Shifrin MV

Runi Malladi

June 24, 2023

1 the derivative

1.1 partial, directional derivatives

1.1. The goal is to understand a function by knowing how it varies with each variable.

Definition 1.2. Consider a function $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, where $U \subset \mathbb{R}^n$ is open. Let $a \in U$. For $j = 1, \ldots, n$, define

$$\frac{\partial f}{\partial x_j}(a) = D_j f(a) = \lim_{t \to 0} \frac{f(a + te_j) - f(a)}{t}$$

provided this limit exists.

1.3. The partial derivatives are measuring rate of change along the coordinate axes e_j. We can generalize the above definition to an arbitrary direction (vector!) v:

Definition 1.4. Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, where $U \subset \mathbb{R}^n$ is open. Let $a \in U$. For a nonzero $v \in \mathbb{R}^n$, define the *directional derivative* of f at a in the direction of v to be

$$D_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t}$$

provided this limit exists.

Remark 1.5. $D_j f(a) = \frac{\partial f}{\partial x_j}(a) = D_{e_j} f(a)$.

1.6. The directional derivative depends not only on the direction of v, but also the magnitude of v:

Proposition 1.7. $D_{cv} f(a) = c D_v f(a)$.

Proof.

$$\begin{aligned} D_{cv} f(a) &= \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t} = c \lim_{t \to 0} \frac{f(a + tv) - f(a)}{ct} \\ &= c \lim_{s \to 0} \frac{f(a + sv) - f(a)}{s} = c D_v f(a). \end{aligned}$$

□

Remark 1.8. So $D_v f(a)$ is not merely the instantaneous rate of change by an observer at a in the direction of v, but perhaps the instantaneous rate of change of an observer at a moving with velocity v (not only the direction of the observer matters, but also their speed).

Remark 1.9. In cases where $f : U \subset \mathbb{R}^n \to \mathbb{R}$ consists of functions we can easily differentiate using single-variable calculus rules, the following substitution can help in computing the $D_v f(a)$:

$$\begin{aligned} \phi &: \mathbb{R} \to \mathbb{R} \\ t &\mapsto f(a+tv). \end{aligned}$$

Then

$$D_v f(a) = \lim_{t\to 0} \frac{f(a+tv) - f(a)}{t} = \lim_{t\to 0} \frac{\phi(t) - \phi(0)}{t} = \phi'(0).$$

1.2 differentiability

Example 1.10. For a function $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, the existence of all partial derivatives, even all directional derivatives, does not imply continuity.

Consider the function

$$\begin{aligned} f &: \mathbb{R}^2 \to \mathbb{R} \\ (x,y) &\mapsto \frac{xy^2}{x^2+y^4}. \end{aligned}$$

finish

1.11. Recall that the derivative of $f : \mathbb{R} \to \mathbb{R}$ is the best (affine) linear approximation to the graph of f at a: for $f'(a) = m$,

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - mh}{h} = 0.$$

This motivates the following:

Definition 1.12. For $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, call f *differentiable* at $a \in U$ if there exists a linear map

$$Df(a) : \mathbb{R}^n \to \mathbb{R}^m$$

such that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - Df(a)h}{\|h\|} = 0.$$

Definition 1.13. Writing $x = a + h$ in the above, define

$$g(x) = f(a) + Df(a)(x-a)$$

to be the *tangent plane* of the graph of f at a. It is the best affine linear approximation to f near a: by definition of $Df(a)$,

$$0 = \lim_{h\to 0} \frac{f(a+h) - f(a) - Df(a)h}{\|h\|} = \lim_{x\to a} \frac{f(x) - g(x)}{\|x-a\|}.$$

Remark 1.14. The tangent plane is obtained by translating the graph $\Gamma(Df(a)) \subset \mathbb{R}^n \times \mathbb{R}^m$ so that it passes through $(a, f(a))$.

Proposition 1.15. If $Df(a)$ exists, it is unique.

Proof. Suppose there exists linear maps $T, T' : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - Th}{\|h\|} = \lim_{h\to 0} \frac{f(a+h) - f(a) - T'h}{\|h\|} = 0.$$

Then subtracting gives

$$\lim_{h\to 0} \frac{(T' - T)(h)}{\|h\|} = 0.$$

Let $h = te_i$ for any $i = 1, \dots, n$. Then the above implies

$$\lim_{t\to 0^+} \frac{(T' - T)(te_i)}{t} = (T' - T)(e_i).$$

So $Te_i = T'e_i$ for all i, implying $T = T'$. □

Proposition 1.16. If f is vector-valued, it is differentiable at a if and only if each component f_i is differentiable at a. Moreover, if say $f = (f_1, \dots, f_n)^T$, then $Df(a) = (Df_1(a), \dots, Df_n(a))^T$.

Proof. □ prove

1.17. As suggseted, differentiability is a stronger condition than the existence of all partial derivatives, and all directional derivatives.

Proposition 1.18. If $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a, then the partials $\frac{\partial f_i}{\partial x_j}$ exist. Furthermore,

$$[Df(a)] = \left[\frac{\partial f_i}{\partial x_j}(a)\right] = [D_j f_i(a)].$$

This matrix is called the *Jacobian* of f.

Proof. Since f is differentiable at a, there exists a linear map $Df(a)$ such that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - Df(a)h}{\|h\|} = 0.$$

For any $j = 1, \dots, n$, consider $h = te_j$ as $t \to 0$. Then

$$\lim_{t\to 0} \frac{f(a+te_j) - f(a) - Df(a)(te_j)}{|t|} = 0.$$

If $t > 0$, by linearity we get

$$0 = \lim_{t\to 0^+} \frac{f(a+te_j) - f(a) - Df(a)(te_j)}{t} = \lim_{t\to 0^+} \frac{f(a+te_j) - f(a)}{t} - Df(a)(e_j),$$
$$\lim_{t\to 0^+} \frac{f(a+te_j) - f(a)}{t} = Df(a)e_j.$$

Similarly, if $t < 0$,

$$0 = \lim_{t\to 0^-} \frac{f(a+te_j) - f(a) - df(a)(te_j)}{-t} = -\left(\lim_{t\to 0^-} \frac{f(a+te_j) - f(a)}{t} - df(a)(e_j)\right),$$

$$\lim_{t\to 0^-} \frac{f(a+te_j)-f(a)}{t} = Df(a)e_j.$$

thus

$$Df(a)e_j = \lim_{t\to 0} \frac{f(a+te_j)-f(a)}{t} = \frac{\partial f}{\partial x_j}(a).$$

So the jth column of $[Df(a)]$ is the row vector $[D_jf(a)]_j$. But also, for a fixed j, we have $D_jf(a)$ is equal to the column vector $[D_jf_i(a)]_i$. So the (i,j) entry of $[Df(a)]$ is $D_jf_i(a)$ as desired. □

Proposition 1.19. If $f : U : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in U$, then f is continuous at a.

