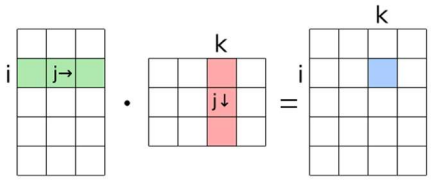
Linear Algebra – Lesson 7

Inner product, projection matrix

1. The dot product

1.1 Algebraic definition

We are already familiar with several operations in linear algebra, like matrix-vector multiplication and matrix-matrix multiplication. For example, let $A_{m\times n}$ and $B_{n\times p}$ be matrices. What are the dimensions of the product AB?



The result will be of size $m \times p$. That is, the number of rows is determined by the left matrix, and the number of columns by the right matrix.

Column vectors are simply $n \times 1$ matrices. To multiply two vectors, we treat one as a row vector, and multiply using the matrix multiplication rules:

$$\bar{v}^T\bar{w} = v_1w_1 + v_2w_2 + \cdots + v_nw_n = \sum_{i=1}^{n} v_iw_i$$

The result is the sum of the products of the corresponding terms in both vectors.
This product is usually called dot product, inner product, or sometimes scalar product (to emphasize that it results in a scalar). It is denoted $\bar{v}^T\bar{w}$ or $\bar{v}\cdot\bar{w}$ or $\langle\bar{v},\bar{w}\rangle$.

From this definition it is easy to see that the dot product is symmetrical:

$$\bar{v}^T\bar{w} = \bar{w}^T\bar{v}$$

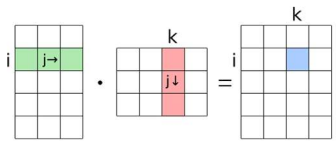
Distributive:

$$\bar{v}^T(\bar{w}+\bar{u}) = \bar{v}^T\bar{w} + \bar{v}^T\bar{u}$$

And satisfies:

$$(k\bar{v})^T\bar{w} = k\bar{v}^T\bar{w}$$

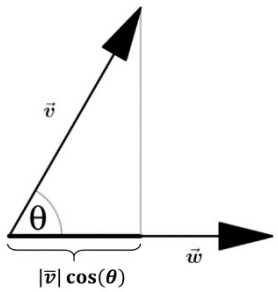
Note: When we calculate the product AB of two matrices, every entry is calculated as a dot product of one row form A and one column from B.



1.2 Geometric definition

The inner product has a geometrical meaning. It is the product of the lengths of the two vectors (denoted $|\bar{v}|$ and $|w|$) times the cosine of the angle θ between them:

$$\bar{v}^T\bar{w} = |\bar{v}||\bar{w}|\cos(\theta)$$



The length of a vector, $|\bar{v}|$, is usually called the norm of the vector. Notice that the inner product can be positive, negative or zero.

Let's examine some extreme cases:

1. $\bar{v}^T\bar{w} = 0$ if and only if $\bar{v}$ and $\bar{w}$ are orthogonal (perpendicular, with $\theta = 90°$).
2. If $\bar{v}$ and $\bar{w}$ are collinear (with $\theta = 0°$), then $\bar{v}^T\bar{w} = |\bar{v}||\bar{w}|$.
3. The dot product of a vector with itself gives its norm squared: $\bar{v}^T\bar{v} = |\bar{v}|^2$. Therefore, $\bar{v}^T\bar{v} \geq 0$.
4. $\bar{v}^T\bar{v} = 0$ if and only if $\bar{v} = 0$.

Length (norm) and angles

The dot product allows us to define notions such as **length** and **angle** for high dimensions.
The **norm** of a vector $\bar{v} \in \mathbb{R}^n$ is defined as the square root of its dot product:

$$|\bar{v}| = \sqrt{\bar{v}^T\bar{v}} = \sqrt{v_1^2 + v_2^2 + .. + v_n^2}$$

The norm is used to calculate the distance between two vectors:

$$|\bar{v} - \bar{w}| = \sqrt{(v_1 - w_1)^2 + (v_2 - w_2)^2 + \cdots + (v_n - w_n)^2}$$

The angle θ between two vectors in $\mathbb{R}^n$ is defined as:

$$\cos(\theta) = \frac{\bar{v}^T\bar{w}}{|\bar{v}||\bar{w}|}$$

Example – Finding an orthogonal vector

To find an orthogonal vector in 2D, we simply switch the x and y entries, and change the sign of one of them. For example, if $\bar{v} = \begin{pmatrix}1\\2\end{pmatrix}$, then $\bar{w} = \begin{pmatrix}-2\\1\end{pmatrix}$ is orthogonal to $\bar{v}$.

