

# Multivariate Calculus

Muchang Bahng

November 6, 2022

Coordinate-dependent calculus of vector-valued functions. Note that the entire concept of calculus is dependent on functions that maps elements between metric spaces. To be more rigorous, we really just need a topology to define the following terms shown, but this course will assume that we are working with Euclidean spaces of  $\mathbb{R}^n$  with metric topologies. The metric in  $\mathbb{R}^n$  will be denoted  $\|\cdot\|$ , defined to be the  $L_2$  metric.

## 1 Differentiation

**Definition 1.1.** A sequence  $(x_k)$  of vectors in  $\mathbb{R}^n$  converges to the vector  $x$  if

$$\lim_{k \rightarrow \infty} \|x_k - x\| = 0$$

Note that this definition of convergence can be defined for any vector in a metric space  $(V, \|\cdot\|)$ . For example, the space of  $n \times n$  matrices with the operator norm or the Frobenius norm. Since this specific definition of convergence is limited in the way that the  $L_2$  norm is dependent on the coordinates of the vector in  $\mathbb{R}^n$ , the entire concept of coordinate-based vector calculus is also limited.

The concepts of continuity and derivatives are dependent on how the output of the function changes as we are changing the input. To measure this change in input, we must define a path function.

**Definition 1.2.** A *path function* is any function

$$p : \mathbb{R} \longrightarrow \mathbb{R}^n$$

Note that  $\text{Im}(p)$  traces out an oriented path in  $\mathbb{R}^n$ . Note also that the parameterization of  $p$  may be different even though the image of  $p$  may not change.

The defining of the path function allows us to construct arbitrary paths traveling through  $\mathbb{R}^n$ . Now, we can define continuity.

**Definition 1.3.** Let  $p : \mathbb{R} \longrightarrow \mathbb{R}^n$  be any path function such that  $p(t_0) = x_0$ , and let  $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$  be a vector-valued function defined on the open set  $U \subset \mathbb{R}^n$ .  $f$  is *continuous* at  $x_0 \in U$  if

$$\lim_{t \rightarrow t_0} \|(f \circ p)(t) - (f \circ p)(t_0)\| = \lim_{t \rightarrow t_0} \|(f \circ p)(t) - f(x_0)\| = 0 \quad (1)$$

This is equivalently written as

$$\lim_{x \rightarrow x_0} \|f(x) - f(x_0)\| = 0 \quad (2)$$

It is important to introduce equation (1) since this allows the reader to realize that the limits and the actual value of  $f$  at  $x_0$  should coincide no matter which path we choose in  $\mathbb{R}^n$ . While equation (2) is more concise, the expression  $x \rightarrow x_0$  does not clearly express the arbitrariness in our choice of path.

Note that the metric  $\|\cdot\|$  in (2) is the metric of  $\mathbb{R}^m$ , not  $\mathbb{R}^n$ .

### 1.1 Derivatives along Paths, Directional Derivatives

Differentiability must also be defined using paths since the change of  $x \in \mathbb{R}^n$  (an element in the domain of  $f$ ) can be modeled using a specific path function.

**Definition 1.4.** Let  $p : \mathbb{R} \rightarrow \mathbb{R}^n$  be a path function such that  $p(t_0) = x_0$ , and let there exist a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . We say that  $f$  is *differentiable* at  $x_0$  if, for every  $p$ , there exists a value  $f'_p(x_0)$  such that

$$\lim_{h \rightarrow 0} \left\| \frac{(f \circ p)(t_0 + h) - (f \circ p)(t_0)}{h} - f'_p(x_0) \right\| = 0$$

This is equivalent to saying that there exists a well-defined linear mapping  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  such that

$$\lim_{x \rightarrow x_0} \frac{\|f(x) - (f(x_0) + Tx)\|}{\|x - x_0\|} = 0$$

This says that there exists an affine linear approximation  $L(x) = f(x_0) + T(x)$  of  $f$  in the neighborhood  $U$  of  $x_0$ .

We now give a visual description of this linear approximation. Let us define the *graph of  $f$*  as the  $m$ -dimensional surface

$$G \equiv \{(x_1, x_2, \dots, x_n, f_1(x), \dots, f_m(x)) \mid (x_1, \dots, x_n) \in \mathbb{R}^n\} \subset \mathbb{R}^n \oplus \mathbb{R}^m$$

We can interpret  $\text{Im } L$  as the  $n$ -dimensional affine linear subspace embedded in  $\mathbb{R}^n \oplus \mathbb{R}^m$ , the extended space where the function is graphed in. That is,

$$\text{Im } L \equiv \{(x_1, \dots, x_n, f_1(x_0) + T_1(x), \dots, f_m(x_0) + T_m(x)) \mid (x_1, \dots, x_n) \in \mathbb{R}^n\}$$

In a way,  $\text{Im } L$  is the affine "tangent space" of  $G$  at point  $x_0$ . We now define the derivative of path functions.

**Definition 1.5.** Let  $p : \mathbb{R} \rightarrow \mathbb{R}^n$  be a path function with parameter  $t$ . Let the coordinate representation of  $p$  be defined

$$p \equiv (p_1, p_2, \dots, p_n)$$

Then, the derivative of  $p$  with respect to  $t$  is defined

$$p'(t) = (p'_1, p'_2, \dots, p'_n)$$

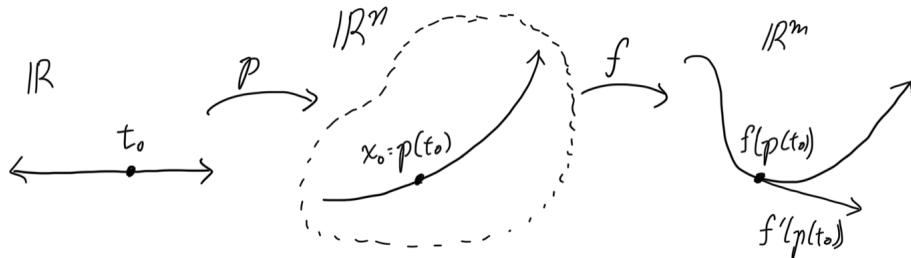
Visually, this outputs a tangent vector that represents the velocity of the particle traveling through  $\text{Im } p \subset \mathbb{R}^n$  as  $t$  travels through  $\mathbb{R}$ . The magnitude of the vector represents the speed of the particle, while the orientation of the vector represents the particle's direction.

The definition of the derivative of the path function now allows us to build on top of it the derivative of a general function.

**Definition 1.6.** Given that the derivative of  $f$  exists at  $x_0$ ,  $f'$  cannot be properly defined unless we specify which path function  $p$  (fully on the graph  $f$  and passing through  $x_0$ ) the particle is traveling through. Therefore, the derivative of the function  $f$  through a path  $p$  at  $x_0$  is defined

$$f'_p(x_0) \equiv \lim_{h \rightarrow 0} \frac{(f \circ p)(t_0 + h) - (f \circ p)(t_0)}{h}$$

We can describe the derivative visually, too.  $p$  defines a path function that describes a particle traveling through  $\mathbb{R}^n$ . Now, think of this entire path being mapped onto  $\mathbb{R}^m$  through  $f$ , which will draw another path in  $\mathbb{R}^m$ . In a way,  $f$  has "warped" the velocity of the particle. With this new velocity curve, the tangent vector of curve  $f(p(t))$  at the point  $t_0$  is the derivative of the  $f$  at point  $x_0$ . (The reader may also realize that we have just described the chain rule, too).



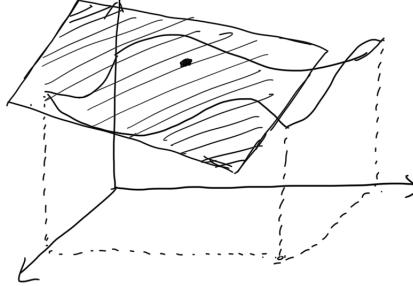
We can also visualize how the affine linear approximation is constructed from the tangent vectors of each path. Given the graph  $G$  of  $f$  in  $\mathbb{R}^n \oplus \mathbb{R}^m$ , the path function  $p$  determines a path on the surface  $G$  that passes through  $(x_0, f(x_0))$  (we are treating  $x_0$  as an  $n$ -tuple and  $f(x_0)$  as an  $m$ -tuple). This path on  $G$  is really just a bigger path function defined as

$$t \mapsto (p_1(t), p_2(t), \dots, p_n(t), f_1(p(t)), \dots, f_m(p(t)))$$

or more simply, let  $\tilde{p}$  be the mapping

$$t \mapsto (p(t), f(p(t)))$$

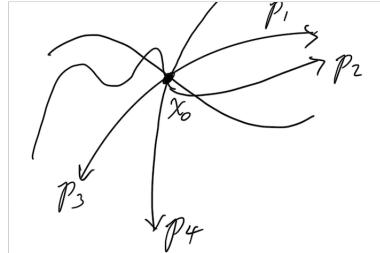
Now, for all paths  $\tilde{p}$  on  $G$  passing through  $(x_0, f(x_0))$ , the set of tangent vectors of every possible  $\tilde{p}$  from  $x_0$  is precisely the affine linear subspace  $\text{Im } L$ . In the  $2 + 1 = 3$  dimensional case, we have our familiar tangent plane at point  $(x, y, f(x, y))$  on the graph of the function  $f$  in  $\mathbb{R}^2 \oplus \mathbb{R}$ .



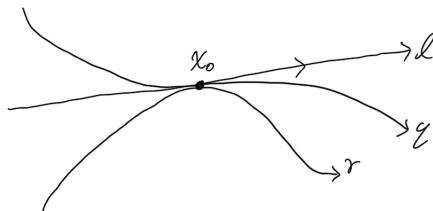
## 1.2 Differential Operators, Total Derivatives

### 1.2.1 Tangent Vectors, Spaces, and Bundles

**Definition 1.7** (Tangent Vectors, Spaces at a Point in  $\mathbb{R}^n$ ). We will construct the geometric tangent space of Euclidean space  $\mathbb{R}^n$ . Denote the set of all possible path functions  $p : \mathbb{R} \rightarrow \mathbb{R}^n$  that passes through  $a \in \mathbb{R}^n$  to be  $\mathcal{P}_a(\mathbb{R}^n)$ . We show some of these path functions in  $\mathcal{P}_a(\mathbb{R}^n)$  below (note that this is merely a set, so there is no operation defined on these paths, as in the case of the fundamental group of a topological space).



We can define a relation on this set: two path functions  $q, r \in \mathcal{P}_a(\mathbb{R}^n)$  are equivalent if the tangent vectors of  $q$  and  $r$  at  $a$  (including magnitude) are equal. Note that we have already defined the tangent vector to a path function above, so this move is completely valid.



That is, given  $q(t_q) = r(t_r) = a$ ,

$$q \sim r \iff q'(t_q) = r'(t_r)$$

which implies that given any smooth function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,

$$(f \circ q)'(a) = (f \circ r)'(a)$$

Note that within each equivalence class, there is exactly one straight-line path that goes through  $a$  with a given magnitude. This fact is extremely significant because this allows us to simplify the concept of derivatives defined on arbitrary paths to paths that are simpler in general. It isn't too hard to see that every arbitrary path  $p$  is equivalent to some line function with constant velocity. That is, given a path  $p$  such that  $p(t_0) = a$  and

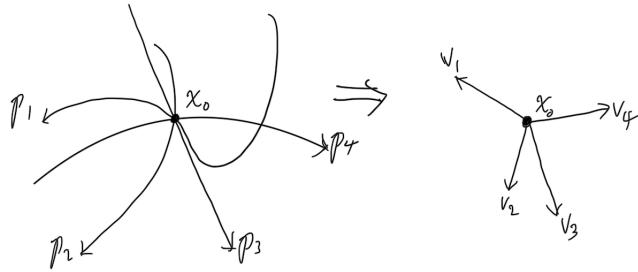
$$p'(t_0) = v \in \mathbb{R}^n$$

we can construct the unique straight-line path that is equivalent to  $p$  under  $\sim$  as

$$l(t) \equiv a + v(t - t_0)$$

and calculate the derivative of  $f$  under the path  $l$  to find  $f'$  under  $p$  at  $a$ . This "simplification" of  $\mathcal{P}_a(\mathbb{R}^n)$  to the quotient space  $\mathcal{P}_a(\mathbb{R}^n)/\sim$  of path functions that draw constant-velocity lines allows us to define directional derivatives. Note that paths that have the same image in  $\mathbb{R}^n$  could be in different equivalence classes under  $\sim$  if their velocities are different at the point  $x_0$ !

Now, since every path function in  $\mathcal{P}_a(\mathbb{R}^n)$  can be simplified to a straight line path represented by a vector protruding from  $a$ ,  $\mathcal{P}_a(\mathbb{R}^n)$  looks a lot like the vector space  $\mathbb{R}^n$ .



We can define vector addition, scalar multiplication, etc. to give this set the structure of a vector space. Therefore, this quotient space with this vector space structure is isomorphic to the geometric tangent space of  $f$  at point  $a$ , denoted  $T_a \mathbb{R}^n$ .

$$T_a \mathbb{R}^n \simeq \frac{\mathcal{P}_a(\mathbb{R}^n)}{\sim}$$

The vectors in here are called the *tangent vectors of  $T_a \mathbb{R}^n$* . So, if we would like to find the directional derivative of  $f$  the point  $a$  in direction of vector  $v$ , (given that  $l(t) \equiv a + v(t - t_0)$ ) we can calculate

$$\left. \frac{d}{dt} f(l(t)) \right|_{t=t_0} = \left. \frac{d}{dt} f(a + vt) \right|_{t=0} = \lim_{t \rightarrow 0} \frac{f(a + vt) - f(a)}{t}$$

Note that while

$$T_a \mathbb{R}^n \simeq T_b \mathbb{R}^n \simeq \mathbb{R}^n$$

for all  $a, b \in \mathbb{R}^n$ , these tangent spaces are not equivalent to each other.

**Theorem 1.1.** *If  $f'(t) = 0$  for all  $t \in \mathbb{R}^n$  and all paths  $p$ , then  $f$  is a constant function.*

**Definition 1.8** (Tangent Bundles in  $\mathbb{R}^n$ ). In  $\mathbb{R}^n$ , the *fiber bundle* of the tangent spaces  $T_x \mathbb{R}^n$  for all  $x \in \mathbb{R}^n$  is the disjoint union of them, called the *tangent bundle*.

$$T \mathbb{R}^n \equiv \bigsqcup_{a \in \mathbb{R}^n} T_a \mathbb{R}^n$$

It has a dimension of  $2n$ . In this chapter, this notation is used when we must identify a geometric tangent space  $T_x \mathbb{R}^n$  but the point  $x$  at which the tangent space is constructed is not specified. Therefore, we refer to *all* of the tangent spaces at once.

Another concept that will pop up is the vector field of  $\mathbb{R}^n$ . A vector field of  $\mathbb{R}^n$  is actually not a function  $V : \mathbb{R}^n \rightarrow \mathbb{R}^n$ .

**Definition 1.9** (Vector Field of Tangent Vectors on  $\mathbb{R}^n$ ). A vector field  $V$  on  $\mathbb{R}^n$  is a mapping

$$V : \mathbb{R}^n \rightarrow T \mathbb{R}^n$$

where  $V(x) \in T_x \mathbb{R}^n$ . We can interpret this visually since every vector is "attached" to a different point in  $\mathbb{R}^n$ , which represents its own tangent space.

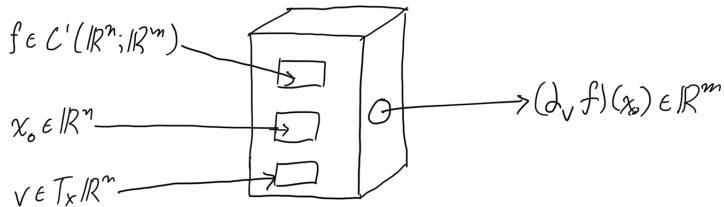
**Definition 1.10** (Differential Operator). Let us denote the set of all smoothly continuous functions mapping from  $\mathbb{R}^n$  to  $\mathbb{R}^m$  as  $C^1(\mathbb{R}^n; \mathbb{R}^m)$ . We have seen that in order to find the derivative of a function, we need

1. a smooth function  $f \in C^1(\mathbb{R}^n; \mathbb{R}^m)$
2. a point  $x_0 \in \mathbb{R}^n$ , where the derivative will be evaluated
3. a geometric tangent vector  $v \in T_{x_0}\mathbb{R}^n$  that represents the direction in which we are evaluating the derivative

In the most abstract sense, we can interpret the differentiation operator as a function  $d$  that takes in these three inputs and outputs a vector that represents the rate of change of  $f$  at  $x_0$  in direction  $v$ .

$$d : C^1(\mathbb{R}^n; \mathbb{R}^m) \times \mathbb{R}^n \times T\mathbb{R}^n \longrightarrow \mathbb{R}^m$$

where  $T\mathbb{R}^n_x$  is the fiber bundle of all geometric tangent spaces at  $\mathbb{R}^n$ .



Note that the differential operator, denoted  $d$ , is

1. linear with respect to the function argument

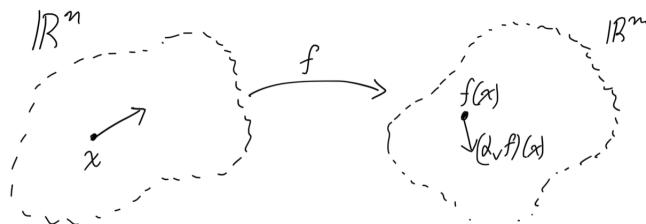
$$(d_v(f+g))(p) = d_v f(p) + d_v g(p), \quad (d_v(cf))(p) = c(d_v f)(p)$$

2. linear with respect to the directional vector argument

$$(d_{v+w}f)(p) = (d_v f)(p) + (d_w f)(p), \quad (d_{cv}f)(p) = c(d_v f)(p)$$

3. not linear with respect to the point argument

In the most abstract sense,  $d$  finds the derivative of any function  $f$  when the input, starting at point  $x$ , moves infinitesimally in the direction of vector  $v$  within the domain. The output is a vector that represents the direction the corresponding output vector travels from point  $f(x)$ .

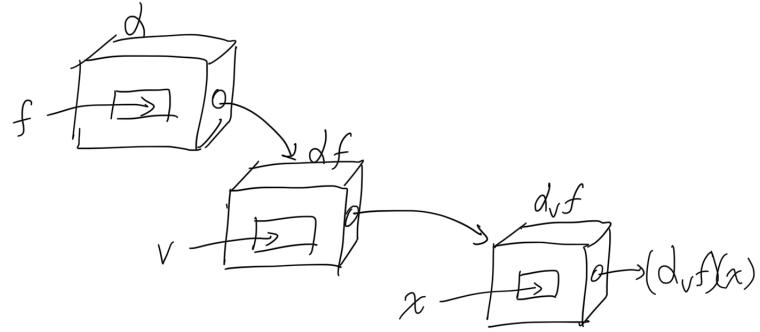


$df(x)$  is also called the *total derivative or differential* at  $x$ .

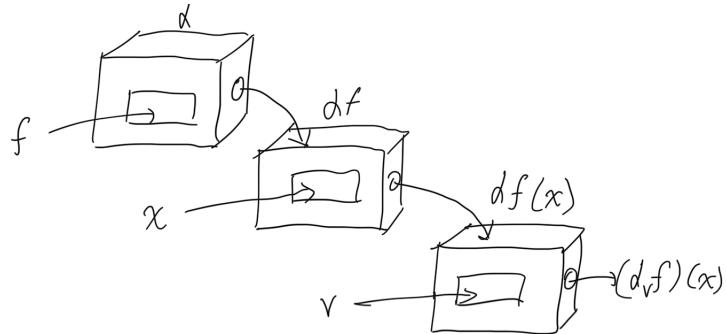
### 1.2.2 Differential Operator as Iterated Mappings

Rather than interpreting  $d$  as a mapping that takes in three inputs all at once to return a vector in  $\mathbb{R}^m$ , we can interpret it as a mapping that takes in some arguments and outputs another function that takes in the remaining arguments. We list some common interpretations, but note that usually, the function argument is given or is inputted first:

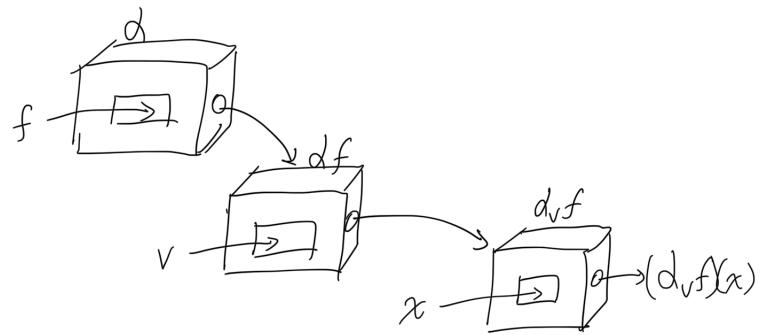
1.  $d$  takes in two inputs  $f \in C^1(\mathbb{R}^n; \mathbb{R}^m)$  and  $v \in T_x \mathbb{R}^n$  and outputs the vector field  $d_v f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . This vector field then takes in a point  $x \in \mathbb{R}^n$  and outputs the corresponding vector.



2.  $d$  takes in function  $f \in C^1(\mathbb{R}^n; \mathbb{R}^m)$  and outputs a function  $df$ .  $df$  then takes in a point  $x \in \mathbb{R}^n$  and outputs another function  $(df)(x)$ .  $(df)(x)$  finally takes in a vector  $v \in T_x \mathbb{R}^n$  (note that this does not have to be a fiber bundle, since  $x$  has been determined) and outputs the derivative vector in  $\mathbb{R}^m$ .



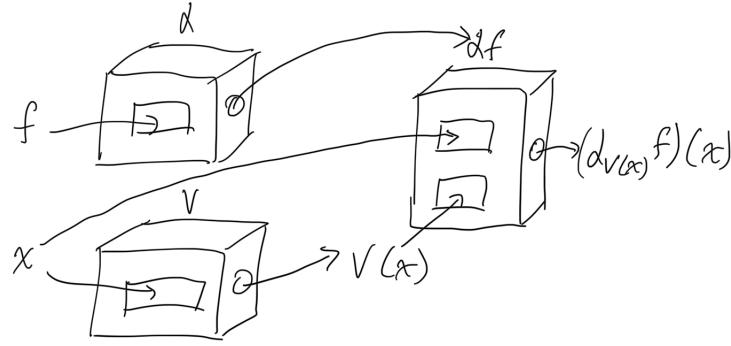
3.  $d$  takes in function  $f \in C^1(\mathbb{R}^n; \mathbb{R}^m)$  and outputs a function  $df$ .  $df$  then takes in a vector  $v \in T_x \mathbb{R}^n$  and outputs another function  $d_v f$ .  $d_v f$  finally takes in a point  $x \in \mathbb{R}^n$  and outputs the derivative vector in  $\mathbb{R}^m$ .



4. This one is quite different from the previous ones, albeit very powerful. We can define the differential operator on  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  as a function that takes in a vector field  $V$  of  $\mathbb{R}^n$  and outputs  $df(V)$ .  $df(V) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is another function that takes in point  $x \in \mathbb{R}^n$  and outputs a derivative vector according to the rule:

$$df(V)(x) \equiv (d_{V(x)} f)(x)$$

This can be visualized as such:



We can use this interpretation to define partial derivatives as

$$\frac{\partial f}{\partial v_i}(x) \equiv (d_{e_i} f)(x) \equiv df(V_{e_i})(x)$$

where  $V_{e_i}$  is the constant vector field that outputs  $e_i$ .

5.  $d$  takes in function  $f \in C^1(\mathbb{R}^n; \mathbb{R}^m)$  and point  $x$  and outputs a function  $df(x) : T_x \mathbb{R}^n \rightarrow \mathbb{R}^m$ . This interpretation will be used later when we view  $df(x)$  as an element of the cotangent space at  $x$ .

### 1.2.3 Cotangent Vectors, Spaces, and Bundles

We will now focus on real valued functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , since it is easier to define certain concepts this way without having to worry about abstract generalizations. Also, every function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  can be decomposed into its component functions  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  for  $i = 1, 2, \dots, n$  such that

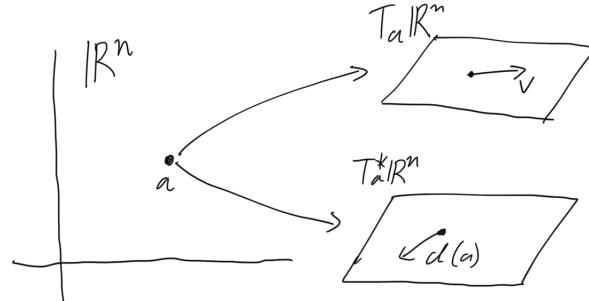
$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}$$

so we do not lose any data with this simplified interpretation. We will also denote  $C^1(\mathbb{R}^n; \mathbb{R})$  as just  $C^1(\mathbb{R}^n)$ .