Proof. Suppose f is differentiable at a. We want to show $\lim_{x\to a} f(x) = f(a)$, i.e. that $\lim_{h\to 0} f(a+h) = f(a)$. Well

$$\lim_{h\to 0} \frac{f(a+h)-f(a)-Df(a)h}{\|h\|} = 0,$$

so

$$\lim_{h\to 0} f(a+h)-f(a)-Df(a)h = \lim_{h\to 0} \frac{f(a+h)-f(a)-Df(a)h}{\|h\|}\|h\| = 0.$$

But also $\lim_{h\to 0} Df(a)h = 0$, since $Df(a)$ is a linear map and hence continuous, hence the result. □

Example 1.20. The following is an example of a function whose partials exist at a point but no other directional derivatives exist at that point and it is not even continuous there:

$$\begin{aligned} f : \mathbb{R}^2 &\to \mathbb{R} \\ (x,y)^T &\mapsto \frac{xy}{x^2+y^2} \\ (0,0)^T &\mapsto 0 \end{aligned}$$

at $(0,0)^T$.

Example 1.21. The following is an example of a function whose partials exist at a point, who is continuous at that point, but not differentiable there:

$$\begin{aligned} f : \mathbb{R}^2 &\to \mathbb{R} \\ (x,y)^T &\mapsto \frac{x^2y}{x^2+y^2} \\ (0,0)^T &\mapsto 0 \end{aligned}$$

at $(0,0)^T$.

Proposition 1.22. Let f be differentiable at a. Then for all $v \in \mathbb{R}^n$,

$$D_vf(a) = Df(a)v.$$

Proof. We know

$$\lim_{h\to 0}\frac{f(a+h)-f(a)-Df(a)h}{\|h\|}=0.$$

Letting $h=tv$ and $t\to 0$, we get

$$0=\lim_{t\to 0}\frac{f(a+tv)-f(a)-Df(a)(tv)}{|t|}$$

(we can pull out and discard the constant $\frac{1}{\|v\|}$). By linearity, as before,

$$\begin{aligned}Df(a)v&=\lim_{t\to 0^+}\frac{f(a+tv)-f(a)}{t}=\lim_{t\to 0^-}\frac{f(a+tv)-f(a)}{t}\\&=\lim_{t\to 0}\frac{f(a+tv)-f(a)}{t}=D_vf(a).\end{aligned}$$

□

Definition 1.23. A function $f:U\subset\mathbb{R}^n\to\mathbb{R}^m$ with partials which are continuous on U is called C^1.

Theorem 1.24. If $f:U\subset\mathbb{R}^n\to\mathbb{R}^m$ is C^1, then f is differentiable.

Proof. The idea is the following. Fundamentally, continuity of the partials allows us to apply the mean value theorem to the partials. First recall that f is differentiable if and only if each component f_i is differentiable. Now if f is C^1 then so it each f_i, so it suffices to prove the following simplified claim: if $f:U\subset\mathbb{R}^n\to\mathbb{R}$ is C^1, then f is differentiable.

We know that if the derivative exists at $a\in U$, then it must be $B=[D_jf(a)]_j$. So what we really want to show is that

$$\lim_{h\to 0}\frac{f(a+h)-f(a)-Bh}{\|h\|}=0.$$

Letting $h=\sum_{j=1}^n h_je_j$, we see $Bh=\sum_{j=1}^n D_jf(a)h_j$. If we can show

$$\lim_{h\to 0}f(a+h)-f(a)=\sum_{j=1}^n D_jf(a)h_j$$

then we would be done.

The sum on the right is "partitioning" the difference on the left dimension-wise. Consider the elements

$$\left\{p_j=a+\sum_{k=1}^j h_ke_k\right\}_{j=0}^n,$$

where $p_0=a$. Then

$$\sum_{j=1}^m f(p_j)-f(p_{j-1})=f(a+h)-f(a)$$

provided f is defined on these points; to ensure this, consider a closed cube $C\subset U\subset\mathbb{R}^n$ centered at a with radius ϵ. Since f is C^1, we have that f is defined for all points in C

and its partials are continuous at all points in C. Now the limit in question takes $h \to 0$, so we may restrict our focus to h such that $\|h\| < \epsilon$. For each such h, let $C_h \subset C$ be the cube of radius $\|h\|$ centered at a. Then each p_j lies on the boundary of C_h, and the above sum makes sense.

Fixing j, consider the function

$$\phi(t) = f(p_{j-1} + te_j) = f(q_j)$$

where $t \in [0, h_j]$. We see that $p_{j-1}+te_j$ ranges in a line over $[p_{j-1}, p_j]$, and this line is in (on the boundary of) C_h. So $\phi(t)$ is well-defined and in fact differentiable: the line is aligned with the coordinate axis e_j, so that $\phi(t) : \mathbb{R} \to \mathbb{R}$ has derivative $\phi'(t) = D_j f(p_{j-1} + te_j)$.

