A Brief Introduction to Matrix Algebra

Konstantin Tretyakov

Institute of Computer Science
University of Tartu, Estonia

Matrix algebra is one of the most useful tools in many disciplines, in particularly those related to data analysis (statistics, signal processing, machine learning, etc.). This short tutorial intends to provide a brief summary of the most important facts relating to matrix algebra for a person that has taken a linear algebra course a long time ago and has forgotten most of the details. Although I did my best to make the text readable even by a complete novice, it is clear that the size of the document does not allow for a comprehensive explanation, so beginners are advised to consult some good fat textbook full of pictures, examples and exercises. This exposition sacrifices a lot of generality for the sake of simplicity, and only covers those topics that were required by a certain course on machine learning.

1 The Euclidian Space $\mathbb{R}^n$

Vector space $\mathbb{R}^n$ In the following we shall denote the set of all real numbers by $\mathbb{R}$. The set of all pairs of real numbers will be denoted by $\mathbb{R}^2$, the set of triples — by $\mathbb{R}^3$, and, in general, the set of n-tuples — by $\mathbb{R}^n$. So,

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n} = \{(r_1, r_2, \ldots, r_n) \,|\, r_i \in \mathbb{R}\}.$$

We call elements of $\mathbb{R}^n$ *vectors* and denote them by small boldface letters. For example:

$$\mathbf{v} := (v_1, v_2, \ldots, v_n).$$

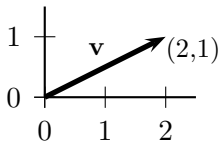
We define *addition of vectors* in a straighforward manner:

$$\mathbf{v} + \mathbf{w} = (v_1, v_2, \ldots, v_n) + (w_1, w_2, \ldots, w_n) := (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n).$$

Similarly we define componentwise *multiplication of a vector by a scalar* (by *scalars* we refer to real numbers). That is, for any $\mathbf{v} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$:

$$\alpha\mathbf{v} := (\alpha v_1, \alpha v_2, \ldots, \alpha v_n).$$

Exercise 1: Consider vectors in $\mathbb{R}^2$. It is useful to regard each vector (v_1, v_2) as an arrow on a plane, pointing from point $(0, 0)$ to point (v_1, v_2). Let $\mathbf{v} = (2, 1)$, $\mathbf{w} = (1, 2)$. Construct the corresponding arrows. Construct arrows $\mathbf{v} + \mathbf{w}$; $2\mathbf{v}$; $-\mathbf{w}$. Do you see how addition and multiplication corresponds to shifting and stretching?



The set of vectors $\mathbb{R}^n$ together with the operation of addition and multiplication by a scalar forms a *vector space*. A vector space is a convenient structure to work with, because for any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$ it holds naturally:

$$\begin{aligned} \mathbf{v} + \mathbf{w} &= \mathbf{w} + \mathbf{v} \\ \alpha(\mathbf{v} + \mathbf{w}) &= \alpha\mathbf{v} + \alpha\mathbf{w} \\ (\alpha + \beta)\mathbf{v} &= \alpha\mathbf{v} + \beta\mathbf{v} \\ \alpha(\beta\mathbf{v}) &= (\alpha\beta)\mathbf{v} \end{aligned}$$

Exercise 2: Prove it.

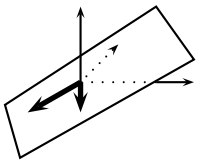
A vector space is a rather simple structure: the *only* thing we can do in a vector space, is form *linear combinations*.

Definition 1.1 (Linear combination) *A vector* $\mathbf{w} \in \mathbb{R}^n$ *is called* a linear combination *of vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ *with coefficients* $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}$ *if*

$$\mathbf{w} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k. \tag{1}$$

Exercise 3: Show that any expression consisting only of vector additions and multiplications by scalars can be transformed to the form (1).

It turns out that the concept of a linear combination is both simple enough to allow thorough mathematical analysis, and powerful enough to be useful for practical data analysis.



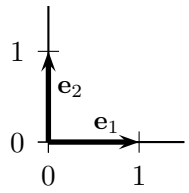
Exercise 4: Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$. Interpret them as points in space. Convince yourself that the set of all linear combinations of these two vectors is in fact the plane passing through points $\mathbf{0}$, $\mathbf{v}_1$ and $\mathbf{v}_2$.

Hint: It is the same to say that you can reach any point $\mathbf{p}$ on this plane from $\mathbf{0}$, by first moving some distance along the direction given by $\mathbf{v}_1$, and then — along the direction given by $\mathbf{v}_2$.

One very good thing about linear combinations, is that they allow us to describe large *subspaces* of $\mathbb{R}^n$ using a small set of vectors (like in the exercise above, we could describe a whole plane by just specifying its two *basis vectors*). Let's consider another example.

Basis in $\mathbb{R}^n$ Note that any vector $\mathbf{v} \in \mathbb{R}^n$ can be expressed as a *unique* linear combination of vectors

$$\begin{aligned}\mathbf{e}_1 &= (1, 0, 0, \ldots, 0)\\ \mathbf{e}_2 &= (0, 1, 0, \ldots, 0)\\ &\ldots\\ \mathbf{e}_n &= (0, 0, 0, \ldots, 1)\end{aligned}$$



with coefficients given by the components of this vector, that is:

$$\mathbf{v} = (v_1, v_2, \ldots, v_n) = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n.$$

This is a remarkable property of the set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$: we *only* need *these n vectors* to *uniquely* express *any other* vector as a linear combination of them. This property may seem obvious for $\{\mathbf{e}_1, \ldots, \mathbf{e}_2\}$, but there are other sets of vectors which allow to express any point in $\mathbb{R}^n$ as a unique linear combination, and we introduce a special name for this phenomenon:

Definition 1.2 (Basis in $\mathbb{R}^n$) *A set of vectors $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ is called a* basis in $\mathbb{R}^n$ *if any other vector $\mathbf{v} \in \mathbb{R}^n$ can be* uniquely *expressed as a linear combination of these vectors:*

$$\mathbf{v} = \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 + \cdots + \alpha_n\mathbf{b}_n.$$

The coefficients α_i are referred to as the coordinates of $\mathbf{v}$ in the basis $\mathcal{B}$.

It is not accidental that a basis in $\mathbb{R}^n$ has always exactly n elements. After all, each vector in $\mathbb{R}^n$ has n components, so it would be unfair if there existed an equivalently good representation that required either less or more information. Not every set of n vectors is a basis, however.

Exercise 5: Show that $\{(1, 2), (3, 0)\}$ is a basis in $\mathbb{R}^2$.

Exercise 6: Show that $\{(1, 0)\}$ is not a basis in $\mathbb{R}^2$.

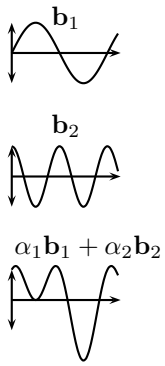
Exercise 7: Show that $\{(1, -1), (-1, 1)\}$ is not a basis in $\mathbb{R}^2$.

Exercise 8: Show that $\{(1, 0), (0, 1), (-1, 1)\}$ is not a basis in $\mathbb{R}^2$.