**Definition 1.11** (Cotangent Vectors, Spaces in at a Point in  $\mathbb{R}^n$ ). The *cotangent space* of  $\mathbb{R}^n$  at point  $x$  is the dual of the tangent space  $T_x \mathbb{R}^n$ . It is denoted

$$T_x^* \mathbb{R}^n \equiv (T_x \mathbb{R}^n)^*$$

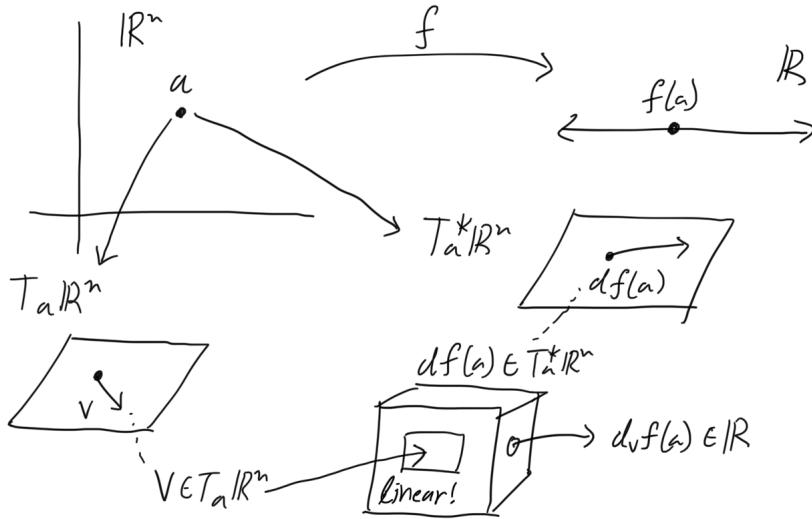
Its elements are *cotangent vectors*, which are functionals from  $T_x \mathbb{R}^n$  to  $\mathbb{R}$ . Therefore, for every point  $a \in \mathbb{R}^n$ , there exists a tangent space  $T_a \mathbb{R}^n$  and a cotangent space  $T_a^* \mathbb{R}^n$ .



**Definition 1.12** (Derivative at a Point as a Cotangent Vector). Let us view the total derivative  $d$  with interpretation 5. The total derivative of  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  at point  $a$  is a linear functional

$$df(a) : T_a \mathbb{R}^n \rightarrow \mathbb{R}$$

which, by definition, implies that it is a cotangent vector in  $T_a^* \mathbb{R}^n$ !



To generalize this to a field  $df$ , we must introduce the cotangent bundle.

**Definition 1.13** (Cotangent Bundles in  $\mathbb{R}^n$ ). The fiber bundle of cotangent spaces is called the *cotangent bundle*.

$$T^*\mathbb{R}^n \equiv \bigsqcup_{a \in \mathbb{R}^n} T_a^*\mathbb{R}^n$$

**Definition 1.14** (Covector Field of Total Derivatives on  $\mathbb{R}^n$ ). The total derivative of  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  can be interpreted as a covector field

$$df : \mathbb{R}^n \rightarrow T^*\mathbb{R}^n = \bigsqcup_{a \in \mathbb{R}^n} T_a^*\mathbb{R}^n$$

To summarize, we have the following analogous ideas:

$$\begin{aligned} \text{Tangent Vectors at a Point} &\iff \text{Cotangent Vectors/Total Derivatives at a Point} \\ \text{Tangent Spaces at a Point} &\iff \text{Cotangent Spaces at a Point} \\ \text{Vector Fields} &\iff \text{Covector Fields/Total Derivatives} \end{aligned}$$

### 1.3 Derivatives with Bases

This abstraction, although useful, doesn't help us actually calculate these derivatives, so let's step away from it for a moment. Fortunately, we are working in  $T_a \mathbb{R}^n \simeq \mathbb{R}^n$  which is already endowed with the structure of the standard orthonormal basis  $\{e_i\}_{i=1}^n$ . Since every vector  $v \in T_a \mathbb{R}^n$  can be represented as a linear combination of the basis vectors,

$$v = v_1 e_1 + v_2 e_2 + \dots + v_n e_n$$

by linearity we can find  $d_v$  for all  $v$  if we can find

$$d_{e_1}, d_{e_2}, \dots, d_{e_n}$$

**Definition 1.15** (Partial Derivatives). The differential operator  $d_v$  that has eaten the basis vector argument  $e_i$  of  $\mathbb{R}^n$  is called the *partial derivative*, denoted  $d_{e_i}$ .

$$d_{e_i} \equiv \frac{d}{dv_i}$$

Note that when referring to partial derivatives, the basis vector  $e_i$  is written in  $d_{e_i}$  while the coefficient  $v_i$  is written in  $d/dv_i$ . They are both referring to the same thing, since the former refers to the point following along the vector  $e_i$  at a rate of 1 while the latter refers to the coefficient of  $e_i$  increasing at a rate of 1. If  $d_{e_i}$  eats function  $f$ ,  $d_{e_i} f$  is called the *partial derivative of f*.

$$d_{e_i} f \equiv \frac{\partial f}{\partial v_i}$$

which can be solved using our familiar differentiation techniques from single variable calculus. If  $d_v$  eats both  $f$  and point  $x$ ,  $(d_v f)(x)$  is called the *partial derivative of  $f$  at  $x$* .

$$(d_{e_i} f)(x) \equiv \frac{d}{dt} f(x + e_i t) \Big|_{t=0} \equiv \frac{\partial f}{\partial v_i}(x)$$

### 1.3.1 Matrix Interpretation of Differentiation

We will derive the Jacobian in full generality for when  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

**Definition 1.16** (Jacobian Matrix). Given that  $v = v_1 e_1 + \dots + v_n e_n$ ,

$$\begin{aligned} d_v &= v_1 d_{e_1} + v_2 d_{e_2} + \dots + v_n d_{e_n} \\ &= v_1 \frac{\partial}{\partial v_1} + v_2 \frac{\partial}{\partial v_2} + \dots + v_n \frac{\partial}{\partial v_n} \end{aligned}$$

We can rewrite this equation in matrix form.

$$d_v = \begin{pmatrix} | & \cdots & | \\ d_{e_1} & \cdots & d_{e_n} \\ | & \cdots & | \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} | & \cdots & | \\ \frac{\partial}{\partial v_1} & \cdots & \frac{\partial}{\partial v_n} \\ | & \cdots & | \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

where  $v_1, \dots, v_n$  are the scalar coefficients of  $v$  and the matrix encodes all of the partial derivative operators within itself. If we plug in an arbitrary function  $f$  to  $d_v$ , we would get

$$d_v f = \begin{pmatrix} | & \cdots & | \\ d_{e_1} f & \cdots & d_{e_n} f \\ | & \cdots & | \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} | & \cdots & | \\ \frac{\partial f}{\partial v_1} & \cdots & \frac{\partial f}{\partial v_n} \\ | & \cdots & | \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

In order to decompose this even further to a matrix form, we can also decompose the function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  into its component functions  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  for  $i = 1, 2, \dots, m$ . So we would have

$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix} \implies d_{e_i} f = \begin{pmatrix} d_{e_i} f_1 \\ d_{e_i} f_2 \\ \vdots \\ d_{e_i} f_m \end{pmatrix}$$

and therefore, we can rewrite the above equation of  $d_v f$  into

$$d_v f = \begin{pmatrix} \frac{\partial f_1}{\partial v_1} & \frac{\partial f_1}{\partial v_2} & \frac{\partial f_1}{\partial v_3} & \cdots & \frac{\partial f_1}{\partial v_n} \\ \frac{\partial f_2}{\partial v_1} & \frac{\partial f_2}{\partial v_2} & \frac{\partial f_2}{\partial v_3} & \cdots & \frac{\partial f_2}{\partial v_n} \\ \frac{\partial f_3}{\partial v_1} & \frac{\partial f_3}{\partial v_2} & \frac{\partial f_3}{\partial v_3} & \cdots & \frac{\partial f_3}{\partial v_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial v_1} & \frac{\partial f_m}{\partial v_2} & \frac{\partial f_m}{\partial v_3} & \cdots & \frac{\partial f_m}{\partial v_n} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{pmatrix}$$

where the matrix is called the *Jacobian matrix*. Notice how this matrix equation is consistent with the abstract mapping with three inputs.

1. the function  $f$  is needed to evaluate all the derivatives as elements of the matrix
2. the point  $x$  is needed to evaluate all  $n \times m$  expressions  $\partial f_i / \partial v_j$ . That is,

$$d_v f(x) = \begin{pmatrix} \frac{\partial f_1}{\partial v_1} & \cdots & \frac{\partial f_1}{\partial v_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial v_1} & \cdots & \frac{\partial f_m}{\partial v_n} \end{pmatrix}(x) = \begin{pmatrix} \frac{\partial f_1}{\partial v_1}(x) & \cdots & \frac{\partial f_1}{\partial v_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial v_1}(x) & \cdots & \frac{\partial f_m}{\partial v_n}(x) \end{pmatrix}$$

3. the directional vector  $v$  is needed to evaluate the vector  $(v_1 \dots v_n)^T$ .

This matrix allows us to calculate all directional derivatives at every point in all differentiable functions  $f$ . The entries of  $d_v f$  are the "building blocks" that generate directional derivatives at the point  $x$ .

**Theorem 1.2** (Chain Rule). *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $g : \mathbb{R}^p \rightarrow \mathbb{R}^n$  be two functions such that  $f \circ g : \mathbb{R}^p \rightarrow \mathbb{R}^m$  is defined. Suppose  $g$  is differentiable at  $x_0 \in \mathbb{R}^p$  and  $f$  is differentiable at  $y_0 = g(x_0) \in \mathbb{R}^n$ . Then  $f \circ g$  is differentiable at  $x_0$  and*

$$D(f \circ g)(x_0) = Df(y_0) \cdot Dg(x_0)$$

where the right hand side is the matrix product of real  $m \times n$  matrix  $Df(y_0)$  and  $n \times p$  matrix  $Dg(x_0)$ .

Therefore, given the composition of function  $f \circ g$ , we have two methods of finding the derivative matrix of  $f \circ g$  at point  $x_0$ . First is to explicitly compute  $f \circ g$  and find its  $m \times p$  derivative matrix  $D(f \circ g)$ , and plug in  $x_0$  to get  $D(f \circ g)(x_0)$ .

The second way is to use the chain rule to find the individual derivative matrices  $Df(g(x_0)) = Df(y_0)$  and  $Dg(x_0)$  and multiply them together to get the derivative matrix  $D(f \circ g)(x_0)$ .

**Theorem 1.3** (Product, Quotient Rules). *Given that  $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$  are differentiable at  $x_0$ . Then given that  $h(x) \equiv f(x)g(x)$  for all  $x$ ,*

$$Dh(x_0) = g(x_0)Df(x_0) + f(x_0)Dg(x_0)$$

Additionally, given that  $g$  never vanishes and letting  $k(x) \equiv f(x)/g(x)$  for all  $x$ ,

$$Dk(x_0) = \frac{g(x_0)Df(x_0) - f(x_0)Dg(x_0)}{(g(x_0))^2}$$

## 1.4 Del Operators, Gradients

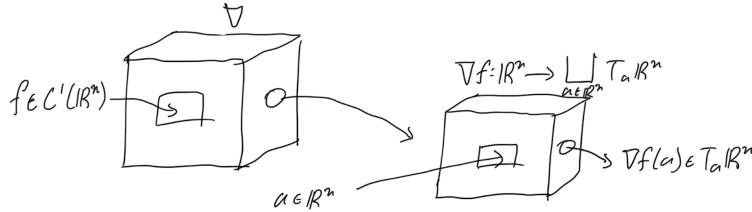
We have successfully defined the total derivative of function  $f$ , denoted  $df$ , as a covector field

$$df : \mathbb{R}^n \rightarrow T^*\mathbb{R}^n$$

on  $\mathbb{R}^n$ , where  $df(a) \in T_a\mathbb{R}^n$ . We now define the gradient, which is the vector field mapping  $\mathbb{R}^n \rightarrow T\mathbb{R}^n$  on  $\mathbb{R}^n$ .

**Definition 1.17** (Del Operator). The *del* operator takes in a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and outputs a vector field. In Cartesian coordinates, it is defined

$$\nabla \equiv \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix}, \quad \nabla f \equiv \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$



**Definition 1.18** (Gradient). The gradient of a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , denoted  $\nabla f$  or  $\text{grad } f$ , is just the del operator done on  $f$ . It is a vector field on  $\mathbb{R}^n$

$$\nabla f : \mathbb{R}^n \rightarrow T\mathbb{R}^n = \bigsqcup_{a \in \mathbb{R}^n} T_a\mathbb{R}^n$$

satisfying one property. The gradient is the unique vector field whose dot product with any vector  $v$  at each point  $a$  is the directional derivative of  $f$  at  $p$  along  $v$ . That is,  $\nabla f$  is the vector field satisfying

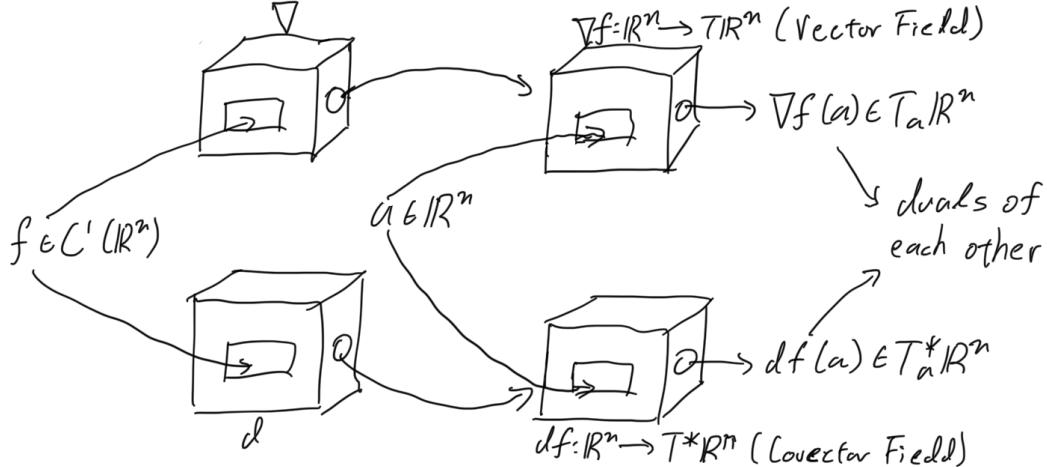
$$\nabla f(a) \cdot v = df_v(a) \text{ for all } a \in \mathbb{R}^n$$

where  $\cdot$  is the dot product of  $T_a\mathbb{R}^n$ .

Note that  $\nabla f$  is a vector field, while  $df$  is a covector field! However, for visual purposes, we can think of the value of the gradient as a vector in the original space  $\mathbb{R}^n$ , while the value of the derivative at a point can be thought of as a covector on the original space.

$$\begin{aligned}\nabla f : \mathbb{R}^n &\longrightarrow T\mathbb{R}^n = \bigsqcup_{a \in \mathbb{R}^n} T_a \mathbb{R}^n, \quad \nabla f(a) \in T_a \mathbb{R}^n \\ df : \mathbb{R}^n &\longrightarrow T^*\mathbb{R}^n = \bigsqcup_{a \in \mathbb{R}^n} T_a^* \mathbb{R}^n, \quad df(a) \in T_a^* \mathbb{R}^n\end{aligned}$$

The following diagram summarizes the relationship between  $d$  and  $\nabla$  quite nicely.



**Theorem 1.4** (Gradient as Direction of Fastest Increase). *Let  $f$  be a real-valued function such that  $\nabla f(x) \neq 0$ . Then, at the point  $x$ ,  $\nabla f(x)$  points in the direction along which  $f$  is increasing the fastest. Equivalently,  $-\nabla f(x)$  points in the direction along which  $f$  is decreasing the fastest.*

*Proof.* Note that this is a coordinate-independent proof. Let us have a tangent vector  $v \in T_x \mathbb{R}^n$ ; since we are only interested in direction, we can normalize  $v$  such that  $\|v\| = 1$ . Evaluating it with the total derivative at  $x$  gives us  $(d_v f)(x)$ . But by definition,

$$(d_v f)(x) = \nabla f(x) \cdot v$$

which means that

$$\begin{aligned}\sup_{\|v\|=1} \{(d_v f)(x)\} &= \sup_{\|v\|=1} \{\nabla f(x) \cdot v\} \\ &= \sup_{\|v\|=1} \{|\nabla f(x)| \|v\| \cos(\theta)\} \\ &= \sup \{|\nabla f(x)| \cos(\theta)\} \\ &= |\nabla f(x)| \text{ when } \theta = 0\end{aligned}$$

Therefore,  $v$  must point in the direction of  $\nabla f(x)$ . ■

Therefore, we can interpret the gradient evaluated at a point as the tangent vector that points in the direction of fastest increase. We can also interpret the gradient  $\nabla f$  itself as the vector field that determines some sort of "flow" in the domain  $\mathbb{R}^n$ . Therefore, if we drop a point in this field, the point will flow through  $\mathbb{R}^n$  through a current determined by  $\nabla f$  and will eventually end up at a local maximum.

#### 1.4.1 Realization in Terms of Bases

Since we are working in  $\mathbb{R}^n$ , we can use the isomorphism  $T_a \mathbb{R}^n \simeq \mathbb{R}^n \simeq T_a^* \mathbb{R}^n$  to induce a basis in every tangent space and cotangent space. This allows us to write all vectors as  $n$ -tuples representing the coefficients of the basis vectors within a linear combination. Our familiar notion of representing vectors as column ( $n$ -tuple) vectors and covectors as row vectors will be used.

**Definition 1.19** (Realization of the Differential and Del Operator). The differential operator  $d$  that eats function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and outputs covector field  $df$  can be realized as a row vector. The del operator  $\nabla$  that eats  $f$  and outputs the (gradient) vector field  $\nabla f$  is realized as a column vector.

$$d = \begin{pmatrix} \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \dots & \frac{\partial}{\partial x_n} \end{pmatrix}, \quad \nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix}$$

**Definition 1.20** (Realization of Total Derivative and Gradients). The realizations of the covector field  $df : \mathbb{R}^n \rightarrow T^*\mathbb{R}^n$  and the vector field  $\nabla f : \mathbb{R}^n \rightarrow T\mathbb{R}^n$  is

$$df = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \dots & \frac{\partial f}{\partial x_n} \end{pmatrix}, \quad \nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

If we feed these mappings the point  $a \in \mathbb{R}^n$ , this is same as evaluating the vectors as such:

$$df(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \dots & \frac{\partial f}{\partial x_n} \end{pmatrix}(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) & \dots & \frac{\partial f}{\partial x_n}(a) \end{pmatrix}, \quad \nabla f(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) \\ \frac{\partial f}{\partial x_2}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{pmatrix}$$

and these resulting vectors  $df(a)$  and  $\nabla f(a)$  are in their respective cotangent and tangent spaces (not the original domain space  $\mathbb{R}^n$ !).

## 1.5 Graphs, Level Surfaces, and Tangent Planes

**Definition 1.21.** Let the graph  $G$  of  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a subset of  $\mathbb{R}^n \oplus \mathbb{R}$ . Then the *extended level surface* of  $f$  is the set of points

$$\tilde{S}_k = \{(x, k) \in \mathbb{R}^n \oplus \mathbb{R} \mid f(x) = k\}$$

We can interpret it as the cross section of the graph  $G$  that forms when we intersect  $G$  with the affine hyperplane  $(0, k) \in \mathbb{R}^n \oplus \mathbb{R}$ . The *level surface* of  $f$  are the points

$$S_k = \{x \in \mathbb{R}^n \mid f(x) = k\}$$

that exist in  $\mathbb{R}^n$ , the domain of  $f$ . Note that there exists a canonical injection between the level surface  $S_k$  and extended level surface  $\tilde{S}_k$ . That is,

$$\rho : \mathbb{R}^n \rightarrow \mathbb{R}^n \oplus \mathbb{R}, \quad \rho(x) \equiv (x, k)$$

To define further theorems, we must now introduce the concept of orthogonality between vectors and surfaces, along with tangent planes.

**Definition 1.22.** Let there be a surface  $S \subset \mathbb{R}^n$  and a vector  $v \in \mathbb{R}^n$  protruding from a point  $x_0 \in S$ .  $v$  is *orthogonal*, or *normal*, to  $S$  at  $x_0$  if for every path function

$$p : \mathbb{R} \rightarrow S \subset \mathbb{R}^n$$

such that  $x_0 = p(t_0)$ , the tangent vector of  $p$  at  $x_0$  is orthogonal to  $v$ . That is,

$$v \cdot p'(t_0) = 0 \text{ for all } p$$

**Definition 1.23.** Let there be a surface  $S \subset \mathbb{R}^n$  with a normal vector  $v(x_0) \neq 0$  at  $x_0$ . Then, the *tangent plane* of  $S$  at  $x_0$  is the set of points

$$\{x \in \mathbb{R}^n \mid v(x_0) \cdot (x - x_0) = 0\}$$

**Theorem 1.5.** Given a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  with a point  $x_0 \in S_k \subset \mathbb{R}^n$ . Then, the gradient vector at  $x_0$  is normal to the surface of  $S_k$  at  $x_0$ .

**Corollary 1.5.1.** The tangent plane of level set  $S_k$  at  $x_0$  is defined

$$\{x \in \mathbb{R}^n \mid \nabla f(x_0) \cdot (x - x_0) = 0\}$$

## 1.6 Iterated Partial Derivatives

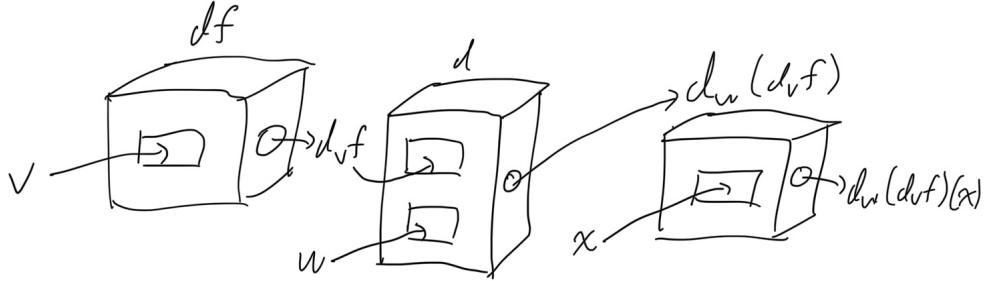
**Definition 1.24** (Iterated Derivatives). Let us refer to interpretation 1 when thinking about the differential operator. Given a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and tangent vector  $v \in T\mathbb{R}^n$ , we have the scalar (since codomain is  $\mathbb{R}$ ) field

$$d_v f : \mathbb{R}^n \rightarrow \mathbb{R}$$

that outputs the derivative of  $f$  in direction  $v$  at an arbitrary input point in  $\mathbb{R}^n$ . But notice that this scalar field is really just another function  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ . So interpreting  $d_v f$  as a function  $g$  (and assuming smoothness), we can choose another vector  $w \in T\mathbb{R}^n$  to find its derivative.

$$d_w(d_v f) : \mathbb{R}^n \rightarrow \mathbb{R}$$

This is called the *2nd derivative of  $f$  in direction  $v$  and  $w$* . Visually,



By assuming smoothness when needed, we can extend this to get

$$d_{v_1}(d_{v_2}(d_{v_3}(\dots(d_{v_k}f)\dots)))$$

called the *kth (iterated) derivative*.

### 1.6.1 Tensor Fields, Tensor Bundles

**Definition 1.25** (Iterated Derivatives at a Point as Tensors). In the 2nd derivative operator  $d_w(d_v f)(x)$  shown previously, let us fix the point  $x$  and function  $f$  and interpret the rest as arguments. We are left with an operator that must take in 2 tangent vectors  $v$  and  $w$ . That is,

$$d.(d.f)(x) : T_x \mathbb{R}^n \times T_x \mathbb{R}^n \rightarrow \mathbb{R}$$

But since  $ddf(x) = (d^2 f)(x)$  is bilinear, this means that

$$(d^2 f)(x) \in T_x^* \mathbb{R}^n \otimes T_x^* \mathbb{R}^n = (T_x^* \mathbb{R}^n)^{\otimes 2}$$

Similarly, the *kth iterated total derivative* of  $f$  at point  $x \in \mathbb{R}^n$  is

$$(d^k f)(x) \in (T_x^* \mathbb{R}^n)^{\otimes k}, \quad (d^k f)(x) : \prod_k T_x \mathbb{R}^n \rightarrow \mathbb{R}$$

**Definition 1.26** (Iterated Total Derivatives as Tensor Fields). We can unfix the point  $x$  and view  $d^k f$  as a mapping receiving point  $x$  and vectors  $v_1, \dots, v_k$  as arguments. Then,  $d^k f$  becomes a *tensor field of rank k*.

$$d^k f : \mathbb{R}^n \rightarrow \bigsqcup_{x \in \mathbb{R}^n} (T_x^* \mathbb{R}^n)^{\otimes k}$$

We can just interpret  $d^k f$  as a field that assigns to every point  $x \in \mathbb{R}^n$  a  $k$ -tensor. Upon selecting a point  $a$ , you are given a  $k$ -tensor living in  $(T_a^* \mathbb{R}^n)^{\otimes k}$ . Then, you choose the  $k$  vectors  $v_1, \dots, v_k$  that define the direction in which you want to take the iterated  $k$ th derivative, input them into the tensor  $d^k f(a)$  and it will output a real number representing the  $k$ th iterated derivative you are looking for.