Example – Finding the angle between two vectors

Let $\bar{v} = \begin{pmatrix}1\\3\end{pmatrix}$ and $\bar{w} = \begin{pmatrix}-1\\0\end{pmatrix}$. The angle between them is $\cos(\theta) = -\frac{1}{1\cdot\sqrt{10}} \rightarrow \theta = \cos^{-1}\left(-\frac{1}{\sqrt{10}}\right)$

What is "$\bar{v}$ multiplied by $\bar{w}$?" – a bad question with good answers

When talking about the product of two vectors, we must define clearly which product we talk about. We already know two kinds of vector products:

1. Column vector times row vector: $\bar{v}\bar{w}^T$
2. Dot product (row vector times column vector): $\bar{v}^T\bar{w}$

The result of each product is fundamentally different. For example: $\bar{v} = \begin{pmatrix}1\\2\end{pmatrix}, \bar{w} = \begin{pmatrix}-1\\3\end{pmatrix}$

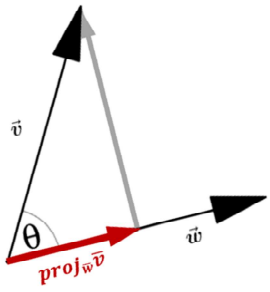
$$\bar{v}\bar{w}^T = \begin{pmatrix}1\\2\end{pmatrix}(-1\ 3) = \begin{pmatrix}-1 & 3\\-2 & 6\end{pmatrix}$$

$$\bar{v}^T\bar{w} = (1\ 2)\begin{pmatrix}-1\\3\end{pmatrix} = 5$$

Whenever you see a series of vector-matrix products, make sure you understand what kind of result you are expecting. For example, what is $\bar{x}^T A\bar{x}$? It is a scalar, which can be seen as the result of a dot product between the vector $\bar{x}$ and the vector $A\bar{x}$.

1.3 The dot product and projection

1.3.1 Projection onto a vector



The dot product is tightly related to the notion of vector projection.
To find the vector projection of $\bar{v}$ onto $\bar{w}$, we can rewrite $\bar{v}$ as the sum of two component vectors, one that is parallel to the $\bar{w}$ and one that is perpendicular to the $\bar{w}$. The parallel vector is the projection of $\bar{v}$ onto $\bar{w}$.

The **magnitude** of the projection is $|\bar{v}|\cos(\theta)$. Since the projection vector is a **vector**, we have to multiply this magnitude by a unit vector (a vector with norm=1), pointing in the direction of $\bar{w}$, which can be written as $\frac{\bar{w}}{|\bar{w}|}$ (and sometimes written as $\hat{w}$). The projection vector of $\bar{v}$ onto $\bar{w}$ is therefore:

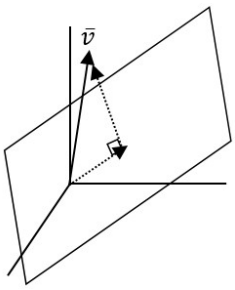
$$proj_{\bar{w}}\bar{v} = (|\bar{v}|\cos(\theta))\frac{\bar{w}}{|\bar{w}|}$$

Using the geometric definition of the dot product, $|\bar{v}||\bar{w}|\cos(\theta) = \bar{v}^T\bar{w}$:=

$$proj_{\bar{w}}\bar{v} = \underbrace{\frac{\bar{v}^T\bar{w}}{|\bar{w}|}}_{\substack{proj.\\magn.}} \overbrace{\frac{\bar{w}}{|\bar{w}|}}^{\substack{proj.\\direction}} = \frac{\bar{v}^T\bar{w}}{|\bar{w}|^2}\bar{w}$$