You might wonder: why bother about different bases, if we have that nice and simple *canonical basis* $\{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots\}$, with which we started. The reason is that often the canonical basis is not the best one — viewing vectors in another basis would gain us much more insight. In fact, a lot of data analysis methods are about transforming the data to another basis.

Exercise 9: Imagine you are an amateur computer musician. You are composing new music by mixing together different instruments. The resulting sound is a long vector $\mathbf{s}$ (say, $\mathbf{s} \in \mathbb{R}^{1000}$) containing the discretized soundwave. Just looking at this soundwave won't tell you much about how does it really sound. You know, however, that you composed your music from a fixed set of sounds $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k$. That is,

$$\mathbf{s} = \alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2 + \cdots + \alpha_k \mathbf{b}_k.$$



That means that for you, the basis $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k\}$ provides a much more convenient way of looking at the data. Explain. Think of more examples like that.

For now, we have defined the *basis in* $\mathbb{R}^n$. But recollect the example of exercise 4. Assume that the vectors $\mathbf{v}_1$ are $\mathbf{v}_2$ are not collinear (i.e. $\mathbf{v}_1 \neq \alpha\mathbf{v}_2$) and consider again the set $\mathcal{L}$ of all possible linear combinations of these vectors — the plane built on $\mathbf{v}_1$ and $\mathbf{v}_2$. It is possible to show, that every point on this plane can be *uniquely* represented as a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2\}$, that is, the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ form a *basis* for $\mathcal{L}$.

Exercise 10: Prove it. Hint: Assume the contrary and show that then necessarily $\mathbf{v}_1 = \alpha\mathbf{v}_2$.

Once again, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is not the only possible basis of $\mathcal{L}$, we could have built the same plane using other basis vectors. However, we shall always need 2 vectors in the basis of $\mathcal{L}$, because $\mathcal{L}$ is a *2-dimensional subspace* of $\mathbb{R}^3$.
Now that we have been talking about subspaces, linear combinations and basis vectors long enough, let us define them in a rigorous manner.

Definition 1.3 (Subspace) *A subset $\mathcal{S}$ of $\mathbb{R}^n$ is called a* subspace of $\mathbb{R}^n$ *if $\mathcal{S}$ is* closed under addition and multiplication by scalar*. That is, for each $\mathbf{v}, \mathbf{w} \in \mathcal{S}$ and $\alpha \in \mathbb{R}$:*

$$\mathbf{v} + \mathbf{w} \in \mathcal{S}, \qquad \alpha\mathbf{v} \in \mathcal{S}.$$

In other words, a subspace *is a subset of $\mathbb{R}^n$ which is itself a vector space.*

Exercise 11: Let $\mathcal{S}$ be a subspace of $\mathbb{R}^n$ and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathcal{S}$. Show that any linear combination of these vectors belongs to $\mathcal{S}$.

It follows, that we could have defined a subspace as something *closed with respect to linear combinations*, which should explain why the notion of a subspace is so important (at least if we are interested in linear combinations). The definition given above is, however, more convenient to use when you need to check whether a given subset is a subspace.

Exercise 12: Prove that the set $\{\mathbf{0}\}$ is a subspace of $\mathbb{R}^n$.

Exercise 13: Prove that the set $\{(x, y) \,|\, x \in [0, 1], y \in [0, 1]\}$ is not a subspace of $\mathbb{R}^2$.

Exercise 14: Prove that the set $\{(x, x^2) \,|\, x \in \mathbb{R}\}$ is not a subspace of $\mathbb{R}^2$.

Exercise 15: Prove that if $\mathcal{S}$ is a subspace of $\mathbb{R}^n$ then $\mathbf{0} = (0, 0, \ldots, 0) \in \mathcal{S}$.

Exercise 16: Prove that any straight line passing through zero is a subspace of $\mathbb{R}^2$.

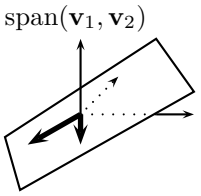
It makes sense to think of subspaces as straight lines or planes passing through zero: these are exactly the subspaces in $\mathbb{R}^3$. Don't forget, however, that in higher dimensions (e.g. $\mathbb{R}^5$), subspaces of dimensionality 4 and 5 are possible, and it's not that simple to visualize those.
How can we construct subspaces? We can take a set of vectors and try building all possible linear combinations of these vectors (their *linear span*). Naturally, such procedure will always produce a subspace.

Definition 1.4 (Linear span) *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$. The set of all linear combinations of these vectors:*

$$\mathcal{L} = \{\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_k \mathbf{v}_k \,|\, \lambda_i \in \mathbb{R}\}$$

is called the linear span *of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ and denoted as*

$$\text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k).$$



Exercise 17: Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$. Show that $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ is a subspace of $\mathbb{R}^n$.

Exercise 18: Show that the linear span of the canonical basis is the space $\mathbb{R}^n$.

Now, given a set of vectors, we can always produce a certain subspace by taking the linear span of these vectors. Conversely, if we are given a subspace, can we find a set of vectors that has this subspace as a linear span? It turns out we can, and we'll be interested in the *smallest* such set.

Exercise 19: Let $\mathcal{S} = \text{span}(\mathbf{v}_1, \mathbf{v}_2)$.
Show that then also $\mathcal{S} = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_1 + \mathbf{v}_2)$.

Definition 1.5 (Basis) *Let $\mathcal{S} \neq \{\mathbf{0}\}$ be a subspace of $\mathbb{R}^n$. A set of vectors $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k\}$ is called a* basis in $\mathcal{S}$ *if $\mathcal{S} = \text{span}(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k)$ and it is not possible to make this set smaller (i.e. it is not possible to construct $\mathcal{S}$ by taking the span of less than k vectors of $\mathcal{B}$).*

Note once more, that a basis is not just *any* "construction set" for a subspace $\mathcal{S}$, but the *smallest* such set. That is, there are no "excessive" vectors in a basis, and it results in a very remarkable property — any vector in a subspace has a *unique* representation as a linear combination of the basis vectors. In fact we could have even *defined* basis as a "set of vectors allowing a unique representation for any other vector in a subspace". Let us demonstrate that this is the case.

Exercise 20: Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k\}$ be a set of vectors. Suppose $\mathbf{b}_k$ can be expressed as a linear combination of other vectors in $\mathcal{B}$. Show that $\mathcal{B}$ is not a basis of $\text{span}(\mathcal{B})$.

Exercise 21: Let $\mathcal{B}$ be a set of vectors and let $\mathcal{S} = \text{span}(\mathcal{B})$. Suppose there exists a vector $\mathbf{v} \in \mathcal{S}$ that can be expressed as two different linear combinations of $\mathcal{B}$:

$$\mathbf{v} = \alpha_1\mathbf{b}_1 + \cdots + \alpha_k\mathbf{b}_k = \beta_1\mathbf{b}_1 + \cdots + \beta_k\mathbf{b}_k.$$

Show that then one of the vectors in $\mathcal{B}$ can be expressed as a linear combination of the others.

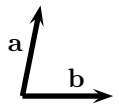
Exercise 22: Let $\mathcal{B}$ be a basis of $\mathcal{S}$. Show that any $\mathbf{v} \in \mathcal{S}$ can be expressed as a *unique* linear combination of the basis vectors. (Hint: use the results of two previous exercises)

We have just seen that a set of vectors can be a basis only if none of these vectors can be expressed via the others. There is a special name for that property.