This is a generalization of  $df$  being a covector field that assigns to every point  $x$  a cotangent vector. After choosing a point  $a$ ,  $df$  outputs the cotangent vector  $df(a)$  which in turn eats a tangent vector  $v$  that you choose and outputs the regular directional derivative of  $f$  in direction  $v$  at  $a$ .

Recall that the tensor algebra of vector space  $T_x \mathbb{R}^n$  is the direct sum of all the tensor product spaces of  $T_x \mathbb{R}^n$ . That is,

$$\mathcal{T}(T_x^* \mathbb{R}^n) \equiv \bigoplus_{i=0}^{\infty} (T_x^* \mathbb{R}^n)^{\otimes i}$$

Therefore, for consistency we can envelop every rank-k tensor space in the tensor algebra

$$d^k f : \mathbb{R}^n \longrightarrow \bigsqcup_{x \in \mathbb{R}^n} \mathcal{T}(T_x^* \mathbb{R}^n) = T\mathcal{T}(T_x^* \mathbb{R}^n)$$

### 1.6.2 Iterated Derivatives with Bases

**Definition 1.27** (Iterated Partial Derivatives). An iterated derivative in which all the tangent vectors are basis vectors  $e_i$  of  $T\mathbb{R}^n$  are called *partial iterated derivatives*. The  $k$ th partial derivative of  $f$  has form

$$d_{e_{i_1}}(\dots(d_{e_{i_k}}f)) = \frac{\partial^k f}{\partial e_{i_1} \dots \partial e_{i_k}}$$

The *2nd partial derivative* has form

$$d_{e_i}(d_{e_j}f) = \frac{\partial}{\partial e_i} \frac{\partial f}{\partial e_j} = \frac{\partial^2 f}{\partial e_i \partial e_j}, \quad i, j = 1, 2, \dots, n$$

**Definition 1.28** (Hessian Matrix as a 2-Tensor Field). The matrix realization of the tensor field

$$d^2 f : \mathbb{R}^n \longrightarrow \bigsqcup_{a \in \mathbb{R}^n} (T_a^* \mathbb{R}^n)^{\otimes 2}$$

is the  $n \times n$  *Hessian matrix*  $H$ , which is of the form

$$(H)_{ij} \equiv \frac{\partial^2 f}{\partial x_i \partial x_j}$$

of partials. Note that even though this matrix looks like it takes in a row vector (left multiplication) and a column vector (right multiplication), it actually takes in *two* row vectors, both in  $T_a \mathbb{R}^n$ .

**Theorem 1.6.** Given function  $f : \mathbb{R}^n \longrightarrow \mathbb{R}$  with existing 2nd derivatives, for all pairs of  $1 \leq i, j \leq n$ ,

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

$\implies$  the Hessian matrix of  $f$  is symmetric.

*Proof.* By abuse of notation, we let us focus on two variables  $x, y$  and ignore the rest. Then, the partial derivatives  $f_{xy}$  and  $f_{yx}$  at a point  $(x_0, y_0)$  can be expressed as double limits:

$$f_{xy}(x_0, y_0) = \lim_{y \rightarrow y_0} \frac{f_x(x_0, y) - f_x(x_0, y_0)}{y - y_0}$$

We can use the two limit definitions of partial derivatives

$$f_x(x_0, y) = \lim_{x \rightarrow x_0} \frac{f(x, y) - f(x_0, y)}{x - x_0}, \quad f_x(x_0, y_0) = \lim_{x \rightarrow x_0} \frac{f(x, y_0) - f(x_0, y_0)}{x - x_0}$$

to get the two partials

$$\begin{aligned} f_{xy}(x_0, y_0) &= \lim_{y \rightarrow y_0} \frac{\lim_{x \rightarrow x_0} \frac{f(x, y) - f(x_0, y)}{x - x_0} - \lim_{y \rightarrow y_0} \frac{f(x_0, y) - f(x_0, y_0)}{x - x_0}}{y - y_0} \\ &= \lim_{y \rightarrow y_0} \lim_{x \rightarrow x_0} \left( \frac{f(x, y) - f(x_0, y) - f(x, y_0) + f(x_0, y_0)}{(x - x_0)(y - y_0)} \right) \\ f_{yx}(x_0, y_0) &= \lim_{x \rightarrow x_0} \frac{\lim_{y \rightarrow y_0} \frac{f(x, y) - f(x, y_0)}{y - y_0} - \lim_{x \rightarrow x_0} \frac{f(x_0, y) - f(x_0, y_0)}{y - y_0}}{x - x_0} \\ &= \lim_{x \rightarrow x_0} \lim_{y \rightarrow y_0} \left( \frac{f(x, y) - f(x, y_0) - f(x_0, y) + f(x_0, y_0)}{(y - y_0)(x - x_0)} \right) \end{aligned}$$

We can see that therefore  $f_{xy} = f_{yx}$  for all  $(x_0, y_0)$ . ■

**Corollary 1.6.1.** Given function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , let

$$(\alpha_1, \alpha_2, \dots, \alpha_k)$$

be any  $k$  distinct numbers from  $\{1, 2, \dots, n\}$ , and let

$$p(\alpha_1, \alpha_2, \dots, \alpha_k) \equiv (p(\alpha_1), p(\alpha_2), \dots, p(\alpha_n))$$

be any permutation of them. Then

$$\frac{\partial^k f}{\partial x_{\alpha_1} \dots \partial x_{\alpha_k}} = \frac{\partial^k f}{\partial x_{p(\alpha_1)} \dots \partial x_{p(\alpha_k)}}$$

for all  $p$ .

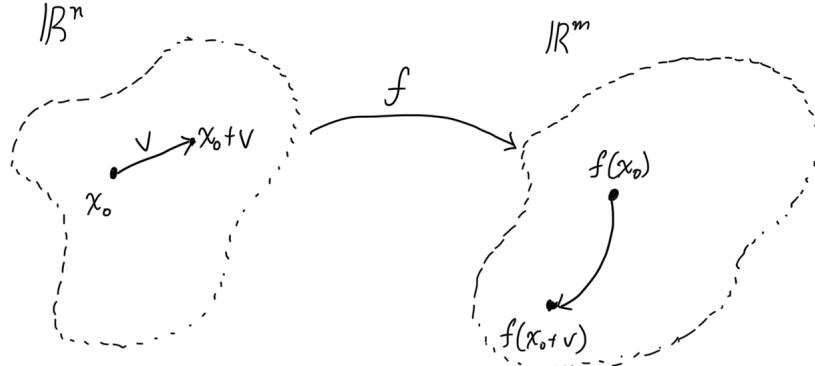
## 1.7 Linear, Quadratic, and Taylor Approximations

### 1.7.1 First Order Approximation

We have defined the total derivative of a function  $f$  at a point  $x_0$ , denoted  $df(x_0)$  as the cotangent vector living in the cotangent space  $T_{x_0}^* \mathbb{R}^n$ . We assume that  $f$  is a vector-valued function when we geometrically describe the construction of the total derivative as a linear approximation.

Most functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  are nonlinear and may behave in all kinds of ways; we can control them by setting the condition that they must be smooth. Now, since these smooth functions are differentiable, we can construct a linear function that best "approximates"  $f$ .

To do this, we first choose an origin point  $x_0 \in \mathbb{R}^n$  and evaluate  $f(x_0)$ . Denote the approximation function as  $P_{x_0}^1$ . Given that we want to evaluate  $f(x_0 + v)$  for some small vector  $v \in T_{x_0} \mathbb{R}^n$ , it turns out that as the point moves from  $x_0 \rightarrow x_0 + v$ ,  $f(x_0)$  doesn't move linearly towards  $f(x_0 + v)$  (marked by the curved line).



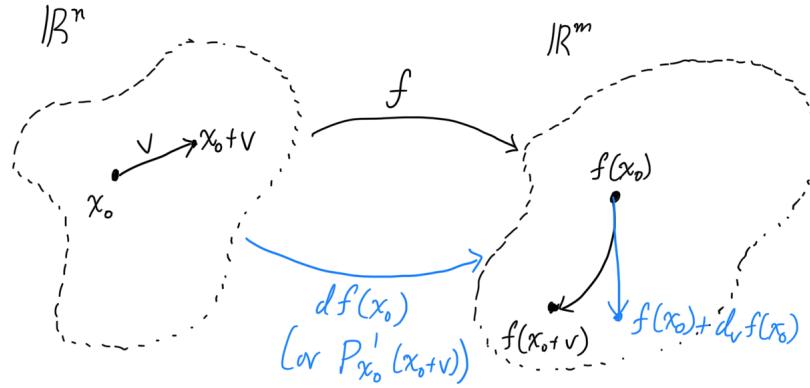
So what is the best way to approximate this? In order to choose the "best" linear approximation to these nonlinear motions, we look at two criterion

1. the approximation  $P_{x_0}^1(x)$  must be equal to the actual function  $f(x)$  at  $x_0$ .
2. It satisfies

$$\lim_{x \rightarrow x_0} \frac{\|f(x) - P_{x_0}^1(x)\|}{\|x - x_0\|} = 0$$

This can be interpreted geometrically by viewing  $P_{x_0}^1$  as a tangent affine subspace on the graph of  $f$  at  $x_0$ , where the directional derivatives of  $f$  and  $P_{x_0}^1$  align at  $x_0$  for all directional vectors  $v \in T_{x_0} \mathbb{R}^n$ .

Conveniently, we can use the total derivative  $df(x_0)$ , which is a linear map, and use it to create the affine linear function  $f(x_0 + v) \approx f(x_0) + (d_v f)(x_0)$ .



It can be clearly seen that as the vector moves from  $x_0 \rightarrow x_0 + v$ , the output vector  $f(x_0) + d_v f(x_0)$  moves in a straight line, which is consistent with our claim of it being an affine linear transformation. Note that by abuse of language, we call this a linear approximation, when it is in fact not linear and rather an *affine* linear approximation.

In order to make this approximation a function of  $x = x_0 + v$ , we change the above formula to

$$f(x) \approx P_{x_0}^1(x) \equiv f(x_0) + (d_{(x-x_0)}f)(x_0)$$

which satisfies both conditions 1 and 2.

**Definition 1.29** (First-Order Approximation of Vector-Valued  $f$  at  $x_0$  and its Matrix Realization). The first order affine approximation of a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  centered at  $x_0$  is

$$P_{x_0}^1(x) \equiv f(x_0) + (d_{(x-x_0)}f)(x_0)$$

The matrix realization of this can be expressed with the Jacobian matrix  $Jf(x_0)$ .

$$\begin{aligned} P_{x_0}^1(x) &\equiv f(x_0) + Jf(x_0)(x - x_0) \\ &= \begin{pmatrix} f_1(x_0) \\ \vdots \\ f_m(x_0) \end{pmatrix} + \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \dots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \dots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix} \begin{pmatrix} x_1 - x_{01} \\ \vdots \\ x_n - x_{0n} \end{pmatrix} \end{aligned}$$

Note that when interpreting  $f$  as a vector valued function, the linear term  $(df)(x_0)$  is *not* a tensor since it takes  $(df)(x_0) : T_{x_0}\mathbb{R}^n \rightarrow \mathbb{R}^m$ .

We can interpret it as a tensor when considering scalar-valued functions  $f$ .

**Definition 1.30** (First-Order Approximation of Real-Valued  $f$  at  $x_0$  and its Matrix Realization). The first order affine approximation of a real-valued function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  centered at  $x_0$  is

$$P_{x_0}^1(x) \equiv f(x_0) + (d_{(x-x_0)}f)(x_0)$$

Where  $(df)(x_0)$  is a 1-tensor taking in tangent vector  $v = x - x_0 \in T_{x_0}\mathbb{R}^n$ , and  $f(x_0)$  is a 0-tensor. The matrix realization of this can be expressed with the Jacobian (row vector) matrix  $Jf(x_0)$ .

$$\begin{aligned} P_{x_0}^1(x) &\equiv f(x_0) + Jf(x_0)(x - x_0) \\ &= f \begin{pmatrix} x_{01} \\ \vdots \\ x_{0n} \end{pmatrix} + \left( \frac{\partial f}{\partial x_1}(x_0) \quad \dots \quad \frac{\partial f}{\partial x_n}(x_0) \right) \begin{pmatrix} x_1 - x_{01} \\ \vdots \\ x_n - x_{0n} \end{pmatrix} \end{aligned}$$

Note that the Jacobian row matrix of  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is the realization of the 1-tensor. In summation form, the approximation expands to

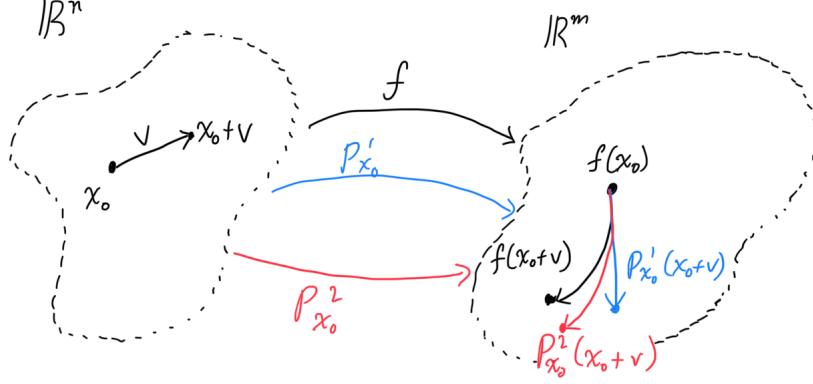
$$P_{x_0}^1(x) = f(x_0) + \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i}(x_0)(x_i - x_{0i}) \right)$$

### 1.7.2 Second Order Approximation

We can improve the linear approximation  $P_{x_0}^1$  to a quadratic approximation  $P_{x_0}^2$  by utilizing higher order derivatives. Using similar logic, given that we know  $f(x_0)$ , for some small vector  $v$  we can approximate  $f(x_0 + v)$  as

$$f(x_0 + v) \approx f(x_0) + (d_v f)(x_0) + \frac{1}{2!}(d_{v,v}^2 f)(x_0)$$

Note that this is a quadratic approximation, and therefore has more "flexibility" than the linear function in approximating  $f$ .



We can make this a function of  $x = x_0 + v$  and define it as such.

**Definition 1.31** (Second-Order Approximation of Vector-Valued  $f$  at  $x_0$ ). The second order affine approximation of a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  centered at  $x_0$  is

$$P_{x_0}^2(x) \equiv f(x_0) + (d_{(x-x_0)} f)(x_0) + \frac{1}{2!}(d_{(x-x_0, x-x_0)} f)(x_0)$$

Unfortunately, we cannot write the matrix realization of this polynomial, since this would require higher dimensional analogies of matrices. Rather, we can just interpret  $f(x_0)$  as a scalar, and the other terms as

$$(df)(x_0) : T_{x_0} \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$(d^2 f)(x_0) : T_{x_0} \mathbb{R}^n \times T_{x_0} \mathbb{R}^n \rightarrow \mathbb{R}^m$$

which take in tangent vectors  $v = x - x_0 \in T_{x_0} \mathbb{R}^n$ .

We can, however, write down the matrix realization of the second order approximation of a real-valued function  $f$ .

**Definition 1.32** (Second-Order Approximation of Real-Valued  $f$  at  $x_0$ ). The second order affine approximation of a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  centered at  $x_0$  is

$$P_{x_0}^2 \equiv f(x_0) + (d_{(x-x_0)} f)(x_0) + \frac{1}{2!}(d_{(x-x_0, x-x_0)}^2 f)(x_0)$$

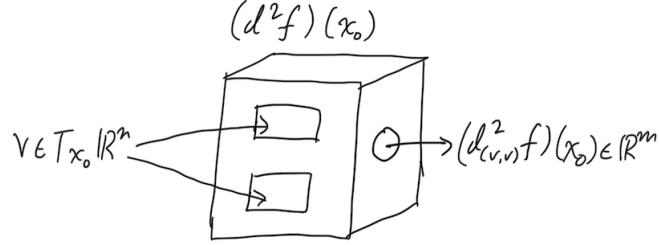
where  $(d^2 f)(x_0)$  is a 2-tensor (taking in tangent vector  $v = x - x_0$ ),  $(df)(x_0)$  is a 1-tensor, and  $f(x_0)$  is a 0-tensor. The matrix realization of this can be expressed with the Jacobian (row vector) matrix  $Jf(x_0)$  and the  $n \times n$  Hessian matrix  $Hf(x_0)$ .

$$\begin{aligned} P_{x_0}^2(x) &\equiv f(x_0) + Jf(x_0)(x - x_0) + \frac{1}{2!}(x - x_0)^T Hf(x_0)(x - x_0) \\ &= f \begin{pmatrix} x_{01} \\ \vdots \\ x_{0n} \end{pmatrix} + \begin{pmatrix} \frac{\partial f}{\partial x_1}(x_0) & \dots & \frac{\partial f}{\partial x_n}(x_0) \end{pmatrix} \begin{pmatrix} x_1 - x_{01} \\ \vdots \\ x_n - x_{0n} \end{pmatrix} \\ &\quad + \frac{1}{2!} (x_1 - x_{01}) \dots (x_n - x_{0n}) \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x_0) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x_0) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x_0) \end{pmatrix} \begin{pmatrix} x_1 - x_{01} \\ \vdots \\ x_n - x_{0n} \end{pmatrix} \end{aligned}$$

In summation form, the multivariate Taylor quadratic expands to

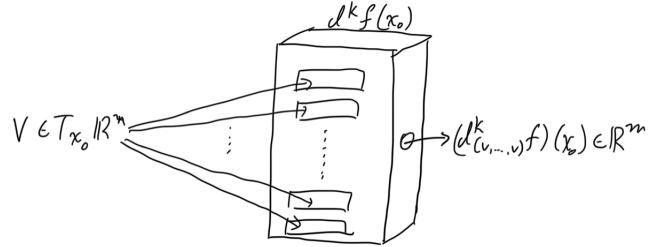
$$P_{x_0}^2(x) = f(x_0) + \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i}(x_0)(x_i - x_{0i}) \right) + \frac{1}{2!} \sum_{i,j=1}^n \left( \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)(x_i - x_{0i})(x_j - x_{0j}) \right)$$

Note that the quadratic term is just a 2-tensor. This is significant because we can notice that every higher order  $n$ th term (cubic, quartic, etc.) is just a multilinear map that accepts one argument  $v$ ,  $n$  times. For example, a quadratic is a 2-tensor accepting 2 copies of  $v \in T_{x_0} \mathbb{R}^n$ .



Therefore, increasing the vector  $v$  by a factor of 2 would, by linearity of the first argument of  $(d^2 f)(x_0)$  increase the output by 2 and by linearity of the second argument also increase the output by 2, resulting in a total increase by 4 times.

Generalizing this to  $(d^k f)(x_0)$ , increasing the input vector  $v$  by a factor of  $\alpha$  would increase the output  $(d_{(v,...,v)}^k f)(x_0)$  of the  $k$ -tensor by  $\alpha^k$ .



### 1.7.3 Higher Order Approximations

We can extend this approximation to higher order multivariate polynomials. The  $k$ th order approximation of function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  centered at  $x_0$  is

$$f(x_0 + v) \approx f(x_0) + (d_v f)(x_0) + \frac{1}{2!}(d_{v,v}^2 f)(x_0) + \frac{1}{3!}(d_{v,v,v}^3 f)(x_0) + \dots + \frac{1}{k!}(d_{v,...,v}^k f)(x_0)$$

where all the  $(d^i f)(x_0)$ 's are multilinear mappings (but not tensors, since codomain is not  $\mathbb{R}$ )

$$(d^i f)(x_0) : \prod_i T_{x_0} \mathbb{R}^n \rightarrow \mathbb{R}^m$$

with respect to  $v$ . Treating this as a function of  $x = x_0 + v$ , we get the following definition.

**Definition 1.33** (Higher-Order Approximation of Vector-Valued  $f$  at  $x_0$ ). The  $k$ th order affine approximation of function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  centered at  $x_0$  is

$$P_{x_0}^k(x) \equiv f(x_0) + (d_{(x-x_0)} f)(x_0) + \frac{1}{2!}(d_{(x-x_0,x-x_0)}^2 f)(x_0) + \dots + \frac{1}{k!}(d_{(x-x_0,...,x-x_0)}^k f)(x_0)$$

Again, this expression encompasses high-dimensional terms that cannot be written in matrices, so it is best to interpret each  $(d^i f)(x_0)$ 's as multilinear mappings

$$\begin{aligned} (df)(x_0) &: T_{x_0} \mathbb{R}^n \rightarrow \mathbb{R}^m \\ (d^2 f)(x_0) &: T_{x_0} \mathbb{R}^n \times T_{x_0} \mathbb{R}^n \rightarrow \mathbb{R}^m \\ &\dots \\ (d^k f)(x_0) &: \prod_k T_{x_0} \mathbb{R}^n \rightarrow \mathbb{R}^m \end{aligned}$$

which take in tangent vectors  $v = x - x_0 \in T_{x_0} \mathbb{R}^n$ .

We distinguish a separate definition for real valued functions  $f$ .

**Definition 1.34** (Higher-Order Approximation of Real-Valued  $f$  at  $x_0$ ). The  $k$ th order affine approximation of function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  centered at  $x_0$  is

$$P_{x_0}^k(x) \equiv f(x_0) + (d_{(x-x_0)}f)(x_0) + \frac{1}{2!}(d_{(x-x_0,x-x_0)}^2f)(x_0) + \dots + \frac{1}{k!}(d_{(x-x_0,\dots,x-x_0)}^k f)(x_0)$$

where  $(d^i f)(x_0)$  is a tensor of rank  $i$ , and the scalar  $f(x_0)$  is a 0-tensor.

$$\begin{aligned} (df)(x_0) &\in T_{x_0}^* \mathbb{R}^n, & (df)(x_0) : T_{x_0} \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ (d^2f)(x_0) &\in (T_{x_0}^* \mathbb{R}^n)^{\otimes 2}, & (d^2f)(x_0) : T_{x_0} \mathbb{R}^n \times T_{x_0} \mathbb{R}^n &\longrightarrow \mathbb{R}^m \\ &\dots & &\dots \\ (d^k f)(x_0) &\in (T_{x_0}^* \mathbb{R}^n)^{\otimes k}, & (d^k f)(x_0) : \prod_k T_{x_0} \mathbb{R}^n &\longrightarrow \mathbb{R}^m \end{aligned}$$

Since tensors of rank greater than 2 cannot be represented in matrix notation, we must embrace this abstract Taylor polynomial.

**Example 1.1.** The summation representation of the third order approximation of function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  centered at  $x_0$  is:

$$\begin{aligned} f(x) \approx f(x_0) &+ \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i}(x_0)(v_i) \right) + \frac{1}{2!} \sum_{i,j=1}^n \left( \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)(v_i)(v_j) \right) \\ &+ \frac{1}{3!} \sum_{i,j,k=1}^n \left( \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(x_0)(v_i)(v_j)(v_k) \right) \end{aligned}$$

where  $v = x - x_0$ , or in components,  $v_i = x_i - x_{0i}$  for  $i = 1, 2, \dots, n$ .

## 1.8 Local/Global Extrema, Lagrange Multipliers

Note that while the approximation isn't exact, the  $n$ th-degree approximation of  $f$  "mimics"  $f$  in the way that the iterated total derivatives, up to the  $n$ th order, are the same as the iterated partial derivatives of  $f$  at the point  $x_0$ . This allows us to analyze the behavior of the function  $f$  up to the  $n$ th order at  $x_0$  by looking only at the components of its  $n$ th degree Taylor expansion.

### 1.8.1 Local Extrema

An application of this is to find the local extrema of  $f$  using the second total derivative. In order for the extrema to be properly defined, the codomain of  $f$  must be ordered, so we will focus on scalar functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ .

**Definition 1.35** (Local Extrema). Given a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , a point  $x_0 \in \mathbb{R}^n$  is a *local minimum* if there exists a neighborhood  $U$  of  $x_0$  such that

$$f(x) \geq f(x_0) \text{ for every } x \in U$$

Similarly,  $x_0$  is a *local maximum* if there exists a neighborhood  $U$  of  $x_0$  such that

$$f(x) \leq f(x_0) \text{ for every } x \in U$$

**Theorem 1.7** (1st Derivative Test). If  $x_0$  is a local extremum of a smooth function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , then  $df(x_0) = 0$ , the zero covector in the cotangent space  $T_{x_0}^* \mathbb{R}^n$ . That is, for every tangent vector  $v \in T_{x_0} \mathbb{R}^n$ , every directional derivative of  $f$  through  $x_0$  in direction  $v$  is 0. That is,

$$x_0 \text{ local extremum} \implies \text{total derivative is 0}$$

However, the converse of this theorem is not true.

In order to determine whether a critical point  $x_0$  is a relative maximum, minimum, or neither, we use the second derivative test.

**Theorem 1.8** (2nd Derivative Test). Let  $x_0$  be a critical point of smooth function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . That is,  $df(x_0) = 0$ . Then,

1.  $x_0$  is a local minimum if  $d^2 f(x_0)$  is a positive definite linear map (note that  $(d^2 f(x_0))$  is technically a 2-tensor with two inputs, but both inputs by definition must be the same)
2.  $x_0$  is a local maximum if  $d^2 f(x_0)$  is a negative definite linear map
3.  $x_0$  is a saddle point if  $d^2 f(x_0)$  is neither positive definite nor negative definite.

