Chapter 0

Mathematics Review

In this chapter we will review relevant notions from linear algebra and multivariable calculus that will figure into 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 safely can be skipped or used as reference by students with stronger background in mathematics.

0.1 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* $\mathbb{Q} = \{a/b : a, b \in \mathbb{Z}\}$[1]
- 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 we think of i as satisfying $i = \sqrt{-1}$.

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.

As with any other sets, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ can be manipulated using generic operations to generate new sets of numbers. In particular, recall that we can 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

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

[1] This is the first of many times that we will use the notation $\{A : B\}$; the braces should 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 $\mathbb{N} = \{n \in \mathbb{Z} : n > 0\}$.

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

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

0.2 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 that all satisfy the formal aspects and thus can benefit from the machinery we will develop.

0.2.1 Defining Vector Spaces

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

Definition 0.1 (Vector space)**.** *A* vector space *is a set $\mathcal{V}$ that is closed under scalar multiplication and addition.*

For our purposes, a scalar is a number in $\mathbb{R}$, and the addition and multiplication operations satisfy the usual axioms (commutativity, associativity, and so on). It is usually straightforward to spot vector spaces in the wild, including the following examples:

Example 0.1 ($\mathbb{R}^n$ 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:*

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

Example 0.2 (Polynomials)**.** *A second important example of a vector space is the "ring" of polynomials with real number inputs, denoted $\mathbb{R}[x]$. A polynomial $p \in \mathbb{R}[x]$ is a function $p : \mathbb{R} \to \mathbb{R}$ taking the form*[2]

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

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.

Elements $\vec{v} \in \mathcal{V}$ of a vector space $\mathcal{V}$ are called *vectors*, and a weighted sum of the form $\sum_i a_i \vec{v}_i$, 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 our second example, the "vectors" are functions, although we do not normally use this language to discuss $\mathbb{R}[x]$. One way to link these two viewpoints would be to identify the polynomial $\sum_k a_k x^k$ with the sequence $(a_0, a_1, a_2, \cdots)$; remember that polynomials have finite numbers of terms, so the sequence eventually will end in a string of zeros.

[2]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.

0.2.2 Span, Linear Independence, and Bases

Suppose we start with vectors $\vec{v}_1, \ldots, \vec{v}_k \in \mathcal{V}$ for vector space $\mathcal{V}$. By Definition 0.1, we have two ways to start with these vectors and construct new elements of $\mathcal{V}$: addition and scalar multiplication. The idea of *span* is that it describes all of the vectors you can reach via these two operations:

Definition 0.2 (Span)**.** *The* span *of a set $S \subseteq \mathcal{V}$ of vectors is the set*

$$\textit{span } S \equiv \{a_1\vec{v}_1 + \cdots + a_k\vec{v}_k : k \geq 0, v_i \in \mathcal{V} \textit{ for all } i, \textit{ and } a_i \in \mathbb{R} \textit{ for all } i\}.$$

Notice that span S is a *subspace* of $\mathcal{V}$, that is, a subset of $\mathcal{V}$ that is in itself a vector space. We can provide a few examples:

Example 0.3 (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 simple well, we can represent drinks as points in $\mathbb{R}^4$, with one slot for each ingredient. For instance, a typical "tequila sunrise" can be represented using the point $(0, 1.5, 6, 0.75)$, representing amounts of vodka, tequila, orange juice, and grenadine (in ounces), resp.*

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

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

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:

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

Example 0.4 (Polynomials)**.** *Define the $p_k(x) \equiv x^k$. Then, it is easy to see that*

$$\mathbb{R}[x] = \textit{span } \{p_k : k \geq 0\}.$$

Make sure you understand notation well enough to see why this is the case.

Adding another item to a set of vectors does not always increase the size of its span. For instance, in $\mathbb{R}^2$ it is clearly the case that

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

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

Definition 0.3 (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 span S = 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 leads to an idea of "redundancy" in a set of vectors. In this sense, it is natural to ask how large a set we can choose before adding another vector cannot possibly increase the span. In particular, 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 leads to 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 the span.

The *dimension* of $\mathcal{V}$ is nothing more than the maximal number of times we can get outcome 1, add $\vec{v}$ to S, and repeat.