Definition 1.6 (Linear independence) *We say that a set of vectors* $\mathbf{v}_1$, $\mathbf{v}_2$, *...,* $\mathbf{v}_k$ *is* linearly independent *if there exists no linear combination of these vectors, apart from the one where all the coefficients are zeros, that is zero. That is,*

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$$

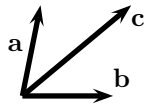
only holds when $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$.



$\mathbf{a}, \mathbf{b}$ are linearly independent

Exercise 23: Show that a set of vectors is not linearly independent iff it is possible to express one of the vectors as a linear combination of the others.

Exercise 24: Show that if $\mathcal{B}$ is a basis of $\mathcal{S}$, then $\mathcal{B}$ is linearly independent.



$\mathbf{a}, \mathbf{b}, \mathbf{c}$ are *not* linearly independent because $\mathbf{c} = \mathbf{a} + \mathbf{b}$

So it turns out that a basis of a subspace $\mathcal{S}$ is a *linearly independent* set of vectors that has this subspace as its span. Importantly, any subspace always has a basis. The size of any basis of a subspace is known as the *dimensionality* of the subspace.

Theorem 1.1 *Any subspace* $\mathcal{S} \neq \{\mathbf{0}\}$ *of* $\mathbb{R}^n$ *has a basis.*

Theorem 1.2 *Let* $\mathcal{S}$ *be a subspace of* $\mathbb{R}^n$. *Let* $\mathcal{B}_1$ *and* $\mathcal{B}_2$ *be two different bases of* $\mathcal{S}$. *Then* $|\mathcal{B}_1| = |\mathcal{B}_2| \leq n$. *We say that* $|\mathcal{B}_1|$ *is the* dimensionality *of* $\mathcal{S}$ *and denote it by* $\dim(\mathcal{S})$. *For the special case of* $\mathcal{S} = \{\mathbf{0}\}$ *we define* $\dim(\mathcal{S}) = 0$.

Exercise 25: Let $\mathcal{L} = \{(x, y, 0) \,|\, x, y \in \mathbb{R}^n\}$. Show that $\mathcal{L}$ is a subspace of $\mathbb{R}^n$ and find some basis of $\mathcal{L}$. What is the dimensionality of $\mathcal{L}$?

Exercise 26: Let $\mathcal{S}$ be a d-dimensional subspace. Show that any set of more than d vectors is not linearly independent.

The notions and properties introduced in this section can all be summarized in one short sentence: *a* d*-dimensional subspace of* $\mathbb{R}^n$ *is a linear span of* d *linearly independent vectors. These vectors form a basis of this subspace.*

Exercise 27: Prove it.

Norm and Inner Product. Let's now get back to reality for a moment. It is most natural to view $\mathbb{R}^n$ as a formalization and generalization of the three-dimensional space we are living in, the elements corresponding to locations or directions. The next best thing to do is therefore to define ways of measuring *angles* and *distances*. We define the *norm* or *length* of a vector $\mathbf{v}$ as:

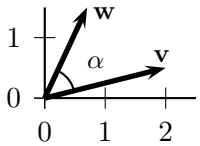
$$\|\mathbf{v}\| := \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Exercise 28: Show that if we naturally map points in our world to elements of $\mathbb{R}^3$, then distance between points $\mathbf{a}$ and $\mathbf{b}$ must be equal to $\|\mathbf{a} - \mathbf{b}\|$.

Exercise 29: Show that $\|\alpha\mathbf{v}\| = \alpha\|\mathbf{v}\|$.

For measuring angles we define the *inner product* of two vectors $\mathbf{v}$ and $\mathbf{w}$ as follows:

$$\langle \mathbf{v}, \mathbf{w} \rangle := v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$



Exercise 30: Consider the plane $\mathbb{R}^2$. Prove that $\langle \mathbf{v}, \mathbf{w} \rangle = \|\mathbf{v}\|\|\mathbf{w}\| \cos\alpha$ where α is the angle between the vectors.

Hint: Use the equation: $\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$.

Exercise 31: Let $\|\mathbf{v}\| = 1$. Show that $\langle \mathbf{w}, \mathbf{v} \rangle$ is the length of the orthogonal projection of $\mathbf{w}$ onto the line defined by $\mathbf{v}$.

Exercise 32: Let $\|\mathbf{v}\| = \|\mathbf{w}\| = 1$ and $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. Show that for any $\mathbf{x} \in \mathbb{R}^2$:

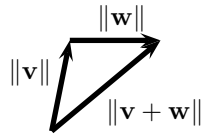
$$\mathbf{x} = \langle \mathbf{x}, \mathbf{v} \rangle \mathbf{v} + \langle \mathbf{x}, \mathbf{w} \rangle \mathbf{w}.$$

That means that $\{\mathbf{v}, \mathbf{w}\}$ is a basis in $\mathbb{R}^2$. It is possible to show that, in general, a set of n *pairwise orthogonal* vectors is always a basis in $\mathbb{R}^n$. Moreover in any d-dimensional subspace of $\mathbb{R}^n$ there exists an orthogonal basis.

Exercise 33: Prove that $\langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2$ for any $\mathbf{v} \in \mathbb{R}^n$.

Exercise 34: Prove that inner product is *symmetric*, *(bi)linear* and *positive definite*:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$$
$$\langle \alpha\mathbf{v} + \mathbf{w}, \mathbf{x} \rangle = \alpha\langle \mathbf{v}, \mathbf{x} \rangle + \langle \mathbf{w}, \mathbf{x} \rangle$$
$$\langle \mathbf{v}, \mathbf{v} \rangle = 0 \Leftrightarrow \mathbf{v} = \mathbf{0}$$



Triangle inequality

Exercise 35: Prove that $\|\mathbf{v} + \mathbf{w}\| \le \|\mathbf{v}\| + \|\mathbf{w}\|$ (*The triangle inequality*).

Exercise 36: Prove the *Cauchy-Schwarz inequality*: $\langle \mathbf{v}, \mathbf{w} \rangle \le \|\mathbf{v}\|\|\mathbf{w}\|$.

Hint: Consider $\left\| \|\mathbf{v}\|\mathbf{w} - \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\|}\mathbf{v} \right\|^2$. Rewrite it as inner product and "open the brackets".

Exercise 37: Let $\mathbf{v} \in \mathbb{R}^n$. Show that $\mathcal{S} = \{\mathbf{w} \mid \langle \mathbf{v}, \mathbf{w} \rangle = 0\}$ is a subspace of $\mathbb{R}^n$. $\mathcal{S}$ is called the *orthogonal space* of $\mathbf{v}$.

The set $\mathbb{R}^n$ together with operations of addition, multiplication by scalar, the norm and the inner product forms a *euclidian space*.