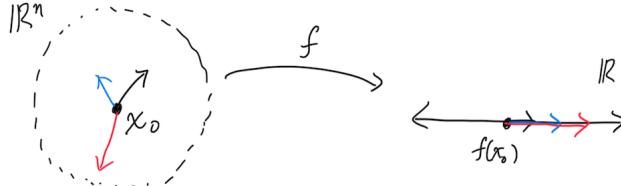
Visually, this makes sense since given a critical point  $x_0$ , the derivative matrix would be 0, meaning that the 2nd degree Taylor expansion of  $f$  near  $x_0$  would be in form

$$f(x) \approx f(x_0) + \frac{1}{2}(d^2_{x-x_0, x-x_0} f)(x_0)$$

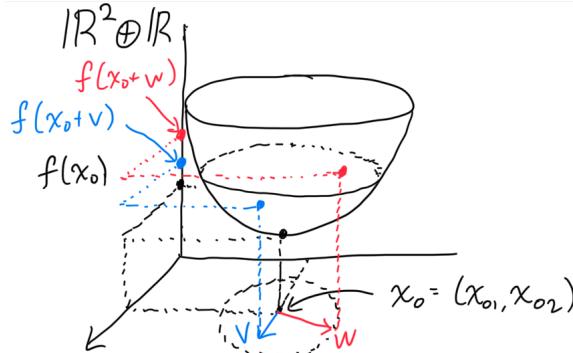
If  $Hf(x_0)$  is positive definite, then by definition

$$\frac{1}{2}(d^2_{x-x_0, x-x_0} f)(x_0) > 0$$

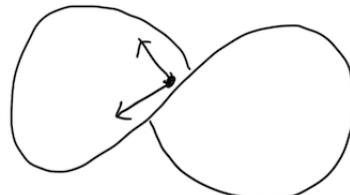
for all  $x$  near  $x_0$ , and so  $f$  would increase in every direction within the neighborhood of  $x_0$ .



or when imagining functions of two variables  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , the neighborhood of  $x_0$  looks like a paraboloid.



The logic follows similarly for negative definite map  $(d^2 f)(x_0)$ . If  $(d^2 f)(x_0)$  is neither positive nor negative definite, then  $\frac{1}{2}(d^2_{(x-x_0, x-x_0)} f)(x_0)$  could be positive or negative, depending on which direction vector  $v = x - x_0$  we choose for computing the directional derivative. Therefore,  $f$  will increase for certain  $h$  and decrease for other  $h$  and is not an extremum. We call this a saddle point since the graph of functions  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  in  $\mathbb{R}^3$  looks like a saddle within the neighborhood of  $x_0$ .



### 1.8.2 Global Extrema

**Definition 1.36** (Global Extrema). Given  $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , a point  $x_0 \in A$  is said to be an *absolute, or global, maximum* if

$$f(x_0) \geq f(x) \text{ for all } x \in A$$

and a *global minimum* if

$$f(x_0) \leq f(x) \text{ for all } x \in A$$

Unfortunately, determining whether a point  $x_0$  is a local extremum requires us to define an open neighborhood  $U$  around  $x_0$  (such that every point  $x \in U$  is greater/smaller than  $x_0$ ). This means that we can only determine local extrema within open sets in  $\mathbb{R}^n$ . Therefore, we must modify our procedure when looking for extrema on functions defined over closed bounded sets.

We now describe a method of computing global extrema.

**Theorem 1.9** (Computing Global Extrema). Let  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  be a function defined on a closed and bounded set  $D = U \cup \partial U$ , where  $U$  is open and  $\partial U$  is the boundary of  $D$ . To find the global extrema on  $D$ , we find all the stationary points of:

1.  $f$  defined over open  $U$
2.  $f$  defined over  $\partial U$ , which can be done by composing the path functions  $p : \mathbb{R}^{n-1} \rightarrow \partial U$  and  $f : \partial U \rightarrow \mathbb{R}$  and finding the stationary points of  $f \circ p$ .

We take all these stationary points and choose the largest to be the global maximum and the smallest to be the global minimum.

### 1.8.3 The Method of Lagrange Multipliers

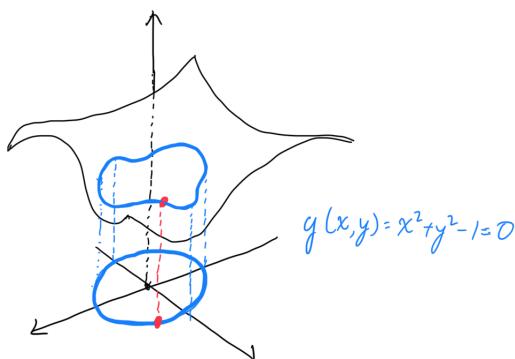
In many cases we are required to find the local extrema of a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  subject to a system of equality constraints (i.e. subject to the condition that one or more equations have to be satisfied exactly by the chosen values of the variables) of the form:

$$g_1(x) = 0, g_2(x) = 0, \dots, g_c(x) = 0$$

which can be summarized into the constraint  $g : \mathbb{R}^n \rightarrow \mathbb{R}^c$

$$g = \begin{pmatrix} g_1 \\ \vdots \\ g_c \end{pmatrix} \implies g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_c(x) \end{pmatrix} = 0$$

Sometimes, these constraints are written as  $g(x) = r$  for some vector  $r$ , but we can just equivalently set the constraint function as  $g(x) - r = 0$ . In physics, these types of "well-behaved" constraints are known as *holonomic constraints*. Here is an example of a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  constrained to the unit circle, where  $g(x, y) = x^2 + y^2 - 1 = 0$ .



To solve this constraint problem, we use the method of Lagrange multipliers. The basic idea is to convert a constrained problem into a form such that the derivative test of an unconstrained problem can still be applied. The relationship between the gradient of the function and gradients of the constraints rather

naturally leads to a reformulation of the original problem, known as the *Lagrangian function*. That is, in order to find the maximum/minimum of  $f$  subjected to the equality constraint  $g(x) = 0$ , we form the Lagrangian function

$$\mathcal{L}(x, \lambda) \equiv f(x) - \lambda g(x)$$

and find the stationary points of  $\mathcal{L}$  considered as a function of  $x \in \mathbb{R}^n$  and the Lagrange multiplier  $\lambda \in \mathbb{R}$ .

The main advantage to this method is that it allows the optimization to be solved without explicit parameterization in terms of the constraints.

**Theorem 1.10** (Lagrange Multipliers Theorem). *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a smooth function and let  $g(x) = 0$ , where  $g : \mathbb{R}^n \rightarrow \mathbb{R}^c$ , be a system of smooth constraint equations.*

$$g \equiv \begin{pmatrix} g_1 \\ \vdots \\ g_c \end{pmatrix}$$

Let  $x^*$  be an optimal solution to the optimization problem of maximizing  $f(x)$  subject to the constraint  $g(x) = 0$  such that  $\text{rank}(dg(x^*)) = c < n$ . Note that  $dg(x^*) : T_{x^*}\mathbb{R}^n \rightarrow \mathbb{R}^c$  is the linear differential map of  $g$  evaluated at  $x^*$ , but it can equivalently be interpreted as a map

$$dg(x^*) : T_{x^*}\mathbb{R}^n \times (\mathbb{R}^c)^* \rightarrow \mathbb{R}$$

or as a map

$$dg(x^*) : (\mathbb{R}^c)^* \rightarrow T_{x^*}^*\mathbb{R}^n$$

Then, there exists a unique vector  $\lambda^* \in (\mathbb{R}^c)^*$  of Lagrange multipliers  $\lambda_1^*, \dots, \lambda_c^*$  such that the two cotangent vectors of  $T_{x^*}^*\mathbb{R}^n$  are equal.

$$df(x^*) = dg(x^*)(\lambda^*)$$

Or equivalently, that the two tangent vectors (which are the gradient vector fields  $\nabla f$  and  $\nabla g(\lambda^*)$  evaluated at  $x^*$ ).

$$\nabla f(x^*) = \nabla g(x^*)(\lambda^*)$$

Conventionally, we use the latter equation comparing the gradients.

**Corollary 1.10.1** (Matrix Realization of the Lagrange Multipliers Theorem). *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be the objective function and let  $g : \mathbb{R}^n \rightarrow \mathbb{R}^c$  be the constraints function with components  $g_i$ , both smooth. Let the vector  $x^*$  be the optimal solution to the optimization problem of maximizing  $f(x)$  subject to the constraint  $g(x) = 0$  such that  $\text{rank}(Jg(x^*)) = c < n$ , where  $Jg(x^*)$  is the Jacobian matrix of partial derivatives of  $g$  evaluated at  $x^*$ . Then, there exists a unique vector  $\lambda^*$  such that*

$$Jf(x^*) = \lambda^{*T} Jg(x^*)$$

which in matrix terms is:

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(x^*) & \dots & \frac{\partial f}{\partial x_c}(x^*) \end{pmatrix} = (\lambda_1^* \quad \dots \quad \lambda_c^*) \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x^*) & \dots & \frac{\partial g_1}{\partial x_n}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_c}{\partial x_1}(x^*) & \dots & \frac{\partial g_c}{\partial x_n}(x^*) \end{pmatrix}$$

But conventionally, we express things in terms of vectors, so we just take the transpose of everything to get the gradient form:

$$\begin{aligned} \nabla f(x^*) &= (Jg(x^*))^T \lambda^* \\ &= \lambda_1^* \nabla g_1(x^*) + \lambda_2^* \nabla g_2(x^*) + \dots + \lambda_c^* \nabla g_c(x^*) \\ &= \sum_{i=1}^c \lambda_i^* \nabla g_i(x^*) \end{aligned}$$

which has a matrix realization of

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(x^*) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x^*) \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x^*) & \dots & \frac{\partial g_c}{\partial x_1}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial x_n}(x^*) & \dots & \frac{\partial g_c}{\partial x_n}(x^*) \end{pmatrix} \begin{pmatrix} \lambda_1^* \\ \vdots \\ \lambda_c^* \end{pmatrix}$$

$$= \lambda_1^* \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x^*) \\ \vdots \\ \frac{\partial g_1}{\partial x_n}(x^*) \end{pmatrix} + \lambda_2^* \begin{pmatrix} \frac{\partial g_2}{\partial x_1}(x^*) \\ \vdots \\ \frac{\partial g_2}{\partial x_n}(x^*) \end{pmatrix} + \dots + \lambda_c^* \begin{pmatrix} \frac{\partial g_c}{\partial x_1}(x^*) \\ \vdots \\ \frac{\partial g_c}{\partial x_n}(x^*) \end{pmatrix}$$

This equation tells us that at any critical points  $x^*$  of  $f$  evaluated under the equality constraints, the gradient of  $f$  at  $x^*$  can be expressed as a linear combination of the gradients of the constraints  $\nabla g_i(x^*)$  (at  $x^*$ ), with the Lagrange multipliers acting as coefficients. Therefore, finding the critical points  $x^*$  of  $f$  constrained with  $g$  is equivalent to solving the system of equations

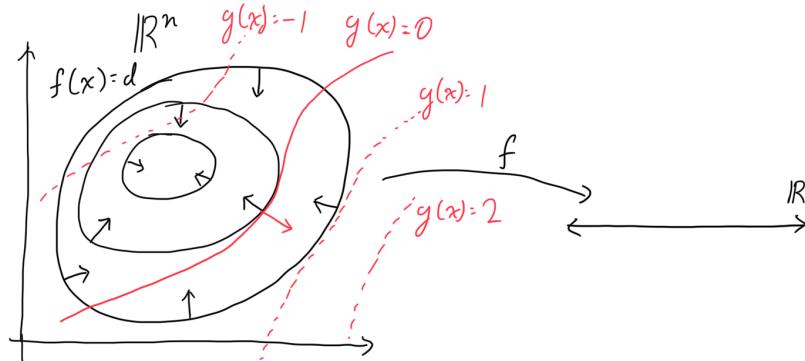
$$g(x) = 0$$

$$\nabla f(x^*) = (Jg(x^*))^T \lambda^*$$

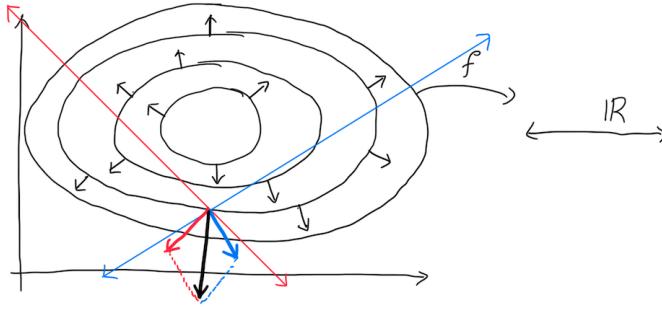
which can be rewritten as

$$\begin{aligned} c \text{ constraint equations} & \left\{ \begin{array}{l} g_1(x) = 0 \\ \dots = 0 \\ g_c(x) = 0 \end{array} \right. \\ n \text{ Lagrangean equations} & \left\{ \begin{array}{l} \frac{\partial f}{\partial x_1}(x^*) = \lambda_1^* \frac{\partial g_1}{\partial x_1}(x^*) + \lambda_2^* \frac{\partial g_2}{\partial x_1}(x^*) + \dots + \lambda_c^* \frac{\partial g_c}{\partial x_1}(x^*) \\ \dots = \dots \\ \frac{\partial f}{\partial x_n}(x^*) = \lambda_1^* \frac{\partial g_1}{\partial x_n}(x^*) + \lambda_2^* \frac{\partial g_2}{\partial x_n}(x^*) + \dots + \lambda_c^* \frac{\partial g_c}{\partial x_n}(x^*) \end{array} \right. \end{aligned}$$

Let us introduce a visualization for when there is a single constraint  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ . From the properties of the gradient,  $\nabla f(x_0)$  is orthogonal to the level set of points satisfying  $f(x) = f(x_0)$  at point  $x_0$ . Note that the constraint function  $g$  also maps  $\mathbb{R}^n \rightarrow \mathbb{R}$ , and so it has its own level surfaces. We can see that the point where the contour line of  $g(x) = 0$  tangentially touches the contours of  $f$  is the maximum. Since it intersects it tangentially, the gradient vector at that point  $\nabla g(x_0)$  is parallel to  $\nabla f(x_0)$ .



We can visualize this for multiple constraints as well, where  $\nabla f(x_0)$  (the gradient vector of  $f$  at  $x^*$ ) can be expressed as a linear combination of  $\nabla g_1(x_0)$  and  $\nabla g_2(x_0)$  (gradient vectors of the constraint functions at  $x^*$ ).



From the properties of the gradient introduced before,  $\nabla f(x_0)$  is orthogonal to the level set of points satisfying  $f(x) = c$  at the point  $x_0$ . But this level set  $f(x) = c$  actually intersects the level set determined by  $g(x) = c$  at the point  $x_0$  and is indistinguishable from each other at  $x_0$ . This means that  $\nabla g(x_0)$  is normal to the level set of  $g(x) = c$  at  $x_0 \iff$  it is normal to the level set of  $f(x) = c$  at  $x_0$ . But  $\nabla f(x_0)$  is also normal at that point, so  $\nabla f(x_0)$  must be parallel to  $\nabla g(x_0)$ .

## 1.9 k-times Continuously Differentiable Functions

So far, we have thrown around the words continuous and differentiable a lot, but note that continuity is a topological property, while differentiability is a property of functions mapping between Euclidean spaces. More specifically, for  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$

1.  $f$  continuous at  $x_0$  means that the preimage of every open neighborhood of  $f(x_0)$  in  $\mathbb{R}^m$  under  $f$  is an open neighborhood of  $x_0$  in  $\mathbb{R}^n$ .
2.  $f$  smooth (i.e. differentiable) at  $x_0$  means that there exists a first order approximation  $P_{x_0}^1$  centered at  $x_0$ . In other words, the covector  $df(x_0)$  is well defined for every input tangent vector  $v \in T_{x_0} \mathbb{R}^n$ .

We will examine the relationship between these properties. First, it is clear that if a function is differentiable, then its partials exist since they are by definition  $d_v f(x_0)$  where  $v \in \{e_1, \dots, e_n\}$ .

**Lemma 1.11** (Differentiability Implies Existence of Partials).

$$\text{Differentiability} \implies \text{Existence of Partials}$$

We remind the reader that differentiability means that the derivative exists at every point in *every path*. The partials are just one of the few paths out of the infinitely many possible ones. We can imagine an example by visualizing a surface with a "crinkle" at a point, which may have well-defined partials but upon a certain path, the derivative may not exist at all. By this logic,

$$\text{Existence of Partials} \not\implies \text{Differentiability}$$

**Example 1.2.**

Furthermore, differentiability implies continuity.

**Lemma 1.12** (Differentiability Implies Continuity).

$$\text{Differentiability} \implies \text{Continuity}$$

But continuity  $\not\implies$  differentiability. We show two examples of this case.

**Example 1.3** (Simple Continuous but not Differentiable Function). *The function  $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto |x|$  is continuous but not differentiable at  $x = 0$ .*

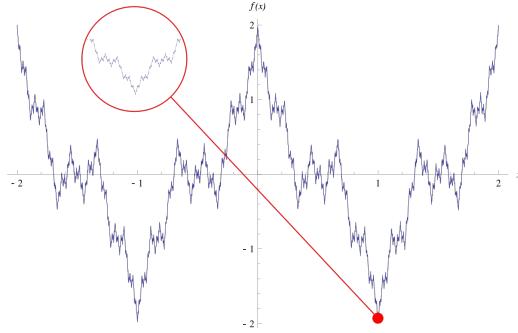
**Example 1.4** (Continuous but nowhere Differentiable Function). *The Weierstrass function is an example of a function that is continuous everywhere but differentiable nowhere. The function is described as a Fourier series*

$$f(x) \equiv \sum_{n=0}^{\infty} a^n \cos(b^n \pi x)$$

where  $0 < a < 1$ ,  $b$  is a positive integer, and

$$ab > 1 + \frac{3}{2}\pi$$

Like other fractals, this function exhibits self-similarity.



We introduce a powerful theorem that allows us to determine smoothness.

**Theorem 1.13** (Differentiability Theorem). *Given function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , if all of its partials exist and are continuous, then  $f$  is differentiable.*

$$\text{Continuous Partial} \implies \text{Differentiability}$$

We can also visualize this theorem. Since the partials are continuous, then the tangent subspace, which is determined by the span of the tangent vectors determined by the partials, also changes continuously. This means that given a  $C^1$  function  $f$ , we can choose *any* directional vector  $v \in \mathbb{R}^n$  and the graph of  $f'_v$  will be well defined. Following this visual, we can interpret the differentiability theorem as:

$$f \in C^k(\mathbb{R}^n) \implies f \text{ can be differentiated } k \text{ times along any } k \text{ paths everywhere}$$

This theorem gives rise to a nice classification of functions based on their smoothness.

**Definition 1.37** (Differentiability Classes). Given a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , if  $f$  is continuous, it is said to be of *class  $C^0$* . If  $f$  has continuous partial derivatives, that is, if

$$d_{e_i} f \equiv \frac{\partial f}{\partial x_i} : \mathbb{R}^n \rightarrow \mathbb{R}^m \text{ for } i = 1, 2, \dots, n$$

are continuous, then  $f$  is said to be of *class  $C^1$* , or  *$C^1$  differentiable*. If it additionally has continuous second partial derivatives, that is, if

$$d_{(e_i, e_j)} f \equiv \frac{\partial^2 f}{\partial x_i \partial x_j} : \mathbb{R}^n \rightarrow \mathbb{R}^m \text{ for } i = 1, 2, \dots, n$$

are continuous, then  $f$  is said to be of *class  $C^2$* , or  *$C^2$  differentiable*. In general, we say that  $f$  is of class  $C^k$ , or  *$C^k$  differentiable*, if its first through  $k$ th partial derivatives are continuous; that is, if

$$d_{(e_{i_1}, \dots, e_{i_k})} f \equiv \frac{\partial^k f}{\partial e_{i_1} \dots \partial e_{i_k}} : \mathbb{R}^n \rightarrow \mathbb{R}^m \text{ for } i = 1, 2, \dots, n$$

are continuous, which implies that  $f$  can be differentiated  $k$  times (i.e. there exists a  $k$ th order Taylor approximation of  $f$ ).

**Lemma 1.14** (Nested  $C^k$  Function Spaces). *The set of all real-valued  $C^k$ -functions defined over  $\mathbb{R}^n$  form an infinite-dimensional vector space, denoted  $C^k(\mathbb{R}^n)$ . Furthermore, this gives rise to the nested space:*

$$C^0(\mathbb{R}^n) \supset C^1(\mathbb{R}^n) \supset C^2(\mathbb{R}^n) \supset \dots \supset C^k(\mathbb{R}^n) \supset \dots \supset C^\infty(\mathbb{R}^n)$$

Note that differentiability does not imply continuous partials!

**Example 1.5** (Differentiable but Not Continuously Differentiable Function). *The function*

$$g(x) \equiv \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

*is differentiable, with derivative*

$$g'(x) \equiv \begin{cases} -\cos\left(\frac{1}{x}\right) + 2x \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

*But because  $\cos(\frac{1}{x})$  oscillates at  $x \rightarrow 0$ ,  $g'(x)$  is not continuous at  $x = 0$ . Therefore  $g(x)$  is differentiable but not in  $C^1(\mathbb{R})$ .*

**Theorem 1.15** (Nested  $C^k$  and  $\mathcal{D}^k$  Function Spaces). *Let the space of all  $k$ -times differentiable functions over  $\mathbb{R}^n$  be denoted  $\mathcal{D}(\mathbb{R}^n)$ . Then,*

$$C^0(\mathbb{R}^n) \supset \mathcal{D}^1(\mathbb{R}^n) \supset C^1(\mathbb{R}^n) \supset \mathcal{D}^2(\mathbb{R}^n) \supset C^2(\mathbb{R}^n) \dots \mathcal{D}^k(\mathbb{R}^n) \supset C^k(\mathbb{R}^n) \dots C^\infty(\mathbb{R}^n)$$

Note that mathematicians throw around the word "smooth" a lot. Usually, it means one of three things

1. it is of class  $C^1$
2. it is of class  $C^\infty$
3. it is of class  $C^k$ , where  $k$  is however high it needs to be to satisfy our assumptions. For example, if I say let us differentiate smooth  $f$  two times, then I am assuming that  $f \in C^2(\mathbb{R}^n)$ .

Visualizing  $C^k$ -functions is easy for low orders. A  $C^0$  function produces a graph that isn't "ripped" or "punctured," since this is exactly what a discontinuity would look like. A  $C^1$  function requires the surface to be smooth in such a way that there is a well defined affine tangent subspace at every point. This means that there cannot be any sharp "points" or "edges" on the graph since a tangent subspace cannot be well defined.

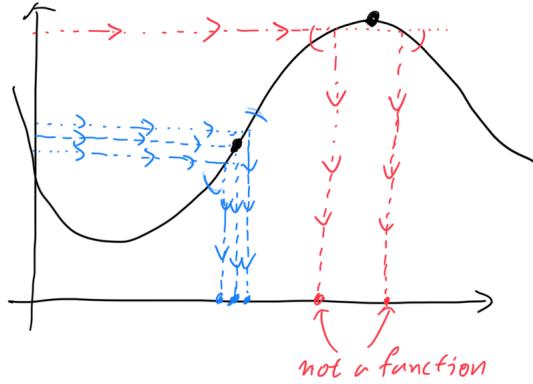
## 1.10 Inverse Function Theorem

A special case of the general implicit function theorem is the inverse function theorem. It gives sufficient condition for a function to be invertible in a neighborhood of a point in its domain.

**Theorem 1.16** (Inverse Function Theorem for Single-Variable  $C^1$  Functions). *If  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a  $C^1$  (continuously differentiable) function with a nonzero derivative at point  $x_0$ , then  $f$  is invertible in a neighborhood of  $x_0$ , the inverse is also  $C^1$ , and the derivative of the inverse function at  $y_0 = f(x_0)$  is the reciprocal of the derivative of  $f$  at  $x_0$ .*

$$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}$$

This can be visualized easily by looking at the graph of any  $C^1$  function.



In high school mathematics, this theorem is informally presented as the horizontal line test.

This can be stated in an alternative form: If  $f : \mathbb{R} \rightarrow \mathbb{R}$  is continuous and injective near  $x_0$ , and differentiable at  $x_0$  such that  $f'(x_0) \neq 0$ , then  $f$  is invertible near  $x_0$  with an inverse that's similarly continuous and injective, and where the above formula would apply as well.

**Corollary 1.16.1** (Inverse Function Theorem for Single-Variable  $C^k$  Functions). *If  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a  $C^k$  functions with a nonzero derivative at point  $x_0$ , then  $f$  is invertible in a neighborhood of  $x_0$ , the inverse is also  $C^k$ , and the derivative of the inverse function at  $y_0 = f(x_0)$  is the reciprocal of the derivative of  $f$  at  $x_0$ .*

**Theorem 1.17** (Inverse Function Theorem for Multivariable Functions and its Matrix Realization). *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a  $C^1$  function defined on an open neighborhood of  $x_0$  in the domain. If the total derivative  $d_{x_0}f$  (i.e. the Jacobian matrix  $Jf(x_0)$ ) at  $x_0$  is invertible, an inverse function of  $f$  is defined on some neighborhood of  $y_0 = f(x_0)$ . Given that we are working with a fixed basis,  $f$  can be modeled by the set of  $n$  equations*

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= y_1 \\ &\dots = \dots \\ f_2(x_1, x_2, \dots, x_n) &= y_2 \end{aligned}$$

This theorem says that this system of  $n$  equations has a unique solution for  $x_1, x_2, \dots, x_n$  in terms of  $y_1, \dots, y_n$ , provided that we restrict  $x$  and  $y$  to small enough neighborhoods of  $x_0$  and  $y_0$ .