Definition 0.4 (Dimension and basis)**.** *The* dimension *of $\mathcal{V}$ is the maximal size $|S|$ of a linearly-independent set $S \subset \mathcal{V}$ such that span $S = \mathcal{V}$. Any set S satisfying this property is called a* basis *for $\mathcal{V}$.*

Example 0.5 ($\mathbb{R}^n$)**.** *The* standard basis *for $\mathbb{R}^n$ is the set of vectors of the form*

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

That is, $\vec{e}_k$ has all zeros except for a single one in the k-th slot. It is clear that these vectors are linearly independent and form a basis; 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 we would expect.

Example 0.6 (Polynomials)**.** *It is clear that the set $\{1, x, x^2, x^3, \ldots\}$ is a linearly independent set of polynomials spanning $\mathbb{R}[x]$. Notice that this set is infinitely large, and thus the dimension of $\mathbb{R}[x]$ is ∞.*

0.2.3 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 course will deal with transformations 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 0.5 (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*

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

Example 0.7 ($\mathbb{R}^2$)**.** *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 simply $\|\vec{b} - \vec{a}\|_2$.

Dot products yield not only notions of lengths and distances but also of angles. Recall the following identity from trigonometry for $\vec{a}, \vec{b} \in \mathbb{R}^3$:

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

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

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

We must do our homework before making such a definition! In particular, recall that 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 $\vec{a} \cdot \vec{b} \leq \|\vec{a}\| \|\vec{b}\|$ 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 vectors to be perpendicular? Let's substitute $\theta = 90°$. Then, we have

$$\begin{aligned} 0 &= \cos 90° \\ &= \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} \end{aligned}$$

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

Definition 0.6 (Orthogonality)**.** *Two vectors are perpendicular, or* orthogonal*, when $\vec{a} \cdot \vec{b} = 0$.*

This definition is somewhat surprising from a geometric standpoint. In particular, we have managed to define what it means to be perpendicular without any explicit use of angles. This construction will make it easier to solve certain problems for which the nonlinearity of sine and cosine might have created headache in simpler settings.

Aside 0.1. *There are many theoretical questions to ponder here, some of which we will address in future chapters when they are more motivated:*

- *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.

0.3 Linearity

A function between vector spaces that preserves structure is known as a *linear* function:

Definition 0.7 (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 generate linear maps between vector spaces, as we can see in the following examples:

Example 0.8 (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:*

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

- *Scalar product preservation:*

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

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 this form does not preserve scalar products.*

Example 0.9 (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 somewhat more abstract example maps polynomials $p(x)$ to real numbers $\mathcal{L}[p(x)]$. For example, we can write

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

Linearity comes from the following well-known facts 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 $\mathbb{R}^n$. Recall that 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. Then, if $\mathcal{L}$ is linear we know:

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

This derivation shows the following important fact:

$\mathcal{L}$ 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 0.10 (Expanding a linear map)**.** *Recall the map in Example 0.8 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:*

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

0.3.1 Matrices

The expansion of linear maps above suggests one of many contexts 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 around separately can be cumbersome notationally, so to simplify matters we simply 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 0.11 (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 0.12 (Identity matrix multiplication)**.** *It is clearly true that 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 0.11.*

Example 0.13 (Linear map)**.** *We return once again to the expression from Example 0.8 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 in $M \in \mathbb{R}^{m \times n}$ and another matrix in $\mathbb{R}^{n \times p}$ by concatenating individual matrix-vector products:

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

Example 0.14 (Mixology)**.** *Continuing Example 0.3, 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:*

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

In general, 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.

0.3.2 Scalars, Vectors, and Matrices

It comes as no surprise that we can write a scalar as a 1×1 vector $c \in \mathbb{R}^{1 \times 1}$. Similar, as we already suggested in §0.2.3, if we write vectors in $\mathbb{R}^n$ in column form, they can be considered $n \times 1$ matrices $\vec{v} \in \mathbb{R}^{n \times 1}$. Notice that matrix-vector products can be interpreted easily 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

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

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

Definition 0.8 (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 0.15 (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 on its diagonal.