2 Matrix Notation

An $n \times m$ *matrix* is a table of real numbers with n rows and m columns. For example the following is a 2×3 matrix:

$$\begin{pmatrix} 2 & 7.1 & 0.82 \\ 8.18 & 2.8 & 4 \end{pmatrix}$$

We denote matrices by capital boldface letters (e.g. $\mathbf{A}, \mathbf{B}$). The element of matrix $\mathbf{A}$ at row i and column j will be denoted as $(\mathbf{A})_{ij}$ or a_{ij} so:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$

It is customary to use the shorthand:

$$\mathbf{A} = (a_{ij}).$$

We denote the set of all $n \times m$ matrices by $\mathbb{R}^{n \times m}$.
Similarly to how addition and multiplication by scalar was defined for the vectors in $\mathbb{R}^n$, we define *matrix addition* and *multiplication of a matrix by a scalar* componentwise:

$$\mathbf{A} + \mathbf{B} = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$$
$$\alpha\mathbf{A} = \alpha(a_{ij}) = (\alpha a_{ij})$$

It is easy to see that the set of $n \times 1$ matrices with addition and multiplication by a scalar corresponds exactly to the vector space $\mathbb{R}^n$. Therefore from now on, we use *single-column matrices* to denote vectors and make no difference between $\mathbb{R}^n$ and $\mathbb{R}^{n \times 1}$. For example:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

The operation of "mirroring" a matrix so that its rows become columns and vice versa is known as *transposition*. That is, the *transpose* $\mathbf{A}^T$ of a $n \times m$ matrix is a $m \times n$ matrix, for which

$$(\mathbf{A}^T)_{ij} = (\mathbf{A})_{ji}$$

The transpose of an n-element vector $\mathbf{v}$ is a $1 \times n$ matrix $\mathbf{v}^T$.
At last, we define the operation of *matrix multiplication*. Let $\mathbf{A} \in \mathbb{R}^{n \times l}$ and $\mathbf{B} \in \mathbb{R}^{l \times m}$. Then the *product* $\mathbf{AB}$ is an $n \times m$ matrix with entries:

$$(\mathbf{AB})_{ij} = \sum_{k=1}^{l} a_{ik} b_{kj}.$$

Let

$$\mathbf{A} = \begin{pmatrix} 2 & 7 & 0 \\ 8 & 2 & 4 \end{pmatrix}.$$

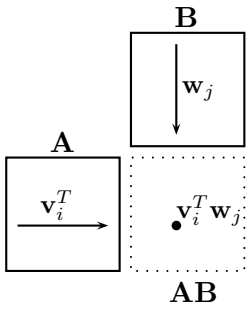
Then

$$\mathbf{A}^T = \begin{pmatrix} 2 & 8 \\ 7 & 2 \\ 0 & 4 \end{pmatrix}.$$

Note that you can't multiply *any* two matrices: the number of columns of the first term in the product must be equal to the number of rows of the second one.

Exercise 38: Show that $\mathbf{Av}$ is a linear combination of the columns of $\mathbf{A}$, with coefficients being the components of $\mathbf{v}$.

Exercise 39: Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. Show that $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \mathbf{w}$. Due to this, in the following we shall avoid the $\langle \cdot, \cdot \rangle$ notation in favor of matrix multiplication.



Exercise 40: Let $\mathbf{v}_1^T, \mathbf{v}_2^T, \ldots, \mathbf{v}_n^T \in \mathbb{R}^{1 \times l}$ be the rows of $\mathbf{A}$ and $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m \in \mathbb{R}^{l \times 1}$ — the columns of $\mathbf{B}$. Show that $(\mathbf{AB})_{ij} = \mathbf{v}_i^T \mathbf{w}_j$.

Exercise 41: Show that matrix multiplication is associative: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

Exercise 42: Show that matrix multiplication is distributive: $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$.

Exercise 43: Show that $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$.

Exercise 44: Find the matrix $\mathbf{I} \in \mathbb{R}^{n \times n}$ for which $\mathbf{IA} = \mathbf{A}$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$. $\mathbf{I}$ is called the *identity matrix*.

Exercise 45: Show that in general $\mathbf{AB} \neq \mathbf{BA}$.

3 Linear Functions

As you remember, the two most basic things we could do with vectors in $\mathbb{R}^n$ were addition and multiplication by scalars. In this part we shall try to do something useful with the vectors without losing this structure. That is, we shall examine all possible functions f that *preserve* addition and scalar multiplication (which means, preserve linear combinations).

Definition 3.1 (Linear function) *A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is called* linear *if for each $\mathbf{v}, \mathbf{w} \in \mathbb{R}^m, \alpha \in \mathbb{R}$:*

$$f(\alpha\mathbf{v} + \mathbf{w}) = \alpha f(\mathbf{v}) + f(\mathbf{w})$$

Exercise 46: Show that $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$ is a linear function.

Exercise 47: Show that $f(\mathbf{x}) = \mathbf{x}$ is a linear function.

Exercise 48: Let $\mathcal{B}$ be some basis in $\mathbb{R}^n$. Let $f(\mathbf{v})$ return the vector of coordinates of $\mathbf{v}$ in basis $\mathcal{B}$. Show that f is a linear function.

Exercise 49: Consider vectors in $\mathbb{R}^2$, interpret them as arrows. Show that rotation around 0 on a fixed angle α is a linear function.

Exercise 50: Show that $f(\mathbf{v}) = \mathbf{v} + \mathbf{c}$ for $\mathbf{c} \neq \mathbf{0}$ is not a linear function.

Exercise 51: Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a linear function. Show that $f(\mathbf{0}) = \mathbf{0}$.

Exercise 52: Let $f : \mathbb{R}^l \to \mathbb{R}^n$ and $g : \mathbb{R}^m \to \mathbb{R}^l$ be linear functions. Show that $f \circ g$ is also a linear function. $(f \circ g(\mathbf{x}) = f(g(\mathbf{x})))$.

Exercise 53: Show that a linear function maps a subspace to a subspace.

The class of linear functions can be described as the class of all possible projections, reflections, rotations and shears: those functions that transform straight lines into straight lines or single points (but never into curves!) and keep the zero intact.
Let $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ be the canonical basis in $\mathbb{R}^n$. Then each vector can be represented as

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n.$$

But then if f is a linear function:

$$f(\mathbf{v}) = v_1 f(\mathbf{e}_1) + v_2 f(\mathbf{e}_2) + \cdots + v_n f(\mathbf{e}_n).$$

Exercise 54: Prove it.

That means, that in order to *completely define* a linear transformation, you *only* need to specify how it transforms the vectors of the canonical basis. Now let us construct a matrix $\mathbf{F}$, with columns being the vectors $f(\mathbf{e}_1), f(\mathbf{e}_2), \ldots, f(\mathbf{e}_n)$. It holds then:

$$f(\mathbf{v}) = \mathbf{F}\mathbf{v}.$$

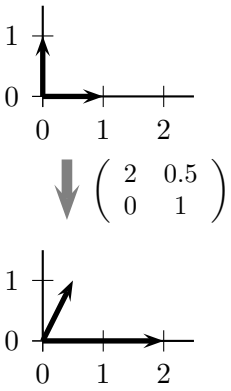
Exercise 55: Prove it. Show that each linear transformation $f : \mathbf{R}^m \to \mathbf{R}^n$ corresponds *uniquely* to a matrix $\mathbf{F} \in \mathbb{R}^{n \times m}$ and vice versa.