This inverse function  $f^{-1}$  is continuously differentiable, and its derivative  $d_{y_0}f^{-1}$  (i.e. the Jacobian matrix  $Jf^{-1}(y_0)$ ) at  $y_0 = f(x_0)$  is the inverse linear map of  $d_{x_0}f$ .

$$d_{y_0}f^{-1} = (d_{x_0}f)^{-1} \iff Jf^{-1}(y_0) = (Jf(x_0))^{-1}$$

**Example 1.6.** Consider the vector-valued function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  defined by

$$f(x, y) = \begin{pmatrix} e^x \cos(y) \\ e^x \sin(y) \end{pmatrix}$$

The Jacobian matrix is

$$Jf(x, y) = \begin{pmatrix} e^x \cos(y) & -e^x \sin(y) \\ e^x \sin(y) & e^x \cos(y) \end{pmatrix} \implies \det Jf(x, y) = e^{2x} \cos^2(y) + e^{2x} \sin^2(y) = e^{2x}$$

Since the determinant  $e^{2x}$  is nonzero everywhere,  $Jf(x, y)$  is nonsingular. Thus, the theorem guarantees that for every point  $x_0 \in \mathbb{R}^2$ , there exists a neighborhood about  $x_0$  over which  $f$  is invertible. However, this does not mean  $f$  is invertible over its entire domain: in this case  $f$  isn't even injective since it is periodic.

## 1.11 Implicit Function Theorem

The implicit function theorem is a tool that allows relations between points in  $\mathbb{R}^n$  to be converted to functions of several real variables. That is, it states that for sufficiently "nice" points on a surface defined as  $f(x) = c$  (where  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , we can locally pretend that this surface is a graph of a function).

That is, let  $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^m$  be a  $C^1$  function. We can think of  $\mathbb{R}^{n+m}$  as the Cartesian product  $\mathbb{R}^n \times \mathbb{R}^m$ , where a point of this product is written

$$(x, y) = (x_1, \dots, x_n, y_1, \dots, y_m)$$

Starting from the given function  $f$ , our goal is to construct a function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  whose graph  $(x, g(x))$  is precisely the set of all  $(x, y)$  such that  $f(x, y) = 0$ .

**Theorem 1.18** (Implicit Function Theorem for 2D, 3D Case). *Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a continuously differentiable function and let there be a point  $(x_0, y_0) \in \mathbb{R}^2$  such that  $f(x_0, y_0) = 0$ . If*

$$\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$$

then there is an open neighborhood  $U$  around  $(x_0, y_0)$  such that we can make  $y$  a function of  $x$  within  $U$  satisfying  $f(x, y(x)) = 0$ .

Let  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  be a continuously differentiable function and let there be a point  $(x_0, y_0, z_0) \in \mathbb{R}^3$  such that  $f(x_0, y_0, z_0) = 0$ . If

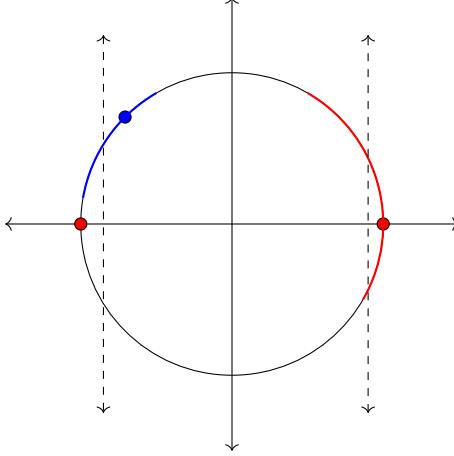
$$\frac{\partial f}{\partial z}(x_0, y_0, z_0) \neq 0$$

then there is an open neighborhood  $U$  around  $(x_0, y_0, z_0)$  such that we can make  $z$  a function of  $x$  and  $y$  within  $U$  satisfying  $f(x, y, z(x, y)) = 0$ .

**Example 1.7.** Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  be defined by  $f(x, y) = x^2 + y^2 - 1$ . The level set at  $z = 0$  would be the set of points satisfying

$$x^2 + y^2 - 1 = 0$$

the unit circle. The derivative of  $f$  with respect to  $y$  is 0 at the points  $(-1, 0)$  and  $(1, 0)$ , meaning that in any neighborhood of these points, we cannot define a function of  $y$  with respect to  $x$ . This is true, indeed, since any such function would fail the vertical line test, which can be seen in the red neighborhood around  $(1, 0)$ . However, the blue neighborhood of the point  $(-\sqrt{2}/2, \sqrt{2}/2)$  does indeed define a function of  $y$  with respect to  $x$  satisfying the vertical line test.



Note that for the 2D and 3D case, the level surface that we dealt with has a codimension of 1; that is, the dimension of the manifold generated by the level surface is  $n - 1$ . The special implicit function theorem generalizes cases such as these.

**Definition 1.38** (Truncated Jacobian Matrix). Given a function  $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^m$  where the variables are

$$x_1, \dots, x_n, y_1, \dots, y_m$$

the *truncated Jacobian matrix* of  $f$  is the  $m \times m$  matrix formed by the rightmost  $m$  columns of  $Jf(x, y)$ . That is, the matrix formed by the elements on the right of the vertical line is the truncated Jacobian matrix.

$$Jf(x, y) = \left( \begin{array}{ccc|ccc} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial y_1} & \dots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} & \frac{\partial f_m}{\partial y_1} & \dots & \frac{\partial f_m}{\partial y_m} \end{array} \right) \mapsto \left( \begin{array}{ccc} \frac{\partial f_1}{\partial y_1} & \dots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \dots & \frac{\partial f_m}{\partial y_m} \end{array} \right) = Jf_y(x, y)$$

**Theorem 1.19** (Special Implicit Function Theorem). Let  $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  be a continuously differentiable function. We can think of  $\mathbb{R}^{n+1}$  as the Cartesian product  $\mathbb{R}^n \times \mathbb{R}$ , where an element of this product is  $(x, y)$  ( $x \in \mathbb{R}^n$ ,  $y \in \mathbb{R}$ ). Let us fix a point  $(x_0, y_0) = (x_{01}, x_{02}, \dots, x_{0n}, y_0)$  with  $f(x_0, y_0) = 0$ .

If the  $(1 \times 1)$  truncated Jacobian matrix defined with the partial derivatives with respect to the  $y$  terms evaluated at  $(x_0, y_0)$

$$Jf_y(x_0, y_0) \equiv \frac{\partial f}{\partial y}(x_0, y_0)$$

which can also be thought as the matrix truncated

$$Jf(x, y) = \left( \begin{array}{ccc|c} \frac{\partial f}{\partial x_1} & \dots & \frac{\partial f}{\partial x_n} & \frac{\partial f}{\partial y} \end{array} \right) \mapsto \left( \begin{array}{c} \frac{\partial f}{\partial y} \end{array} \right)$$

is invertible, then there exists an open neighborhood  $U \subset \mathbb{R}^n$  containing  $x_0$  and a unique  $C^1$  function  $g : U \rightarrow \mathbb{R}$  such that  $g(x_0) = y_0$ , and

$$f(x, g(x)) = 0 \text{ for all } x \in U$$

Moreover, the partial derivatives of  $g$  in  $U$  (represented by its Jacobian matrix) are given by the matrix product

$$Jg(x) = -(Jf_y(x, g(x)))^{-1} Jf(x, g(x))$$

or represented

$$\begin{pmatrix} \frac{\partial g}{\partial x_1} & \cdots & \frac{\partial g}{\partial x_n} \end{pmatrix} = - \left( \frac{\partial f}{\partial y}(x, g(x)) \right)^{-1} \begin{pmatrix} \frac{\partial f}{\partial x_1}(x, g(x)) & \cdots & \frac{\partial f}{\partial x_n}(x, g(x)) \end{pmatrix}$$

**Example 1.8** (Circle Example). Let  $n = m = 1$  and

$$f(x, y) = x^2 + y^2 - 1$$

The matrix of partial derivatives is just a  $1 \times 2$  matrix, given by

$$Jf(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{pmatrix} = (2x_0 \quad 2y_0)$$

The truncated Jacobian matrix is just the  $1 \times 1$  matrix  $(2y_0)$ , which is invertible if and only if  $y_0 \neq 0$ . By the implicit function theorem we see that we can locally write the circle in the form  $y = g(x)$  for all points where  $y \neq 0$ . For  $(\pm 1, 0)$  we cannot (as seen before by the inverse function theorem). But by writing  $x$  as a function of  $y$  ( $x = h(y)$ ), we can write it at these points. The derivative of this function  $g : \mathbb{R} \rightarrow \mathbb{R}$  can be written using the formula below, where  $\frac{\partial f}{\partial y}(x, g(x)) = 2y_{(x, g(x))} = 2g(x)$ .

$$\frac{\partial g}{\partial x} = -(2g(x))^{-1}(2x) = -\frac{x}{g(x)}$$

But since the value of  $g(x)$  is the  $y$  value, we can rewrite this as

$$\frac{dy}{dx} = -\frac{x}{y}$$

and similarly, get the formula for the derivative of  $h$  as

$$\frac{dx}{dy} = -\frac{y}{x}$$

**Theorem 1.20** (General Implicit Function Theorem). Let  $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^m$  be a continuously differentiable function, and let us interpret  $\mathbb{R}^{n+m}$  as the Cartesian product  $\mathbb{R}^n \times \mathbb{R}^m$ , where an element of this product is  $(x, y)$  ( $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ ). Let us fix a point  $(x_0, y_0) = (x_{01}, \dots, x_{0n}, y_{01}, \dots, y_{0m})$ , with  $f(x_0, y_0) = 0$ .

If the  $(m \times m)$  truncated Jacobian matrix constructed by truncating the matrix below as such

$$Jf = \left( \begin{array}{ccc|ccc} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} & \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{array} \right) \mapsto \left( \begin{array}{ccc} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{array} \right) (x_0, y_0) = Jf_y(x_0, y_0)$$

evaluated at  $(x_0, y_0)$  is invertible (i.e. determinant is nonzero), then there exists an open neighborhood  $U \subset \mathbb{R}^n$  containing  $x_0$  and a unique  $C^1$  function  $g : U \rightarrow \mathbb{R}^m$  such that  $g(x_0) = y_0$ , or in component terms, the smooth functions  $k_i$  exist such that

$$y_{0i} = k_i(x_{01}, x_{02}, \dots, x_{0n}), \quad i = 1, 2, \dots, m$$

and

$$f(x, g(x)) = 0 \text{ for all } x \in U$$

Moreover, the partial derivatives of  $g$  in  $U$  (represented by its Jacobian matrix) are given by the matrix product

$$Jg(x) = -(Jf_y(x, g(x)))^{-1} Jf(x, g(x))$$

or equivalently,

$$\begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix} = - \left( \begin{array}{ccc|ccc} \frac{\partial f_1}{\partial y_1}(x, g(x)) & \cdots & \frac{\partial f_1}{\partial y_m}(x, g(x)) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial y_1}(x, g(x)) & \cdots & \frac{\partial f_m}{\partial y_m}(x, g(x)) \end{array} \right)^{-1} \left( \begin{array}{ccc} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{array} \right) (x, g(x))$$

The derivatives may also be computed by implicit differentiation.

## 1.12 Divergence and Curl

### 1.12.1 Tensor Fields

Note that a scalar field is a set of scalars associated with every point in space. Similarly, a vector field is a set of vectors associated with every point in space. Going further, a tensor field is a set of tensors associated with every point in space. It follows that:

1. A scalar field is a rank 0 tensor field
2. A vector field is a rank 1 tensor field

Therefore, a rank 2 tensor field would be:

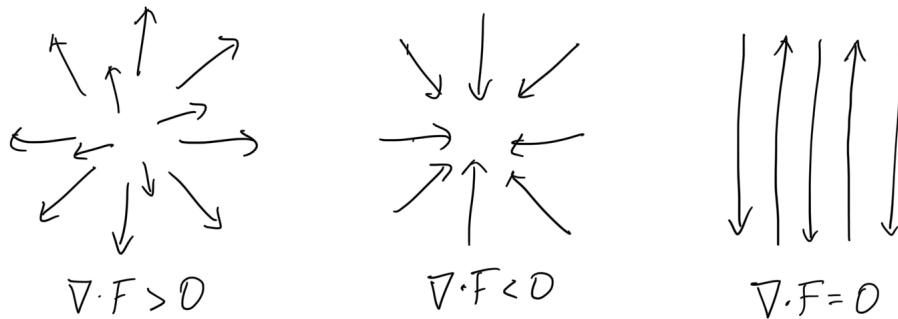
$$F : \mathbb{R}^n \longrightarrow \mathbb{R}^n \otimes \mathbb{R}^n$$

where  $F(x_0)$  would be a rank 2 tensor (or a matrix, given that coordinates are well defined).

### 1.12.2 Coordinate Definition of Divergence

Colloquially, the divergence is an operator  $\text{div}$  that operates on a vector field and produces a scalar field which provides the quantity of the vector field's source at each point. Technically, the divergence represents the volume density of the outward flux of a vector field from an infinitesimal volume around a given point.

There is a very nice geometric interpretation for divergence. Imagine that the vector field  $F$  represents fluid flow in  $\mathbb{R}^n$ . Divergence is then the "measure" of the net amount of fluid flowing in and out of an infinitesimally small region, labeled at each point. If the net fluid flow is positive (i.e. more fluid is flowing in than out) at point  $x_0$ , then  $\text{div } F(x_0) > 0$ . If the net fluid flow is negative (i.e. more fluid is flowing out than in) at point  $x_0$ , then  $\text{div } F(x_0) < 0$ . This measure assigns a number to every point in the space (creating a scalar field). Therefore, each point either acts as a "source" of fluid emanating from it or as a "sink" that sucks in more fluid than it puts out.

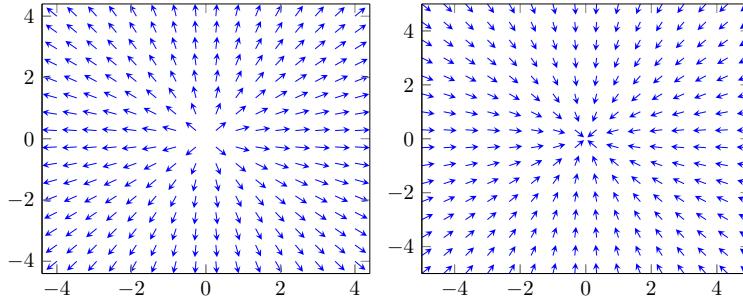


**Definition 1.39** (Divergence in Coordinates). In Cartesian coordinates, the *divergence* of a  $C^1$  vector field  $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$  is defined

$$\text{div } F \equiv \nabla \cdot F \equiv \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} \cdot \begin{pmatrix} F_1 \\ \vdots \\ F_n \end{pmatrix} = \sum_{i=1}^n \frac{\partial F_i}{\partial x_i}$$

where  $\cdot$  is the Euclidean dot product and  $\nabla$  is the del operator.

**Example 1.9.** The divergence of the origin in the left graph is clearly negative since the net flow is out of the point, while the divergence of the origin in the right graph is positive since the net fluid flow is in.



**Definition 1.40** (Divergence of Tensor Fields). Let  $A$  be a  $C^1$  second-order tensor field; that is, it assigns a tensor to every point in Euclidean space, defined as

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$$

the divergence in a Cartesian coordinate system is a first-order tensor field and can be defined

$$\operatorname{div} A = \left( \frac{\partial A_{11}}{\partial x_1} + \frac{\partial A_{12}}{\partial x_2} + \frac{\partial A_{13}}{\partial x_3} \right) \\ \left( \frac{\partial A_{21}}{\partial x_1} + \frac{\partial A_{22}}{\partial x_2} + \frac{\partial A_{23}}{\partial x_3} \right) \\ \left( \frac{\partial A_{31}}{\partial x_1} + \frac{\partial A_{32}}{\partial x_2} + \frac{\partial A_{33}}{\partial x_3} \right)$$

**Definition 1.41** (Divergence in Cylindrical, Spherical Coordinates). For vector field  $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  expressed in cylindrical coordinates as

$$F = \begin{pmatrix} F_r \\ F_\theta \\ F_z \end{pmatrix}$$

the divergence is

$$\operatorname{div} F = \nabla \cdot F = \frac{1}{r} \frac{\partial}{\partial r} (r F_r) + \frac{1}{r} \frac{\partial F_\theta}{\partial \theta} + \frac{\partial F_z}{\partial z}$$

Note that the condition of locality is important, since in general a global cylindrical coordinate system would be inconsistent.

In spherical coordinates  $(r, \theta, \phi)$ , the divergence is

$$\operatorname{div} F = \nabla \cdot F = \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 F_r) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta F_\theta) + \frac{1}{r \sin \theta} \frac{\partial F_\phi}{\partial \phi}$$

### 1.12.3 Coordinate Definition of Curl

Colloquially, the curl is a vector operator that describes the infinitesimal circulation of a vector field in 3-dimensional Euclidean space, where the curl at each point is represented by a vector whose length and direction denote the magnitude and axis as the maximum circulation.

That is, if one drops a twig or a ball with its center of mass at a certain point, the curl measures how much it will spin. In physics, the rotation of a rigid body in 3-dimensions can be described by a vector  $\omega$  along the axis of rotation.  $\omega$  is called the *angular velocity vector*, with  $\|\omega\|$  denoting the angular speed of the body. The curl of this vector field measured at the center of mass of the body is measured as  $2\omega$ . That is, the curl outputs *twice* the angular velocity vector of any rigid body.

Furthermore, if  $\operatorname{curl} F(x_0) = 0$ , then this indicates that there are no "whirlpools" at  $x_0$ , meaning that any rigid body placed at  $x_0$ , while it may travel along a path, will not rotate around its own axis. Such a vector field  $F$  with this property is called *irrotational*.

**Definition 1.42.** The *curl* of a 3-dimensional  $C^k$  vector field  $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is an operator

$$\operatorname{curl} : C^k(\mathbb{R}^3; \mathbb{R}^3) \rightarrow C^{k-1}(\mathbb{R}^3; \mathbb{R}^3)$$

defined

$$\operatorname{curl} F \equiv \nabla \times F \equiv \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \equiv \begin{pmatrix} \frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z} \\ \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x} \\ \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \end{pmatrix}$$

Note that unlike the gradient and divergence operators, curl does not generalize as simply to other dimensions.

**Definition 1.43.** A vector field  $F$  is *irrotational* if

$$\operatorname{curl} F = 0$$

It has been shown that fluid draining from a tub is usually irrotational except for right at the center, which is surprising since the fluid itself is "rotating" around the drain.

**Theorem 1.21.** For any  $C^2$  vector field  $F$ ,

$$\operatorname{div} \operatorname{curl} F = \nabla \cdot (\nabla \times F) = 0$$

That is, the divergence of any curl is 0.

*Proof.* Also proved by equality of mixed partials. ■

**Definition 1.44.** The *Laplace operator*, or *Laplacian*, of a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is the divergence of the gradient.

$$\nabla^2 f \equiv \nabla \cdot (\nabla f) \equiv \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}$$

#### 1.12.4 Conservative, Solenoidal Vector Fields

**Definition 1.45** (Conservative Vector Fields). A vector field  $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a *conservative vector field* if and only if there exists a scalar field  $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$  such that

$$F = \nabla f$$

on  $U$ .

Conservative vector fields appear naturally in mechanics: they are vector fields representing forces of physical systems in which energy is conserved.

**Theorem 1.22.** Given a  $C^2$ -function  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ ,

$$\nabla \times (\nabla f) = 0$$

That is, the curl of any gradient vector field is the zero vector.

*Proof.*  $\nabla \times \nabla f$  can be expanded to

$$\left( \frac{\partial^2 f}{\partial y \partial z} - \frac{\partial^2 f}{\partial z \partial y}, \frac{\partial^2 f}{\partial z \partial x} - \frac{\partial^2 f}{\partial x \partial z}, \frac{\partial^2 f}{\partial x \partial y} - \frac{\partial^2 f}{\partial y \partial x} \right) = (0, 0, 0)$$

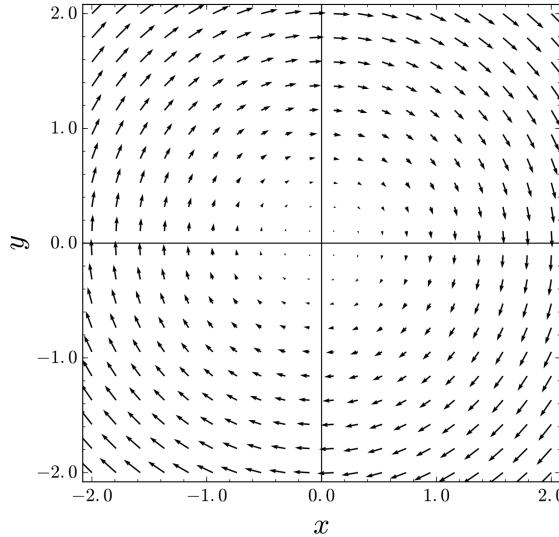
by equality of mixed partials. ■

**Definition 1.46** (Solenoidal Vector Fields). A *solenoidal*, or *incompressible*, vector field is a vector field  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that

$$\operatorname{div} F = \nabla \cdot F = 0$$

at all point in the field. That is, the field has no sources or sinks.

**Example 1.10.** The vector field  $F : (x, y) \mapsto (y, -x)$  is solenoidal.



## 1.13 Differentials of Functions

Note that differentials of functions are not to be confused with differential forms.

### 1.13.1 One Variable Functions

Given a continuous real valued function  $f : \mathbb{R} \rightarrow \mathbb{R}$ , let the parameter for the input be labeled  $x$  and the output parameter be labeled with the variable  $y$ . This is conventional, leading to the well known expression

$$y = f(x)$$

Now, note that even though the function itself is  $f$ , we can choose to represent its output values with either  $f$  or  $y$ .

**Definition 1.47.** A given change in  $x$  is denoted  $\Delta x$ . Given that the input value changes by  $\Delta x$  (i.e. the input changes from  $x \rightarrow x + \Delta x$ ), the change in the output is denoted with  $\Delta y$ . That is,

$$\Delta y \equiv f(x + \Delta x) - f(x)$$

Note that the behavior of  $\Delta y$  is completely independent from  $\Delta x$ . We cannot assume that, for example, an increase in  $x$  (i.e.  $\Delta x > 0$ ) corresponds to an increase in  $y$ . Note that  $\Delta y$  is really a function of two independent variables, which is  $x$  (the initial point where the change happens) and  $\Delta x$  (how much  $x$  is changing). That is,

$$\Delta y(x, \Delta x) \equiv f(x + \Delta x) - f(x)$$

Now we continue to introduce the differential.

**Definition 1.48.** An *infinitesimal* is a real number  $\epsilon > 0$  such that  $\epsilon$  is less than any real positive number. It can also be defined as a positive real number  $\epsilon$  such that

$$\epsilon < \frac{1}{n} \text{ for all } n \in \mathbb{N}$$

Since  $f \in C^0\mathbb{R}$ , we can assume that an infinitesimal change in input value  $\Delta x$  corresponds to an infinitesimal change in output value  $\Delta y$ . This infinitesimal change is now denoted  $dx$  and  $dy$ , respectively. Therefore, the instantaneous rate of change of  $y$  (or  $f$ ) with respect to the change of  $x$  is really just the ratio of the infinitesimal changes of both  $x$  and  $y$ . This is denoted by the fraction

$$\frac{dy}{dx} \equiv \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}$$

This is precisely the *Leibniz Notation* of the derivative of a function  $f$ . Note that the quotient  $dy/dx$  is not infinitesimally small, but it is rather a real number.

**Definition 1.49.** The *differential* of a function  $f$  of a single real variable  $x$  is the linear function  $df$  of two independent variables  $x$  and  $\Delta x$  given by

$$dy \equiv df \equiv df(x, \Delta x) \equiv f'(x) \Delta x$$

Note that since  $y$  and  $f$  both denote the output of the function  $f$ , we can use them interchangeably in this notation. Note also that  $f'(x)$  is just the derivative of  $f$  with respect to  $x$ . Note that since

$$dx(x, \Delta x) \equiv \Delta x$$

it is conventional to substitute  $\Delta x = dx$  to get the equivalent expression

$$df \equiv f'(x) dx \text{ or } dy \equiv \frac{dy}{dx} dx$$

Note that the differential of  $f$  of a point  $x$  is a linear approximation of  $f$  at the point  $x$  since  $df$  is linear and the error bound  $\epsilon$  of the equality

$$\Delta y = f'(x) \Delta x + \epsilon = df(x) + \epsilon$$

satisfies

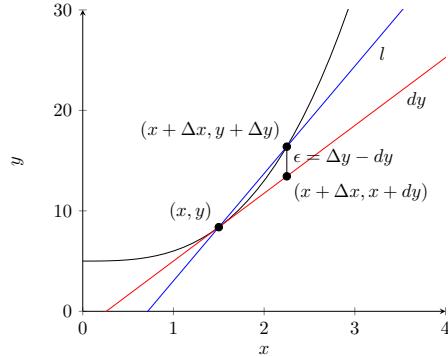
$$\lim_{\Delta x \rightarrow 0} \frac{\epsilon}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y - dy}{\Delta x} = 0$$

$\Rightarrow dy \approx \Delta y$ , where we can make the approximation arbitrarily small by constraining  $\Delta x$  to be sufficiently small. Therefore, the differential of a function is known as the *principal (linear) part* in the increment of a function. That is,  $df$  is linear with respect to  $dx$ , the increment (but not with respect to  $x$  itself). Although the error  $\epsilon$  may be nonlinear, it tends to 0 rapidly as  $\Delta x$  tends to 0.