1.3.2. Projection onto a plane

We can also project a vector $\bar{v}$ onto a plane. In this case too, we decompose $\bar{v}$ into a component inside the plane and a component orthogonal to the plane:



The orthogonal component is also called the residual (because this is what is left of $\bar{v}$ after subtracting the projection onto the plane: $\bar{v} = \bar{v}_{proj} + \bar{v}_{orth} \rightarrow \bar{v}_{orth} = \bar{v} - \bar{v}_{proj}$).

1.3.3 Representing a vector using an orthogonal basis

We already talked about examples in which we want to represent a vector as a linear combination of some new basis vectors (in the Fibonacci Numbers, for example):

$$\bar{w} = a_1\bar{u}_1 + a_2\bar{u}_2 + \cdots + a_n\bar{u}_n$$

How do we find the coefficients a_n?

To find the coefficients a_n, we write a matrix equation $A\bar{x} = \bar{w}$, with the columns of A being the vectors $\bar{u}_1, \bar{u}_2, \dots, \bar{u}_n$. A solution $\bar{x} = \begin{pmatrix} a_1 \\ \cdots \\ a_n \end{pmatrix}$ gives us the correct coefficients. But there are important cases in which the procedure is much simpler:

Let $\bar{u}_1, \bar{u}_2, \dots, \bar{u}_n \in \mathbb{R}^n$ be nonzero pairwise orthogonal vectors:

$$\bar{u}_i^T \bar{u}_j \begin{cases} = 0 \; for \; i \neq j \\ > 0 \; for \; i = j \end{cases}$$

Then for any $\bar{w} \in \mathbb{R}^n$:

$$\bar{w} = \left(\frac{\bar{w}^T \bar{u}_1}{\bar{u}_1^T \bar{u}_1}\right)\bar{u}_1 + \left(\frac{\bar{w}^T \bar{u}_2}{\bar{u}_2^T \bar{u}_2}\right)\bar{u}_2 + \cdots + \left(\frac{\bar{w}^T \bar{u}_n}{\bar{u}_n^T \bar{u}_n}\right)\bar{u}_n$$

In other words, the coefficient for $\bar{u}_i$ are simply the projection of $\bar{w}$ onto $\bar{u}_i$. To **prove** this, start with the general form:

$$\bar{w} = a_1\bar{u}_1 + a_2\bar{u}_2 + \cdots + a_n\bar{u}_n$$

Multiply the equation by $\bar{u}_i^T$ from the left (all the terms of the form $\bar{u}_i^T \bar{u}_j$ will disappear):

$$\bar{u}_i^T \bar{w} = a_i \bar{u}_i^T \bar{u}_i$$

$$a_i = \frac{\bar{u}_i^T \bar{w}}{\bar{u}_i^T \bar{u}_i}$$

Note: Multiplying an entire expression by one vector to obtain a simpler term is a very common a useful trick you should know.

Symmetric *matrices have an orthogonal eigenbasis*

Why is this even useful? Where can we find a basis in which all vectors are orthogonal to each other? Well, there is a family of matrices whose eigenvectors are all orthogonal to each other: the symmetric matrices.

For a real $n \times n$ symmetric matrix:

- All eigenvalues are real
- You can find n linearly independent eigenvectors, which are orthogonal to each other
- And remember: this means that any real symmetric matrix is diagonalizable.

We will prove this later in the course.

2. Orthogonality

We defined the term orthogonality for pairs of vectors. We can also talk about orthogonal subspaces. If V and W are two orthogonal spaces, then:

$$\forall \; (\bar{v} \in V, \bar{w} \in W): \;\; \bar{v}^T \bar{w} = 0$$

Question: Are the XY plane and the Z axis two orthogonal subspaces?

Answer: Yes. For any two vector of the form $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}$ we get a dot product of 0.

Question: Are the XY plane and XZ plane orthogonal?