Exercise 56: If you were wondering why matrix multiplication was defined like it was — here is the answer. Let $\mathbf{F} \in \mathbb{R}^{n \times l}$ correspond to linear function f and let $\mathbf{G} \in \mathbb{R}^{l \times m}$ correspond to linear transformation g. Show that matrix $\mathbf{FG}$ corresponds to linear transformation $f \circ g$.

Exercise 57: Show that $(f \circ g) \circ h = f \circ (g \circ h)$ for any transformations f, g, h. Show how it follows that matrix multiplication is associative.

Exercise 58: Let $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ be the identity transformation, ie $\mathrm{id}(\mathbf{x}) = \mathbf{x}$. Show that the matrix corresponding to it is the identity matrix $\mathbf{I}$.

We have just discovered a very remarkable fact: *linear functions and matrices are equivalent*. Every time someone is talking about matrices you may think about linear functions. Every time someone says "linear function" — you know he means "matrix". It also provides a useful way of interpreting matrices. You know how to interpret vectors as arrows, right? Now you interpret a matrix as a *set of arrows* corresponding to the columns of the matrix. This set of arrows shows how the matrix transforms the canonical basis.



Invertible Transformations

Definition 3.2 (Invertible function) *We say that a function $f : \mathbb{R}^m \to \mathbb{R}^n$ is* invertible *if there exists a transformation $f^{-1} : \mathbb{R}^n \to \mathbb{R}^m$ such that for any $\mathbf{v} \in \mathbb{R}^m$ $f^{-1}(f(\mathbf{v})) = \mathbf{v}$. That is, f^{-1} can "undo" what f does. We call f^{-1} the* inverse *of f.*

Exercise 59: If f is invertible, is f^{-1} necessarily also invertible? When is it?

Exercise 60: Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be invertible. Show that f maps any d-dimensional subspace into a d-dimensional subspace.
Hint: First show that f maps linearly independent sets to linearly independent sets.

Exercise 61: Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be invertible. Show that f^{-1} is then also invertible and $(f^{-1})^{-1} = f$.

Exercise 62: Let $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ be the identity transformation and let $f : \mathbb{R}^n \to \mathbb{R}^n$ be invertible. Show that $f \circ f^{-1} = f^{-1} \circ f = \mathrm{id}$.

It follows from above that it makes most sense to speak about invertible transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$. Otherwise we can only "undo" transformations that map points from a lower dimensional space into a higher dimensional space, but not vice versa. So next time, when we say invertible transformation, we mean the case of $\mathbb{R}^n \to \mathbb{R}^n$.

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation. We call the $n \times n$ matrix $\mathbf{F}$, corresponding to this transformation *invertible*, and the matrix $\mathbf{F}^{-1}$, corresponding to f^{-1} is called the *inverse* of $\mathbf{F}$.

Exercise 63: Let $\mathbf{F}$ be an invertible $n \times n$ matrix.
Show that $\mathbf{F}\mathbf{F}^{-1} = \mathbf{F}^{-1}\mathbf{F} = \mathbf{I}$.

Exercise 64: Let $\mathbf{A}$ be an invertible matrix and let $\mathbf{A}\mathbf{x} = \mathbf{y}$.
Show that $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$.

Exercise 65: Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis in $\mathbb{R}^n$. Let f be a function that, given a vector $\mathbf{v} \in \mathbb{R}^n$ returns its coordinates in basis $\mathcal{B}$. Show that $f(\mathbf{v}) = \mathbf{B}^{-1}\mathbf{v}$, where $\mathbf{B}$ is a matrix with columns $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$.
Hint: Show that if $\mathbf{w} = f(\mathbf{v})$ then $\mathbf{v} = \mathbf{B}\mathbf{w}$.

Exercise 66: Find the inverse of $\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$.

Exercise 67: Find the inverse of $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

Invertible transformations turn out to be very useful, so it's important to know how they differ from the noninvertible ones. Clearly, a function is invertible if and only if it maps different arguments to different values. (Otherwise, if for some $\mathbf{v}_1 \neq \mathbf{v}_2$ we have $f(\mathbf{v}_1) = f(\mathbf{v}_2) = \mathbf{w}$, we can't really tell by only looking at $\mathbf{w}$, whether it came from $\mathbf{v}_1$ or $\mathbf{v}_2$.) Linear functions are even more interesting in that respect. Let f be a noninvertible linear function. That is, let there exist $\mathbf{v}_1 \neq \mathbf{v}_2$ such that $f(\mathbf{v}_1) = f(\mathbf{v}_2)$. It easily follows from it, that there must be a whole subspace that f maps to $\mathbf{0}$.

Exercise 68: Let $f(\mathbf{v}_1) = f(\mathbf{v}_2)$ for $\mathbf{v}_1 \neq \mathbf{v}_2$. Show that there exists $\mathbf{s} \neq \mathbf{0}$ such that $f(\alpha\mathbf{s}) = \mathbf{0}$ for any $\alpha \in \mathbb{R}$.

Exercise 69: Show that the set of all $\mathbf{v} \in \mathbb{R}^n$ for which $f(\mathbf{v}) = \mathbf{0}$ is a subspace.

In the following we shall denote the subspace of all vectors that f maps to $\mathbf{0}$ by $\mathrm{Ker}(f)$. That is,

$$\mathrm{Ker}(f) = \{\mathbf{v} \,|\, f(\mathbf{v}) = \mathbf{0}\}$$

If the only vector that f maps to $\mathbf{0}$ is $\mathbf{0}$ (i.e. $\mathrm{Ker}(f) = \{\mathbf{0}\}$), then f is certainly invertible. !

Exercise 70: Prove it.

Here's another important observation. An invertible function maps the whole space $\mathbb{R}^n$ onto the whole space. Which is the same to say that it maps the basis of $\mathbb{R}^n$ to some other basis of $\mathbb{R}^n$.

Exercise 71: Show that if f maps some basis of $\mathbb{R}^n$ to a non-basis then f is not invertible.

Therefore the columns of the corresponding matrix must be linearly independent.

Exercise 72: Prove it.

Conversely, let the columns of a matrix be linearly independent. Then the corresponding f is certainly invertible.

Exercise 73: Prove it.

Which makes us conclude: *f is invertible iff the columns of $\mathbf{F}$ are linearly independent.* !

Suppose now that f is not invertible. As we've noted above, f should map $\mathbb{R}^n$ to some lower-dimensional subspace of $\mathbb{R}^n$. The dimensionality of this space is an important parameter as it tells us how much information f "preserves".

Definition 3.3 (Rank) *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and let $\mathcal{S} = \{f(\mathbf{v}) \,|\, \mathbf{v} \in \mathbb{R}^n\}$. We call $\dim(\mathcal{S})$ the* rank *of f and denote by* $\text{rank}(f)$.

Note that rank is defined for any linear transformations, not only $\mathbb{R}^n \to \mathbb{R}^n$.

Exercise 74: Find $\text{rank}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

Exercise 75: Show that f is invertible iff $\text{rank}(f) = n$.

Exercise 76: Consider $f : \mathbb{R}^3 \to \mathbb{R}^3$. Show that: a) if f is invertible then $\text{rank}(f) = 3$; b) if f projects all points to some plane, then $\text{rank}(f) = 2$; c) if f projects all points to some line then $\text{rank}(f) = 1$; d) if f maps every point to $\mathbf{0}$ then $\text{rank}(f) = 0$.