There are two cases. If $h_j = 0$, then

$$f(p_j) - f(p_{j-1}) = f(p_{j-1}) - f(p_{j-1}) = 0 = D_j(q_j)h_j$$

where, say, $q_j = a$. If $h_j \neq 0$, then $\phi(t)$ is continuous (it is differnetiable) on $[0, h_j]$ and differentiable on $(0, h_j)$ so by the mean value theorem there exists $c_j \in (0, h_j)$ such that $\phi(h_j) - \phi(0) = \phi'(c_j)h_j$. Letting $q_j = p_{j-1} + c_j e_j$, this is equivalent to

$$f(p_j) - f(p_{j-1}) = D_j(q_j)h_j.$$

In both cases, $c_j \in C_h$. We can now write

$$f(a + h) - f(a) = \sum_{j=1}^{m} D_j f(q_j)h_j.$$

As $h \to 0$, the radius of C_h goes to 0 as well, so that $q_j \to a$. Thus

$$\lim_{h\to 0} f(a + h) - f(a) = \sum_{j=1}^{m} D_j f(a)h_j,$$

where we have used the continuity of $D_j f$ on C_h. □

Example 1.25. The following is a function which is differentiable but not C^1:

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} x^2 \sin(\frac{1}{x}), & x \neq 0 \\ 0, & x = 0 \end{cases} \end{aligned}$$

Then $f'(0) = 0$ but $f'(x)$ is not continuous at 0.

1.3 differentiation rules

Proposition 1.26. Let $U \subset \mathbb{R}^n$ be open. Let $f, g : U \to \mathbb{R}^m$ and $k : U \to \mathbb{R}$. Suppose f, g, k are differnetiable at $a \in U$. Then, for any $v \in \mathbb{R}^n$,

1. $D(f + g)(a) = Df(a) + Dg(a)$.
2. $D(kf)(a)v = (Dk(a)v)f(a) + k(a)Df(a)v$.

3. $D(fg)(a)v = (Df(a)v)g(a) + f(a)(Dg(a)v)$.

Proof. We plug in the candidate and check that it satisfies the definition.

(2) We calculate

$$\begin{aligned}
&\lim_{h\to 0}\frac{(kf)(a+h)-(kf)(a)-((Dk(a)h)f(a)+k(a)(Df(a)h))}{\|h\|}\\
&=\lim_{h\to 0}\frac{(k(a+h)-k(a))f(a+h)+k(a)(f(a+h)-f(a))}{\|h\|}\\
&\quad-\lim_{h\to 0}\frac{((Dk(a)h)f(a)+k(a)(Df(a)h))}{\|h\|}\\
&=\lim_{h\to 0}\frac{(k(a+h)-k(a))f(a+h)-(Dk(a)h)f(a)}{\|h\|}\\
&\quad+k(a)\lim_{h\to 0}\frac{f(a+h)-f(a)-Df(a)h}{\|h\|}
\end{aligned}$$

where the second term is 0 by definition of the derivative, and

$$\begin{aligned}
&\lim_{h\to 0}\frac{(k(a+h)-k(a))f(a+h)-(Dk(a)h)f(a)}{\|h\|}\\
=&f(a+h)\lim_{h\to 0}\frac{(k(a+h)-k(a)-(Dk(a)h))}{\|h\|}+\lim_{h\to 0}\frac{(Dk(a)h)(f(a+h)-f(a))}{\|h\|}
\end{aligned}$$

where the first term is 0, and

$$\lim_{h\to 0}\frac{(Dk(a)h)(f(a+h)-f(a))}{\|h\|}=\lim_{h\to 0}\left(Dk(a)\frac{h}{\|h\|}\right)(f(a+h)-f(a)).$$

Now

$$\begin{aligned}
0\leq&\lim_{h\to 0}\left\|Dk(a)\frac{h}{\|h\|}\right\|\cdot\|f(a+h)-f(a)\|\\
\leq&\lim_{h\to 0}\|Dk(a)\|\cdot\|f(a+h)-f(a)\|\\
=&0
\end{aligned}$$

where we have used the boundedness of the linear operator $Dk(a)$. □

Corollary 1.27. Differntion is a linear operator.

Theorem 1.28 (chain rule)**.** Suppose

$$\mathbb{R}^n \xrightarrow{g} \mathbb{R}^m \xrightarrow{f} \mathbb{R}^\ell$$

and g is differentiable at a and f is differentiable at $g(a)$. Then $f\circ g$ is differentiable at a and

$$D(f\circ g)(a) = Df(g(a))Dg(a).$$

Proof. We want to show

$$\lim_{h\to 0} \frac{f(g(a+h)) - f(g(a)) - (Df(g(a))Dg(a))h}{\|h\|} = 0. \tag{$*$}$$

Letting $b = g(a)$, we know

$$\begin{aligned} \lim_{h\to 0} \frac{g(a+h) - g(a) - Dg(a)h}{\|h\|} &= 0 \\ \lim_{k\to 0} \frac{f(b+k) - f(b) - Df(b)k}{\|k\|} &= 0. \end{aligned}$$

Given $\epsilon > 0$, there exists $\delta_1, \eta > 0$ such that

$$\begin{aligned} \|h\| < \delta_1 \quad &\Rightarrow \quad \|g(a+h) - g(a) - Dg(a)h\| < \epsilon\|h\|, \\ \|k\| < \eta \quad &\Rightarrow \quad \|f(b+k) - f(b) - Df(b)k\| < \epsilon\|k\|. \end{aligned} \tag{$**$}$$

Setting $k = g(a+h) - g(a)$,

$$\|h\| < \delta_1 \quad \Rightarrow \quad \|k - Dg(a)h\| < \epsilon\|h\|.$$

By the reverse triangle inequality,

$$\|k\| - \|Dg(a)h\| \leq \big|\|k\| - \|Dg(a)h\|\big| \leq \|k - Dg(a)h\|$$

so

$$\|k\| < \|Dg(a)h\| + \epsilon\|h\| < (\|Dg(a)\| + \epsilon)\|h\|, \tag{$***$}$$

where we have used the fact that $\|Dg(a)h\| \leq \|Dg(a)\| \cdot \|h\|$. Now let

$$\delta_2 = \frac{\eta}{\|Dg(a)\| + \epsilon}, \qquad \delta = \min(\delta_1, \delta_2).$$

