CHAPTER 1

Mathematics Review

CONTENTS

IN this chapter, we will outline notions from linear algebra and multivariable calculus that will be relevant to our discussion of computational techniques. It is intended as a review of background material with a bias toward ideas and interpretations commonly encountered in practice; the chapter can be safely skipped or used as reference by students with stronger background in mathematics.

1.1 PRELIMINARIES: NUMBERS AND SETS

Rather than considering algebraic (and at times philosophical) discussions like "What is a number?," we will rely on intuition and mathematical common sense to define a few sets:

- The *natural numbers* $\mathbb{N} = \{1, 2, 3, \ldots\}$
- The *integers* $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
- The *rational numbers* $\mathbb{Q} = \{a/b : a, b \in \mathbb{Z}, b \neq 0\}$
- The *real numbers* $\mathbb{R}$ encompassing $\mathbb{Q}$ as well as *irrational* numbers like π and $\sqrt{2}$
- The *complex numbers* $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$, where $i \equiv \sqrt{-1}$.

The definition of $\mathbb{Q}$ is the first of many times that we will use the notation $\{A : B\}$; the braces denote a set and the colon can be read as "such that." For instance, the definition of $\mathbb{Q}$ can be read as "the set of fractions a/b *such that* a and b are integers." As a second example, we could write $\mathbb{N} = \{n \in \mathbb{Z} : n > 0\}$. It is worth acknowledging that our definition

of $\mathbb{R}$ is far from rigorous. The construction of the real numbers can be an important topic for practitioners of cryptography techniques that make use of alternative number systems, but these intricacies are irrelevant for the discussion at hand.

$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ can be manipulated using generic operations to generate new sets of numbers. In particular, we define the "Euclidean product" of two sets A and B as

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

We can take *powers* of sets by writing

$$A^n = \underbrace{A \times A \times \cdots \times A}_{n \text{ times}}.$$

This construction yields what will become our favorite set of numbers in chapters to come:

$$\mathbb{R}^n = \{(a_1, a_2, \ldots, a_n) : a_i \in \mathbb{R} \text{ for all } i\}.$$

1.2 VECTOR SPACES

Introductory linear algebra courses easily could be titled "Introduction to Finite-Dimensional Vector Spaces." Although the definition of a vector space might appear abstract, we will find many concrete applications expressible in vector space language that can benefit from the machinery we will develop.

1.2.1 Defining Vector Spaces

We begin by defining a vector space and providing a number of examples:

Definition 1.1 (Vector space over $\mathbb{R}$)**.** A *vector space* over $\mathbb{R}$ is a set $\mathcal{V}$ closed under addition and scalar multiplication satisfying the following axioms:

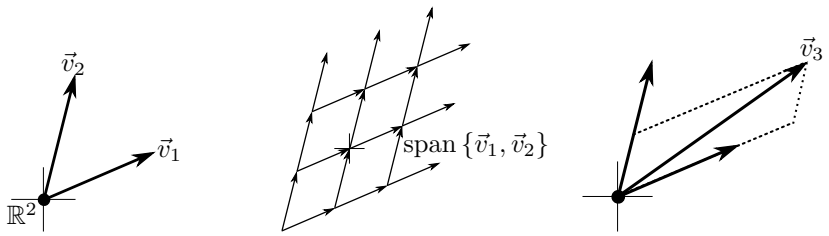
- *Additive commutativity and associativity:* For all $\vec{u}, \vec{v}, \vec{w} \in \mathcal{V}$, $\vec{v} + \vec{w} = \vec{w} + \vec{v}$ and $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.
- *Distributivity:* For all $\vec{v}, \vec{w} \in \mathcal{V}$ and $a, b \in \mathbb{R}$, $a(\vec{v}+\vec{w}) = a\vec{v}+a\vec{w}$ and $(a+b)\vec{v} = a\vec{v}+b\vec{v}$.
- *Additive identity:* There exists $\vec{0} \in \mathcal{V}$ with $\vec{0} + \vec{v} = \vec{v}$ for all $\vec{v} \in \mathcal{V}$.
- *Additive inverse:* For all $\vec{v} \in \mathcal{V}$, there exists $\vec{w} \in \mathcal{V}$ with $\vec{v} + \vec{w} = \vec{0}$.
- *Multiplicative identity:* For all $\vec{v} \in \mathcal{V}$, $1 \cdot \vec{v} = \vec{v}$.
- *Multiplicative compatibility:* For all $\vec{v} \in \mathcal{V}$ and $a, b \in \mathbb{R}$, $(ab)\vec{v} = a(b\vec{v})$.

A member $\vec{v} \in \mathcal{V}$ is known as a *vector*; arrows will be used to indicate vector variables.

For our purposes, a scalar is a number in $\mathbb{R}$; a *complex* vector space satisfies the same definition with $\mathbb{R}$ replaced by $\mathbb{C}$. It is usually straightforward to spot vector spaces in the wild, including the following examples:

Example 1.1 ($\mathbb{R}^n$ as a vector space)**.** The most common example of a vector space is $\mathbb{R}^n$. Here, addition and scalar multiplication happen component-by-component:

$$\begin{aligned}(1, 2) + (-3, 4) &= (1 - 3, 2 + 4) = (-2, 6)\\ 10 \cdot (-1, 1) &= (10 \cdot -1, 10 \cdot 1) = (-10, 10).\end{aligned}$$



(a) $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^2$ (b) span $\{\vec{v}_1, \vec{v}_2\}$ (c) span $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$

Figure 1.1 (a) Vectors $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^2$; (b) their span is the plane $\mathbb{R}^2$; (c) span $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ = span $\{\vec{v}_1, \vec{v}_2\}$ because $\vec{v}_3$ is a linear combination of $\vec{v}_1$ and $\vec{v}_2$.

Example 1.2 (Polynomials)**.** A second example of a vector space is the ring of polynomials with real-valued coefficients, denoted $\mathbb{R}[x]$. A polynomial $p \in \mathbb{R}[x]$ is a function $p : \mathbb{R} \to \mathbb{R}$ taking the form*

$$p(x) = \sum_k a_k x^k.$$

Addition and scalar multiplication are carried out in the usual way, e.g., if $p(x) = x^2+2x-1$ and $q(x) = x^3$, then $3p(x) + 5q(x) = 5x^3 + 3x^2 + 6x - 3$, which is another polynomial. As an aside, for future examples note that functions like $p(x) = (x-1)(x+1) + x^2(x^3-5)$ are still polynomials even though they are not explicitly written in the form above.

A weighted sum $\sum_i a_i \vec{v}_i$, where $a_i \in \mathbb{R}$ and $\vec{v}_i \in \mathcal{V}$, is known as a *linear combination* of the $\vec{v}_i$'s. In the second example, the "vectors" are polynomials, although we do not normally use this language to discuss $\mathbb{R}[x]$; unless otherwise noted, we will assume variables notated with arrows $\vec{v}$ are members of $\mathbb{R}^n$ for some n. One way to link these two viewpoints would be to identify the polynomial $\sum_k a_k x^k$ with the sequence $(a_0, a_1, a_2, \ldots)$; polynomials have finite numbers of terms, so this sequence eventually will end in a string of zeros.

1.2.2 Span, Linear Independence, and Bases

Suppose we start with vectors $\vec{v}_1, \ldots, \vec{v}_k \in \mathcal{V}$ in vector space $\mathcal{V}$. By Definition 1.1, we have two ways to start with these vectors and construct new elements of $\mathcal{V}$: addition and scalar multiplication. *Span* describes all of the vectors you can reach via these two operations:

Definition 1.2 (Span)**.** The *span* of a set $S \subseteq \mathcal{V}$ of vectors is the set

$$\text{span}\, S \equiv \{a_1\vec{v}_1 + \cdots + a_k\vec{v}_k : \vec{v}_i \in S \text{ and } a_i \in \mathbb{R} \text{ for all } i\}.$$

Figure 1.1(a-b) illustrates the span of two vectors. By definition, span S is a *subspace* of $\mathcal{V}$, that is, a subset of $\mathcal{V}$ that is itself a vector space. We provide a few examples:

Example 1.3 (Mixology)**.** The typical well at a cocktail bar contains at least four ingredients at the bartender's disposal: vodka, tequila, orange juice, and grenadine. Assuming we have this well, we can represent drinks as points in $\mathbb{R}^4$, with one element for each ingredient. For instance, a tequila sunrise can be represented using the point $(0, 1.5, 6, 0.75)$,

*The notation $f : A \to B$ means f is a function that takes as input an element of set A and outputs an element of set B. For instance, $f : \mathbb{R} \to \mathbb{Z}$ takes as input a real number in $\mathbb{R}$ and outputs an integer $\mathbb{Z}$, as might be the case for $f(x) = \lfloor x \rfloor$, the "round down" function.

representing amounts of vodka, tequila, orange juice, and grenadine (in ounces), respectively.