We can visually compare the secant line and the differential of an arbitrary function. We can plot the graph of  $f$  as the set

$$l \equiv \{(x, f(x)) \in \mathbb{R} \oplus \mathbb{R}\} \subset \mathbb{R}^2$$

and observe the unique secant line that connects the two points  $(x, f(x))$  and  $(x + \Delta x, y + \Delta y) = (x + \Delta x, f(x + \Delta x))$ . This line is labeled in blue. The differential, on the other hand, is labeled in red.



### 1.13.2 Multivariable Functions

**Definition 1.50.** For functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  in the form

$$y = f(x_1, x_2, \dots, x_n)$$

the *partial differential* of  $y$  (or  $f$ ) with respect to any one of the variables  $x_i$  is the principal part of change in  $y$  resulting in the change  $dx_1$ . Therefore, the partial differential is

$$\frac{\partial f}{\partial x_1} dx_1$$

where  $\partial f / \partial x_1$  is the partial derivative of  $f$ . The sum of all the partial differentials is called the *total differential* of  $f$ .

$$dy = \frac{\partial y}{\partial x_1} dx_1 + \frac{\partial y}{\partial x_2} dx_2 + \dots + \frac{\partial y}{\partial x_n} dx_n$$

More precisely, in vector calculus, if  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a differentiable function, then by definition of differentiability (i.e. a linear approximation exists),

$$\begin{aligned} y &\equiv f(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n) - f(x_1, x_2, \dots, x_n) \\ &\equiv \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \dots + \frac{\partial f}{\partial x_n} dx_n + \epsilon_1 \Delta x_1 + \dots + \epsilon_n \Delta x_n \end{aligned}$$

where the  $\epsilon_i$  can be made arbitrary close to 0 by constraining  $\Delta x_i$  to 0. Similar to the one variable case, we can therefore see that  $dy \approx \Delta y$ .

## 2 Integration

### 2.1 Geometric Interpretations of Integration

The concept of integration in one variable calculus limits the applicability of the operation to finding only areas of functions under curves. We will replace the reader's intuition of integration with the following description. Given a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , we interpret it as a scalar field that assigns a "weight" to each point in  $\mathbb{R}^n$ . Now, given any "shape"  $B$  in  $\mathbb{R}^n$  that is closed (but not necessarily bounded), an integral can calculate the "weighed" volume of  $B$  by cutting  $B$  into infinitesimally small points, multiplying them by their respective weights determined by  $f$ , and then summing up the weighed points. Integrating  $B$  in  $\mathbb{R}$ ,  $\mathbb{R}^2$ , and  $\mathbb{R}^3$  with a constant scalar field equal to 1 is equivalent to finding the length, area, and volume of  $B$ , respectively. We deconstruct specific types of iterated integrals.

#### 2.1.1 Single Integral as Weighed Length or Area

A single integral is calculated from a function  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Given some intervals (or a collection of intervals)  $B \subset \mathbb{R}$ , the integration notation is familiar to us.

$$\int_B f(x) dx$$

We can interpret this integration in two ways. First, we imagine that the function  $f$  is a scalar field in  $\mathbb{R}$ . Therefore, every point  $x$  in  $\mathbb{R}$  has a certain real number  $f(x)$  associated to it. Therefore, the interval  $B \in \mathbb{R}$  now consists of points that now have different densities each (which can be negative). The entire  $B$  can now be thought of as a 1-dimensional "rod" in  $\mathbb{R}$  that has an uneven distribution of mass determined by  $f$ . The total mass of the rod  $B$  is calculated by the integral. In the diagram below, we use different "thickness" to represent different densities.



Secondly, we can visualize the entire graph of  $f$  in  $\mathbb{R} \oplus \mathbb{R}$ . This represents a curve in the  $xy$ -axis that most beginner calculus students are familiar with. Note that the interval  $B$  exists in the  $x$ -axis, and in this case, the "weight" of each point  $x$  in  $B$  is represented as the "height" of the infinitesimally thin bar at  $x$ . It is easy to see that the weight of the rod at point  $x$  and the height of the bar at point  $x$  are really the same measure determined by  $f$ . Therefore, the density distribution in the rod is now modeled as the height of the function at each point. Calculating the integral of this function now calculates the "area" under the curve.

It is important to point out that  $B$  does not necessarily need to be a length in the form of  $[a, b]$ . It can be any union of disjoint lengths, too. However, adding a finite number of single points to  $B$  will not affect the integral. It is also customary for  $B$  to be closed.

#### 2.1.2 Double Integral as Weighed Area or Volume

The double integral is calculated from function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Given a certain closed subset  $B \subset \mathbb{R}^2$ , the integration notation is

$$\iint_B f(x, y) dA$$

Again, we can interpret double integration in two ways. First, we think of  $f$  as a scalar field assigning a density to every point in  $\mathbb{R}^2$ . Then, the 2-dimensional shape  $B$  itself should have a certain density distribution on it determined by  $f$ . The double integral above then determines the mass of  $B$ .

The second way to interpret this is to imagine the 2-dimensional shape  $B$  lying in the extended space  $\mathbb{R}^2 \oplus \mathbb{R}$ . We then model the density distribution as merely the height of the infinitesimally thin bar at each point  $x$ . Again, the height of this bar at  $x$  is precisely its density described in the first interpretation. Therefore, integrating this shape is equivalent to finding the volume of the infinite union of the infinitesimally thin bars at each point in  $B$ .

Note again that  $B$  need not be one solid region. It can be a union of multiple disjoint ones. However, adding a single point  $p$  or a 1-dimensional path  $p$  to  $B$  will not affect the integral since they have an area of 0.

### 2.1.3 Triple Integral as Weighed Volumes or Hyper-Volumes

The double integral is calculated from function  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ . Given a certain closed subset  $B \subset \mathbb{R}^3$ , the integration notation is

$$\iiint_B f(x, y, z) dV$$

Following the logic of the previous two examples, the function  $f$ , interpreted as a scalar field, assigns a scalar at each point  $x$  in the solid  $B$ . Therefore, we can visualize  $B$  as a solid, 3-dimensional object in  $\mathbb{R}^3$  with a certain density distribution defined by  $f$ . The total mass of  $B$  is therefore determined by the triple integral above.

Following similar logic, we can interpret this integral as the hypervolume of a 4 dimensional object, but this is not often used.

## 2.2 Reduction to Iterated Integrals

We first state a basic condition of integration.

**Theorem 2.1.** *Any function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  that is continuous over a certain region  $B \subset \mathbb{R}^n$  can be integrated over  $B$ .*

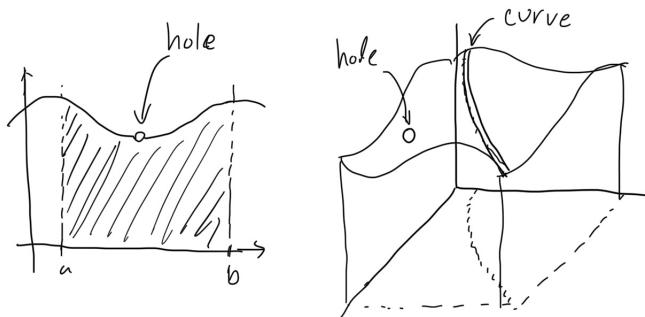
That is, if  $f$  is discontinuous at a certain subset  $D \subset B$ , then the infinitesimal neighborhoods around each point  $d \in D$  is not well defined, since they would always contain two values of  $f$  that do not converge to each other at  $d$ .

However, there are some discontinuous functions that are in fact integrable. Assuming  $B \subset \mathbb{R}^n$  is the region that we are integrating over,

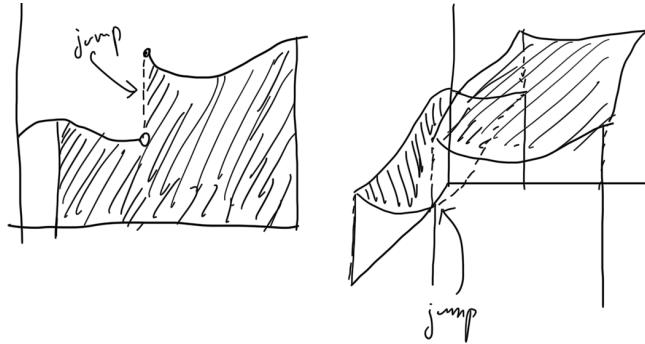
- Given that there is a subset  $N$  in  $B$  with volume 0 over which  $f$  is not defined, we can integrate over  $B \setminus N$ . In the one and two dimensional cases,

$$\int_{B \setminus N} f(x) dx \text{ and } \iint_{B \setminus N} f(x) dA$$

are well-defined. Visually,



2. The function is defined for all values in the region, but there is a jump in the value of the function.



Informally, if we can visualize the Riemann sum converging to a well-defined area as the rectangles get thinner and thinner, then a discontinuous function is integrable. Indeed, all continuous functions (over a bounded set) are integrable since their Riemann sums are well defined.

### 2.2.1 Integration over Intervals, Rectangles, Boxes

The simplest region that we can integrate over is a single interval

$$B \equiv [a, b] \subset \mathbb{R}$$

a rectangle

$$B \equiv [a, b] \times [c, d] \subset \mathbb{R}^2$$

and a box

$$B \equiv [a, b] \times [c, d] \times [e, f] \subset \mathbb{R}^3$$

Clearly, this extends to integration over any dimension.

$$B \equiv \prod_{i=1}^n [\alpha_i, \beta_i] \subset \mathbb{R}^n$$

Solving these integrals are quite simple. However, to rigorously define the methodology, we must use the following theorems.

**Theorem 2.2** (Cavalieri's Principle). *Let  $S$  be a bounded  $n$ -dimensional solid in  $\mathbb{R}^n$  (note that  $S$  can be an interval in  $\mathbb{R}$ ). Define an  $n - 1$  subspace  $P$  in  $\mathbb{R}^n$  and given the quotient space  $\mathbb{R}^n/P$  with elements  $P_x$ , let*

$$S \subset \bigcap_{a \leq x \leq b} P_x$$

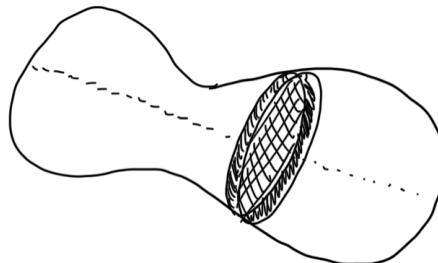
*That is,  $S$  is "in between" affine subspaces  $P_a$  and  $P_b$ . The cross section of  $S$  cut by  $P_x$  is the intersection of it with  $S$*

$$\text{Cross Section at } P_x \equiv P_x \cap S$$

*Denote the area of this cross section as  $A(x)$ . Then,*

$$\text{Volume of } S = \int_a^b A(x) dx$$

This theorem basically says that the volume of  $S$  is the sum of the areas of its infinitesimal cross sections.



This clearly works for an interval in  $\mathbb{R}$ , which is computed by the sum of all its "points" (rigorously speaking, infinitesimally thin intervals). The integral works for a shape in  $\mathbb{R}^2$ , which is computed by the sum of its "line segments" (rigorously speaking, infinitesimally thin rectangles) that add up to the shape. In  $\mathbb{R}^3$ , the solid is computed by the sum of its cross sections (infinitesimally thin "molded" cylinders). This analogy continues into higher dimensions.

Given a solid  $S \subset \mathbb{R}^n$ , it is easy to see that no matter what subspace  $P$  we choose—that is, no matter what orientation we choose to "cut" the solid—the sum of all of its cross sections should be equal to the true volume of  $S$ . In the case when  $S$  is a box in  $\mathbb{R}^n$ , Fubini's theorem states that whether we cut  $S$  up along the  $x_1$ -axis,  $x_2$ -axis, ..., or the  $x_n$ -axis, the symmetry in volume is always preserved. This theorem is really just a specific case of this general symmetry in volume.

**Theorem 2.3** (Fubini's Theorem). *Given a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , let*

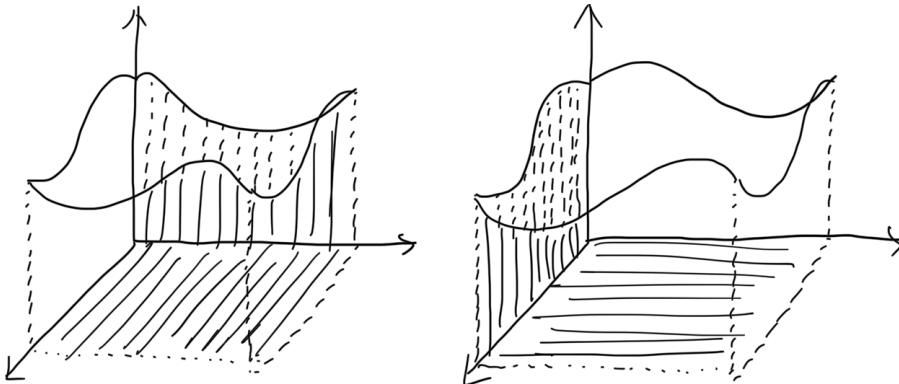
$$B \equiv \prod_{i=1}^n [\alpha_i, \beta_i]$$

and let  $p$  be any permutation of the elements  $\{x_1, x_2, \dots, x_n\}$ . Then

$$\begin{aligned} \int_B f \, dV &= \int_{\alpha_n}^{\beta_n} \cdots \int_{\alpha_1}^{\beta_1} f(x_1, x_2, \dots, x_n) \, dx_1 \dots dx_n \\ &= \int_{p(\alpha_n)}^{p(\beta_n)} \cdots \int_{p(\alpha_1)}^{p(\beta_1)} f(x_1, x_2, \dots, x_n) \, dp(x_1) \dots dp(x_n) \end{aligned}$$

In the two dimensional case, we have

$$\iint_B f \, dA = \int_c^d \int_a^b f(x, y) \, dx \, dy = \int_a^b \int_c^d f(x, y) \, dy \, dx$$



In the three dimensional case, we have

$$\begin{aligned} \iiint_B f \, dV &= \int_e^f \int_c^d \int_a^b f(x, y, z) \, dx \, dy \, dz = \int_e^f \int_a^b \int_c^d f(x, y, z) \, dy \, dx \, dz \\ &= \int_c^d \int_a^b \int_e^f f(x, y, z) \, dz \, dx \, dy = \int_c^d \int_e^f \int_a^b f(x, y, z) \, dx \, dz \, dy \\ &= \int_a^b \int_e^f \int_c^d f(x, y, z) \, dy \, dz \, dx = \int_a^b \int_c^d \int_e^f f(x, y, z) \, dz \, dy \, dx \end{aligned}$$

Computation of these integrals is simple. You do the innermost integral first with respect to the corresponding variable, while treating the rest of the variables constant. Evaluating each integral outputs a formula for a higher dimensional cross section of the solid  $S$ . It is clear that computing iterated integrals is really just doing Cavalieri's principle repeatedly.

### 2.2.2 Integration over Solids Bounded by Curves

We must first define the different types of *elementary regions* first.

**Definition 2.1.** A bounded region  $D$  in  $\mathbb{R}^n$  is said to be  $x_i$ -simple if it is bounded by the graphs of two continuous functions  $u_1, u_2 : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$  of the variables

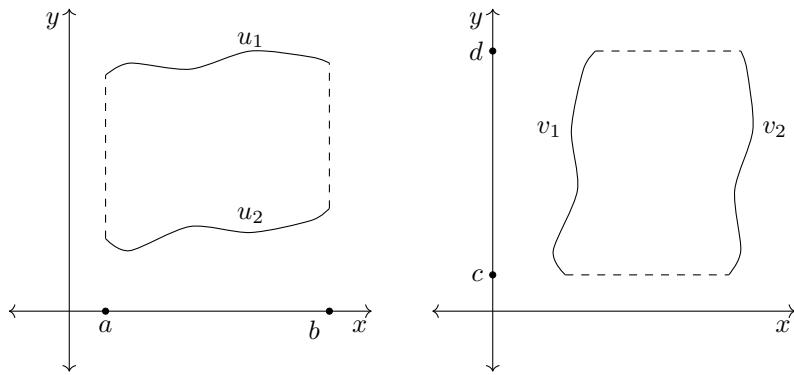
$$x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n$$

That is,  $D$  can be expressed in the form

$$\{x \in \mathbb{R}^n \mid u_1(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \leq x_i \leq u_2(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)\}$$

If a region is simple in all of its variables, it is simply called *simple*. Note that  $n$ -dimensional boxes are simple regions.

**Example 2.1.** In  $\mathbb{R}^2$ , the region on the left graph is an  $y$ -simple region and the region on the right is a  $x$ -simple region.



We now describe the method of calculating double integrals over elementary regions.

**Theorem 2.4.** The double integral over a  $y$ -simple region  $D$  bounded by functions  $u_1$  and  $u_2$  in  $\mathbb{R}^2$  and the  $x$ -values  $a$  and  $b$  (as shown in the left graph of example 2.1) is

$$\iint_D f(x, y) = \int_a^b \int_{u_2(x)}^{u_1(x)} f(x, y) dy dx$$

The double integral over an  $x$ -simple region  $D$  bounded by functions  $v_1$  and  $v_2$  in  $\mathbb{R}^2$  and the  $y$ -values  $c$  and  $d$  (shown in the right of graph of example 2.1) is

$$\iint_D f(x, y) = \int_c^d \int_{v_2(y)}^{v_1(y)} f(x, y) dx dy$$

**Example 2.2.** Integrating  $f(x, y)$  over the unit disk would have the form

$$\int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x, y) dy dx \text{ or } \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} f(x, y) dx dy$$

Note that the unit disk is both  $x$  and  $y$  simple.

### 2.3 Change of Basis

Sometimes, integrating a region over a different basis would make the integral computation much more simpler. In this case, we may be able to transform more complicated regions into elementary regions. We first introduce a change of basis in 2 dimensions and then generalize it into higher dimensions.

Let  $\mathbb{R}^2$  have the standard orthonormal basis  $e_1, e_2$ , commonly known as the  $x, y$  basis. Now, let us

construct new basis vectors of  $\mathbb{R}^2$ , denoted  $f_1, f_2$  such that  $f_1, f_2$  are functions of  $e_1, e_2$ . Since they are both bases that span  $\mathbb{R}^2$ , we can equally represent  $e_1, e_2$  as functions of  $f_1, f_2$ .

$$\begin{aligned} e_1 &= g(f_1, f_2) \\ e_2 &= h(f_1, f_2) \end{aligned}$$

Note that this change of basis does not necessarily have to be linear, as in the context of passive transformation in linear algebra. Then, every point  $(x, y)$  in the  $(e_1, e_2)$ -basis can be rewritten as

$$\begin{aligned} (x, y) &= xe_1 + ye_2 \\ &= xg(f_1, f_2) + yh(f_1, f_2) \\ &= uf_1 + vf_2 \end{aligned}$$

Note that it is customary to denote  $x, y$  as the coefficients in the  $e_1, e_2$  basis and  $u, v$  as the coefficients in the new  $f_1, f_2$  basis. This way, we can not only write  $e_1$  and  $e_2$  as functions of  $f_1$  and  $f_2$ , but we can also write the coefficients  $x, y$  as functions of the coefficients  $u, v$ ! That is,

$$\begin{aligned} x &= x(u, v) \\ y &= y(u, v) \end{aligned}$$

which is really just a function

$$B : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad B(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \end{pmatrix}$$

Notice that  $B$  changes the  $u, v$  coordinates to the  $x, y$  coordinates, and  $B^{-1}$  changes the  $x, y$  coordinates to the  $u, v$  coordinates.

$$B^{-1} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad B^{-1}(x, y) = \begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix}$$

Note that these coefficients actually change *contravariantly*, that is, they change inversely with respect to how the basis vectors are changed. In vector calculus, it is conventional to represent a change of basis with functions that relate the coefficients  $x, y$  with  $u, v$ , rather than the bases  $f_1, f_2$  with  $e_1, e_2$ .

**Theorem 2.5** (Integration over Change of Bases in  $\mathbb{R}^2$ ). *Let  $\mathbb{R}^2$  have the standard orthonormal basis  $e_1, e_2$ . Now, let us construct new basis vectors of  $\mathbb{R}^2$ , denoted  $f_1, f_2$  such that the coefficients of the vectors in  $\mathbb{R}^2$  are related by the change of basis function*

$$B = \begin{pmatrix} x \\ y \end{pmatrix} \implies B(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \end{pmatrix}$$

*Given region  $D \subset \mathbb{R}^2$  and  $S = B(D)$  is the region transformed by  $B$ , the integral of function  $f(x, y)$  over region  $D$  can be expressed as*

$$\iint_D f(x, y) dA = \iint_S f(x(u, v), y(u, v)) |JB(u, v)| d\bar{A}$$

*where  $|JB(u, v)|$  is the determinant of the Jacobian matrix of  $B$ . Expanding the Jacobian determinant gives*

$$|JB(u, v)| = \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u}$$

**Theorem 2.6** (Integration over Change of Bases in  $\mathbb{R}^3$ ). *Given that we have the change of basis function*

$$B : \mathbb{R}^3 \longrightarrow \mathbb{R}^3, \quad B(u, v, w) = \begin{pmatrix} x(u, v, w) \\ y(u, v, w) \\ z(u, v, w) \end{pmatrix}$$

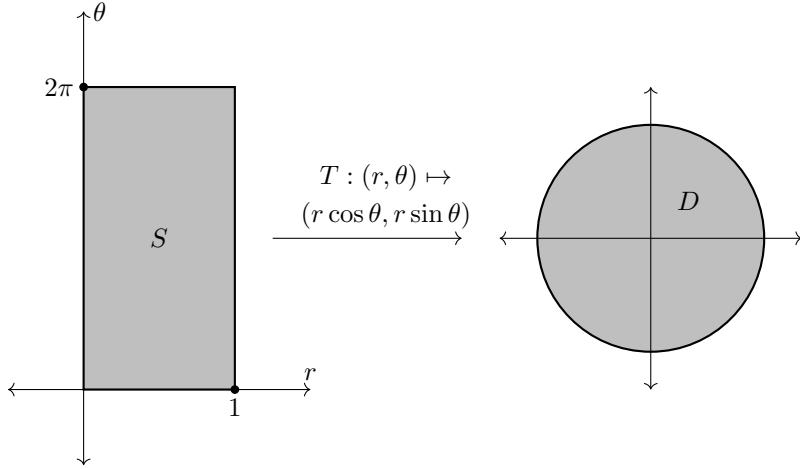
*a region  $D \in \mathbb{R}^3$  and  $S = B(D)$ , the region transformed by  $B$ , the integral of  $f(x, y, z)$  over region  $D$  can be expressed as*

$$\iiint_D f(x, y, z) dV = \iiint_S f(x(u, v, w), y(u, v, w), z(u, v, w)) |JB(u, v, w)| d\bar{V}$$

*where  $|JB(u, v, w)|$  is the Jacobian determinant of  $B$ .*

**Example 2.3.** Given a real-valued function  $f$  defined over the region  $D \subset \mathbb{R}^2$ , we can perform a change of basis of the  $x, y$  coordinates into polar ones within a new region  $S$ . The change of basis

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \end{aligned}$$



**Theorem 2.7** (Integration over Change of Bases in  $\mathbb{R}^n$ ). Let  $\mathbb{R}^n$  have the standard orthonormal basis  $e_1, e_2, \dots, e_n$ , and let us construct a new basis  $f_1, f_2, \dots, f_n$  such that the coefficients of the vectors in  $\mathbb{R}^n$  are related with the functions

$$B : \mathbb{R}^n \longrightarrow \mathbb{R}^n, \quad B(u_1, u_2, \dots, u_n) = \begin{pmatrix} x_1(u_1, \dots, u_n) \\ x_2(u_1, \dots, u_n) \\ \vdots \\ x_n(u_1, \dots, u_n) \end{pmatrix}$$

Given that the region  $D \subset \mathbb{R}^n$  is transformed into a new region  $S = B(D) \subset \mathbb{R}^n$  under this basis transformation, the integral of function  $f(x_1, \dots, x_n)$  over region  $D$  can be expressed as

$$\int_D f(x) dH = \int_S f(x_1(u), x_2(u), \dots, x_n(u)) |JB(u_1, \dots, u_n)| d\bar{H}$$

where the integral on both the left and right hand side represents integration over an  $n$ -dimensional region,  $x$  represents the  $n$ -tuple  $(x_1, \dots, x_n)$ ,  $u$  represents the  $n$ -tuple  $(u_1, \dots, u_n)$ , and  $|JB(u_1, \dots, u_n)|$  represents the Jacobian determinant of function  $B$ .

We now describe some common change of basis formulas for polar, cylindrical, and spherical coordinates.

