as $\nabla f(x, y, z) = [3x^2 + 2y + 3z^2, 2x, 6x]$ by renaming the variables.
The gradient $\nabla f(t_1, ..., t_d)$ has an important geometric interpretation. Imagine that we are standing at the point $p(t_1, ..., t_d)$. Then the gradient points to the direction we should move in order to increase the value of function $f(x_1, ..., x_d)$ the most. Next, we will formalize the intuition.

Suppose that we decide to move from $p$ towards the direction of a unit vector $u$ by a distance $\Delta s$. Let $q$ be the point we will reach, as shown below:

We now prove an important lemma:

**Lemma 1.**

$$\lim_{\Delta s \to 0} \frac{f(q) - f(p)}{\Delta s} = \left( \nabla f(p) \right) \cdot u.$$  \hspace{1cm} (1)

**Proof.** Suppose that $u = [u_1, u_2, ..., u_d]$, and the coordinates of $p$ are $(t_1, t_2, ..., t_d)$.

Let $\ell$ be the line that passes $p$ and $q$. We know that we can represent any point on $\ell$ as $(x_1(s), x_2(s), ..., x_d(s))$, where for all $i \in [1, d]$:

$$x_i(s) = t_i + s \cdot u_i.$$  

In particular, if $s = 0$, the above representation gives $p$, whereas if $s = \Delta s$, the above representation gives $q$.

Define $g(s)$ to be the $f(x_1(s), ..., x_d(s))$. We can re-write the left hand side of (1) as:

$$\lim_{\Delta s \to 0} \frac{f(q) - f(p)}{\Delta s} = \lim_{\Delta s \to 0} \frac{g(\Delta s) - g(0)}{\Delta s} = (\text{by def. of derivative}) = g'(0).$$

On the other hand, applying the chain rule, we know:

$$g'(s) = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(x_1(s), ..., x_d(s)) \frac{dx_i}{ds}$$

$$= \left[ \frac{\partial f}{\partial x_1}(x_1(s), ..., x_d(s)), ..., \frac{\partial f}{\partial x_d}(x_1(s), ..., x_d(s)) \right] \cdot [x'_1(s), ..., x'_d(s)]$$

$$= (\nabla f(x_1(s), ..., x_d(s))) \cdot [u_1, ..., u_d]$$

$$= (\nabla f(x_1(s), ..., x_d(s))) \cdot u.$$  \hspace{1cm} \Box

Therefore, $g'(0) = (\nabla f(x_1(0), ..., x_d(0))) \cdot u = (\nabla f(p)) \cdot u$.

As a corollary of the above lemma, we obtain

$$\lim_{\Delta s \to 0} \frac{f(q) - f(p)}{\Delta s} = \left| \nabla f(p) \right| |u| \cos \gamma.$$  

where $\gamma$ is the angle between the directions of $\nabla f(p)$ and $u$. Hence, the limit is maximized if $\gamma = 0$, namely, $u$ has the same direction as $\nabla f(p)$.

It is worth mentioning that the limit on the left hand side of (1) is called the *directional derivative* in the direction of $u$, and is denoted as $D_u f$. Note that this is a function of $p$. In other words, $D_u f(p)$ gives the directional derivative in the direction of $u$ at point $p$. 