Exercise 77: Show that $\text{rank}(\mathbf{FG}) \leq \min(\text{rank}(\mathbf{F}), \text{rank}(\mathbf{G}))$.

Exercise 78: Show that $\text{rank}(\mathbf{F}) = \text{rank}(\mathbf{F}^T)$.

Hint:

First show that $\text{Ker}(\mathbf{F}) = \text{Ker}(\mathbf{F}^T\mathbf{F})$ by using the fact that $\|\mathbf{Fv}\|^2 = \mathbf{v}^T\mathbf{F}^T\mathbf{Fv}$.

Follow from it that $\text{rank}(\mathbf{F}) = \text{rank}(\mathbf{F}^T\mathbf{F})$ and hence $\text{rank}(\mathbf{F}) \leq \text{rank}(\mathbf{F}^T)$.

At last, note that by symmetry $\text{rank}(\mathbf{F}^T) \leq \text{rank}(\mathbf{F})$.

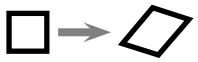
We conclude this with one more insightful observation. We've noted above that a noninvertible f maps some subspace of $\mathbb{R}$ to $\mathbf{0}$. The dimensionality of this subspace therefore indicates how much information f "loses". Now, if $\dim(\text{Ker}(f))$ is the dimensionality "lost" by f and $\text{rank}(f)$ is the dimensionality "preserved" by f, it would make sense to have their sum equal to n. This actually holds true.

Theorem 3.1 *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then*

$$\dim(\text{Ker}(f)) + \text{rank}(f) = n.$$

Orthogonal Transformations A particularly interesting class of linear transformations is formed by those that *preserve angles and distances*. These are precisely all the rotations and mirrorings.


Preserves angles


Doesn't preserve angles

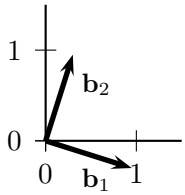
Definition 3.4 (Orthogonal transformation) *We say that a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^n$ is* orthogonal*, if it preserves the inner product. That is, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle f(\mathbf{x}), f(\mathbf{y}) \rangle.$$

It turns out that an orthogonal transformation is always invertible, and the matrix $\mathbf{F}$ of an orthogonal transformation f has a nice property: $\mathbf{F}^{-1} = \mathbf{F}$ (which can sometimes be very convenient: while inverting matrices is in general rather complicated, for orthogonal functions it's trivial!) . To show that, we first note that an orthogonal transformation maps an *orthonormal basis* to an *orthonormal basis*.

Definition 3.5 (Orthonormal basis) *Let $\mathcal{S}$ be a subspace of $\mathbb{R}^n$ and $\mathcal{B}$ be a basis in it. We say that $\mathcal{B}$ is* orthonormal *if all the vectors in $\mathcal{B}$ are of unit length and pairwise orthogonal. That is:*

$$\|\mathbf{b}_i\| = 1, \qquad \textit{and} \qquad \langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0, \ \textit{if } i \neq j.$$



Exercise 79: Let $\mathbf{B}$ be a matrix, the columns of which are orthonormal. Show that $\mathbf{B}^T\mathbf{B} = \mathbf{I}$.

Exercise 80: Let f be an orthogonal transformation. Show that f maps an orthonormal basis to an orthonormal basis.

Exercise 81: Let $\mathbf{F}$ be an orthogonal matrix. Show that $\mathbf{F}^T\mathbf{F} = \mathbf{I}$. (Hint: Show that columns of $\mathbf{F}$ are orthogonal.)

To summarize, *orthogonal transformations* correspond to rotations and mirrorings, they preserve angles and distances, the columns of an orthogonal matrix form an orthonormal basis, and the inverse of an orthogonal matrix is simply its transpose.

!

Symmetric Transformations Another interesting class of transformations are *symmetric* transformations.

Definition 3.6 (Symmetric transformation) *We call a linear transformation f* symmetric *if its matrix $\mathbf{F}$ is symmetric, i.e. $\mathbf{F} = \mathbf{F}^T$.*

Exercise 82: Let f be a symmetric transformation. Show that $\langle f(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, f(\mathbf{y}) \rangle$.

One simple example of a symmetric transformation is given by a *diagonal matrix* — a matrix whose nonzero elements are only on the diagonal. Such a transformation performs a *scaling* along the coordinate axes.

Exercise 83: Examine the symmetric transformation given by the matrix $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$. How does it transform the points of the plane?

It turns out that all the symmetric transformations are in a sense similar to such diagonal scaling, with the only difference that in general the scaling is not performed along the coordinate axes, but maybe along some other orthogonal set of directions.

Exercise 84: Let $\mathcal{B}$ be an orthonormal basis in $\mathbb{R}^n$. Let $\mathbf{B}$ be a matrix that transforms the canonical basis to $\mathcal{B}$. Let $\mathbf{D}$ be some diagonal matrix with entries d_1, d_2, ..., d_n on the diagonal. Show that the matrix $\mathbf{F} = \mathbf{BDB}^T$ is a transformation, that scales all the vectors along direction $\mathbf{b}_1$ by d_1, all the vectors along $\mathbf{b}_2$ by d_2, etc. Show that $\mathbf{F}$ is a symmetric matrix.

Theorem 3.2 *Any symmetric matrix* $\mathbf{F}$ *can be represented as:*

$$\mathbf{F} = \mathbf{BDB}^T, \tag{2}$$

where $\mathbf{B}$ *is orthogonal and* $\mathbf{D}$ *a diagonal matrix.*

Eigenvalues and Eigenvectors We have just seen that a symmetric matrix in fact performs a *scaling* along certain orthogonal directions. These directions, and the factors of scaling are often of special interest, hence they have a special name.

Definition 3.7 (Eigenvalues and Eigenvectors) *Let* $f : \mathbb{R}^n \to \mathbb{R}^n$ *be a linear transformation. If for some* $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} \neq \mathbf{0}$ *there exists* $\alpha \in \mathbb{R}$ *such that*

$$f(\mathbf{v}) = \alpha\mathbf{v},$$

we call $\mathbf{v}$ *an* eigenvector *of* f*, and* α *— the corresponding* eigenvalue*.*

Exercise 85: Let $\mathbf{v}$ be an eigenvector of f. Show that $c\mathbf{v}$ is also an eigenvector for any $c \neq 0$.

Exercise 86: Show that all the eigenvectors of f corresponding to a given eigenvalue α form a subspace.

Exercise 87: Suppose you know eigenvectors and eigenvalues of f. What can you say about eigenvectors and eigenvalues of f^{-1}?

Exercise 88: Let $\mathbf{F}$ be symmetric, and let $\mathbf{F} = \mathbf{BDB}^T$ the decomposition given in theorem 3.2. Show that each column of $\mathbf{B}$ is an eigenvector of $\mathbf{F}$, and the corresponding eigenvalues are given by the diagonal elements of $\mathbf{D}$.

Exercise 89: Show that it is possible to construct an orthonormal basis from the eigenvectors of a symmetric matrix.

Although the notion of eigenvalues and eigenvectors is most often used in the context of symmetric matrices, it is not necessarily always the case. Eigenvectors of nonsymmetric matrices may sometimes also be of interest.