This unified treatment of scalars, vectors, and matrices combined with operations like transposition and multiplication can lead to slick derivations of well-known identities. For instance,

we can compute the dot products of vectors $\vec{a}, \vec{b} \in \mathbb{R}^n$ by making the following series of steps:

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

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

$$\begin{aligned}
(A^\top)^\top &= A \\
(A+B)^\top &= A^\top + B^\top \\
(AB)^\top &= B^\top A^\top
\end{aligned}$$

Example 0.16 (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 might use the norm $\|\vec{r}\|_2$ as a proxy for the relationship between $A\vec{x}$ and $\vec{b}$. We can use the identities above to simplify:*

$$\begin{aligned}
\|\vec{r}\|_2^2 &= \|\vec{b} - A\vec{x}\|_2^2 \\
&= (\vec{b} - A\vec{x}) \cdot (\vec{b} - A\vec{x}) \text{ as explained in §0.2.3} \\
&= (\vec{b} - A\vec{x})^\top (\vec{b} - A\vec{x}) \text{ by our expression for the dot product above} \\
&= (\vec{b}^\top - \vec{x}^\top A^\top)(\vec{b} - A\vec{x}) \text{ by properties of transposition} \\
&= \vec{b}^\top \vec{b} - \vec{b}^\top A\vec{x} - \vec{x}^\top A^\top \vec{b} + \vec{x}^\top A^\top A\vec{x} \text{ after multiplication}
\end{aligned}$$

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

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

This allows us to simplify our expression 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 intermediate steps above will prove useful in our later discussion.

0.3.3 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 §0.3.1 allow 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 linear systems 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 clearly 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 task:

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

Why? Recall from §0.3.1 that the product $A\vec{x}$ is encoding a linear combination of the columns of A with weights contained in elements of $\vec{x}$. So, the equation $A\vec{x} = \vec{b}$ asks that the linear combination $A\vec{x}$ equal 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 has a solution:

Definition 0.9 (Column space)**.** *The* column space *of a matrix $A \in \mathbb{R}^{m\times n}$ is the span of the columns of A. We can write as*

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

One important case is somewhat easier to consider. 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 $\vec{b}$. The only condition on $\vec{b}$ is that it is a member of $\mathbb{R}^n$, so by our interpretation above of $A\vec{x} = \vec{b}$ we can conclude that the columns of A span $\mathbb{R}^n$.

In this case, since the linear system is always solvable suppose we plug in the standard basis $\vec{e}_1, \ldots, \vec{e}_n$ to yield vectors $\vec{x}_1, \ldots, \vec{x}_n$ satisfying $A\vec{x}_k = \vec{e}_k$ for each k. Then, we can "stack" these expressions to show:

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

where $I_{n\times n}$ is the identity matrix from Example 0.11. We will call the matrix with columns $\vec{x}_k$ the *inverse* A^{-1}, which satisfies

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

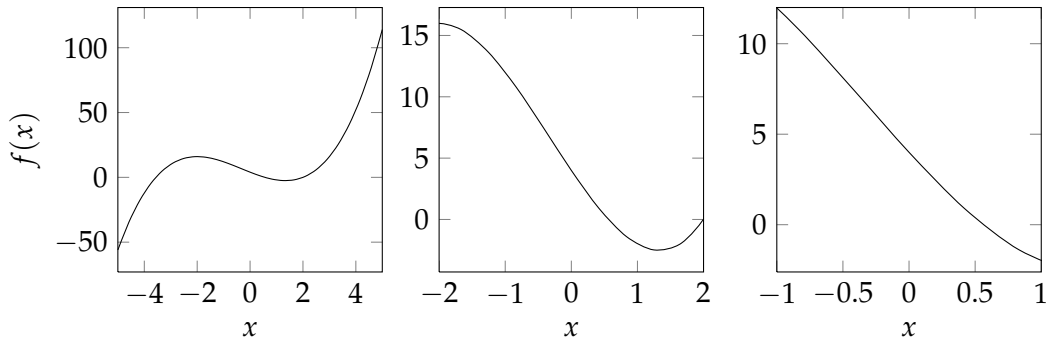

Figure 1: The closer we zoom into $f(x) = x^3 + x^2 - 8x + 4$, the more it looks like a line.

It is also easy to check that $(A^{-1})^{-1} = A$. When such an inverse exists, it is easy to solve the system $A\vec{x} = \vec{b}$. In particular, we find:

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

0.4 Non-Linearity: Differential Calculus

While the beauty and applicability of linear algebra makes it a key target of study, nonlinearities abound in nature and we often must design computational systems that can deal with this fact of life. After all, at the most basic level the square in the famous relationship $E = mc^2$ makes it less than amenable to linear analysis.

0.4.1 Differentiation

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. For instance, Figure 1 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 nothing more than the slope of the approximating line, computed by finding the slope of lines through closer and closer points to x:

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

We can express local linearity by writing $f(x + \Delta x) = f(x) + \Delta x \cdot f'(x) + O(\Delta x^2)$.