The set of drinks that can be made with our well is contained in

$$\text{span}\{(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)\},$$

that is, all combinations of the four basic ingredients. A bartender looking to save time, however, might notice that many drinks have the same orange juice-to-grenadine ratio and mix the bottles. The new simplified well may be easier for pouring but can make fundamentally fewer drinks:

$$\text{span}\{(1,0,0,0),(0,1,0,0),(0,0,6,0.75)\}.$$

For example, this reduced well cannot fulfill orders for a screwdriver, which contains orange juice but not grenadine.

Example 1.4 (Cubic polynomials)**.** Define $p_k(x) \equiv x^k$. With this notation, the set of cubic polynomials can be written in two equivalent ways

$$\{ax^3 + bx^2 + cx + d \in \mathbb{R}[x] : a, b, c, d \in \mathbb{R}\} = \text{span}\{p_0, p_1, p_2, p_3\}.$$

Adding another item to a set of vectors does not always increase the size of its span, as illustrated in Figure 1.1(c). For instance, in $\mathbb{R}^2$,

$$\text{span}\{(1,0),(0,1)\} = \text{span}\{(1,0),(0,1),(1,1)\}.$$

In this case, we say that the set $\{(1,0),(0,1),(1,1)\}$ is *linearly dependent*:

Definition 1.3 (Linear dependence)**.** We provide three equivalent definitions. A set $S \subseteq \mathcal{V}$ of vectors is *linearly dependent* if:

1. One of the elements of S can be written as a linear combination of the other elements, or S contains zero.
2. There exists a non-empty linear combination of elements $\vec{v}_k \in \mathcal{S}$ yielding $\sum_{k=1}^{m} c_k \vec{v}_k = 0$ where $c_k \neq 0$ for all k.
3. There exists $\vec{v} \in S$ such that $\text{span}\, S = \text{span}\, S \backslash \{\vec{v}\}$. That is, we can remove a vector from S without affecting its span.

If S is not linearly dependent, then we say it is *linearly independent*.

Providing proof or informal evidence that each definition is equivalent to its counterparts (in an "if and only if" fashion) is a worthwhile exercise for students less comfortable with notation and abstract mathematics.

The concept of linear dependence provides an idea of "redundancy" in a set of vectors. In this sense, it is natural to ask how large a set we can construct before adding another vector cannot possibly increase the span. More specifically, suppose we have a linearly independent set $S \subseteq \mathcal{V}$, and now we choose an additional vector $\vec{v} \in \mathcal{V}$. Adding $\vec{v}$ to S has one of two possible outcomes:

1. The span of $S \cup \{\vec{v}\}$ is *larger* than the span of S.
2. Adding $\vec{v}$ to S has no effect on its span.

The *dimension* of $\mathcal{V}$ counts the number of times we can get the first outcome while building up a set of vectors:

Definition 1.4 (Dimension and basis)**.** The *dimension* of $\mathcal{V}$ is the maximal size $|S|$ of a linearly independent set $S \subset \mathcal{V}$ such that $\operatorname{span} S = \mathcal{V}$. Any set S satisfying this property is called a *basis* for $\mathcal{V}$.

Example 1.5 ($\mathbb{R}^n$)**.** The *standard basis* for $\mathbb{R}^n$ is the set of vectors of the form

$$\vec{e}_k \equiv (\underbrace{0, \ldots, 0}_{k-1 \text{ elements}}, 1, \underbrace{0, \ldots, 0}_{n-k \text{ elements}}).$$

That is, $\vec{e}_k$ has all zeros except for a single one in the k-th position. These vectors are linearly independent and form a basis for $\mathbb{R}^n$; for example in $\mathbb{R}^3$ any vector (a, b, c) can be written as $a\vec{e}_1 + b\vec{e}_2 + c\vec{e}_3$. Thus, the dimension of $\mathbb{R}^n$ is n, as expected.

Example 1.6 (Polynomials)**.** The set of monomials $\{1, x, x^2, x^3, \ldots\}$ is a linearly independent subset of $\mathbb{R}[x]$. It is infinitely large, and thus the dimension of $\mathbb{R}[x]$ is ∞.

1.2.3 Our Focus: $\mathbb{R}^n$

Of particular importance for our purposes is the vector space $\mathbb{R}^n$, the so-called *n-dimensional Euclidean space*. This is nothing more than the set of coordinate axes encountered in high school math classes:

- $\mathbb{R}^1 \equiv \mathbb{R}$ is the number line.
- $\mathbb{R}^2$ is the two-dimensional plane with coordinates (x, y).
- $\mathbb{R}^3$ represents three-dimensional space with coordinates (x, y, z).

Nearly all methods in this book will deal with transformations of and functions on $\mathbb{R}^n$.

For convenience, we usually write vectors in $\mathbb{R}^n$ in "column form," as follows:

$$(a_1, \ldots, a_n) \equiv \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

This notation will include vectors as special cases of *matrices* discussed below.

Unlike some vector spaces, $\mathbb{R}^n$ has not only a vector space structure, but also one additional construction that makes all the difference: the *dot product*.

Definition 1.5 (Dot product)**.** The dot product of two vectors $\vec{a} = (a_1, \ldots, a_n)$ and $\vec{b} = (b_1, \ldots, b_n)$ in $\mathbb{R}^n$ is given by

$$\vec{a} \cdot \vec{b} \equiv \sum_{k=1}^{n} a_k b_k.$$

Example 1.7 ($\mathbb{R}^2$)**.** The dot product of $(1, 2)$ and $(-2, 6)$ is $1 \cdot -2 + 2 \cdot 6 = -2 + 12 = 10$.

The dot product is an example of a *metric*, and its existence gives a notion of geometry to $\mathbb{R}^n$. For instance, we can use the Pythagorean theorem to define the *norm* or *length* of a vector $\vec{a}$ as the square root

$$\|\vec{a}\|_2 \equiv \sqrt{a_1^2 + \cdots + a_n^2} = \sqrt{\vec{a} \cdot \vec{a}}.$$

Then, the distance between two points $\vec{a}, \vec{b} \in \mathbb{R}^n$ is $\|\vec{b} - \vec{a}\|_2$.

Dot products provide not only lengths and distances but also angles. The following trigonometric identity holds for $\vec{a}, \vec{b} \in \mathbb{R}^3$:

$$\vec{a} \cdot \vec{b} = \|\vec{a}\|_2 \|\vec{b}\|_2 \cos\theta,$$

where θ is the angle between $\vec{a}$ and $\vec{b}$. When $n \geq 4$, however, the notion of "angle" is much harder to visualize in $\mathbb{R}^n$. We might *define* the angle θ between $\vec{a}$ and $\vec{b}$ to be

$$\theta \equiv \arccos \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|_2 \|\vec{b}\|_2}.$$

We must do our homework before making such a definition! In particular, cosine outputs values in the interval $[-1, 1]$, so we must check that the input to arc cosine (also notated $\cos^{-1}$) is in this interval; thankfully, the well-known Cauchy-Schwarz inequality $|\vec{a} \cdot \vec{b}| \leq \|\vec{a}\|_2 \|\vec{b}\|_2$ guarantees exactly this property.

When $\vec{a} = c\vec{b}$ for some $c \in \mathbb{R}$, we have $\theta = \arccos 1 = 0$, as we would expect: The angle between parallel vectors is zero. What does it mean for (nonzero) vectors to be perpendicular? Let's substitute $\theta = 90°$. Then, we have

$$0 = \cos 90° = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|_2 \|\vec{b}\|_2}.$$

Multiplying both sides by $\|\vec{a}\|_2 \|\vec{b}\|_2$ motivates the definition:

Definition 1.6 (Orthogonality)**.** Two vectors $\vec{a}, \vec{b} \in \mathbb{R}^n$ are perpendicular, or *orthogonal*, when $\vec{a} \cdot \vec{b} = 0$.

This definition is somewhat surprising from a geometric standpoint. We have managed to define what it means to be perpendicular without any explicit use of angles.

Aside 1.1. There are many theoretical questions to ponder here, some of which we will address in future chapters:

- Do all vector spaces admit dot products or similar structures?
- Do all finite-dimensional vector spaces admit dot products?
- What might be a reasonable dot product between elements of $\mathbb{R}[x]$?

Intrigued students can consult texts on real and functional analysis.