The numerator of ($*$) is

$$\begin{aligned} f(b+k) - f(b) - &(Df(b)Dg(a))h \\ &= (f(b+k) - f(b) - Df(b)k) + (Df(b)k - Df(b)Dg(a)h) \\ &= (f(b+k) - f(b) - Df(b)k) + Df(b)(k - Dg(a)h). \end{aligned}$$

Considering $0 < \|h\| < \delta$, we get

$$\begin{aligned} \|f(b+k) - f(b) - &Df(b)Dg(a)h\| \\ &\leq \|f(b+k) - f(b) - Df(b)k\| + \|Df(b)\| \cdot \|k - Dg(a)k\| \\ &\leq \epsilon\|k\| + \|Df(b)\| \cdot \epsilon\|h\| \end{aligned}$$

where we have used ($**$) and the definitions of k and $Dg(a)$. Continuing,

$$\begin{aligned} \|f(b+k) - f(b) - &Df(b)Dg(a)h\| \\ &\leq \epsilon\|k\| + \|Df(b)\| \cdot \epsilon\|h\| \\ &< \epsilon\|h\| \left(\frac{\|k\|}{\|h\|} + \|Df(b)\| \right). \end{aligned}$$

Now

$$\|k\| \leq \epsilon\|h\| + \|Dg(a)h\| \leq \epsilon\|h\| + \|Dg(a)\| \cdot \|h\|$$

by $(***)$, so

$$\begin{aligned}
&\|f(b+k) - f(b) - Df(b)Dg(a)h\| \\
&\quad <\epsilon\|h\| \left(\frac{\|k\|}{\|h\|} + \|Df(b)\|\right) \\
&\quad \leq\epsilon\|h\|\big(\|Dg(a)\| + \epsilon + \|Df(b)\|\big).
\end{aligned}$$

So whenever $0 < \|h\| < \delta$, we have

$$\frac{\|f(g(a+h)) - f(g(a)) - Df(g(a))Dg(a)h\|}{\|h\|} < \epsilon(\|Dg(a)\| + \epsilon + \|Df(b)\|).$$

Since ϵ is arbitrary, the limit goes to 0 as $h \to 0$. □

Remark 1.29. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a, and we want to evaluation $D_v f(a)$ for some $v \in \mathbb{R}^n$.

Define $g : \mathbb{R} \to \mathbb{R}^n$ by $g(t) = a + tv$, and consider $\phi(t) = (f \circ g)(t)$. Then by definition $D_v f(a) = \phi'(0)$. By the chain rule,

$$D_v f(a) = \phi'(0) = (f \circ g)'(0) = Df(a)g'(0) = Df(a)v$$

confirming our earlier result.

1.4 gradient

Definition 1.30. Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at a. The *gradient* of f at a is the vector

$$\nabla f(a) = (Df(a))^T = \begin{pmatrix} D_1 f(a) \\ \vdots \\ D_n f(a) \end{pmatrix}.$$

1.31. Now the directional derivative is

$$D_v f(a) = Df(a)v = \nabla f(a) \cdot v.$$

In particular, this is saying that $\nabla f(a)$ is orthogonal to v where $D_v f(a) = 0$, i.e. $\nabla f(a)$ is orthogonal to the v along which f remains constant (instantaneously).

If v is a unit vector, the Cauchy-Schwartz inequality implies

$$D_v f(a) \leq \|\nabla f(a)\|,$$

and there is equality if and only if $\nabla f(a)$ is a positive scalar multiple of v. This implies the following:

Proposition 1.32. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at a. Then

$$\|\nabla f(a)\| = \max_{\|v\|=1} D_v f(a),$$

i.e.

- $\nabla f(a)$ points in the direction in which f increases at the greatest rate.
- $\|\nabla f(a)\|$ is the greates possible (instantaneous) rate of change.

1.5 curves

Proposition 1.33. Suppose $g : (a,b) \to \mathbb{R}^n$ is a differentiable parametric curve with the property that g has constant length (i.e. $\|g(t)\|$ is constant for all $t \in (a,b)$). Then $g(t) \cdot g'(t) = 0$ for all $t \in (a,b)$.

Proof. We know $g(t) \cdot g(t) = c$ for some $c \in \mathbb{R}$. Then, utilizing the product rule,

$$g'(t) \cdot g(t) + g(t) \cdot g'(t) = 2g(t) \cdot g'(t) = 0.$$

□

Remark 1.34. The condition above is that the curve lies on a sphere centered at the origin.

Definition 1.35. Suppose $g : [a,b] \to \mathbb{R}^n$ is continuous, except perhaps at finitely many points. Define

$$\int_a^b g(t)dt = \begin{pmatrix} \int_a^b g_1(t)dt \\ \cdots \\ \int_a^b g_n(t)dt \end{pmatrix}.$$

Lemma 1.36. Suppose $g : [a,b] \to \mathbb{R}^n$ is continuous, except perhaps at finitely many points. Then

$$\left\| \int_a^b g(t)dt \right\| \leq \int_a^b \|g(t)\|dt.$$

Proof. Let $v = \int_a^b g(t)dt$. If $v = 0$ then we are done. Otherwise, first note that $|v \cdot g(t)| \leq \|v\| \cdot \|g(t)\|$ by Cauchy Schwartz. Then

$$\begin{aligned} \|v\|^2 &= v \cdot \int_a^b g(t)dt = \int_a^b v \cdot g(t)dt \\ &\leq \int_a^b \|v\| \cdot \|g(t)\|dt = \|v\| \int_a^b \|g(t)\|dt. \end{aligned}$$

Since $v \neq 0$, we can divide by $\|v\|$ to get the result. □

Definition 1.37. Let $g : [a,b] \to \mathbb{R}^n$ be a continuous parameterized curve. Given a partition

$$\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_k = b\}$$

of $[a,b]$, let

$$\ell_{\mathcal{P}}(g) = \sum_{i=1}^k \|g(t_i) - g(t_{i-1})\|.$$

Define the *arclength* of g to be

$$\ell(g) = \sup_{\mathcal{P}} \ell_{\mathcal{P}}(g),$$

provided this quantity is finite.