**Theorem 2.8** (Integration in Polar Coordinates).

$$\iint_D f(x, y) dx dy = \iint_S f(r \cos \theta, r \sin \theta) r dr d\theta$$

**Definition 2.2** (Cylindrical, Spherical Coordinates). In  $\mathbb{R}^3$ , cylindrical coordinates have the following relation to rectangular coordinates.

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \\ z &= z \end{aligned}$$

In  $\mathbb{R}^3$ , spherical coordinates have the following relation to rectangular coordinates.

$$\begin{aligned} x &= \rho \sin \phi \cos \theta \\ y &= \rho \sin \phi \sin \theta \\ z &= \rho \cos \phi \end{aligned}$$

**Corollary 2.8.1** (Integration in Cylindrical Coordinates).

$$\iiint_D f(x, y, z) dx dy dz = \iiint_S f(r \cos \theta, r \sin \theta, z) r dr d\theta dz$$

**Corollary 2.8.2** (Integration in Spherical Coordinates).

$$\iiint_D f(x, y, z) dx dy dz = \iiint_S f(\rho \sin \phi \cos \theta, \rho \sin \phi \sin \theta, \rho \cos \phi) \rho^2 \sin \theta d\rho d\theta d\phi$$

**Example 2.4** (Gaussian Integral). *The following is the (un-normalized) probability distribution function of the Gaussian distribution.*

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

## 2.4 Average Values, Centers of Mass

**Definition 2.3.** The *average value* of a function defined over a region  $D \subset \mathbb{R}^n$  is

$$[f]_{av} = \left( \int_D f(x) dV \right) / \left( \int_D dV \right)$$

where both integrals represent integration over the  $n$ -dimensional region  $D$ . Informally, the integral above represents the infinitesimal sum of all the values of the function  $f$  over  $D$  and divides it by the hypervolume of  $D$  to average it out. More specifically, the average value of  $f : \mathbb{R} \rightarrow \mathbb{R}$  in the interval  $[a, b]$  is defined

$$[f]_{av} = \left( \int_a^b f(x) dx \right) / \left( \int_a^b dx \right) = \frac{1}{b-a} \int_a^b f(x) dx$$

For a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  over a two dimensional region  $D$ , we have

$$[f]_{av} = \left( \iint_D f(x, y) dx dy \right) / \left( \iint_D dx dy \right)$$

For  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  over a three dimensional region  $V$ , we have

$$[f]_{av} = \left( \iiint_V f(x, y, z) dx dy dz \right) / \left( \iiint_V dx dy dz \right)$$

It is quite easy to get the center of mass of a system of  $n$ -distinct points in  $\mathbb{R}^n$ . We can solve each  $x_i$  coordinate for the center of mass by averaging out the  $x_i$  coordinates scaled by their respective masses. That is, given points  $x_1, x_2, \dots, x_n$  with respective masses  $m_1, \dots, m_n$ , the center of mass is defined as

$$\bar{x} = \frac{\sum m_i x_i}{\sum m_i}$$

**Definition 2.4.** Given an  $n$ -dimensional continuous mass density distribution, denoted  $\delta(x)$ , defined over a region  $D \subset \mathbb{R}^n$ , the center of mass of  $D$  can be determined through coordinates.

$$\bar{x}_i = \left( \int_D x_i \delta(x) dV \right) / \left( \int_D \delta(x) dV \right), \quad i = 1, 2, 3, \dots, n$$

Note that  $\delta$  must be continuous (in order for it to be integrable). More specifically, the center of mass of a one dimensional interval  $I \subset \mathbb{R}$  is

$$\bar{x} = \left( \int_I x \delta(x) dx \right) / \left( \int_I \delta(x) dx \right)$$

For a two dimensional region (which we can visualize as a "disk" or "plate," the  $x$  and  $y$  coordinates for the center of mass is

$$\begin{aligned} \bar{x} &= \left( \iint_D x \delta(x, y) dx dy \right) / \left( \iint_D \delta(x, y) dx dy \right) \\ \bar{y} &= \left( \iint_D y \delta(x, y) dx dy \right) / \left( \iint_D \delta(x, y) dx dy \right) \end{aligned}$$

For a three dimensional mass, the  $x, y, z$  coordinates of the center of mass of volume  $V$  can be found with

$$\begin{aligned}\bar{x} &= \left( \iiint_V x \delta(x, y, z) dx dy dz \right) / \left( \iiint_V \delta(x, y, z) dx dy dz \right) \\ \bar{y} &= \left( \iiint_V y \delta(x, y, z) dx dy dz \right) / \left( \iint_V \delta(x, y, z) dx dy dz \right) \\ \bar{z} &= \left( \iiint_V z \delta(x, y, z) dx dy dz \right) / \left( \iint_V \delta(x, y, z) dx dy dz \right)\end{aligned}$$

## 2.5 Improper Integrals

There are generally two types of improper integrals.

1. The region  $D$  integrated over is unbounded.
2. The function  $f$  that is integrated is unbounded within the region  $D$ .

### 2.5.1 Single Variable Improper Integrals

These types of improper integrals are usually evaluated using a limiting process. When the interval  $I$  is unbounded, say  $(1, \infty)$ , the integral can be evaluated as

$$\int_1^\infty \frac{1}{x^2} dx = \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^2} dx = \lim_{b \rightarrow \infty} \left( 1 - \frac{1}{b} \right) = 1$$

In case 2, we can add a limit at the point where the function  $f$  diverges as such.

$$\int_0^1 \frac{1}{\sqrt{x}} dx = \lim_{a \rightarrow 0} \int_a^1 \frac{1}{\sqrt{x}} dx = \lim_{a \rightarrow 0} (2 - 2\sqrt{a}) = 2$$

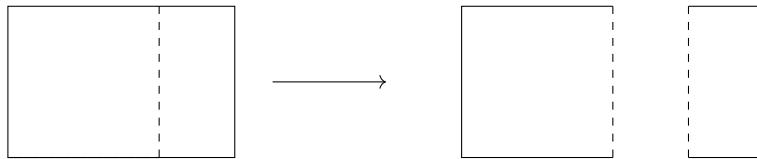
We now describe how to integrate over a certain path  $p$  embedded in a higher dimensional space  $\mathbb{R}^n$ , possibly with a scalar or vector field  $f$ . We must first go over oriented paths.

### 2.5.2 Two Variable Improper Integrals

Extending the previous case, we use a multivariate limiting process in  $\mathbb{R}^2$ . We will first work with case 2, when  $f$  is unbounded within the region  $D$ . Let us define an elementary region  $D$  in  $\mathbb{R}^2$ ; without loss of generality, we will make it  $y$ -simple, meaning that  $D$  can be expressed as

$$D \equiv \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, \phi_1(x) \leq y \leq \phi_2(x)\}$$

We can actually assume that the region in which  $f$  is unbounded lies in the boundary  $\partial D$ . This is because if it lied in the interior of  $D$ , we could split  $D$  into pieces across a path that intersects this region with divergent values, evaluate the integrals over the pieces separately, and then sum the integrals. For example, in the rectangular region below, let the dashed line represent the values where the function  $f$  diverges. Then, we can split the region into two rectangular regions shown in the right.



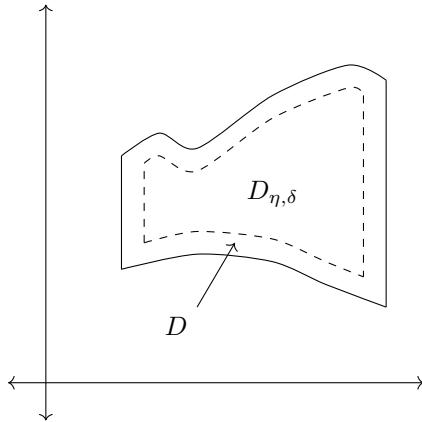
Therefore, assuming that  $f$  is unbounded in  $\partial D$ , we can construct a new region

$$D_{\eta, \delta} \equiv \{(x, y) \in \mathbb{R}^2 \mid a + \eta \leq x \leq b - \eta, \phi_1(x) + \delta \leq y \leq \phi_2(x) - \delta\}$$

for some arbitrarily small numbers  $\eta, \delta > 0$ , meaning that the integral (reduced to iterated integrals using Fubini's theorem)

$$F(\eta, \delta) \equiv \iint_{D_{\eta, \delta}} f(x, y) dA = \int_{a+\eta}^{b-\eta} \int_{\phi_1(x)+\delta}^{\phi_2(x)-\delta} f(x, y) dy dx$$

is well defined.



Clearly, the function  $F(\eta, \delta)$  is a function of two variables  $\eta$  and  $\delta$ . So, if the limit

$$\lim_{(\eta, \delta) \rightarrow (0, 0)} F(\eta, \delta)$$

is well defined, then so is the improper integral. For it to exist, the iterated limits must both equal to a well-defined real number  $\mathcal{L}$  (and to each other). That is,

$$\lim_{\eta \rightarrow 0} \lim_{\delta \rightarrow 0} F(\eta, \delta) = \lim_{\delta \rightarrow 0} \lim_{\eta \rightarrow 0} F(\eta, \delta) = \mathcal{L} \implies \lim_{(\eta, \delta) \rightarrow (0, 0)} F(\eta, \delta) = \mathcal{L}$$

It is also worthwhile to note that functions unbounded at isolated points can be evaluated using the methods above using a change of basis. Consider the example below.

**Example 2.5.** In the unit disk  $D \subset \mathbb{R}^2$ , let the function  $f$  be defined as

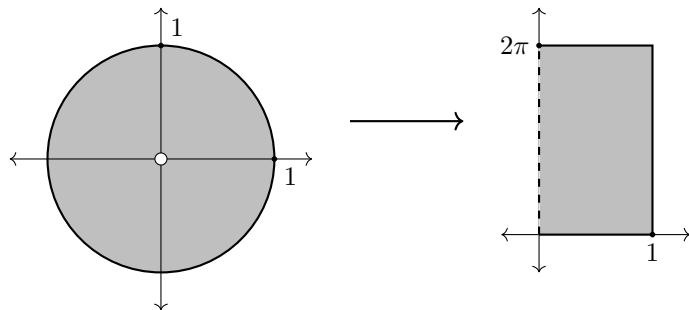
$$f(x, y) \equiv \frac{1}{\sqrt{x^2 + y^2}}$$

Clearly,  $f$  is continuous at every point except  $0 = (0, 0)$ , meaning that

$$\iint_{D \setminus \{0\}} f(x, y) dA$$

is well-defined. In order to solve the integral over the entire disk, we convert to polar coordinates and evaluate the limit

$$\iint_{D \setminus \{0\}} f(x, y) dA = \lim_{\delta \rightarrow 0} \int_{\delta}^1 \int_0^{2\pi} r f(r \cos \theta, r \sin \theta) d\theta dr$$



If we are given an unbounded region  $D \subset \mathbb{R}^2$ , we can first create a bounded region and expand that region using a limit to cover all of  $D$ .

## 2.6 Line Integrals

**Definition 2.5** (Orientations, Simple Curves, Closed Curves). A path function  $p : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^n$  determines a curve in  $\mathbb{R}^n$  with endpoints  $p(a)$  and  $p(b)$ . The direction the curve  $p$  takes, that is from  $p(a)$  to  $p(b)$  in  $\mathbb{R}^n$  is called the *orientation* of  $p$ . A path or a curve with a defined orientation is called an *oriented curve*.

A *simple curve*  $C$  to be the image of an injective piecewise  $C^1$  map  $c : I \subset \mathbb{R} \rightarrow \mathbb{R}^3$ . Since it is injective, it does not intersect itself, and  $C$  is piecewise smooth in  $\mathbb{R}^n$ . If  $I = [a, b]$ , then  $c(a)$  and  $c(b)$  are the endpoints of the curve. A simple curve with an orientation is called an *oriented simple curve*.

A closed curve  $C$  is the image of piecewise  $C^1$  map  $c : [a, b] \rightarrow \mathbb{R}^n$  such that  $c(a) = c(b)$ . That is, the endpoints of  $C$  are equal. A *simple closed curve* is a closed curve that is injective over the interval  $[a, b]$ . Note that a closed curve has two possible orientations.

If  $C$  is an oriented simple curve or an oriented simple closed curve, then we can unambiguously define line integrals along them.

**Definition 2.6.** Let  $h$  be an injective function that takes  $[\alpha, \beta] \subset \mathbb{R}$  to the interval  $[a, b] \subset \mathbb{R}$ . Given an oriented simple path function  $p : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^n$ , the composition

$$\rho = p \circ h : [\alpha, \beta] \rightarrow \mathbb{R}^n$$

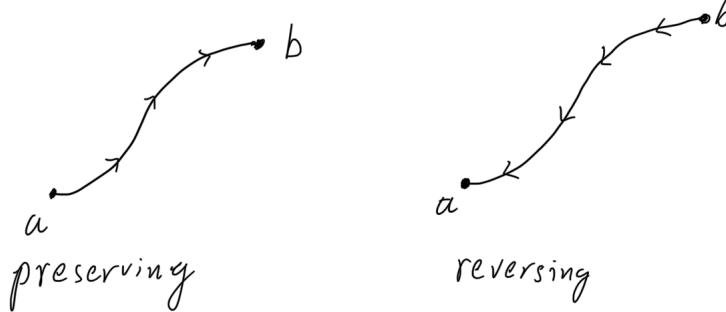
is called a *reparamaterization* of  $p$ . Note that since  $h$  is injective, it takes endpoints to endpoints. If  $h$  preserves the direction in which the path travels, that is, if

$$(p \circ h)(\alpha) = a \text{ and } (p \circ h)(\beta) = b$$

then  $h$  is *orientation preserving*. If

$$(p \circ h)(\alpha) = b \text{ and } (p \circ h)(\beta) = a$$

then  $h$  is *orientation reversing*. Note that a path  $c$  having the same image as  $p$  does not imply that  $c$  is a reparamaterization of  $p$ , since  $c$  may not be injective.



**Definition 2.7** (Scalar Line Integral). Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , which can be interpreted as a scalar field. Now define a  $C^1$  path function

$$c : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^n$$

such that the composition of functions

$$f \circ c : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^n$$

is continuous. Then, the *path integral*, or *scalar line integral*, of  $f$  along the path  $c$ . is defined

$$\begin{aligned} \int_c f \, ds &= \int_a^b f(c(t)) \|c'(t)\| \, dt \\ &= \int_a^b f(x_1(t), x_2(t), \dots, x_n(t)) \|c'(t)\| \, dt \end{aligned}$$

If  $c(t)$  is only piece-wise  $C^1$ , we can define the path integral by breaking  $[a, b]$  into pieces over with  $f(c(t))\|c'(t)\|$  is continuous and then summing the integrals over the pieces. That is,

$$\int_a^b f(c(t)) \|c'(t)\| \, dt = \sum_{i=0}^{n-1} \int_{\alpha_i}^{\alpha_{i+1}} f(c(t)) \|c'(t)\| \, dt$$

Note that since  $f$  is a scalar-valued function, we can interpret a path integral as the sum of infinitesimal segments of the path  $c$  having a weight determined by  $f$  at each section. If  $f$  is a constant function outputting 1 at every point, then the path integral just outputs the length of the path  $c$  in  $\mathbb{R}^n$ .

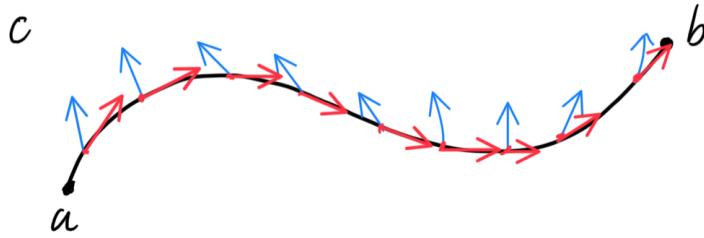
$$L = \int_a^b f(c(t)) \|c'(t)\| dt = \int_a^b \|c'(t)\| dt$$

**Definition 2.8** (Vector Line Integral). Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a vector field on  $\mathbb{R}^n$  that is continuous on the  $C^1$  oriented path  $c : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^n$ . The *line integral* of  $F$  along  $c$  is defined by the formula

$$\int_c F \cdot ds = \int_a^b F(c(t)) \cdot c'(t) dt$$

where  $\cdot$  represents the dot product of  $F$  with  $c'$  over the interval  $[a, b]$ . It is also commonly written in differential notation,

$$\int_c F \cdot ds = \int_c F \cdot (dx_1, \dots, dx_n) = \int_c F_1 dx_1 + F_2 dx_2 + \dots + F_n dx_n$$



Similarly with path integrals, we can also define line integrals as the sum of integrals over piece-wise continuous sections of  $c$ . That is, given an oriented curve  $C$  made up of several oriented component curves  $C_i$ ,  $i = 1, 2, \dots, k$ , we can parameterize  $C$  by parameterizing the pieces  $C_i$ 's separately. Thus, we can treat  $C = C_1 + \dots + C_k$  and get

$$\int_C F \cdot ds = \sum_{i=1}^k \int_{C_i} F \cdot ds$$

Note that a vector line integral is a generalization of scalar line integrals, so any results holding for vector line integrals also holds for their scalar counterpart.

**Example 2.6** (Work). In mechanics, work  $W$  is defined as

$$W = F \cdot d$$

where  $F$  is force and  $d$  is displacement. With this knowledge, the reader can easily see that the work done by vector field  $F$  on a particle traveling along a path  $c$  from time  $a$  to time  $b$  can be calculated by the line integral

$$\begin{aligned} W &= \int_a^b F(c(t)) \cdot c'(t) dt \\ &= \int_c F_1 dx + F_2 dy + F_3 dz \end{aligned}$$

**Theorem 2.9** (Invariance of Path Parameterizations on Vector Line Integrals). Let  $F$  be a vector field and  $f$  be a scalar field, both continuous on the  $C^1$  path function  $p : [a, b] \rightarrow \mathbb{R}^n$  and let  $q : [\alpha, \beta] \rightarrow \mathbb{R}^n$  be a reparameterization of  $p$ . Then,

$$\begin{aligned} q \text{ is orientation preserving} &\implies \int_p F \cdot ds = \int_q F \cdot ds \\ q \text{ is orientation reversing} &\implies \int_p F \cdot ds = - \int_q F \cdot ds \end{aligned}$$

### 2.6.1 Conservative Vector Fields

We now introduce a fundamental theorem about line integrals over gradient fields. Recall the fundamental theorem of calculus and its equivalent form.

**Theorem 2.10** (Fundamental Theorem of Single Variable Calculus). *Let function  $\nabla g : \mathbb{R} \rightarrow \mathbb{R}$  be the gradient of the single variable  $C^1$  function  $g : \mathbb{R} \rightarrow \mathbb{R}$ ; that is,  $\nabla g$  is a conservative vector field on  $\mathbb{R}$ . Then,*

$$\int_a^b \nabla g(x) dx = g(b) - g(a)$$

Note that in the single variable case,

$$\frac{d}{dx} g(x) = \nabla g(x)$$

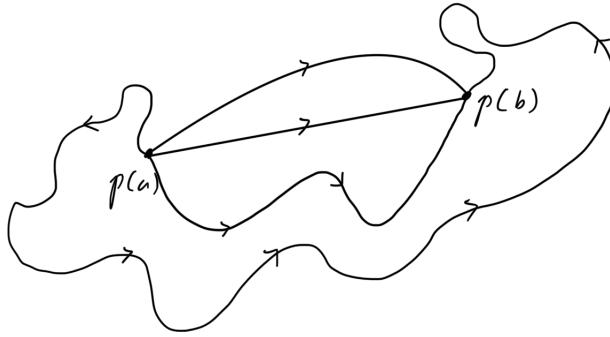
This means that the value of the integral of  $\nabla g$  only depends on the value of  $g$  at the endpoints of the interval  $[a, b]$ .

We can extend this to line integrals for functions mapping  $\mathbb{R}^n$  to  $\mathbb{R}$ .

**Theorem 2.11** (Invariance of Line Integrals in Conservative Vector Fields). *Given that  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a  $C^1$  conservative vector field with  $\nabla f = F$  for  $C^2$  function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and path function  $p : [a, b] \rightarrow \mathbb{R}^n$  is a piecewise  $C^1$  path, then*

$$\int_p F \cdot ds = \int_p \nabla f \cdot ds = f(p(b)) - f(p(a))$$

That is, the line integral of any path in a conservative vector field is dependent on the value of  $f$  at the endpoints  $p(a)$  and  $p(b)$ .



In physics, calculating the work done by a force represented by a vector field requires us to know the path that it travels through.

$$W = \int_p F \cdot ds$$

However, in many cases  $F$  is assumed to be conservative, so it is only necessary that we find the displacement of the particle from its endpoints, resulting in the simplification of the formula.

$$W = \int_p \nabla f \cdot ds = f(p(b)) - f(p(a))$$

**Corollary 2.11.1** (Equivalent Conditions for Vector Field to be Conservative). *The following conditions are equivalent:*

1.  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a conservative vector field.
2. The line integral of  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  in curve  $C$  is path independent; that is, if  $C_1$  and  $C_2$  are two parameterizations of  $C$ ,

$$\int_{C_1} F \cdot ds = \int_{C_2} F \cdot ds$$

3. Given that  $C$  is a closed loop, the line integral of  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  across  $C$  is 0.

$$\oint_C F \cdot ds = 0$$

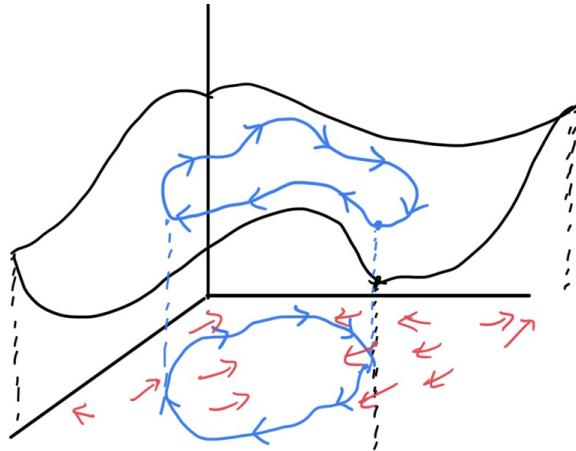
4. The curl of  $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  vanishes

$$\operatorname{curl} F = \nabla \times F = \begin{pmatrix} \frac{\partial F}{\partial z} \\ \frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial x} \end{pmatrix} \times \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = 0$$

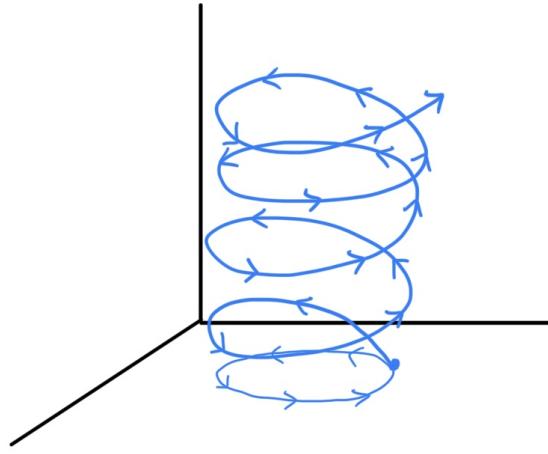
5. The following partial derivatives of  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  are equal

$$\frac{\partial F_1}{\partial y} = \frac{\partial F_2}{\partial x}$$

We can develop a bit of intuition to determine whether a vector field is conservative or not. If vector field  $F$  is conservative, then there exists a smooth scalar field  $f$  such that  $\nabla f = F$ . For each latitude and longitude on a certain map, we can give it an altitude as a function of those coordinates (picture a map with a bunch of hills and valleys). The gradient and thus the vector field is all the vectors that point in the direction of highest ascent. The vector field is all the vectors that point in the direction of highest descent. Extending the metaphor the path integral is like starting on at a point and climbing the hills and valleys, creating work as you go up a hill (proportional to the steepness and thus the dot product of your motion vector with the gradient vector field in the path integral) and decreasing the work you put in by going down a hill. Since the path is closed, it is like you are going up and down the same amount overall, so the path integral is zero. Following this analogy, the vector field determined by this function (marked as arrows in the  $x, y$  plane) is conservative.



If we can construct a closed loop around  $F$  where the line integral is nonzero, then it means that we have ended up at a "higher" or "lower" (altitude) at the same point. This means that rather than being a certain landscape, there exist different "levels" of values at one point, like a spiraling staircase. For example, look at the solenoidal vector field below, where we can construct a closed loop (a circle going around the origin counterclockwise). There is no "surface" that can be defined such that it contains the solenoid.



Clearly, as a particle travels through the vector field along the path, it does positive work while it has zero displacement, and clearly, there exists no function that can output both these values as determined by vector field  $F$ .

**Theorem 2.12** (Helmholtz Decomposition). *Let  $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be a  $C^2$  vector field. Then,  $F$  can be decomposed into a curl-free component and a divergence-free component. That is, there exists vector fields  $A$  and  $\Phi$*

$$F = -\nabla \cdot \Phi + \nabla \times A$$

## 2.6.2 Curvature

**Definition 2.9** (Curvature at a Point). Let  $c : [a, b] \rightarrow C \subset \mathbb{R}^3$  be a unit-speed parameterization of  $C$ , meaning that  $\|c'(t)\| = 1$  for all  $t \in [a, b]$ , and let  $p = c(t_0)$  be a point in  $C$ . The *curvature*  $\kappa(p)$  at  $p$  is a mapping defined

$$\kappa : C \rightarrow \mathbb{R}, \quad \kappa(p) \equiv \|c''(t_0)\|$$

Notice that since we require a unit speed parameterization of  $C$ , we do not need to worry about how a given curve is parameterized.