1.3 LINEARITY

A function from one vector space to another that preserves linear structure is known as a *linear* function:

Definition 1.7 (Linearity)**.** Suppose $\mathcal{V}$ and $\mathcal{V}'$ are vector spaces. Then, $\mathcal{L} : \mathcal{V} \to \mathcal{V}'$ is *linear* if it satisfies the following two criteria for all $\vec{v}, \vec{v}_1, \vec{v}_2 \in \mathcal{V}$ and $c \in \mathbb{R}$:

- $\mathcal{L}$ *preserves sums:* $\mathcal{L}[\vec{v}_1 + \vec{v}_2] = \mathcal{L}[\vec{v}_1] + \mathcal{L}[\vec{v}_2]$
- $\mathcal{L}$ *preserves scalar products:* $\mathcal{L}[c\vec{v}] = c\mathcal{L}[\vec{v}]$

It is easy to express linear maps between vector spaces, as we can see in the following examples:

Example 1.8 (Linearity in $\mathbb{R}^n$)**.** The following map $f : \mathbb{R}^2 \to \mathbb{R}^3$ is linear:

$$f(x, y) = (3x, 2x + y, -y).$$

We can check linearity as follows:

- *Sum preservation:*

$$\begin{aligned} f(x_1 + x_2, y_1 + y_2) &= (3(x_1 + x_2), 2(x_1 + x_2) + (y_1 + y_2), -(y_1 + y_2)) \\ &= (3x_1, 2x_1 + y_1, -y_1) + (3x_2, 2x_2 + y_2, -y_2) \\ &= f(x_1, y_1) + f(x_2, y_2) \quad \checkmark \end{aligned}$$

- *Scalar product preservation:*

$$\begin{aligned} f(cx, cy) &= (3cx, 2cx + cy, -cy) \\ &= c(3x, 2x + y, -y) \\ &= cf(x, y) \quad \checkmark \end{aligned}$$

Contrastingly, $g(x, y) \equiv xy^2$ is not linear. For instance, $g(1, 1) = 1$, but $g(2, 2) = 8 \neq 2 \cdot g(1, 1)$, so g does not preserve scalar products.

Example 1.9 (Integration)**.** The following "functional" $\mathcal{L}$ from $\mathbb{R}[x]$ to $\mathbb{R}$ is linear:

$$\mathcal{L}[p(x)] \equiv \int_0^1 p(x)\, dx.$$

This more abstract example maps polynomials $p(x) \in \mathbb{R}[x]$ to real numbers $\mathcal{L}[p(x)] \in \mathbb{R}$. For example, we can write

$$\mathcal{L}[3x^2 + x - 1] = \int_0^1 (3x^2 + x - 1)\, dx = \frac{1}{2}.$$

Linearity of $\mathcal{L}$ is a result of the following well-known identities from calculus:

$$\int_0^1 c \cdot f(x)\, dx = c \int_0^1 f(x)\, dx$$
$$\int_0^1 [f(x) + g(x)]\, dx = \int_0^1 f(x)\, dx + \int_0^1 g(x)\, dx.$$

We can write a particularly nice form for linear maps on $\mathbb{R}^n$. The vector $\vec{a} = (a_1, \ldots, a_n)$ is equal to the sum $\sum_k a_k \vec{e}_k$, where $\vec{e}_k$ is the k-th standard basis vector from Example 1.5. Then, if $\mathcal{L}$ is linear we can expand:

$$\begin{aligned}\mathcal{L}[\vec{a}] &= \mathcal{L}\left[\sum_k a_k \vec{e}_k\right] \text{ for the standard basis } \vec{e}_k \\ &= \sum_k \mathcal{L}\left[a_k \vec{e}_k\right] \text{ by sum preservation} \\ &= \sum_k a_k \mathcal{L}\left[\vec{e}_k\right] \text{ by scalar product preservation.}\end{aligned}$$

This derivation shows:

A linear operator $\mathcal{L}$ on $\mathbb{R}^n$ is completely determined by its action on the standard basis vectors $\vec{e}_k$.

That is, for any vector $\vec{a} \in \mathbb{R}^n$, we can use the sum above to determine $\mathcal{L}[\vec{a}]$ by linearly combining $\mathcal{L}[\vec{e}_1], \ldots, \mathcal{L}[\vec{e}_n]$.

Example 1.10 (Expanding a linear map)**.** Recall the map in Example 1.8 given by $f(x, y) = (3x, 2x + y, -y)$. We have $f(\vec{e}_1) = f(1, 0) = (3, 2, 0)$ and $f(\vec{e}_2) = f(0, 1) = (0, 1, -1)$. Thus, the formula above shows:

$$f(x, y) = xf(\vec{e}_1) + yf(\vec{e}_2) = x \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

1.3.1 Matrices

The expansion of linear maps above suggests a context in which it is useful to store multiple vectors in the same structure. More generally, say we have n vectors $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{R}^m$. We can write each as a column vector:

$$\vec{v}_1 = \begin{pmatrix} v_{11} \\ v_{21} \\ \vdots \\ v_{m1} \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} v_{12} \\ v_{22} \\ \vdots \\ v_{m2} \end{pmatrix}, \cdots, \vec{v}_n = \begin{pmatrix} v_{1n} \\ v_{2n} \\ \vdots \\ v_{mn} \end{pmatrix}.$$

Carrying these vectors around separately can be cumbersome notationally, so to simplify matters we combine them into a single $m \times n$ matrix:

$$\begin{pmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{pmatrix}.$$

We will call the space of such matrices $\mathbb{R}^{m \times n}$.

Example 1.11 (Identity matrix)**.** We can store the standard basis for $\mathbb{R}^n$ in the $n \times n$ "identity matrix" $I_{n\times n}$ given by:

$$I_{n\times n} \equiv \begin{pmatrix} | & | & & | \\ \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Since we constructed matrices as convenient ways to store sets of vectors, we can use multiplication to express how they can be combined linearly. In particular, a matrix in $\mathbb{R}^{m\times n}$ can be multiplied by a column vector in $\mathbb{R}^n$ as follows:

$$\begin{pmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \equiv c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n.$$

Expanding this sum yields the following explicit formula for matrix-vector products:

$$\begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} c_1v_{11} + c_2v_{12} + \cdots + c_nv_{1n} \\ c_1v_{21} + c_2v_{22} + \cdots + c_nv_{2n} \\ \vdots \\ c_1v_{m1} + c_2v_{m2} + \cdots + c_nv_{mn} \end{pmatrix}.$$

Example 1.12 (Identity matrix multiplication)**.** For any $\vec{x} \in \mathbb{R}^n$, we can write $\vec{x} = I_{n\times n}\vec{x}$, where $I_{n\times n}$ is the identity matrix from Example 1.11.

Example 1.13 (Linear map)**.** We return once again to the function $f(x, y)$ from Example 1.8 to show one more alternative form:

$$f(x, y) = \begin{pmatrix} 3 & 0 \\ 2 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

We similarly define a product between a matrix $M \in \mathbb{R}^{m\times n}$ and another matrix in $\mathbb{R}^{n\times p}$ with columns $\vec{c}_i$ by concatenating individual matrix-vector products:

$$M \begin{pmatrix} | & | & & | \\ \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_p \\ | & | & & | \end{pmatrix} \equiv \begin{pmatrix} | & | & & | \\ M\vec{c}_1 & M\vec{c}_2 & \cdots & M\vec{c}_p \\ | & | & & | \end{pmatrix}.$$

Example 1.14 (Mixology)**.** Continuing Example 1.3, suppose we make a tequila sunrise and second concoction with equal parts of the two liquors in our simplified well. To find out how much of the basic ingredients are contained in each order, we could combine the recipes for each column-wise and use matrix multiplication:

$$\begin{array}{r} \\ \text{Vodka} \\ \text{Tequila} \\ \text{OJ} \\ \text{Grenadine} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Well 1} & \text{Well 2} & \text{Well 3} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \\ 0 & 0 & 0.75 \end{pmatrix} \end{array} \begin{array}{c} \begin{array}{cc} \text{Drink 1} & \text{Drink 2} \end{array} \\ \begin{pmatrix} 0 & 0.75 \\ 1.5 & 0.75 \\ 1 & 2 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cc} \text{Drink 1} & \text{Drink 2} \end{array} \\ \begin{pmatrix} 0 & 0.75 \\ 1.5 & 0.75 \\ 6 & 12 \\ 0.75 & 1.5 \end{pmatrix} \end{array} \begin{array}{l} \\ \text{Vodka} \\ \text{Tequila} \\ \text{OJ} \\ \text{Grenadine} \end{array}$$

We will use capital letters to represent matrices, like $A \in \mathbb{R}^{m \times n}$. We will use the notation $A_{ij} \in \mathbb{R}$ to denote the element of A at row i and column j.

1.3.2 Scalars, Vectors, and Matrices

If we wish to unify notation completely, we can write a scalar as a 1×1 vector $c \in \mathbb{R}^{1 \times 1}$. Similarly, as suggested in §1.2.3, if we write vectors in $\mathbb{R}^n$ in column form, they can be considered $n \times 1$ matrices $\vec{v} \in \mathbb{R}^{n \times 1}$. Matrix-vector products also can be interpreted in this context. For example, if $A \in \mathbb{R}^{m \times n}$, $\vec{x} \in \mathbb{R}^n$, and $\vec{b} \in \mathbb{R}^m$, then we can write expressions like

$$\underbrace{A}_{m \times n} \cdot \underbrace{\vec{x}}_{n \times 1} = \underbrace{\vec{b}}_{m \times 1}.$$

We will introduce one additional operator on matrices that is useful in this context:

Definition 1.8 (Transpose)**.** The *transpose* of a matrix $A \in \mathbb{R}^{m \times n}$ is a matrix $A^\top \in \mathbb{R}^{n \times m}$ with elements $(A^\top)_{ij} = A_{ji}$.