Exercise 90: Let $\mathbf{V}$ be an invertible transformation and $\mathbf{D}$ a diagonal matrix. Find the eigenvectors and eigenvalues of $\mathbf{F} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$.

Finding the eigenvectors of a matrix really means trying to represent it in the form $\mathbf{F} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$. Such representation is often very insightful, as it immediately shows the vectors that are "most important" with respect to the transformation $\mathbf{F}$. The eigenvectors (given by the columns of $\mathbf{V}$) can provide us with a very convenient *basis* for our data: when we represent the data in this basis, the transformation $\mathbf{F}$ is nothing more than a coordinate-wise scaling given by the diagonal matrix $\mathbf{D}$.

By just looking at the eigenvalues of $\mathbf{F}$ (the *spectrum* of $\mathbf{F}$) we can immediately see how $\mathbf{F}$ works: which directions are "amplified", which are preserved and which are mapped to 0 (these correspond to zero eigenvalues).

Exercise 91: Let $\mathbf{F} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ with $\mathbf{D}$ a diagonal matrix. Show that the number of nonzero entries on the diagonal of $\mathbf{D}$ is equal to $\text{rank}(\mathbf{F})$.

If the spectra of two different matrices are equal, such matrices may often be considered equivalent, in a certain sense. Of course, not every matrix can be represented like that. For example, 2D rotations have no eigenvectors.

Exercise 92*: Have you ever wondered, why *sine-waves* are considered so important in sound processing? Well, the answer is, that sine-waves are precisely the *eigensignals* of time-invariant linear transformations.

Exercise 93*: *Covariance matrix* is a matrix, commonly used in data analysis. It's a symmetric matrix $\mathbf{\Sigma}$, with the property that for a data point $\mathbf{x}$, the product $\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}$ indicates how "interesting" (improbable) the point $\mathbf{x}$ is — the larger, the better. Show how the eigenvalue decomposition (2) of $\mathbf{\Sigma}$ can provide a more convenient basis for the data.

Determinant As we know, a linear transformation can be described by saying what it does to the canonical basis. Consider the parallelogram built on the vectors of the canonical basis. That is, consider the "unit-hypercube"

$$\mathcal{U} = \{(v_1, v_2, \ldots, v_n) \,|\, v_i \in [0,1]\}.$$

In one-dimensional case this is just the segment $[0,1]$, in two-dimensional case — the square $[0,1] \times [0,1]$, in three-dimensional case — the cube $[0,1] \times [0,1] \times [0,1]$ and in higher dimensions it's some analogous object. It corresponds to our notion of a "hypercube" and it's "volume" (which is *length* in $\mathbb{R}^1$ and *area* in $\mathbb{R}^2$) is 1, hence the name "unit hypercube".
Let's ask, for a given linear transformation f, how does it transform this unit hypercube. Does it stretch it or shrink it? Does it mirror it? Make it "flat"? Such knowledge often turns out to be useful: for example, if we know that f shrinks the unit hypercube to half the size, we can immediately deduce that f performs such shrinking for *any* set of points; if we know that f "flattens" the n-dimensional hypercube to a lower-dimensional hypercube, we can conclude that f is not invertible, etc.

Exercise 94: Examine the statements above in detail and prove them.

So, the question we're interested in is, *given a linear transformation f, what is the volume of $f(\mathcal{U})$*? If $\mathcal{U}$ is a hyper-parallelogram built on the vectors of the canonical basis, then $f(\mathcal{U})$ is a hyper-parallelogram built on the vectors $f(\mathbf{e}_i)$, so another way to ask the same question is: *given a matrix $\mathbf{F}$, what is the volume of the parallelogram built on the columns of this matrix*?

Exercise 95: Consider $\mathbf{F} = \begin{pmatrix} 1 & 0.5 \\ 0 & 0.5 \end{pmatrix}$. Draw the parallelogram to which $\mathbf{F}$ transforms the unit square. What is the area of this parallelogram?

A last twist before the final definition: not only shall we be interested in the volume of $f(\mathcal{U})$, but rather the *signed volume* of it. Intuitively, we shall say that the volume of $f(\mathcal{U})$ is negative, if f had to "mirror" the unit hypercube inbetween in order to obtain $f(\mathcal{U})$.

Exercise 96: Consider $\mathbf{F} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Show how $\mathbf{F}$ transforms the unit square. Do you see that although the square itself is left intact, its *orientation* has changed (i.e. it has been mirrored)? In this case we say that the signed volume of $f(\mathcal{U})$ is -1 rather than 1.

Definition 3.8 (Determinant) *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and let the corresponding matrix be $\mathbf{F}$. We define the* determinant *of f as the* signed volume *of the parallelogram built on the columns of $\mathbf{F}$. The determinant is denoted as* $\det(f)$ *(or* $\det(\mathbf{F})$*).*
Sometimes it is convenient to speak of the determinant of a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *(which is still the signed area of a parallelogram built on these vectors). Then we use the notation* $\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$.

If $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ *are the columns of* $\mathbf{F}$*, then by definition*

$$\det(\mathbf{F}) = \det(\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n).$$

Determinant satisfies four important properties: !

1. The identity transformation does not change the volume of the hypercube:
$$\det(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = \det(\mathrm{id}) = 1.$$

2. By exchanging any two columns, we change the orientation of the parallelogram, but not its area:
$$\det(\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n) = -\det(\mathbf{f}_2, \mathbf{f}_1, \ldots, \mathbf{f}_n).$$

3. The area of a sum of two parallelograms is the sum of their areas. In other words, determinant is *linear* in every column:
$$\det(\alpha\mathbf{f}_1 + \mathbf{g}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n) = \alpha \det(\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n) + \det(\mathbf{g}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n).$$

4. If two columns are equal, the corresponding area is 0 (i.e. the corresponding parallelogram is "flat"):
$$\det(\mathbf{f}_1, \mathbf{f}_1, \mathbf{f}_3, \mathbf{f}_4, \ldots, \mathbf{f}_n) = 0.$$

It is remarkable that these four properties define determinant uniquely.

Exercise 97*: Express $\det(\mathbf{F})$ in terms of the matrix elements f_{ij}.
Hint: First express each column of $\mathbf{F}$ as $\mathbf{f}_i = f_{1i}\mathbf{e}_1 + f_{2i}\mathbf{e}_2 + \cdots + f_{ni}\mathbf{e}_n$. Use linearity (property 3) to "open the brackets" and rewrite $\det(\mathbf{F})$ as a big sum. At last, apply properties 1 and 4 to evaluate each element of the sum.

There are lots of things which you can discover about the determinant, here are some of them:

Exercise 98: Let $\mathbf{D}$ be diagonal. Show that $\det(\mathbf{D})$ is the product of the diagonal elements of $\mathbf{D}$.

Exercise 99: Show that $\det(\alpha\mathbf{F}) = \alpha^n \det(\mathbf{F})$.

Exercise 100: Show that $\det(\mathbf{FG}) = \det(\mathbf{F}) \det(\mathbf{G})$.

Exercise 101: Show that $\det(\mathbf{F}) = 0$ iff the columns of $\mathbf{F}$ are linearly independent (i.e. f is not invertible).

Exercise 102: Let $\mathbf{F}$ be orthogonal. Show that $|\det(\mathbf{F})| = 1$.