If the 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 $f(x_1, \ldots, x_n)$. Our idea of local linearity breaks down somewhat here, because lines are one-dimensional objects. However, fixing all but one variable reduces back to the case of single-variable calculus. For instance, we could write $g(t) = f(t, x_2, \ldots, x_n)$, where we simply fix constants $x_2, \ldots, x_n$. Then, $g(t)$ is a differentiable function of a single variable. Of course, we could have put t in any of the input slots for f, so in general we make the following definition of the *partial derivative* of f:

Definition 0.10 (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:*

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

The notation "$|_{t=x_k}$" should be read as "evaluated at $t = x_k$."

Example 0.17 (Relativity)**.** *The relationship $E = mc^2$ can be thought of as a function from m and c to E. Thus, we could write $E(m, c) = mc^2$, yielding the derivatives*

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

Using single-variable calculus, we can write:

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

where we define the *gradient* of f as

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

From this relationship, it is easy to see that f can be differentiated in any direction $\vec{v}$; we can evaluate this derivative $D_{\vec{v}} f$ as follows:

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

Example 0.18 ($\mathbb{R}^2$)**.** *Take $f(x, y) = x^2y^3$. Then,*

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

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

Example 0.19 (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 0.20 (Quadratic forms)**.** *Take any matrix $A \in \mathbb{R}^{n\times n}$, and define $f(\vec{x}) \equiv \vec{x}^\top A\vec{x}$. Expanding this function element-by-element shows*

$$f(\vec{x}) = \sum_{ij} A_{ij}x_i x_j;$$

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

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

With this factorization, it is easy to see

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

This sum is nothing more than the definition of matrix-vector multiplication! Thus, we can write

$$\nabla f(\vec{x}) = A\vec{x} + A^\top\vec{x}.$$

We have generalized from $f : \mathbb{R} \to \mathbb{R}$ to $f : \mathbb{R}^n \to \mathbb{R}$. To reach full generality, we would like to consider $f : \mathbb{R}^n \to \mathbb{R}^m$. In other words, f takes in n numbers and outputs m numbers. Thankfully, 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 together into a single vector. That is, we write:

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

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

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

Example 0.21 (Simple function)**.** *Suppose $f(x, y) = (3x, -xy^2, x + y)$. Then,*

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

Make sure you can derive this computation by hand.

Example 0.22 (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

$$Df = \nabla f^\top.$$

0.4.2 Optimization

Recall from single variable calculus minima and maxima of $f : \mathbb{R} \to \mathbb{R}$ must occur at points x satisfying $f'(x) = 0$. Of course, this condition is *necessary* rather than *sufficient*: there may exist points x with $f'(x) = 0$ that are not maxima or minima. That said, finding such *critical points* of f can be a step of a function minimization algorithm, so long as the next step ensures that the resulting x 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, resp. By the discussion in §0.4.1, this means we must find points for which $\nabla f = 0$.

Example 0.23 (Simple function)**.** *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:*

$$\begin{aligned} 2x + 2y &= 0 \\ 2x + 8y &= 0 \end{aligned}$$

Clearly this system is solved at $(x, y) = (0, 0)$. Indeed, this is the minimum of f, as can be seen more clearly by writing $f(x, y) = (x + y)^2 + 3y^2$.

Example 0.24 (Quadratic functions)**.** *Suppose $f(\vec{x}) = \vec{x}^\top A\vec{x} + \vec{b}^\top \vec{x} + c$. Then, from the examples in the previous section we can write $\nabla f(\vec{x}) = (A^\top + A)\vec{x} + \vec{b}$. Thus, critical points $\vec{x}$ of f satisfy $(A^\top + A)\vec{x} + \vec{b} = 0$.*

Unlike single-variable calculus, when we do calculus on $\mathbb{R}^n$ we can add *constraints* to our optimization. The most general form of such a problem looks like:

$$\begin{aligned} &\text{minimize } f(\vec{x}) \\ &\text{such that } g(\vec{x}) = \vec{0} \end{aligned}$$

Example 0.25 (Rectangle areas)**.** *Suppose a rectangle has width w and height h. A classic geometry problem is to maximize area with a fixed perimeter 1:*

$$\begin{aligned} &\textit{maximize } wh \\ &\textit{such that } 2w + 2h - 1 = 0 \end{aligned}$$