Proposition 1.38. Let $g : [a, b] \to \mathbb{R}^n$ be a piecewise-C^1 paramaterized curve. Then

$$\ell(g) = \int_a^b \|g'(t)\| dt.$$

Remark 1.39. This says the distance a particle travels is the integral of its speed.

Proof. By the lemma, for any $\mathcal{P}$ we have

$$\begin{aligned}
\ell_{\mathcal{P}}(g) &= \sum_{i=1}^{k} \|g(t_i) - g(t_{i-1})\| \\
&= \sum_{i=1}^{k} \left\| \int_{t_{i-1}}^{t_i} g'(t) dt \right\| \\
&\leq \sum_{i=1}^{k} \int_{t_{i-1}}^{t_i} \|g'(t)\| dt = \int_a^b \|g'(t)\| dt,
\end{aligned}$$

where we have use the fundamental theorem of calculus which requires the g_i to be continuous on $[t_i, t_{i-1}]$ and differentiable on (t_i, t_{i-1}) which is implied by the piecewise C^1 hypothesis, and so

$$\ell(g) \leq \int_a^b \|g'(t)\| dt.$$

Now for $a \leq t \leq b$, define $s(t)$ to be the arclength of the curve g on the interval $[a, t]$. Then, for $h > 0$,

$$\frac{\|g(t+h) - g(t)\|}{h} \leq \frac{s(t+h) - s(t)}{h} \leq \frac{1}{h} \int_t^{t+h} \|g'(u)\| du;$$

where the first inequality utilizes the fact that $s(t + h) - s(t)$ is the arclength of g on $[t, t+h]$, which is at least the shortest possible path length $\|g(t+h) - g(t)\|$. The second inequality additionally uses the previous inequality above.

First suppose $t < b$. Taking limits, as $h \to 0^+$ the left is

$$\lim_{h \to 0^+} \frac{\|g(t+h) - g(t)\|}{h} = \|g'(t)\|.$$

The right is

$$\lim_{h \to 0^+} j \frac{1}{h} \int_t^{t+h} \|g'(u)\| du = \|g'(t)\|,$$

which can be deduced, for example, by L'Hopital's rule. Hence

$$\lim_{h \to 0^+} \frac{s(t+h) - s(t)}{h} = \|g'(t)\|$$

for all $t \in [a, b)$. Similarly, for $t \in (a, b]$, we can show

$$\lim_{h \to 0^-} \frac{s(t+h) - s(t)}{h} = \|g'(t)\|.$$

Thus on $t \in (a, b)$ we have

$$\lim_{h\to 0} \frac{s(t+h)-s(t)}{h} = s'(t) = \|g'(t)\|, \quad s(t) = \int_a^t \|g'(u)\| du.$$

But g is piecewise continuous on $[a, b]$, hence so is $s(t)$ (ref). Since limits are unique, it follows

$$\lim_{t\to a} s(t) = \lim_{t\to a} \int_a^t \|g'(u)\| du = 0,$$
$$\lim_{t\to b} s(t) = \lim_{t\to a} \int_a^t \|g'(u)\| du = \int_a^b \|g'(u)\| du.$$

□

Definition 1.40. We say a parameterized curve $g(t)$ is *arclength-parameterized* if $\|g'(t)\| = 1$ for all t.

Remark 1.41. If g is arclength-parameterized, then $s'(t) = \|g'(t)\| = 1$, so $s(t) = t + c$ for some constant c. In this case, we often use s as the parameter. But note this is just a matter of notation; if given an arclength-parameterized curve $g : [a, b] \to \mathbb{R}^n$ you can assume that s varies over the domain $[a, b]$ and not over the domain plus some constant.

In what follows, let g be arclength-parameterized.

Definition 1.42. Call $g'(s)$ the *unit tangent vector*, and denote it $T(s)$. If g is twice differentiable, let the *curvature* be $\kappa(s) = \|T'(s)\|$. If $g : [0, L] \to \mathbb{R}^n$ is closed, i.e. $g(0) = g(L)$, then the *total curvature* is $\int_0^L \kappa(s)ds$.
If $T'(s) \neq 0$, define the *principal normal vector* to be $N(s) = T'(s)/\|T'(s)\|$.

Remark 1.43. By (ref), $T(s) \cdot T'(s) = 0$.

Proposition 1.44. For any convex plane curve, total curvature is 2π.

1.6 higher order partial derivatives

Theorem 1.45. Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ be C^2. Then, for any i, j, we have

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

Proof. It suffices to show the result for $m = 1$, since we go component-by-component, and for $n = 2$, since we can repeat the argument for any two i, j and fixing all other variables as partials do. Without loss of generality, let $i = 1, j = 2$. Define

$$\delta \begin{pmatrix} h \\ k \end{pmatrix} = f \begin{pmatrix} a+h \\ b+k \end{pmatrix} - f \begin{pmatrix} a+h \\ b \end{pmatrix} - f \begin{pmatrix} a \\ b+k \end{pmatrix} + f \begin{pmatrix} a \\ b \end{pmatrix},$$
$$q(s) = f \begin{pmatrix} s \\ b+k \end{pmatrix} - f \begin{pmatrix} s \\ b \end{pmatrix},$$

$$r(t) = f\begin{pmatrix} a+h \\ t \end{pmatrix} - f\begin{pmatrix} a \\ t \end{pmatrix}.$$

The picture is as follows:

By the mean value theorem, there exists $\xi \in (a, a+h)$ and $\eta \in (b, b+k)$ such that

$$\begin{aligned}
\delta\begin{pmatrix} h \\ k \end{pmatrix} =& q(a+h) - q(a) = hq'(\xi) \\
=& h\left(\frac{\partial f}{\partial x}\begin{pmatrix} \xi \\ b+k \end{pmatrix} - \frac{\partial f}{\partial x}\begin{pmatrix} \xi \\ b \end{pmatrix}\right) \\
=& hk\frac{\partial^2 f}{\partial y \partial x}\begin{pmatrix} \xi \\ \eta \end{pmatrix}.
\end{aligned}$$

Our application of the mean value theorem required f to be C^1 and twice differentiable. Likewise, there exists $\tau \in (b, b+k)$ and $\sigma \in (a, a+h)$ such that

$$\delta\begin{pmatrix} h \\ k \end{pmatrix} = hk\frac{\partial^2 f}{\partial x \partial y}\begin{pmatrix} \sigma \\ \tau \end{pmatrix}.$$

As $h, k \to 0$ we have $\zeta, \sigma \to a$ and $\eta, \tau \to b$. By the continuity of the second partials (using the fact that f is C^2), we get the desired equality. □

2 implict/explicit solutions

2.1. Any subspace $V \subset \mathbb{R}^n$ can be represented in two ways:

- (explicit) as the span of a collection of vectors.
- (implicit) as the space of solutions to an equation $Ax = 0$.