Example 1.15 (Transposition)**.** The transpose of the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

is given by

$$A^\top = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Geometrically, we can think of transposition as flipping a matrix over its diagonal.

This unified notation combined with operations like transposition and multiplication yields slick expressions and derivations of well-known identities. For instance, we can compute the dot products of vectors $\vec{a}, \vec{b} \in \mathbb{R}^n$ via the following sequence of equalities:

$$\vec{a} \cdot \vec{b} = \sum_{k=1}^{n} a_k b_k = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \vec{a}^\top \vec{b}.$$

Identities from linear algebra can be derived by chaining together these operations with a few rules:

$$(A^\top)^\top = A, \qquad (A + B)^\top = A^\top + B^\top, \qquad \text{and} \qquad (AB)^\top = B^\top A^\top.$$

Example 1.16 (Residual norm)**.** Suppose we have a matrix A and two vectors $\vec{x}$ and $\vec{b}$. If we wish to know how well $A\vec{x}$ approximates $\vec{b}$, we might define a *residual* $\vec{r} \equiv \vec{b} - A\vec{x}$; this residual is zero exactly when $A\vec{x} = \vec{b}$. Otherwise, we can use the norm $\|\vec{r}\|_2$ as a proxy for the similarity of $A\vec{x}$ and $\vec{b}$. We can use the identities above to simplify:

$$\begin{aligned} \|\vec{r}\|_2^2 &= \|\vec{b} - A\vec{x}\|_2^2 \\ &= (\vec{b} - A\vec{x}) \cdot (\vec{b} - A\vec{x}) \text{ as explained in §1.2.3} \end{aligned}$$

```
function MULTIPLY(A, x⃗)
  ▷ Returns b⃗ = Ax⃗, where
  ▷ A ∈ ℝ^{m×n} and x⃗ ∈ ℝ^n
  b⃗ ← 0⃗
  for i ← 1, 2, ..., m
    for j ← 1, 2, ..., n
      b_i ← b_i + a_ij x_j
  return b⃗
```

(a)

```
function MULTIPLY(A, x⃗)
  ▷ Returns b⃗ = Ax⃗, where
  ▷ A ∈ ℝ^{m×n} and x⃗ ∈ ℝ^n
  b⃗ ← 0⃗
  for j ← 1, 2, ..., n
    for i ← 1, 2, ..., m
      b_i ← b_i + a_ij x_j
  return b⃗
```

(b)

Figure 1.2 Two implementations of matrix-vector multiplication with different loop ordering.

$$= (\vec{b} - A\vec{x})^\top (\vec{b} - A\vec{x}) \text{ by our expression for the dot product above}$$
$$= (\vec{b}^\top - \vec{x}^\top A^\top)(\vec{b} - A\vec{x}) \text{ by properties of transposition}$$
$$= \vec{b}^\top \vec{b} - \vec{b}^\top A\vec{x} - \vec{x}^\top A^\top \vec{b} + \vec{x}^\top A^\top A\vec{x} \text{ after multiplication}$$

All four terms on the right-hand side are scalars, or equivalently 1×1 matrices. Scalars thought of as matrices enjoy one additional nice property $c^\top = c$, since there is nothing to transpose! Thus,

$$\vec{x}^\top A^\top \vec{b} = (\vec{x}^\top A^\top \vec{b})^\top = \vec{b}^\top A\vec{x}.$$

This allows us to simplify even more:

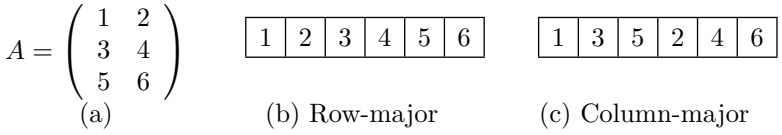
$$\begin{aligned}\|\vec{r}\|_2^2 &= \vec{b}^\top \vec{b} - 2\vec{b}^\top A\vec{x} + \vec{x}^\top A^\top A\vec{x} \\ &= \|A\vec{x}\|_2^2 - 2\vec{b}^\top A\vec{x} + \|\vec{b}\|_2^2.\end{aligned}$$

We could have derived this expression using dot product identities, but the intermediate steps above will prove useful in later discussion.

1.3.3 Matrix Storage and Multiplication Methods

In this section, we take a brief detour from mathematical theory to consider practical aspects of implementing linear algebra operations in computer software. Our discussion considers not only faithfulness to the theory we have constructed but also the *speed* with which we can carry out each operation. This is one of relatively few points at which we will consider computer architecture and other engineering aspects of how computers are designed. This consideration is necessary given the sheer number of times typical numerical algorithms call down to linear algebra routines; a seemingly small improvement in implementing matrix-vector or matrix-matrix multiplication has the potential to increase the efficiency of numerical routines by a large factor.

Figure 1.2 shows two possible implementations of matrix-vector multiplication. The difference between these two algorithms is subtle and seemingly unimportant: The order of the two loops has been switched. Rounding error aside, these two methods generate the same output and do the same number of arithmetic operations; classical "big-O" analysis from computer science would find these two methods indistinguishable. Surprisingly, however,



(a) (b) Row-major (c) Column-major

Figure 1.3 Two possible ways to store (a) a matrix in memory: (b) row-major ordering and (c) column-major ordering.

considerations related to computer architecture can make one of these options much faster than the other!

A reasonable model for the memory or RAM in a computer is as a long line of data. For this reason, we must find ways to "unroll" data from matrix form to something that could be written completely horizontally. Two common patterns are illustrated in Figure 1.3:

- A *row-major* ordering stores the data row-by-row; that is, the first row appears in a contiguous block of memory, then the second, and so on.
- A *column-major* ordering stores the data column-by-column, moving vertically first rather than horizontally.

Consider the matrix multiplication method in Figure 1.2(a). This algorithm computes all of b_1 before moving to b_2, b_3, and so on. In doing so, the code moves along the elements of A row-by-row. If A is stored in row-major order, then the algorithm in Figure 1.2(a) proceeds linearly across its representation in memory (Figure 1.3(b)), whereas if A is stored in column-major order (Figure 1.3(c)), the algorithm effectively jumps around between elements in A. The opposite is true for the algorithm in Figure 1.2(b), which moves linearly through the column-major ordering.

In many hardware implementations, loading data from memory will retrieve not just the single requested value but instead a block of data near the request. The philosophy here is that common algorithms move linearly though data, processing it one element at a time, and anticipating future requests can reduce the communication load between the main processor and the RAM. By pairing, e.g., the algorithm in Figure 1.2(a) with the row-major ordering in Figure 1.3(b), we can take advantage of this optimization by moving linearly through the storage of the matrix A; the extra loaded data anticipates what will be needed in the next iteration. If we take a nonlinear traversal through A in memory, this situation is less likely, leading to a significant loss in speed.

1.3.4 Model Problem: $A\vec{x} = \vec{b}$

In introductory algebra class, students spend considerable time solving linear systems such as the following for triplets (x, y, z):

$$\begin{aligned} 3x + 2y + 5z &= 0 \\ -4x + 9y - 3z &= -7 \\ 2x - 3y - 3z &= 1. \end{aligned}$$

Our constructions in §1.3.1 allows us to encode such systems in a cleaner fashion:

$$\begin{pmatrix} 3 & 2 & 5 \\ -4 & 9 & -3 \\ 2 & -3 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ -7 \\ 1 \end{pmatrix}.$$

More generally, we can write any linear system of equations in the form $A\vec{x} = \vec{b}$ by following the same pattern above; here, the vector $\vec{x}$ is unknown while A and $\vec{b}$ are known. Such a system of equations is *not* always guaranteed to have a solution. For instance, if A contains only zeros, then no $\vec{x}$ will satisfy $A\vec{x} = \vec{b}$ whenever $\vec{b} \neq \vec{0}$. We will defer a general consideration of when a solution exists to our discussion of linear solvers in future chapters.

A key interpretation of the system $A\vec{x} = \vec{b}$ is that it addresses the task:

Write $\vec{b}$ as a linear combination of the columns of A.