Since the curvature is defined pointwise for each point in curve  $C$ , we can integrate over all the curvatures in  $C$  to define the total curvature.

**Definition 2.10** (Total Curvature). The *total curvature* of a curve  $c : [a, b] \rightarrow C \subset \mathbb{R}^3$  is the scalar line integral

$$\int_C \kappa ds$$

We now present an important theorem in differential geometry.

**Theorem 2.13** (Fary-Milnor Theorem). *Given a unit speed parameterization  $c : [a, b] \rightarrow C \subset \mathbb{R}^3$ , if  $C$  is closed (that is,  $c(a) = c(b)$ ), then*

$$\oint_C \kappa ds \geq 2\pi$$

and equals  $2\pi$  only when  $C$  is a circle. Furthermore, if  $C$  is a closed space curve with

$$\oint_C \kappa ds \leq 4\pi$$

then  $C$  is "unknotted." That is,  $C$  can be continuously deformed without every intersecting itself into a planar circle. Therefore, for knotted curves  $C$ , we have

$$\oint_C \kappa ds > 4\pi$$

## 2.7 Surface Integrals

Surface integrals are the 2-dimensional analogue, or the double integral version, of line integrals. It is the integration of surfaces.

### 2.7.1 2-Dimensional Parameterizations of Surfaces

Just like how we create path functions using a parameterization function  $p : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^n$ , we can parameterize surfaces by defining a function

$$\varphi : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}^n, \quad \varphi(u, v) \equiv \begin{pmatrix} x_1(u, v) \\ \vdots \\ x_n(u, v) \end{pmatrix}$$

The surface

$$S = \varphi(D)$$

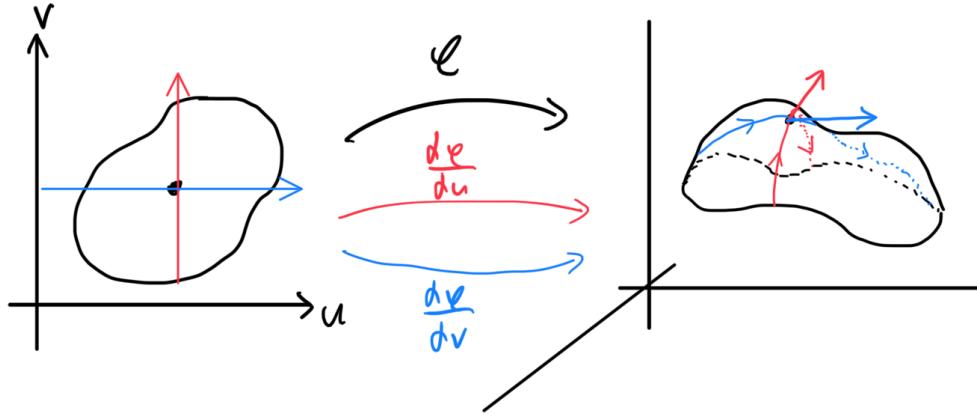
corresponding to the function  $\varphi$  is its image. If  $\varphi$  is differentiable or is of class  $C^1$ , then we call  $S$  a *differentiable* or  $C^1$  surface, respectively.

For those that are familiar with differential geometry, this makes every parameterized surface a 2-manifold induced by the single homeomorphism  $\varphi$ . In fact, it is more than just locally homeomorphic; it is *globally* homeomorphic.

**Definition 2.11** (Tangent Vectors of Surfaces Embedded in  $\mathbb{R}^3$ ). Given surface parameterization

$$\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^3, \quad \varphi(u, v) \equiv \begin{pmatrix} x(u, v) \\ y(u, v) \\ z(u, v) \end{pmatrix}$$

it is visually clear that there can be up to two linearly independent tangent vectors at a point on the surface  $S$ . We can calculate these two vectors by embedding two nonparallel paths in  $D \subset \mathbb{R}^2$  and taking the derivative with respect to a point traveling through these paths, which would give us a tangent vector on  $S$ . To keep things simple, we take the partial derivatives with respect to  $u$  and  $v$ .



Clearly, these paths are functions

$$\frac{\partial \varphi}{\partial u} \equiv \begin{pmatrix} \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial u} \\ \frac{\partial z}{\partial u} \end{pmatrix} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

$$\frac{\partial \varphi}{\partial v} \equiv \begin{pmatrix} \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial v} \end{pmatrix} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

where

$$\frac{\partial \varphi}{\partial u}(u_0, v_0), \quad \frac{\partial \varphi}{\partial v}(u_0, v_0)$$

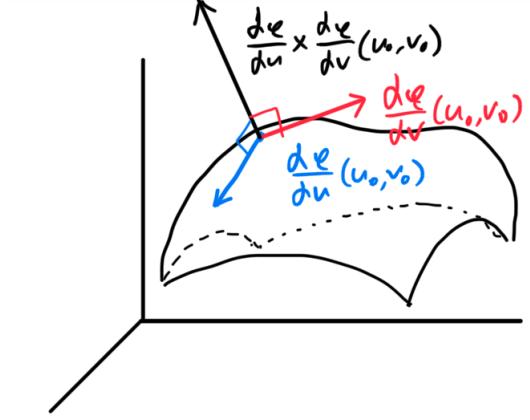
represent two vectors in  $\mathbb{R}^3$  that are tangent to  $S$  at the point  $\varphi(u_0, v_0) \in \mathbb{R}^3$ .

We must make sure that the surface  $S$  is smooth in the sense that (informally) there aren't any wrinkles, points, folds, or self-intersections in such a way that the tangent plane to the surface is not well-defined.

**Definition 2.12** (Regular Surfaces). To formalize this concept, we say that  $S$  is *regular*, or *smooth*, at point  $(u_0, v_0)$  if

$$\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \neq 0$$

where  $\times$  is the Euclidean cross product. That is, if the vector that is orthogonal to the two tangent vectors is well defined at a point, the surface is said to be smooth at that point. Note that  $\frac{\partial \varphi}{\partial u}$  is parallel to  $\frac{\partial \varphi}{\partial v}$  if and only if their cross product is 0.



It is quite clear that  $(\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v})(u_0, v_0) \neq 0 \implies \frac{\partial \varphi}{\partial u}$  and  $\frac{\partial \varphi}{\partial v}$  are linearly independent. This means that an entire span of tangent vectors, i.e. a tangent plane, of the surface  $S$  at  $\varphi(u_0, v_0)$  exists.  $S$  is said to be *regular* if it is regular at all points  $\varphi(u_0, v_0) \in S$ .

In fact, the tangent plane at  $\varphi(u_0, v_0)$  is the set of points

$$\{\varphi(u_0, v_0) + \frac{\partial \varphi}{\partial u}(u_0, v_0)c_1 + \frac{\partial \varphi}{\partial v}(u_0, v_0)c_2 \mid c_1, c_2 \in \mathbb{R}\}$$

which is precisely the affine tangent plane spanned by  $T_u$  and  $T_v$ . Note also that the vector  $T_u \times T_v$ , if nonzero, is normal to this plane, which leads to this equivalent definition.

**Definition 2.13** (Tangent Planes of Surfaces). Given a parameterized surface  $\varphi : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$  that is regular at  $\varphi(u_0, v_0)$ , the tangent plane of the surface  $S$  at  $\varphi(u_0, v_0) = (x_0, y_0, z_0)$  is defined

$$\{(x, y, z) \in \mathbb{R}^3 \mid (x - x_0, y - y_0, z - z_0) \cdot n = 0\}$$

where  $n = (\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v})(u_0, v_0)$ .

We finally construct the concept of signed areas before defining surface integration. We have all the tools we need to calculate surface areas, but remember that integration also covers the concept of *signed areas*, which could be negative. In order to define this, we define the concept of orientation on surfaces.

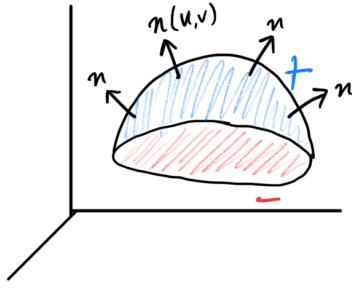
## 2.7.2 Orientation of Surfaces

**Definition 2.14** (Oriented Surfaces). An *oriented surface* is a two-sided surface with one side specified as the *outside/positive* side and the other side as the *inside/negative* side. Note that an oriented surface is not guaranteed to have two sides (e.g. a Möbius strip). To ensure that there exist two sides,  $S$  must be regular.

Surprisingly, a parameterization does not have an intrinsic orientation. Rather, we determine the orientation ourselves by choosing a unit vector that generally points towards the outside of the surface  $S$ . Again, this choice is arbitrary, but it is customary to choose a vector that generally points "out." Either way, the orientation (unit) vector at every point  $\varphi(u, v) \in S$ , denoted as  $n$ , is

$$n(\varphi(u, v)) = \pm \frac{\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v}}{\left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\|}$$

which can be visually calculated using the right hand rule.



**Definition 2.15** (Orientation Preserving, Reversing Parameterizations). Given an oriented surface  $S$  with its positive side determined by the direction of unit vector  $n(\varphi(u, v))$ , the parameterization  $\varphi$  is said to be *orientation preserving* if

$$n(\varphi(u, v)) = \frac{\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v}}{\left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\|}$$

and *orientation reversing* if

$$n(\varphi(u, v)) = -\frac{\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v}}{\left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\|}$$

So, to find whether a parameterization is orientation preserving or reversing, it suffices to find the cross product  $T_u \times T_v$  and see if it points in the same direction of the normal vector  $n$  (which should have already been determined when deciding the orientation of  $S$ ).

Given a parameterization  $\varphi$  and an un-oriented surface  $S$ , we can also just construct  $\varphi$  to be orientation-preserving (or reversing) by *defining* the normal vector  $n$  to be

$$n(\varphi(u, v)) = \frac{\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v}}{\left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\|} \quad \left( \text{or } n(\varphi(u, v)) = -\frac{\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v}}{\left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\|} \right)$$

So rather than finding out whether a parameterization  $\varphi$  is orientation preserving or reversing by comparing  $T_u \times T_v$  with  $n$ , we have defined  $n$  in a way such that  $\varphi$  must be orientation preserving (or reversing). We can utilize these tools of parameterization to now define the surface integral.

### 2.7.3 Scalar, Vector Surface Integrals

A physical interpretation of a scalar surface integral is the weighted surface area of a certain surface.

**Definition 2.16** (Scalar Surface Integrals). Let  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  be a  $C^1$  scalar field defined on a parameterized surface  $S \subset \mathbb{R}^3$  with parameterization  $\varphi : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$ . That is,  $\varphi(D) = S$ . We define the integral  $f$  over  $S$  to be

$$\begin{aligned} \iint_S f \, dS &= \iint_S f(x, y, z) \, dS \\ &= \iint_D f(\varphi(u, v)) \left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\| du \, dv \end{aligned}$$

Note that this will require us to transform  $f$ , a function of  $x, y, z$ , into the function  $f \circ \varphi$  of  $u, v$ . Additionally, if the parameterization of the surface  $S$  is not defined, then it one must be constructed. It is also clear that if  $S$  is a union of surfaces  $S_i$ , then its surface integral is the sum of the surface integrals of the  $S_i$ 's.

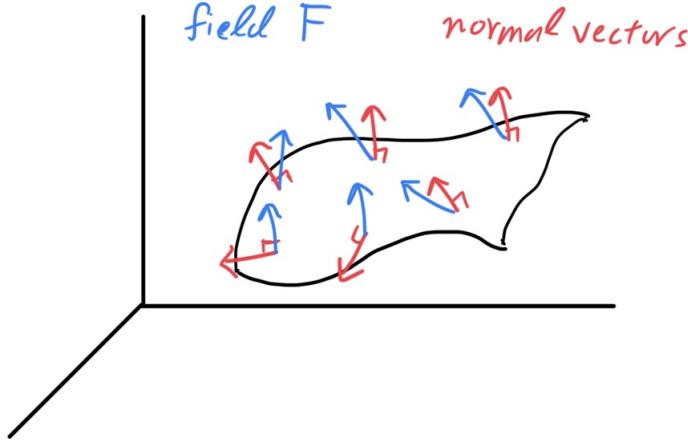
Letting the scalar field  $f$  be the constant field equal to 1, the scalar surface integral measures the surface area of  $S$ .

$$A(S) = \iint_S dS = \iint_D \left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\| du \, dv$$

It is easy to see that the orientation of the parameterization  $\varphi$  does not affect scalar surface integrals, since the sign of the orientation gets nullified by the absolute value sign over  $\left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\|$ .

Its physical interpretation is to measure the rate at which a fluid (determined by a vector field  $F$ ) is crossing a given surface  $S$ . It also has many applications in electromagnetism.

**Definition 2.17** (Vector Surface Integrals). Let  $F$  be a vector field defined on surface  $S$ , the image of a parameterized surface  $\varphi$ . The *surface integral* of  $F$  over  $S$  is defined below, which is equivalent to summing up the dot product of the vector field and the normal vector to the surface.



It can be calculated with the following formulas by converting it into a scalar surface integral where the scalar field is the value of the dot product of the vector field with the normal vectors of the surface.

$$\begin{aligned} \iint_S F \cdot dS &= \iint_S (F \cdot n) dS \\ &= \iint_D \left( F(\varphi(u, v)) \cdot \frac{\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v}}{\left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\|} \right) \left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\| du dv \\ &= \iint_D F(\varphi(u, v)) \cdot \left( \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right) du dv \end{aligned}$$

Since we are now talking about vector fields, the orientation of the parameterization is now significant. Visually, if the orientation of the surface  $S$  generally aligns with the vector field  $F$ , then the integral will be positive (since two vectors  $\alpha, \beta$  generally pointing in the same direction implies that  $\alpha \cdot \beta > 0$ ). The orientation of the parameterization, which is dependent on  $\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v}$ , determines the direction of the normal vector  $n$  (since it is defined to be  $(\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v}) / \left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\|$ ). Therefore, changing the orientation of  $\varphi$  will reverse the direction of  $n$ , which will then reverse the sign of the integral since  $n$  now points in the opposite direction of the vector field  $F$  than it previously did (by reversing the sign of the dot products). This is formalized in the theorem below.

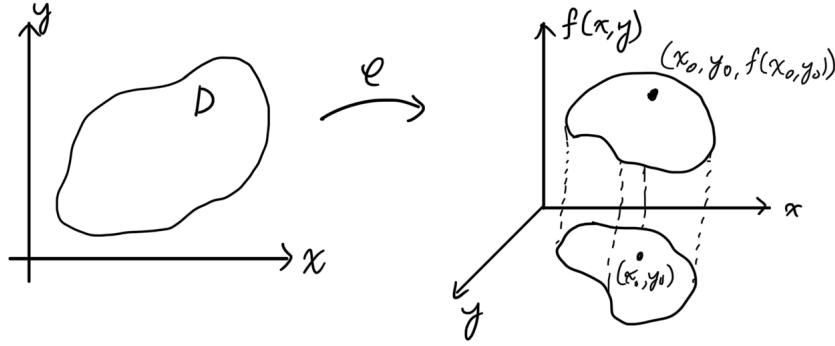
**Theorem 2.14** (Invariance of Surface Parameterizations on Vector Surface Integrals). *Let  $S$  be an oriented surface and let  $\varphi_1$  and  $\varphi_2$  be two regular parameterizations with  $F$  a continuous vector field defined on  $S$ . Then, assuming  $\varphi_1$  is orientation preserving,*

$$\begin{aligned} \varphi_2 \text{ is orientation preserving} &\implies \iint_{\varphi_1} F \cdot dS = \iint_{\varphi_2} F \cdot dS \\ \varphi_2 \text{ is orientation reversing} &\implies - \iint_{\varphi_1} F \cdot dS = \iint_{\varphi_2} F \cdot dS \end{aligned}$$

#### 2.7.4 Surface Integrals over Graphs

Given that we have the graph of a function  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  rather than a general surface, we can parameterize it simply as

$$\varphi(u, v) \equiv (u, v, g(u, v))$$



This means that

$$\frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} = \begin{pmatrix} -\frac{\partial g}{\partial u} \\ -\frac{\partial g}{\partial v} \\ 1 \end{pmatrix} \Rightarrow \left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\| = \sqrt{1 + \left( \frac{\partial g}{\partial u} \right)^2 + \left( \frac{\partial g}{\partial v} \right)^2}$$

So we can simplify the equation for the surface area  $S$  of the graph of  $g$  over the region  $D$  in the  $xy$ -plane, as

$$\begin{aligned} A(S) &= \iint_S dS = \iint_D \left\| \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right\| dA \\ &= \iint_D \sqrt{1 + \left( \frac{\partial g}{\partial u} \right)^2 + \left( \frac{\partial g}{\partial v} \right)^2} du dv \end{aligned}$$

With the same  $g$ , we can find the weighed surface area of  $S$  over the scalar function  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  with the formula

$$\iint_S f \, dS = \iint_D f(u, v, g(u, v)) \sqrt{1 + \left( \frac{\partial g}{\partial u} \right)^2 + \left( \frac{\partial g}{\partial v} \right)^2} du dv$$

Finally, with the same graph  $g$ , the surface integral over the vector field  $F$  is

$$\begin{aligned} \iint_S F \cdot dS &= \iint_D F(\varphi(u, v)) \cdot \left( \frac{\partial \varphi}{\partial u} \times \frac{\partial \varphi}{\partial v} \right) du dv \\ &= \iint_D \left( F_1(u, v) \left( -\frac{\partial g}{\partial u} \right) + F_2(u, v) \left( -\frac{\partial g}{\partial v} \right) + F_3(u, v) \right) du dv \end{aligned}$$

## 2.8 Integral Theorems

Recall the differential notation for writing line integrals. For 2 and 3 dimensions, it is written as

$$\begin{aligned} \int_C F \cdot ds &= \int_C F \cdot (dx, dy) = \int_C F_1 dx + F_2 dy \\ \int_C F \cdot ds &= \int_C F \cdot (dx, dy, dz) = \int_C F_1 dx + F_2 dy + F_3 dz \end{aligned}$$

### 2.8.1 Green's Theorem

Green's Theorem gives the relationship between a line integral around a simple closed curve  $C$  and a double integral over the plane region  $D$  bounded by  $C$ .

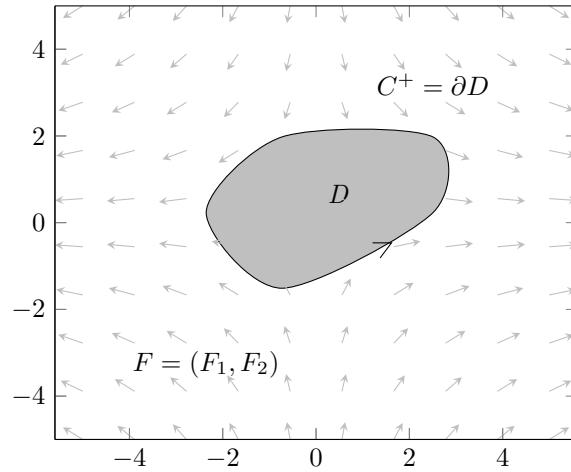
**Theorem 2.15** (Green's Theorem in  $\mathbb{R}^2$ ). *Let there be a 2-dimensional  $C^1$  vector field  $F$  on  $\mathbb{R}^2$  defined on a simple oriented closed piecewise-smooth curve  $C$  and its bounded region  $D \subset \mathbb{R}^2$  (that is,  $C = \partial D$ ). Let the orientation of the path of  $C$  be such that it is traveling counterclockwise, i.e. a point traveling through  $C$  would see the region  $D$  to its left, denoted as  $C^+$  and the clockwise orientation as  $C^-$ . Then,*

$$\oint_{C^+} F_1 dx + F_2 dy = \iint_D \left( \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) dx dy$$

By reversing the orientation, it is clear that we have

$$\oint_{C^-} F_1 dx + F_2 dy = - \iint_D \left( \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) dx dy$$

Note that this theorem is expressed in terms of the components of the vector field  $F$ .



Green's theorem has many applications in physics. For example, in order to solve two-dimensional flow integrals measuring the sum of fluid outflowing from a volume, Green's theorem allows us to calculate the total outflow summed about an enclosing area .

**Corollary 2.15.1.** Let  $D$  be a region for which Green's theorem applies with positively oriented boundary  $\partial D$ . Then, the area of  $D$  can be computed with the formula

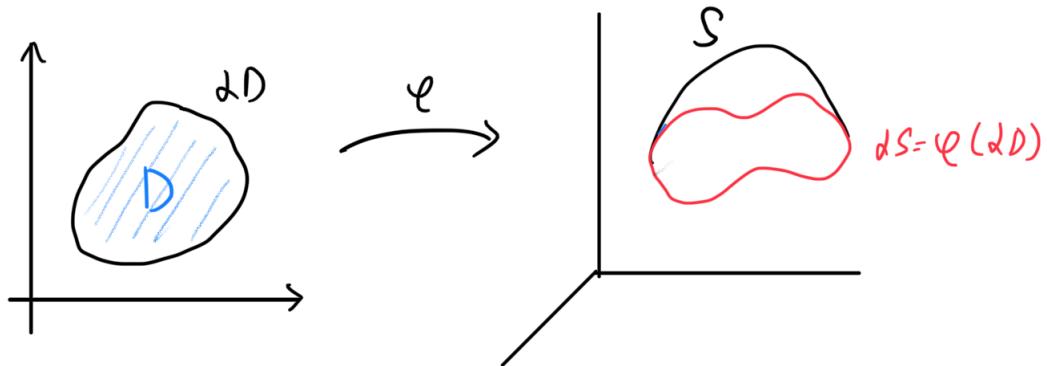
$$A(D) = \frac{1}{2} \oint_{\partial D} x dy - y dx$$

Green's theorem can be used to determine the area of centroid of plane figures solely by integrating over the perimeter.

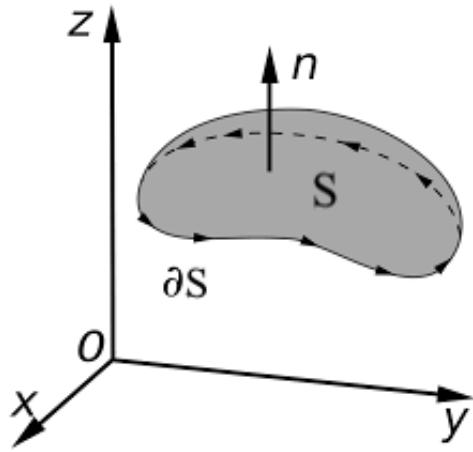
### 2.8.2 Stokes' Theorem

Green's theorem relates line integrals to double integrals. Stokes' theorem generalizes Green's theorem by relating line integrals to surface integrals of 2-dimensional surfaces embedded in  $\mathbb{R}^3$ .

**Theorem 2.16** (Stokes' Theorem). Let  $S$  be an oriented regular surface defined by parameterization  $\varphi : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$ , and let the image of the boundary  $\partial D$  under  $\varphi$  be the boundary  $\partial S$  of  $S$ . We can interpret  $\partial S$  as a path mapping from  $\mathbb{R} \rightarrow S \subset \mathbb{R}^3$ .



The orientation unit vector  $n$  of  $S$  induces the positive orientation of  $\partial S$ , denoted  $\partial S^+$ . Visually, if you are walking along the curve with your head is pointing in the same direction as the unit normal vectors while the surface is on the left then you are walking in the positive direction on  $\partial S$ .



Given that \$F\$ is a \$C^1\$ vector field defined on \$S\$, then

$$\iint_S \operatorname{curl} F \cdot dS = \iint_S (\nabla \times F) \cdot dS = \oint_{\partial S^+} F \cdot ds$$

If \$S\$ has no boundary, that is, if the image of \$p' = \partial S\$ is not a simple closed curve, then the integral is 0.

The above theorem implies that the vector surface integral of a surface without a boundary (i.e. a closed graph, such as a sphere) is always 0 along the curl of any \$C^1\$ field. Geometrically, this means that given a closed solid \$S\$ with field \$\nabla \times F\$, the rate of flow of the vector field into \$S\$ is equal to the flow out of \$S\$.

### 2.8.3 Gauss' Theorem

The divergence theorem relates the flux of a vector field through a closed surface to the divergence of the field in the volume enclosed.

**Theorem 2.17** (Gauss' Divergence Theorem). *Let \$V\$ be a subset of \$\mathbb{R}^3\$. Denote by \$\partial V\$ the oriented closed surface that bounds \$V\$ (with outward pointing normal orientation vectors), and let \$F\$ be a \$C^1\$ vector field defined on a neighborhood of \$V\$. Then,*

$$\iiint_V \operatorname{div} F \, dV = \iiint_V (\nabla \cdot F) \, dV = \iint_{\partial V} F \cdot dS = \iint_{\partial V} (F \cdot n) \, dS$$

where the two left-most integrals are volume integrals, and the two right-most integrals are surface integrals. Intuitively, this makes sense; the volume integrals represent the total of the sources in volume \$V\$, and the right hand side represents the total flow across the boundary \$\partial V\$.