Let's look at the second in more detail. Note x is a solution of $Ax = 0$ if and only if $A_{i,\bullet} \cdot x = 0$ for all i. The equation $A_{i,\bullet} \cdot x = 0$ defines a hyperplane in $\mathbb{R}^n$ with normal $(A_{i,\bullet})^T$. Hence the solution space to $Ax = 0$ is the intersection of the hyperplanes $A_i \cdot x = 0$.

2.1 Gaussian elimination

Definition 2.2. We call the first nonzero entroy of a row (reading left to right) its *leading entry.*

Definition 2.3. A matrix is in *echelon form* if

- the leading entries move right in successive rows.
- the entries in the column of each leading entry, below the leading entry, are all 0.
- all rows of zeros are at the bottom of the matrix.

It is in *reduced echelon form* if it is in echelon form and, additionally,

- every leading entry is 1.
- all entries of the column above each leading entry are 0 (as well as below).

Definition 2.4. A leading entry is called a *pivot* if there is no leading entry above it in the same column. Columns where (a single) pivot appears are called *pivot columns*, and the corresponding variable in the system of equations is called a *pivot variable*. A variable which is not a pivot variable is called a *free variable*.

Remark 2.5. When a matrix is in reduced echelon form, we can determine the general solution by expressing each of the pivot variables in terms of the free variables.

Theorem 2.6. Suppose A and B are echelon forms of the same nonzero matrix M. Then all of their pivots appear in the same positions. Therefore, if they are in reduced echelon form, then they are equal, i.e. the reduced echelon form of the matrix M is unique.

2.7. Consider $Ax = b$, where A is $m \times n$. Then a solution $c = (c_1 \cdots c_n)^T$ has the property that $Ac = b$. As discussed, this means $A_{i,\bullet} \cdot c = b$ for all i. Combining this into aa single equation we can also write

$$b = c_1 A_{\bullet,1} + \cdots + c_n A_{\bullet,n}.$$

Thus a solution to $Ax = b$ gives a representation of b as a linear combination of column vectors of A.

Example 2.8. Suppose we want to express b as a linear combination of vectors v_1, v_2, v_3, where

$$b = \begin{pmatrix} 4 \\ 3 \\ 1 \\ 2 \end{pmatrix}, \quad v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 2 \\ 1 \\ 1 \\ 2 \end{pmatrix}.$$

We want to see if there are solutions (x_1, x_2, x_3) to the equation

$$x_1 v_1 + x_2 v_2 + x_3 v_3 = b.$$

This is the same as solving $Ax = b$ where

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 2 \end{pmatrix}.$$

Definition 2.9. Consider a system $Ax = b$. There are two possibilities:
- (*inconsistent*) no solutions.
- (*consistent*) solutions exist.

Corollary 2.10. A system in echelon form is inconsistent if and only if there is a row of the form

$$\begin{bmatrix} 0 & \cdots & 0 & | & c \end{bmatrix}$$

for $c \neq 0$. Similarly, a system in echelon form is consistent if and only if any zero row has $c = 0$.

Corollary 2.11. If $\operatorname{rank}(A) = m$ then $Ax = b$ is consistent for all $b \in \mathbb{R}^m$.

Remark 2.12. Corresponding to our previous viewpoint:

- (implicit) $Ax = b$ is consistent when the intersection of hyperplanes $A_{i,\bullet} \cdot x = b_i$ is nonempty.
- (explicit) $Ax = b$ is consistent when b is in the span of the column vectors $A_{\bullet,j}$.

Example 2.13. (constraint equations) We wish to find all b for which $Ax = b$ is consistent. Suppose

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 3 & 2 & -1 \\ 1 & 4 & -3 \\ 3 & -3 & 3 \end{pmatrix}.$$

By the Corollary, we can let $b = (b_1\ b_2\ b_3\ b_4)^T$ can calculate the echelon form of the augmented matrix:

$$[A|b] = \begin{pmatrix} 1 & -1 & 1 & b_1 \\ 3 & 2 & -1 & b_2 \\ 1 & 4 & -3 & b_3 \\ 3 & -3 & 3 & b_4 \end{pmatrix} \quad \to \quad \begin{pmatrix} 1 & -1 & 1 & b_1 \\ 0 & 5 & -4 & b_2 - 3b_1 \\ 0 & 0 & 0 & b_3 - b_2 + 2b_1 \\ 0 & 0 & 0 & b_4 - 3b_1 \end{pmatrix}.$$

So b must satisfy as many constraints as there are rows of zeros in the echelon form (recall by theorem this number is unique).

2.2 existence and uniqueness of solutions

Definition 2.14. The rank r of a matrix is the number of nonzero rows (i.e. the number of pivots) in its echelon form.

Remark 2.15. Note $r \leq n$ and $r \leq m$. We get that $r \leq n$ since there cannot be more pivots than columns, and $r \leq m$ since there cannot be more nonzero rows than there are total rows.

We now turn to how many solutions a given consistent system has.

Definition 2.16. A system $Ax = b$ is *inhomogeneous* when $b \neq 0$ and *homogeneous* when $b = 0$.