Why? Recall from §1.3.1 that the product $A\vec{x}$ encodes a linear combination of the columns of A with weights contained in elements of $\vec{x}$. So, the equation $A\vec{x} = \vec{b}$ sets the linear combination $A\vec{x}$ equal to the given vector $\vec{b}$. Given this interpretation, we define the *column space* of A to be the space of right-hand sides $\vec{b}$ for which the system $A\vec{x} = \vec{b}$ has a solution:

Definition 1.9 (Column space and rank)**.** The *column space* of a matrix $A \in \mathbb{R}^{m\times n}$ is the span of the columns of A. It can be written as

$$\text{col } A \equiv \{A\vec{x} : \vec{x} \in \mathbb{R}^n\}.$$

The *rank* of A is the dimension of col A.

$A\vec{x} = \vec{b}$ is solvable exactly when $\vec{b} \in$ col A.

One case will dominate our discussion in future chapters. Suppose A is square, so we can write $A \in \mathbb{R}^{n\times n}$. Furthermore, suppose that the system $A\vec{x} = \vec{b}$ has a solution *for all* choices of $\vec{b}$, so by our interpretation above the columns of A must span $\mathbb{R}^n$. In this case, we can substitute the standard basis $\vec{e}_1, \ldots, \vec{e}_n$ to solve equations of the form $A\vec{x}_i = \vec{e}_i$, yielding vectors $\vec{x}_1, \ldots, \vec{x}_n$. Combining these $\vec{x}_i$'s horizontally into a matrix shows:

$$A\begin{pmatrix} | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ A\vec{x}_1 & A\vec{x}_2 & \cdots & A\vec{x}_n \\ | & | & & | \end{pmatrix}$$
$$= \begin{pmatrix} | & | & & | \\ \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \\ | & | & & | \end{pmatrix} = I_{n\times n},$$

where $I_{n\times n}$ is the identity matrix from Example 1.11. We will call the matrix with columns $\vec{x}_k$ the *inverse* A^{-1}, which satisfies

$$AA^{-1} = A^{-1}A = I_{n\times n}.$$

By construction, $(A^{-1})^{-1} = A$. If we can find such an inverse, solving any linear system $A\vec{x} = \vec{b}$ reduces to matrix multiplication, since:

$$\vec{x} = I_{n\times n}\vec{x} = (A^{-1}A)\vec{x} = A^{-1}(A\vec{x}) = A^{-1}\vec{b}.$$

1.4 NON-LINEARITY: DIFFERENTIAL CALCULUS

While the beauty and applicability of linear algebra makes it a key target for study, nonlinearities abound in nature, and we must design machinery that can deal with this reality.

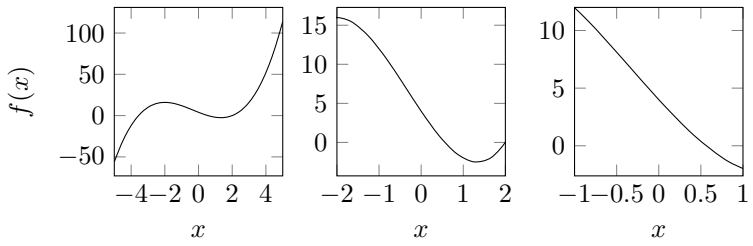


Figure 1.4 The closer we zoom into $f(x) = x^3 + x^2 - 8x + 4$, the more it looks like a line.

1.4.1 Differentiation in One Variable

While many functions are *globally* nonlinear, *locally* they exhibit linear behavior. This idea of "local linearity" is one of the main motivators behind differential calculus. Figure 1.4 shows that if you zoom in close enough to a smooth function, eventually it looks like a line. The derivative $f'(x)$ of a function $f(x) : \mathbb{R} \to \mathbb{R}$ is the slope of the approximating line, computed by finding the slope of lines through closer and closer points to x:

$$f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x}.$$

In reality, taking limits as $y \to x$ may not be possible on a computer, so a reasonable question to ask is how well a function $f(x)$ is approximated by a line through points that are a finite distance apart. We can answer these types of questions using infinitesimal analysis. Take $x, y \in \mathbb{R}$. Then, we can expand:

$$\begin{aligned}
f(y) - f(x) &= \int_x^y f'(t)\,dt \text{ by the Fundamental Theorem of Calculus} \\
&= yf'(y) - xf'(x) - \int_x^y tf''(t)\,dt, \text{ after integrating by parts} \\
&= (y-x)f'(x) + y(f'(y) - f'(x)) - \int_x^y tf''(t)\,dt \\
&= (y-x)f'(x) + y\int_x^y f''(t)\,dt - \int_x^y tf''(t)\,dt \\
&\qquad \text{again by the Fundamental Theorem of Calculus} \\
&= (y-x)f'(x) + \int_x^y (y-t)f''(t)\,dt.
\end{aligned}$$

Rearranging terms and defining $\Delta x \equiv y - x$ shows:

$$\begin{aligned}
|f'(x)\Delta x - [f(y) - f(x)]| &= \left| \int_x^y (y-t)f''(t)\,dt \right| \text{ from the relationship above} \\
&\leq |\Delta x| \int_x^y |f''(t)|\,dt, \text{ by the Cauchy-Schwarz inequality} \\
&\leq D|\Delta x|^2, \text{ assuming } |f''(t)| < D \text{ for some } D > 0.
\end{aligned}$$

We can introduce some notation to help express the relationship we have written:

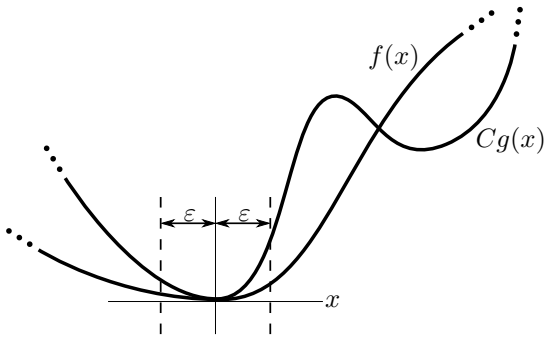


Figure 1.5 Big-O notation; in the ε neighborhood of the origin, $f(x)$ is dominated by $Cg(x)$; outside this neighborhood, $Cg(x)$ can dip back down.

Definition 1.10 (Infinitesimal big-O)**.** We will say $f(x) = O(g(x))$ if there exists a constant $C > 0$ and some $\varepsilon > 0$ such that $|f(x)| \leq C|g(x)|$ for all x with $|x| < \varepsilon$.

This definition is illustrated in Figure 1.5. Computer scientists may be surprised to see that we are defining "big-O notation" by taking limits as $x \to 0$ rather than $x \to \infty$, but since we are concerned with infinitesimal approximation quality, this definition will be more relevant to the discussion at hand.

Our derivation above shows the following relationship for smooth functions $f : \mathbb{R} \to \mathbb{R}$:

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + O(\Delta x^2).$$

This is an instance of Taylor's theorem, which we will apply copiously when developing strategies for integrating ordinary differential equations. More generally, this theorem shows how to approximate differentiable functions with polynomials:

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + f''(x)\frac{\Delta x^2}{2!} + \cdots + f^{(k)}(x)\frac{\Delta x^k}{k!} + O(\Delta x^{k+1}).$$

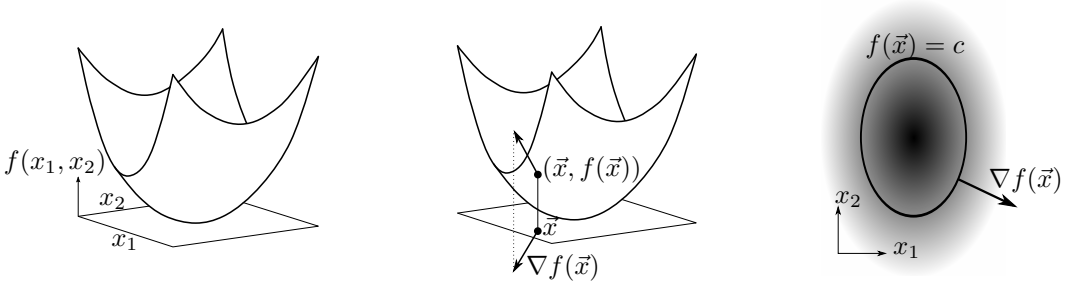
1.4.2 Differentiation in Multiple Variables

If a function f takes multiple inputs, then it can be written $f(\vec{x}) : \mathbb{R}^n \to \mathbb{R}$ for $\vec{x} \in \mathbb{R}^n$. In other words, to each point $\vec{x} = (x_1, \ldots, x_n)$ in n-dimensional space, f assigns a single number $f(x_1, \ldots, x_n)$.