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

For now, suppose $g : \mathbb{R}^n \to \mathbb{R}$. Consider the set of points $S_0 \equiv \{\vec{x} : g(\vec{x}) = 0\}$. Obviously, any two $\vec{x}, \vec{y} \in S_0$ satisfy the relationship $g(\vec{y}) - g(\vec{x}) = 0 - 0 = 0$. Suppose $\vec{y} = \vec{x} + \Delta\vec{x}$ for small $\Delta\vec{x}$. Then, $g(\vec{y}) - g(\vec{x}) = \nabla g(\vec{x}) \cdot \Delta\vec{x} + O(\|\Delta\vec{x}\|^2)$. In other words, if we start at $\vec{x}$ satisfying $g(\vec{x}) = 0$, then if we displace in the $\Delta\vec{x}$ direction $\nabla g(\vec{x}) \cdot \Delta\vec{x} \approx 0$ to continue to satisfy this relationship.

Now, recall that the derivative of f in the direction $\vec{v}$ at $\vec{x}$ is given by $\nabla f \cdot \vec{v}$. If $\vec{x}$ is a minimum of the constrained optimization problem above, then any small displacement $\vec{x}$ to $\vec{x} + \vec{v}$ should cause an increase from $f(\vec{x})$ to $f(\vec{x} + \vec{v})$. 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 $\lambda \in \mathbb{R}$.

Define

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

Then, critical points of Λ without constraints satisfy:

$$\begin{aligned} 0 &= \frac{\partial \Lambda}{\partial \lambda} = -g(\vec{x}) \\ 0 &= \nabla_{\vec{x}} \Lambda = \nabla f(\vec{x}) - \lambda \nabla g(\vec{x}) \end{aligned}$$

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

Extending to multivariate constraints yields the following:

Theorem 0.1 (Method of Lagrange multipliers)**.** *Critical points of the 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 $\vec{\lambda}$.

Example 0.26 (Maximizing area)**.** *Continuing Example 0.25, we define the Lagrange multiplier function $\Lambda(w, h, \lambda) = wh - \lambda(2w + 2h - 1)$. Differentiating, we find:*

$$\begin{aligned} 0 &= \frac{\partial \Lambda}{\partial w} = h - 2\lambda \\ 0 &= \frac{\partial \Lambda}{\partial h} = w - 2\lambda \\ 0 &= \frac{\partial \Lambda}{\partial \lambda} = 1 - 2w - 2h \end{aligned}$$

So, critical points of the system 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 0.27 (Eigenproblems). *Suppose that A is a symmetric positive definite matrix, meaning $A^\top = A$ (symmetry) and $\vec{x}^\top A\vec{x} > 0$ for all $\vec{x} \in \mathbb{R}^n \backslash \{\vec{0}\}$ (positive definite). Often we wish to minimize $\vec{x}^\top A\vec{x}$ subject to $\|x\|_2^2 = 1$ for a given matrix $A \in \mathbb{R}^{n\times n}$; notice that without the constraint the minimum trivially takes place at $\vec{x} = \vec{0}$. We define the Lagrange multiplier function*

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

Differentiating with respect to $\vec{x}$, we find

$$0 = \nabla_{\vec{x}}\Lambda = 2A\vec{x} - 2\lambda\vec{x}$$

In other words, $\vec{x}$ is an eigenvector *of the matrix A:*

$$A\vec{x} = \lambda\vec{x}.$$

0.5 Problems

Problem 0.1. *Take $C^1(\mathbb{R})$ to be the set of functions $f : \mathbb{R} \to \mathbb{R}$ that admit a first derivative $f'(x)$. Why is $C^1(\mathbb{R})$ a vector space? Prove that $C^1(\mathbb{R})$ has dimension ∞.*

Problem 0.2. *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.

Problem 0.3. *Give a linear system of equations satisfied by minima of the energy $f(\vec{x}) = \|A\vec{x} - \vec{b}\|_2^2$ with respect to $\vec{x}$, for $\vec{x} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, and $\vec{b} \in \mathbb{R}^m$. This system is called the "normal equations" and will appear elsewhere in these notes; even so, it is worth working through and fully understanding the derivation.*

Problem 0.4. *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 $\|A\vec{x}\|_2^2$ subject to $\|Bx\|_2^2 = 1$. Also, give an alternative form for the optimal values of $\|A\vec{x}\|_2^2$.*

Problem 0.5. *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 $\|\vec{x}\| = 1$.*