Proposition 2.17. Suppose $Ax = b$ is consistent. Let u_1 be an arbitrary solution. Then all solutions are of the form

$$u = u_1 + v$$

for some solution v (which is dependent on u) of the associated homogeneous system $Ax = 0$.

Proof. First let's show u is a solution:

$$Au = A(u_1 + v) = Au_1 + Av = b + 0 = b.$$

Now let's show every solution has this form: we need to show $u - u_1$ is always a solution to $Ax = 0$:

$$Av = A(u - u_1) = Au - Au_1 = b - b = 0.$$

$\square$

Remark 2.18. This is saying that when the inhomogeneous system $Ax = b$ is constent, its solutions are obtained by translating the set of solutions of the associated homogeneous equation by a "particular solution" u.

2.19. Note a homogeneous system is always consistent, since it has the trivial solution $x = 0$. If $\operatorname{rank}(A) = n$, then there are n pivot variables and $n - n = 0$ free variables. To summarize: the system $Ax = 0$ has

- ($\operatorname{rank}(A) = n$) a unique solution, since there are 0 free variables.
- ($\operatorname{rank}(A) < n$) infinite solutions, since there are > 0 free variables.
- ($n > m$) infinite solutions, since $\operatorname{rank}(A) \leq m < n$ and above.

Corollary 2.20. Suppose $Ax = b$ is consistent. It has a unique solution if and only if $Ax = 0$ has a unique (the trivial) solution, which happens if and only if $\operatorname{rank}(A) = n$.

Definition 2.21. An $n \times n$ matrix of rank n is called *nonsingular*. Otherwise, it is called *singular*.

Recall that if $\operatorname{rank}(A) = m$ then $Ax = b$ is consistent for all $b \in \mathbb{R}^m$.

Theorem 2.22. Let A be $n \times n$. The following are equivalent:

- A is nonsingular.
- $Ax = 0$ has only the trivial solution.
- $Ax = b$ has a unique solution for all $b \in \mathbb{R}^n$.

2.3 elementary matrices, inverses

2.23. There are two interpretations of matrix multiplication AB:

- $(AB)_{\bullet,j} = A(B_{\bullet,j})$, i.e. the jth column of AB is A times the jth column of B.
- $(AB)_{i,\bullet} = (A_{i,\bullet})B$, i.e. the ith row of AB is the ith row of A times B.

In particular,

$$\begin{bmatrix} \vdots & & \vdots \\ a_1 & \cdots & a_n \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 a_1 + \cdots + x_n a_n,$$

$$\begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix} \begin{bmatrix} \cdots & A_1 & \cdots \\ & \vdots & \\ \cdots & A_m & \cdots \end{bmatrix} = x_1 A_1 + \cdots + x_m A_m.$$

Proposition 2.24. Consider the elementary row operations

1. interchangerows i and j
2. multiply row i by a nonzero scalar c
3. row j plus equals c times row i.

To apply these to an $m \times n$ matrix A, multiply A on the left by I'_m, where I'_m is the result of having applied the desired row operations to the identity matrix I_m.

Corollary 2.25. Consider the augmented matrix $[A|b]$ with echelon form $[U|c]$. Then there exists a matrix E such that

$$[EA|Eb] = [U|c],$$

and moreover E is a product of elementary matrices.

Theorem 2.26. An $n \times n$ matrix is *nonsingular* if and only if it is invertible.

Proof. If A is invertible and $Ax = b$, then x can only be $A^{-1}b$.

Conversely, suppose A is nonsingular. Then $Ax = c$ has a unique solution for all c. In particular, there exists a unique b_j such that $Ab_j = e_j$. Let $B = (b_1 \cdots b_n)$. Then $AB = I_n$, so that B is a solution to the equation $AX = I_n$. Then the reduced echelon form of $[A|I_n]$ is $[I_n|B]$, where we have used the fact that A is nonsingular to determine that its reduced echelon form is I_n. By the corollary, there exists a product of elementary matrices E such that $E[A|I_n] = [I_n|B]$. Then $EA = I_n$ and $E = B$, so $BA = I_n$. □

Corollary 2.27. If A, B are $n \times n$ matrices such that $BA = I_n$, then $B = A^{-1}$ and $A = B^{-1}$, i.e. for square matrices every left inverse is an inverse.

Proof. We first claim the homogeneous equation $Ax = 0$ has only the trivial solution. To see this, consider that if x is a solution then $BAx = x = 0 = B(0)$. Now this implies by the proposition that A is nonsingular, hence invertible by the theorem. Then since $BA = I_n$ and inverses are unique, it must be that $B = A^{-1}$. □

Corollary 2.28. Given a square matrix A, its inverse, if it exists, is the matrix E such that

$$E[A|I_n] = [I_n|C]$$

for some matrix C. In particular, one can apply Gaussian elimination

$$[A|I_n] \quad \rightarrow \quad [I_n|C]$$

and then $A^{-1} = C$, provided we can actually perform such a reduction.

Corollary 2.29. For

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

A^{-1} exists if and only if $ad - bc \neq 0$. In that case,

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} b & -b \\ -c & a \end{pmatrix}.$$

Proposition 2.30. A nonsquare matrix never has both a left and right inverse.

2.4 linear independence, dimension

2.31. We are interested in whether $v \in \operatorname{span}\{v_1, \ldots, v_k\}$. Recall that this is asking where the system $Ax = v$ has a solution, where $A_{\bullet,j} = v_j$. We are also interested in whether this solution is unique. Equivalently, this is asking whether (the linear map represented by) A is injective.

Definition 2.32. The $\{v_1, \ldots, v_k\}$ are *linearly independent* if

$$c_1 v_1 + \cdots c_k v_k = 0$$

implies $c_1 = \cdots = c_k = 0$. Otherwise, they are called *linearly dependent.*

Proposition 2.33. Let $v_1, \ldots, v_k \in \mathbb{R}^n$ and let $V = \operatorname{span}\{v_1, \ldots, v_k\}$. Then an arbitrary $v \in V$ has a unique expression as a linear combination of $v_1, \ldots, v_k$ if and only if the v_i are linearly independent.