The idea of local linearity must be repaired in this case, because lines are one- rather than n-dimensional objects. Fixing all but one variable, however, brings a return to single-variable calculus. For instance, we could isolate x_1 by studying $g(t) \equiv f(t, x_2, \ldots, x_n)$, where we think of $x_2, \ldots, x_n$ as constants. Then, $g(t)$ is a differentiable function of a single variable that we can characterize using the machinery in §1.4.1. We can do the same for any of the x_k's, so in general we make the following definition of the *partial derivative* of f:

Definition 1.11 (Partial derivative)**.** The k-th *partial derivative* of f, notated $\frac{\partial f}{\partial x_k}$, is given by differentiating f in its k-th input variable:

$$\frac{\partial f}{\partial x_k}(x_1, \ldots, x_n) \equiv \frac{d}{dt} f(x_1, \ldots, x_{k-1}, t, x_{k+1}, \ldots, x_n)|_{t=x_k}.$$



Graph of $f(\vec{x})$ | Steepest ascent | Level sets of $f(\vec{x})$

Figure 1.6 We can visualize a function $f(x_1, x_2)$ as a three-dimensional graph; then $\nabla f(\vec{x})$ is the direction on the (x_1, x_2) plane corresponding to the steepest ascent of f. Alternatively, we can think of $f(x_1, x_2)$ as the *brightness* at (x_1, x_2) (dark indicates a low value of f), in which case ∇f points perpendicular to level sets $f(\vec{x}) = c$ in the direction where f is increasing and the image gets lighter.

The notation used in this definition and elsewhere in our discussion "$|_{t=x_k}$" should be read as "evaluated at $t = x_k$."

Example 1.17 (Relativity)**.** The relationship $E = mc^2$ can be thought of as a function mapping pairs (m, c) to a scalar E. Thus, we could write $E(m, c) = mc^2$, yielding the partial derivatives

$$\frac{\partial E}{\partial m} = c^2 \qquad\qquad \frac{\partial E}{\partial c} = 2mc.$$

Using single-variable calculus, for a function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\begin{aligned} f(\vec{x} + \Delta\vec{x}) &= f(x_1 + \Delta x_1, x_2 + \Delta x_2, \ldots, x_n + \Delta x_n) \\ &= f(x_1, x_2 + \Delta x_2, \ldots, x_n + \Delta x_n) + \frac{\partial f}{\partial x_1}\Delta x_1 + O(\Delta x_1^2) \\ &\qquad \text{by single-variable calculus in } x_1 \\ &= f(x_1, \ldots, x_n) + \sum_{k=1}^{n}\left[\frac{\partial f}{\partial x_k}\Delta x_k + O(\Delta x_k^2)\right] \\ &\qquad \text{by repeating this } n-1 \text{ times in } x_2, \ldots, x_n \\ &= f(\vec{x}) + \nabla f(\vec{x}) \cdot \Delta\vec{x} + O(\|\Delta\vec{x}\|_2^2), \end{aligned}$$

where we define the *gradient* of f as

$$\nabla f(\vec{x}) \equiv \left(\frac{\partial f}{\partial x_1}(\vec{x}), \frac{\partial f}{\partial x_2}(\vec{x}), \cdots, \frac{\partial f}{\partial x_n}(\vec{x})\right) \in \mathbb{R}^n.$$

Figure 1.6 illustrates interpretations of the gradient of a function, which we will reconsider in our discussion of optimization in future chapters.

We can differentiate f in any direction $\vec{v}$ via the *directional derivative* $D_{\vec{v}}f$:

$$D_{\vec{v}}f(\vec{x}) \equiv \frac{d}{dt}f(\vec{x} + t\vec{v})|_{t=0} = \nabla f(\vec{x}) \cdot \vec{v}.$$

We allow $\vec{v}$ to have any length, with the property $D_{c\vec{v}}f(\vec{x}) = cD_{\vec{v}}f(\vec{x})$.

Example 1.18 ($\mathbb{R}^2$)**.** Take $f(x, y) = x^2y^3$. Then,

$$\frac{\partial f}{\partial x} = 2xy^3 \qquad \frac{\partial f}{\partial y} = 3x^2y^2.$$

Equivalently, $\nabla f(x, y) = (2xy^3, 3x^2y^2)$. So, the derivative of f at $(x, y) = (1, 2)$ in the direction $(-1, 4)$ is given by $(-1, 4) \cdot \nabla f(1, 2) = (-1, 4) \cdot (16, 12) = 32$.

There are a few derivatives that we will use many times. These formulae will appear repeatedly in future chapters and are worth studying independently:

Example 1.19 (Linear functions)**.** It is obvious but worth noting that the gradient of $f(\vec{x}) \equiv \vec{a} \cdot \vec{x} + \vec{c} = (a_1x_1 + c_1, \ldots, a_nx_n + c_n)$ is $\vec{a}$.

Example 1.20 (Quadratic forms)**.** Take any matrix $A \in \mathbb{R}^{n\times n}$, and define $f(\vec{x}) \equiv \vec{x}^\top A\vec{x}$. Writing this function element-by-element shows

$$f(\vec{x}) = \sum_{ij} A_{ij}x_ix_j.$$

Expanding f and checking this relationship explicitly is worthwhile. Take some $k \in \{1, \ldots, n\}$. Then, we can separate out all terms containing x_k:

$$f(\vec{x}) = A_{kk}x_k^2 + x_k\left(\sum_{i\neq k} A_{ik}x_i + \sum_{j\neq k} A_{kj}x_j\right) + \sum_{i,j\neq k} A_{ij}x_ix_j.$$

With this factorization,

$$\frac{\partial f}{\partial x_k} = 2A_{kk}x_k + \left(\sum_{i\neq k} A_{ik}x_i + \sum_{j\neq k} A_{kj}x_j\right) = \sum_{i=1}^{n}(A_{ik} + A_{ki})x_i.$$

This sum looks a lot like the definition of matrix-vector multiplication! Combining these partial derivatives into a single vector shows $\nabla f(\vec{x}) = (A + A^\top)\vec{x}$. In the special case when A is symmetric, that is, when $A^\top = A$, we have the well-known formula $\nabla f(\vec{x}) = 2A\vec{x}$.

We generalized differentiation from $f : \mathbb{R} \to \mathbb{R}$ to $f : \mathbb{R}^n \to \mathbb{R}$. To reach full generality, we should consider $f : \mathbb{R}^n \to \mathbb{R}^m$. That is, f inputs n numbers and outputs m numbers. This extension is straightforward, because we can think of f as a collection of single-valued functions $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ smashed into a single vector. Symbolically,

$$f(\vec{x}) = \begin{pmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{pmatrix}.$$

Each f_k can be differentiated as before, so in the end we get a matrix of partial derivatives called the *Jacobian* of f:

Definition 1.12 (Jacobian)**.** The *Jacobian* of $f : \mathbb{R}^n \to \mathbb{R}^m$ is the matrix $Df(\vec{x}) \in \mathbb{R}^{m\times n}$ with entries

$$(Df)_{ij} \equiv \frac{\partial f_i}{\partial x_j}.$$

Example 1.21 (Jacobian computation)**.** Suppose $f(x, y) = (3x, -xy^2, x + y)$. Then,

$$Df(x, y) = \begin{pmatrix} 3 & 0 \\ -y^2 & -2xy \\ 1 & 1 \end{pmatrix}.$$

Example 1.22 (Matrix multiplication)**.** Unsurprisingly, the Jacobian of $f(\vec{x}) = A\vec{x}$ for matrix A is given by $Df(\vec{x}) = A$.

Here, we encounter a common point of confusion. Suppose a function has vector input and scalar output, that is, $f : \mathbb{R}^n \to \mathbb{R}$. We defined the gradient of f as a column vector, so to align this definition with that of the Jacobian we must write $Df = \nabla f^\top$.

1.4.3 Optimization

A key problem in the study of numerical algorithms is optimization, which involves finding points at which a function $f(\vec{x})$ is maximized or minimized. A wide variety of computational challenges can be posed as optimization problems, also known as variational problems, and hence this language will permeate our derivation of numerical algorithms. Generally speaking, optimization problems involve finding extrema of a function $f(\vec{x})$, possibly subject to constraints specifying which points $\vec{x} \in \mathbb{R}^n$ are *feasible*. Recalling physical systems that naturally seek low- or high-energy configurations, $f(\vec{x})$ is sometimes referred to as an *energy* or *objective*.

From single-variable calculus, the minima and maxima of $f : \mathbb{R} \to \mathbb{R}$ must occur at points x satisfying $f'(x) = 0$. This condition is *necessary* rather than *sufficient*: there may exist saddle points x with $f'(x) = 0$ that are not maxima or minima. That said, finding such critical points of f can be part of a function minimization algorithm, so long as a subsequent step ensures that the resulting x is actually a minimum/maximum.

If $f : \mathbb{R}^n \to \mathbb{R}$ is minimized or maximized at $\vec{x}$, we have to ensure that there does not exist a single direction Δx from $\vec{x}$ in which f decreases or increases, respectively. By the discussion in §1.4.1, this means we must find points for which $\nabla f = 0$.

