1 Recap

- Composition law for limits. $x^y = e^{y\ln(x)}$.
- Derivatives of scalar fields along vectors $\vec{v}$, partial derivatives.

2 Differentiability

Example: Let $f(x,y) = \frac{xy^2}{x^2+y^4}$ if $x \neq 0$ and $f(0,y) = 0$. A much stronger definition of differentiability is necessary. Just for the future, we observe that $f_x = \frac{y^6 - x^2y^2}{(x^2+y^4)^2}$, $f_y = \frac{2x^3y - 2xy^5}{(x^2+y^4)^2}$ when $x \neq 0$ and $f_x = f_y = 0$ when $x = 0$. However, when $(x,y) \to (0,0)$, we see that f_x is not continuous.

The reason differentiable 1-variable functions are continuous is the linear/tangent-line approximation $f(x+h) \approx f(x) + hf'(x)$. Note that $h \to hf'(x)$ is a linear map.
Here is another example: Consider $f(x,y) = ||x| - |y|| - |x| - |y|$. We can see that $\nabla f((0,0)) = 0$ exists. f is also continuous at $(0,0)$. Yet, if we sketch $z = f(x,y)$ we see that in no sense does a tangent plane exist at the origin. (If $y = x > 0$, $z = -2x \neq 0$. The "tangent vector" $(1,1,-2)$ is not in the same plane as $(1,0,0)$ and $(0,1,0)$.)
Even if we have all $\nabla_{\vec{v}} f$ and $\nabla_{\vec{v}} f$ is *linear* in $\vec{v}$, this can still fail: $f(x,y) = \frac{x^3y}{x^4+y^2}$ for $(x,y) \neq (0,0)$ and $f(0,0) = 0$ is continuous, has a linear $\nabla_{\vec{v}} f((0,0))$, and still does not have a tangent-plane.
We want to define differentiability to mean that the linear approximation holds, i.e., $f(\vec{a}+\vec{h}) \approx f(\vec{a}) + Df_{\vec{a}}\vec{h}$ where $Df_{\vec{a}}$ is a linear map. How should the $\approx$ be made rigorous? In one-variable calculus, for twice-differentiable functions one has the Taylor theorem: $f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(\theta)$. In other words, the "Error term" is of second-order. Thus, $\frac{f(x+h)-f(x)-hf'(x)}{h} \to 0$. (Even if f is merely differentiable, this limit still holds.)
Def: Let $f : S \subset \mathbb{R}^n \to \mathbb{R}$ be a scalar field and let $\vec{a} \in S$ be an interior point, i.e., there is an open ball $B(\vec{a}, r) \subset S$. f is said to be *differentiable* at $\vec{a}$ if there exists a linear map $Df_{\vec{a}} : \mathbb{R}^n \to \mathbb{R}$ such that for every $\|\vec{h}\| < r$, $\lim_{\vec{h}\to\vec{0}} \frac{f(\vec{a}+\vec{h})-f(\vec{a})-Df_{\vec{a}}(\vec{h})}{\|\vec{h}\|} = 0$. Alternatively, the error term is of the form $|\vec{h}|E(\vec{a},\vec{h}) = f(\vec{a}+\vec{h}) - f(\vec{a}) - Df_{\vec{a}}(\vec{h})$, where $E(\vec{a},\vec{h}) \to 0$ as $\vec{h} \to \vec{0}$. The linear map $Df_{\vec{a}}$ is called the *total derivative* of f at $\vec{a}$ or simply, the derivative map of f at $\vec{a}$.
Theorem: Let f be differentiable at the interior point $\vec{a}$ with total derivative $Df_{\vec{a}}$. Then $\nabla_{\vec{v}} f(\vec{a})$ exists for all $\vec{v}$ and $Df_{\vec{a}}(\vec{h}) = \nabla_{\vec{h}} f(\vec{a})$. Moreover, $\nabla_{\vec{v}} f(\vec{a}) = \langle \nabla f(\vec{a}), \vec{h} \rangle$, where ∇f is the gradient of f. As a consequence, $Df_{\vec{a}}$ is unique if it exists. In other words, for a differentiable function, $\nabla_{\vec{v}} f(\vec{a})$ is *linear* in $\vec{v}$, and hence one simply needs to know the finitely many numbers $\frac{\partial f}{\partial x_i}(\vec{a})$ to compute all (infinitely many) directional derivatives at the interior point $\vec{a}$. Proof: Since f is differentiable, $f(\vec{a}+\vec{h}) = f(\vec{a}) + Df_{\vec{a}}(\vec{h}) + \|\vec{h}\|E(\vec{a},\vec{h})$ where $E \to 0$ as $\vec{h} \to \vec{0}$. Take $\vec{h} = h\vec{v}$ for a small enough h. Then $\frac{f(\vec{a}+h\vec{v})-f(\vec{a})}{h} - Df_{\vec{a}}(\vec{v}) = \frac{\||h|\|\vec{v}\|\|E(\vec{a},h\vec{v})}{h}$. By the Sandwich law, the RHS goes to 0. Hence the limit of the LHS exists and equals 0. This means that $\nabla_{\vec{v}} f$ exists and equals $Df(\vec{v})$. Since Df is linear, $Df(\vec{v}) = \sum_i v_i Df(\vec{e}_i) = \sum_i v_i \nabla_{\vec{e}_i} f = \langle \nabla f, \vec{v} \rangle$. □

Theorem: If a scalar field f is differentiable at an interior point $\vec{a}$ it is continuous at $\vec{a}$.
Proof: $|f(\vec{a}+\vec{h}) - f(\vec{a})| = |\langle \nabla f, \vec{h}\rangle + \|\vec{h}\| E(\vec{a}, \vec{h})|$ which by the triange inequality is less than $|\langle \nabla f, \vec{h}\rangle| + \|\vec{h}\| |E(\vec{a}, \vec{h})|$. Now the Cauchy-Schwarz inequality implies that the first term is less than $\|\nabla f\| \|\vec{h}\|$. Therefore by limit laws, the RHS goes to 0. Hence by the Sandwich law so does the LHS. $\square$

The meaning of the gradient: Let $\|\vec{v}\| = 1$. The directional derivative $\nabla_{\vec{v}} f$ is $\langle \nabla f, \vec{v}\rangle$. Hence by Cauchy-Schwarz, $|\nabla_{\vec{v}} f| \leq \|\nabla f\|$ with equality if and only if $\vec{v} = \lambda \nabla f$. In other words, the direction of steepest *ascent* is ∇f and that of steepest *descent* is $-\nabla f$. For instance, if $z = 25 - (x^2 + y^2)$ is the height of a mountain, and if we are located at $(3, 4)$ then $\nabla f = (-6, -8)$ is the direction of steepest ascent.
This property leads to a nice algorithm in machine learning called "gradient descent" to minimise a function.