Proof. ($\Leftarrow$) Suppose v is not uniquely expressed, i.e.

$$v = c_1 v_1 + \cdots c_k v_k = d_1 v_1 + \cdots d_k v_k$$

where $\{c_i\} \neq \{d_i\}$. Then

$$0 = (c_1 - d_1)v_1 + \cdots + (c_k - d_k)v_k$$

and at least one of the coefficients is nonzero. Thus the $\{v_i\}$ are linearly dependent.

($\Rightarrow$) Suppose $\{v_k\}$ is not linearly independent. Then there exists a nontrivial linear combination

$$s_1 v_1 + \cdots + s_k v_k = 0.$$

For any $v \in V$, we can write

$$v = c_1 v_1 + \cdots c_k v_k.$$

But also

$$v = (c_1 + s_1)v_1 + \cdots (c_k + s_k)v_k$$

and there exists i such that $s_i \neq 0$ so $c_i + s_i \neq c_i$. So the representation is not unique. $\square$

Remark 2.34. Recall that a solution to $Ax = b$ is unique if and only if $Ax = 0$ has only the trivial solution. By definition, $v \in V$ if and only if $Ax = v$ is consistent. Then uniqueness follows if and only if $Ax = 0$ has the trivial solution, which is equivalent to the columns being linearly independent.

Proposition 2.35. Suppose $v_1, \ldots, v_k \in \mathbb{R}^n$ are linearly independent, and let $x \in \mathbb{R}^n$. Then $\{v_1, \ldots, v_k, x\}$ is linearly independent if and only if $x \notin \operatorname{span}\{v_1, \ldots, v_k\}$.

Proof. It suffices to show that $\{v_1, \ldots, v_k\}$ is linearly dependent if and only if $x \in \operatorname{span}\{v_1, \ldots, v_k\}$.

($\Leftarrow$) Suppose $x \in \operatorname{span}\{v_1, \ldots, v_k\}$. Then $x = c_1 v_1 + \cdots + c_k v_k$ for some scalaras $c_1, \ldots, c_k$. Then

$$c_1 v_1 + \cdots + c_k v_k - x = 0$$

which is a nontrivial linearly combination, hence $\{v_1, \ldots, v_k, x\}$ is linearly dependent.

($\Rightarrow$) Suppose $\{v_1, \ldots, v_k, x\}$ is linearly dependent. Then there exists a nontrivial collection $\{c_k\}$ such that

$$c_1 v_1 + \cdots c_k v_k + cx = 0.$$

Note $c \neq 0$, for otherwise the $v_1, \ldots, v_k$ would be linearly dependent which is a contradiction to our assumptions. Thus

$$x = \frac{-1}{c}(c_1 v_1 + \cdots + c_k v_k)$$

and $x \in \operatorname{span}\{v_1, \ldots, v_k\}$. □

Definition 2.36. Let $V \subset \mathbb{R}^n$ be a subspace. The set $\{v_1, \ldots, v_k\}$ is a basis for V if

- $v_1, \ldots, v_k$ span V, and
- $v_1, \ldots, v_k$ are linearly independent.

Corollary 2.37. Let $V \subset \mathbb{R}^n$ and $v_1, \ldots, v_k \in V$. Then $\{v_i\}$ is a basis for V if and only if every vector in V can be written uniquely as a linear combination of $v_1, \ldots, v_k$.

Definition 2.38. When we write

$$v = c_1 v_1 + \cdots + c_k v_k,$$

we call the $\{c_i\}$ the *coordinates* of V with respect to the ordered basis $\{v_1, \ldots, v_k\}$.

Corollary 2.39. Let A be $n \times n$. Then A is nonsingular if and only if its column vectors form a basis for $\mathbb{R}^n$.

Theorem 2.40. Any nonzero subspace $V \subset \mathbb{R}^n$ has a basis.

Proposition 2.41. Let $V \subset \mathbb{R}^n$ be a subspace with a basis $\{v_1, \ldots, v_k\}$. Let $w_1, \ldots, w_l \in V$. If $l > k$, then $\{w_1, \ldots, w_l\}$ is linearly dependent.

Proof. Each w_j can be written uniquely as

$$w_j = \sum_{i=1}^{k} a_{ij} v_i.$$

Now for any $c \in \mathbb{R}^l$, we have

$$\sum_{j=1}^{l} c_j w_j = \sum_{j=1}^{l} c_j (\sum_{i=1}^{k} c_{ij} v_i) = \sum_{i=1}^{k} (\sum_{j=1}^{l} a_{ij} c_j) v_i = 0.$$

Since $l > k$, there exists a nonzero c such that $Ac = 0$. □

Theorem 2.42. Let $V \subset \mathbb{R}^n$ be a subspace. Let $\{v_1, \ldots, v_k\}$ and $\{w_1, \ldots, w_l\}$ be two bases for V. Then $k = l$.

Definition 2.43. The *dimension* of a subspace $V \subset \mathbb{R}^n$ is the number of vectors in any basis for V. By convention, $\dim\{0\} = 0$.

Lemma 2.44. Suppose V, W are subspaces of $\mathbb{R}^n$ with the property that $W \subset V$. If $\dim(V) = \dim(W)$, then $V = W$.

Proposition 2.45. Let $V \subset \mathbb{R}^n$ be a k-dimensional subspace. Then any k vectors that span V must be linearly independent, and any k linearly independent vectors must span V.

3 appendix

Given a plane $n \cdot x = d$ and a point x_0, the vector $x_0 - x$ is a vector from the plane to the point. To get the perpindicular distance, we project $x_0 - x$ onto n:

$$\mathrm{proj}_n(x_0 - x) = \frac{n \cdot (x_0 - x)}{n \cdot n} n = \frac{n \cdot x_0 - d}{n \cdot n} n.$$

Taking signed magnitude,

$$D = \left\| \frac{n \cdot x_0 - d}{n \cdot n} n \right\| = \frac{n \cdot x_0 - d}{n \cdot n} \|n\| = \frac{n \cdot x_0 - d}{\|n\|}.$$