Example 1.23 (Critical points)**.** Suppose $f(x, y) = x^2 + 2xy + 4y^2$. Then, $\frac{\partial f}{\partial x} = 2x + 2y$ and $\frac{\partial f}{\partial y} = 2x + 8y$. Thus, critical points of f satisfy:

$$2x + 2y = 0 \qquad \text{and} \qquad 2x + 8y = 0.$$

This system is solved by taking $(x, y) = (0, 0)$. Indeed, this is the minimum of f, as can be seen by writing $f(x, y) = (x + y)^2 + 3y^2 \geq 0 = f(0, 0)$.

Example 1.24 (Quadratic functions)**.** Suppose $f(\vec{x}) = \vec{x}^\top A\vec{x} + \vec{b}^\top\vec{x} + c$. Then, from Examples 1.19 and 1.20 we can write $\nabla f(\vec{x}) = (A^\top + A)\vec{x} + \vec{b}$. Thus, critical points $\vec{x}$ of f satisfy $(A^\top + A)\vec{x} + \vec{b} = 0$.

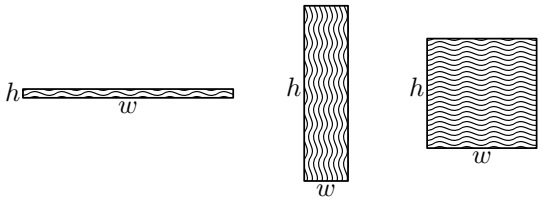


Figure 1.7 Three rectangles with the same perimeter $2w + 2h$ but unequal areas wh; the square on the right with $w = h$ maximizes wh over all possible choices with prescribed $2w + 2h = 1$.

Unlike single-variable calculus, on $\mathbb{R}^n$ we can add nontrivial *constraints* to our optimization. For now, we will consider the *equality-constrained* case, given by

$$\begin{aligned} &\text{minimize } f(\vec{x}) \\ &\text{subject to } g(\vec{x}) = \vec{0}. \end{aligned}$$

When we add the constraint $g(\vec{x}) = 0$, we no longer expect that minimizers $\vec{x}$ satisfy $\nabla f(\vec{x}) = 0$, since these points might not satisfy $g(\vec{x}) = \vec{0}$.

Example 1.25 (Rectangle areas)**.** Suppose a rectangle has width w and height h. A classic geometry problem is to maximize area with a fixed perimeter 1:

$$\begin{aligned} &\text{maximize } wh \\ &\text{subject to } 2w + 2h - 1 = 0. \end{aligned}$$

This problem is illustrated in Figure 1.7.

For now, suppose $g : \mathbb{R}^n \to \mathbb{R}$, so we only have one equality constraint; an example for $n = 2$ is shown in Figure 1.8. We define the set of points satisfying the equality constraint as $S_0 \equiv \{\vec{x} : g(\vec{x}) = 0\}$. Any two $\vec{x}, \vec{y} \in S_0$ satisfy the relationship $g(\vec{y}) - g(\vec{x}) = 0 - 0 = 0$. Applying Taylor's theorem, if $\vec{y} = \vec{x} + \Delta\vec{x}$ for small $\Delta\vec{x}$, then

$$g(\vec{y}) - g(\vec{x}) = \nabla g(\vec{x}) \cdot \Delta\vec{x} + O(\|\Delta\vec{x}\|_2^2).$$

In other words, if $g(\vec{x}) = 0$ and $\nabla g(\vec{x}) \cdot \Delta\vec{x} = 0$, then $g(\vec{x} + \Delta\vec{x}) \approx 0$.

If $\vec{x}$ is a minimum of the constrained optimization problem above, then any small displacement $\vec{x}$ to $\vec{x} + \vec{v}$ still satisfying the constraints should cause an increase from $f(\vec{x})$ to $f(\vec{x} + \vec{v})$. On the infinitesimal scale, since we only care about displacements $\vec{v}$ preserving the $g(\vec{x} + \vec{v}) = c$ constraint, from our argument above we want $\nabla f \cdot \vec{v} = 0$ for all $\vec{v}$ satisfying $\nabla g(\vec{x}) \cdot \vec{v} = 0$. In other words, ∇f and ∇g must be parallel, a condition we can write as $\nabla f = \lambda \nabla g$ for some $\lambda \in \mathbb{R}$, illustrated in Figure 1.8(c).

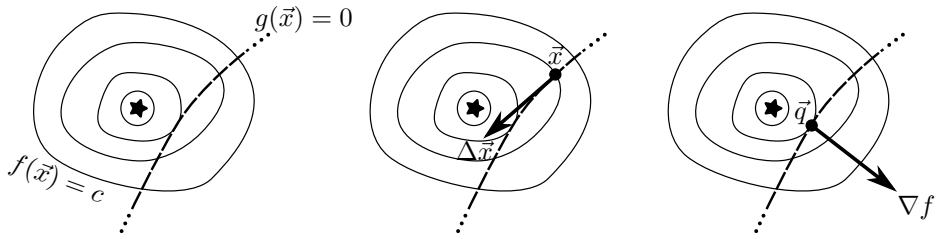
Define

$$\Lambda(\vec{x}, \lambda) \equiv f(\vec{x}) - \lambda g(\vec{x}).$$

Then, critical points of Λ without constraints satisfy:

$$\begin{aligned} \frac{\partial \Lambda}{\partial \lambda} &= -g(\vec{x}) = 0, \text{ by the constraint } g(\vec{x}) = 0. \\ \nabla_{\vec{x}} \Lambda &= \nabla f(\vec{x}) - \lambda \nabla g(\vec{x}) = 0, \text{ as argued above.} \end{aligned}$$

In other words, critical points of Λ with respect to both λ and $\vec{x}$ satisfy $g(\vec{x}) = 0$ and $\nabla f(\vec{x}) = \lambda \nabla g(\vec{x})$, exactly the optimality conditions we derived!



(a) Constrained optimization (b) Suboptimal $\vec{x}$ (c) Optimal $\vec{q}$

Figure 1.8 (a) An equality-constrained optimization. Without constraints, $f(\vec{x})$ is minimized at the star; solid lines show isocontours $f(\vec{x}) = c$ for increasing c. Minimizing $f(\vec{x})$ subject to $g(\vec{x}) = 0$ forces $\vec{x}$ to be on the dashed curve. (b) The point $\vec{x}$ is suboptimal since moving in the $\Delta\vec{x}$ direction decreases $f(\vec{x})$ while maintaining $g(\vec{x}) = 0$. (c) The point $\vec{q}$ is optimal since decreasing f from $f(\vec{q})$ would require moving in the $-\nabla f$ direction, which is perpendicular to the curve $g(\vec{x}) = 0$.

Extending our argument to $g : \mathbb{R}^n \to \mathbb{R}^k$ yields the following theorem:

Theorem 1.1 (Method of Lagrange multipliers)**.** Critical points of the equality-constrained optimization problem above are (unconstrained) critical points of the Lagrange multiplier function

$$\Lambda(\vec{x}, \vec{\lambda}) \equiv f(\vec{x}) - \vec{\lambda} \cdot g(\vec{x}),$$

with respect to both $\vec{x}$ and $\vec{\lambda}$.

Some treatments of Lagrange multipliers equivalently use the opposite sign for $\vec{\lambda}$; considering $\bar{\Lambda}(\vec{x}, \vec{\lambda}) \equiv f(\vec{x}) + \vec{\lambda} \cdot g(\vec{x})$ leads to an analogous result above.

This theorem provides an analog of the condition $\nabla f(\vec{x}) = \vec{0}$ when equality constraints $g(\vec{x}) = \vec{0}$ are added to an optimization problem and is a cornerstone of variational algorithms we will consider. We conclude with a number of examples applying this theorem; understanding these examples is crucial to our development of numerical methods in future chapters.

Example 1.26 (Maximizing area)**.** Continuing Example 1.25, we define the Lagrange multiplier function $\Lambda(w, h, \lambda) = wh - \lambda(2w + 2h - 1)$. Differentiating Λ with respect to w, h, and λ provides the following optimality conditions:

$$0 = \frac{\partial \Lambda}{\partial w} = h - 2\lambda \qquad 0 = \frac{\partial \Lambda}{\partial h} = w - 2\lambda \qquad 0 = \frac{\partial \Lambda}{\partial \lambda} = 1 - 2w - 2h.$$

So, critical points of the area wh under the constraint $2w + 2h = 1$ satisfy

$$\begin{pmatrix} 0 & 1 & -2 \\ 1 & 0 & -2 \\ 2 & 2 & 0 \end{pmatrix} \begin{pmatrix} w \\ h \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Solving the system shows $w = h = 1/4$ (and $\lambda = 1/8$). In other words, for a fixed amount of perimeter, the rectangle with maximal area is a square.