Exercise 103: Let $\mathbf{F} = \mathbf{VDV}^T$ be symmetric. Show that $\det(\mathbf{F}) = \det(\mathbf{D}) =$ the product of the eigenvalues of $\mathbf{F}$.

4 Final Bonus

Although not even close to being complete, the text above hopefully covers most of the important linear algebra notions required in everyday data analysis. Before letting you go, though, I'd like to mention yet some other objects, which, at least to my experience, pop up quite often in practice. As they don't fit too well in the theory of linear functions presented above, they are presented just as a free addon.

Trace

Definition 4.1 (Trace) *Let* $\mathbf{A}$ *be a* $n \times n$ *matrix. The* trace *of* $\mathbf{A}$*, denoted by* $\mathrm{tr}(\mathbf{A})$*, is the sum of the diagonal elements of* $\mathbf{A}$*:*

$$\mathrm{tr}(\mathbf{A}) = \sum_i a_{ii}$$

The typical context when trace shows itself is the following: suppose we have two $n \times m$ matrices $\mathbf{A}$ and $\mathbf{B}$, and we wish to define an *inner product* $\langle \mathbf{A}, \mathbf{B} \rangle$ of these matrices, by simply interpreting them as long vectors with nm elements, and using the inner product of these vectors. That is,

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j} a_{ij} b_{ij}.$$

It turns out that in this case, $\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{tr}(\mathbf{A}^T \mathbf{B}) = \mathrm{tr}(\mathbf{B}\mathbf{A}^T)$.

Exercise 104: Prove it.

Exercise 105: Prove that $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}^T)$.

Exercise 106: Prove that $\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CAB})$.

Bilinear functionals

Definition 4.2 (Bilinear functional) *A function* $b : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ *is called a* bilinear functional *if it's linear in both its arguments. That is:*

$$b(\alpha \mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = \alpha b(\mathbf{x}_1, \mathbf{y}) + b(\mathbf{x}_2, \mathbf{y})$$
$$b(\mathbf{x}, \alpha \mathbf{y}_1 + \mathbf{y}_2) = \alpha b(\mathbf{x}, \mathbf{y}_1) + b(\mathbf{x}, \mathbf{y}_2)$$

Similarly to linear functions, bilinear functionals turn out to be quite common and useful. The simplest example of a bilinear functional is the inner product.

Exercise 107: Show that inner product is a bilinear functional.

Similarly to linear functions, it is possible to describe bilinear functionals using matrices, although in a slightly different way.

Theorem 4.1 *Any bilinear functional b can be represented as:*

$$b(\mathbf{x}, \mathbf{y}) = \sum_{i,j} a_{ij} x_i y_j = \mathbf{x}^T \mathbf{A} \mathbf{y}$$

for a certain matrix $\mathbf{A}$*.*

Exercise 108: Prove it. Hint: Repeat the logic used for linear functions. Represent $\mathbf{x}$ and $\mathbf{y}$ in the canonical basis and use the linearity.

Make note of the equality $\sum_{i,j} a_{ij} x_i y_j = \mathbf{x}^T \mathbf{A} \mathbf{y}$. It's sometimes useful.

Quadratic forms

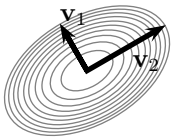
Definition 4.3 (Quadratic form) *Let $b : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a bilinear functional. The function $q(\mathbf{x}) = b(\mathbf{x}, \mathbf{x})$ is called a* quadratic form *corresponding to b.*

Exercise 109: Prove that any quadratic form q can be represented as $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ where $\mathbf{A}$ is a symmetric matrix.

Exercise 110: Let q be a quadratic form with $\mathbf{A}$ the corresponding matrix. Show that $q(\mathbf{x} + \mathbf{y}) = q(\mathbf{x})^2 + 2\mathbf{x}^T \mathbf{A} \mathbf{y} + q(\mathbf{y})^2$.

Exercise 111: Show that $q(\mathbf{x}) = \|\mathbf{x}\|^2$ is a quadratic form.

Exercise 112: Let's examine a quadratic form defined for vectors in $\mathbb{R}^2$. Suppose $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ with $\mathbf{A}$ symmetric and $\mathbf{x} \in \mathbb{R}^2$. As $\mathbf{A}$ is symmetric, it has two unit orthogonal eigenvectors, call them $\mathbf{v}_1$ and $\mathbf{v}_2$. Draw them on the plane. Let the corresponding eigenvalues be λ_1 and λ_2. Examine the values of q on the lines defined by these vectors (show that $q(\alpha \mathbf{v}_1) = \lambda_1 \alpha^2$ and $q(\alpha \mathbf{v}_2) = \lambda_2 \alpha^2$). Try to imagine how q looks in general. Do you see that it is either a paraboloid or a saddle-like surface (depending on the signs of λ_1 and λ_2?



Definition 4.4 (Positive definiteness) *The quadratic form q is said to be* positive definite *if $q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$. We say that q is* negative definite *if $q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$. If the equalities are not strict, we say that q is* positive semidefinite *or* negative semidefinite *correspondingly.*
The same terms are applicable to any matrix $\mathbf{A}$. For example, we say that $\mathbf{A}$ is positive definite if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for any $\mathbf{x} \neq \mathbf{0}$.

Positive definiteness or negative definiteness can be easily described via the eigenvalues of $\mathbf{A}$ (and in most cases positive/negative definiteness is discussed for symmetric matrices, which, as we know, have all the eigenvectors).

Exercise 113: Consider a diagonal matrix $\mathbf{D}$ and the corresponding quadratic form $q(\mathbf{x}) = \mathbf{x}^T\mathbf{D}\mathbf{x}$. Show that q is positive definite iff all the diagonal elements of $\mathbf{D}$ are positive and q is negative definite iff all the diagonal elements are negative.

Exercise 114: Now consider a symmetric matrix $\mathbf{A}$. Use the eigenvalue decomposition (2) to show that $\mathbf{A}$ is positive definite iff all its eigenvalues are positive, and negative definite iff all the eigenvalues are negative.

Exercise 115: Show that if $\mathbf{A}$ is positive definite, it is invertible.

5 Summary

To summarize, here's the contents of this brief guide all in four paragraphs.

Very often *data* can be represented in terms of *vectors*. We introduce addition and multiplication by scalar for vectors, and this allows us to form *linear combinations*. These, in turn, allow to examine *subspaces*. Every subspace has a *basis*, that allows to represent vectors in a given subspace in a *unique* manner.

Norm and *inner product* formalize the notions of *distances* and *angles* between vectors. Of particular interest are *orthogonal* vectors — these being in a sense "completely independent".

Linear functions are functions defined on vectors, that preserve *linear combinations*. Linear functions correspond uniquely to *matrices* and can be analyzed quite thoroughly. For example, the *rank* of a linear function indicates how much information it preserves, its *determinant* indicates how it scales the vectors, and its *eigenvectors* and *eigenvalues* show the directions in which the scaling is performed. Of particular interest are subclasses of linear functions such as *orthogonal* or *symmetric* functions. The former ones preserve angles and distances and correspond to rotations, and the latter ones correspond to scalings along orthogonal directions.

At last, bilinear functionals and quadratic forms can be often met in practice, so it's important to know what are they and how they relate to matrices.