exists it is unique. It also tells us how to compute $T_a(y)$ for every $y \in \mathbb{R}^n$.

Theorem 1.8.1 Assume f is differentiable at a with total derivative T_a. Then the derivative $f'(a; y)$ exists for every y in $\mathbb{R}^n$ and we have

$$T_a(y) = f'(a; y) \qquad \cdots (1)$$

Moreover, $f'(a, y)$ is a linear combination of the components of y. In fact, if $y = (y_1, y_2, \cdots, y_n)$, we have

$$f'(a; y) = \sum_{k=1}^{n} D_k f(a)\, y_k . \qquad \cdots (2)$$

Proof: Equation (1) holds trivially if $y = 0$ since $T_a(0) = 0$ and $f'(a; 0) = 0$. Therefore we can assume that $y \neq 0$

Since f is differentiable at a we have a Taylor formula

$$f(a+v) = f(a) + T_a(v) + \|v\| E(a, v) \cdots (3)$$

for $\|v\| < r$ for some $r > 0$, where $E(a, v) \to 0$ as $\|v\| \to 0$. In this formula, we take $v = hy$, where $h \neq 0$ and $|h| \|y\| < r$. Then $\|v\| < r$.

Since T_a is linear, we have $T_a(v) = T_a(hy) = h T_a(y)$.

So, (3) gives us

$$\frac{f(a+hy) - f(a)}{h} = T_a(y) + \frac{|h| \|y\|}{h} E(a, v) \qquad \cdots (4)$$

Since $\|v\| \to 0$ as $h \to 0$ and since $|h|/h = \pm 1$, the right hand member of (4) tends to the limit $T_a(y)$ as $h \to 0$. So, the left hand member tends to the same limit. This proves (1)

Now we use the linearity of T_a to deduce (2)

If $y = (y_1, y_2, \ldots, y_n)$ we have $y = \sum_{k=1}^{n} y_k e_k$, hence

$$T_a(y) = T_a\left(\sum_{k=1}^{n} y_k e_k\right) = \sum_{k=1}^{n} y_k T_a(e_k)$$

$$= \sum_{k=1}^{n} y_k f'(a; e_k)$$

$$= \sum_{k=1}^{n} y_k D_k f(a)$$

1.9 The gradient of a function $f : S \to \mathbb{R}$, where $S \subseteq \mathbb{R}^n$

The formula in the Theorem 1.8.1, which expresses $f'(a; y)$ as a linear combination of the components of y, can be written as a dot product,

$$f'(a; y) = \sum_{k=1}^{n} D_k f(a)\, y_k = \nabla f(a) \cdot y$$

where $\nabla f(a)$ is the vector whose components are the partial derivatives of f at a, i.e.,

$$\nabla f(a) = (D_1 f(a), D_2 f(a), \ldots, D_n f(a))$$

This is called the gradient of f. The gradient ∇f is a vector field defined at each point a where the partial derivatives $D_1 f(a), D_2 f(a), \ldots D_n f(a)$ exist. We also write grad f for ∇f. The symbol ∇ is pronounced 'del'.

The first order Taylor formula (3) of Theorem 1.8.1 can now be written in the form

$$f(a+v) = f(a) + \nabla f(a) \cdot v + \|v\| E(a,v), \quad \ldots (5)$$

where $E(a,v) \to 0$ as $\|v\| \to 0$. In this form it resembles the one dimensional Taylor formula, with gradient $\nabla f(a)$ playing the role of the derivative $f'(a)$.

From the Taylor formula, we can easily prove that differentiability implies continuity

<u>Theorem 1.9.1</u> If $f: S \to \mathbb{R}$ be a function $S \subseteq \mathbb{R}^n$ such that f is differentiable at a, then f is continuous at a.

<u>Proof</u>: From (5), we have

$$|f(a+v) - f(a)| = |\nabla f(a) \cdot v + \|v\| E(a,v)|.$$

Applying the triangle inequality and the Cauchy-Schwarz inequality, we find that

$$0 \leq |f(a+v) - f(a)| \leq \|\nabla f(a)\| \|v\| + \|v\| |E(a,v)|.$$

This shows that $f(a+v) \to f(a)$ as $\|v\| \to 0$, so f is continuous at a

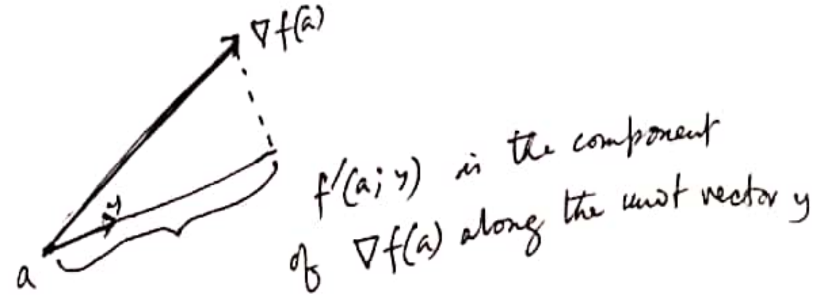

Figure 1

When y is a unit vector, the directional derivative $f'(a;y)$ has a simple geometric relation to the gradient vector. Assume $\nabla f(a) \neq 0$ and let θ denote the angle between y and $\nabla f(a)$.

Then we have

$$f'(a;y) = \nabla f(a) \cdot y = \|\nabla f(a)\| \, \|y\| \cos\theta = \|\nabla f(a)\| \cos\theta$$

This shows that ~~that~~ the ~~derivative~~ directional derivative is simply the component of the gradient vector in the direction of y. Figure 1 shows the vectors $\nabla f(a)$ and y attached to the point a. The derivative is largest when $\cos\theta = 1$, that is, when y has the same direction as $\nabla f(a)$.

In other words, at a given point a, the function f undergoes its maximum rate of change in the direction of the gradient vector; moreover, this maximum is equal to the length of the gradient vector. When $\nabla f(a)$ is orthogonal to y, the directional derivative $f'(a;y)$ is 0.