Example 1.27 (Eigenproblems)**.** Suppose that A is a symmetric positive definite matrix, meaning $A^\top = A$ (symmetric) and $\vec{x}^\top A\vec{x} > 0$ for all $\vec{x} \in \mathbb{R}^n \backslash \{\vec{0}\}$ (positive definite). We may wish to minimize $\vec{x}^\top A\vec{x}$ subject to $\|\vec{x}\|_2^2 = 1$ for a given matrix $A \in \mathbb{R}^{n\times n}$; without the constraint the function is minimized at $\vec{x} = \vec{0}$. We define the Lagrange multiplier function

$$\Lambda(\vec{x}, \lambda) = \vec{x}^\top A\vec{x} - \lambda(\|\vec{x}\|_2^2 - 1) = \vec{x}^\top A\vec{x} - \lambda(\vec{x}^\top\vec{x} - 1).$$

Differentiating with respect to $\vec{x}$, we find $0 = \nabla_{\vec{x}}\Lambda = 2A\vec{x} - 2\lambda\vec{x}$. In other words, critical points of $\vec{x}$ are exactly the *eigenvectors* of the matrix A:

$$A\vec{x} = \lambda\vec{x}, \text{ with } \|\vec{x}\|_2^2 = 1.$$

At these critical points, we can evaluate the objective function as $\vec{x}^\top A\vec{x} = \vec{x}^\top\lambda\vec{x} = \lambda\|\vec{x}\|_2^2 = \lambda$. Hence, the minimizer of $\vec{x}^\top A\vec{x}$ subject to $\|\vec{x}\|_2^2 = 1$ is the eigenvector $\vec{x}$ with minimum eigenvalue λ; we will provide practical applications and solution techniques for this optimization problem in detail in Chapter 6.

1.5 EXERCISES

[SC] 1.1 Illustrate the gradients of $f(x, y) = x^2 + y^2$ and $g(x, y) = \sqrt{x^2 + y^2}$ on the plane, and show that $\|\nabla g(x, y)\|_2$ is constant away from the origin.

[DH] 1.2 Compute the dimensions of each of the following sets:

(a) $\text{col}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

(b) $\text{span}\{(1, 1, 1), (1, -1, 1), (-1, 1, 1), (1, 1, -1)\}$

(c) $\text{span}\{(2, 7, 9), (3, 5, 1), (0, 1, 0)\}$

(d) $\text{col}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

1.3 Which of the following functions is linear? Why?

(a) $f(x, y, z) = 0$

(b) $f(x, y, z) = 1$

(c) $f(x, y, z) = (1 + x, 2z)$

(d) $f(x) = (x, 2x)$

(e) $f(x, y) = (2x + 3y, x, 0)$

1.4 Suppose that $\mathcal{U}_1$ and $\mathcal{U}_2$ are subspaces of vector space $\mathcal{V}$. Show that $\mathcal{U}_1 \cap \mathcal{U}_2$ is a subspace of $\mathcal{V}$. Is $\mathcal{U}_1 \cup \mathcal{U}_2$ always a subspace of $\mathcal{V}$?

1.5 Suppose $A, B \in \mathbb{R}^{n\times n}$ and $\vec{a}, \vec{b} \in \mathbb{R}^n$. Find a (nontrivial) linear system of equations satisfied by any $\vec{x}$ minimizing the energy $\|A\vec{x} - \vec{a}\|_2^2 + \|B\vec{x} - \vec{b}\|_2^2$.

1.6 Take $C^1(\mathbb{R})$ to be the set of continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$. Why is $C^1(\mathbb{R})$ a vector space? Show that $C^1(\mathbb{R})$ has dimension ∞.

1.7 Suppose the rows of $A \in \mathbb{R}^{m\times n}$ are given by the transposes of $\vec{r}_1, \ldots, \vec{r}_m \in \mathbb{R}^n$ and the columns of $A \in \mathbb{R}^{m\times n}$ are given by $\vec{c}_1, \ldots, \vec{c}_n \in \mathbb{R}^m$. That is,

$$A = \begin{pmatrix} - & \vec{r}_1^\top & - \\ - & \vec{r}_2^\top & - \\ & \vdots & \\ - & \vec{r}_m^\top & - \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \\ | & | & & | \end{pmatrix}.$$

Give expressions for the elements of $A^\top A$ and $AA^\top$ in terms of these vectors.

1.8 Give a linear system of equations satisfied by minima of the energy $f(\vec{x}) = \|A\vec{x} - \vec{b}\|_2$ with respect to $\vec{x}$, for $\vec{x} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, and $\vec{b} \in \mathbb{R}^m$.

1.9 Suppose $A, B \in \mathbb{R}^{n\times n}$. Formulate a condition for vectors $\vec{x} \in \mathbb{R}^n$ to be critical points of $\|A\vec{x}\|_2$ subject to $\|B\vec{x}\|_2 = 1$. Also, give an alternative expression for the value of $\|A\vec{x}\|_2$ at these critical points, in terms a Lagrange multiplier for this optimization problem.

1.10 Fix some vector $\vec{a} \in \mathbb{R}^n \backslash \{\vec{0}\}$ and define $f(\vec{x}) = \vec{a} \cdot \vec{x}$. Give an expression for the maximum of $f(\vec{x})$ subject to $\|\vec{x}\|_2 = 1$.

1.11 Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric, and define the *Rayleigh quotient* function $R(\vec{x})$ as

$$R(\vec{x}) \equiv \frac{\vec{x}^\top A\vec{x}}{\|\vec{x}\|_2^2}.$$

Show that minimizers of $R(\vec{x})$ subject to $\vec{x} \neq \vec{0}$ are eigenvectors of A.

1.12 Show that $(A^\top)^{-1} = (A^{-1})^\top$ when $A \in \mathbb{R}^{n\times n}$ is invertible. If $B \in \mathbb{R}^{n\times n}$ is also invertible, show $(AB)^{-1} = B^{-1}A^{-1}$.

1.13 Suppose $A(t)$ is a function taking a parameter t and returning an invertible square matrix $A(t) \in \mathbb{R}^{n\times n}$; we can write $A : \mathbb{R} \to \mathbb{R}^{n\times n}$. Assuming each element $a_{ij}(t)$ of $A(t)$ is a differentiable function of t, define the derivative matrix $\frac{dA}{dt}(t)$ as the matrix whose elements are $\frac{da_{ij}}{dt}(t)$. Verify the following identity:

$$\frac{d(A^{-1})}{dt} = -A^{-1}\frac{dA}{dt}A^{-1}.$$

Hint: Start from the identity $A^{-1}(t) \cdot A(t) = I_{n\times n}$.

1.14 Derive the following relationship stated in §1.4.2:

$$\frac{d}{dt}f(\vec{x} + t\vec{v})|_{t=0} = \nabla f(\vec{x}) \cdot \vec{v}.$$

1.15 A matrix $A \in \mathbb{R}^{n\times n}$ is *idempotent* if it satisfies $A^2 = A$.

(a) Suppose $B \in \mathbb{R}^{m\times k}$ is constructed so that $B^\top B$ is invertible. Show that the matrix $B(B^\top B)^{-1}B^\top$ is idempotent.

(b) If A is idempotent, show that $I_{n\times n} - A$ is also idempotent.

(c) If A is idempotent, show that $\frac{1}{2}I_{n\times n} - A$ is invertible and give an expression for its inverse.

(d) Suppose A is idempotent and that we are given $\vec{x} \neq \vec{0}$ and $\lambda \in \mathbb{R}$ satisfying $A\vec{x} = \lambda\vec{x}$. Show that $\lambda \in \{0, 1\}$.

1.16 Show that it takes at least $O(n^2)$ time to find the product AB of two matrices $A, B \in \mathbb{R}^{n\times n}$. What is the runtime of the algorithms in Figure 1.2? Is there room for improvement?

1.17 ("Laplace approximation," [13]) Suppose $p(\vec{x}) : \mathbb{R}^n \to [0, 1]$ is a *probability distribution*, meaning that $p(\vec{x}) \geq 0$ for all $\vec{x} \in \mathbb{R}^n$ and

$$\int_{\mathbb{R}^n} p(\vec{x})\, d\vec{x} = 1.$$

In this problem, you can assume $p(\vec{x})$ is infinitely differentiable.

One important type of probability distribution is the *Gaussian distribution*, also known as the normal distribution, which takes the form

$$G_{\Sigma,\vec{\mu}}(\vec{x}) \propto e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^\top \Sigma^{-1}(\vec{x}-\vec{\mu})}.$$

Here, $f(\vec{x}) \propto g(\vec{x})$ denotes that there exists some $c \in \mathbb{R}$ such that $f(\vec{x}) = cg(\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$. The covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ and mean $\vec{\mu} \in \mathbb{R}^n$ determine the particular bell shape of the Gaussian distribution.

Suppose $\vec{x}^* \in \mathbb{R}^n$ is a mode, or local maximum, of $p(\vec{x})$. Propose a Gaussian approximation of $p(\vec{x})$ in a neighborhood of $\vec{x}^*$.

Hint: Consider the *negative log likelihood function*, given by $\ell(\vec{x}) \equiv -\ln p(\vec{x})$.