Answer: No. For example, the vectors $\begin{pmatrix}1\\1\\0\end{pmatrix} \in XY$ and $\begin{pmatrix}1\\0\\1\end{pmatrix} \in XZ$ are not orthogonal. Moreover, some vectors are shared by these two subspace. For example, $\hat{\imath}$ is in both planes, and $\hat{\imath}^T\hat{\imath} = 1 > 0$.

Example – two orthogonal subspaces

When we dealt with systems of equations, we talked about several subspaces: $Im(A) = colsp(A), rowsp(A), \ker(A)$. Which of the two spaces – the column space of A or the row space of A – do you think is orthogonal to $\ker(A)$?

Well, the row space is orthogonal to the kernel. Let $\bar{x} \in \ker(A)$. Now:

$$A\bar{x} = 0 \quad \rightarrow \quad \begin{pmatrix} row_1\ of\ A \\ row_2\ of\ A \\ \dots \\ row_n\ of\ A \end{pmatrix}\bar{x} = \begin{pmatrix}0\\0\\\dots\\0\end{pmatrix}$$

We see that every row of A is orthogonal to $\bar{x}$:

$$(row_1 A)\bar{x} = 0$$
$$\dots$$
$$(row_n A)\bar{x} = 0$$

But this means that any linear combination of the rows is also orthogonal to $\bar{x}$:

$$\big(c_1(row_1 A) + \cdots + c_n(row_n A)\big)\bar{x} = 0$$

And this means that any vector in the row space of A is orthogonal to any vector $\bar{x} \in \ker(A)$.

Finally, we say that $rowsp(A)$ and $\ker(A)$ are orthogonal complements in $\mathbb{R}^n$: $\ker(A)$ contains all the vectors that are orthogonal to $rowsp(A)$.

Note: Using the same proof, we can show that **$\ker(A^T)$ is orthogonal to $colsp(A)$** ((which is $Im(A)$).

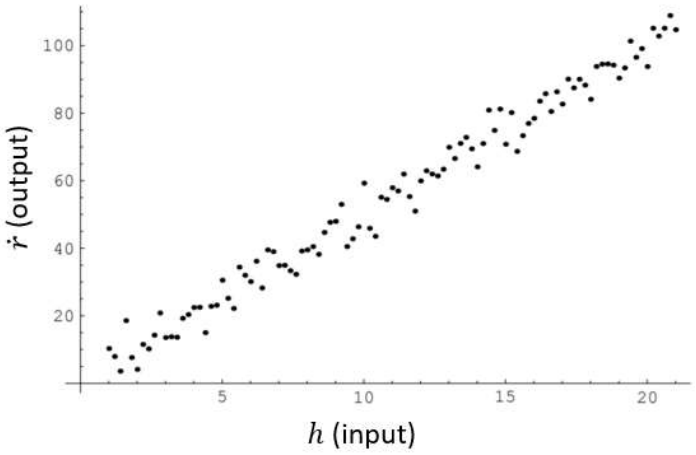
Note: The fact that these two subspaces are orthogonal to each other is related to the rank-nullity theorem, which dealt with the dimensions of these spaces.

3. Least squares: Solving $A\bar{x} = \bar{b}$ when it has no solution

3.1 The problem

We are given a system of m independent equations in n unknowns, and $m > n$ (we have more equations than unknowns – such a system is called overdetermined). In general, we expect there to be no solution to the system (why? Because exactly n equation should suffice for finding a solution. If we have more equations, and they are independent of the previous ones, then they hint at a different relationship between the variables).

Where could we encounter such a case? Data fitting is a very common example. We often collect data and try to model it using simple equations.



For example, let's say you are recording electrophysiological data from a neuron, and you are interested in studying how its firing rate changes with the input[1]. By changing the input current many times and measuring the change in firing rate, you can try and estimate the exact parameters that describe the linear relationship between the input (the x-axis) and the rate of change in the firing rate (the y-axis).

But since any measurement is noisy, in fact there is no line that passes through all the measures points. We must therefore look for a solution that will be "best" in some sense. One common choice is to choose the line for which the prediction error (for example, the sum of the differences between the measured points and their predicted values) is smallest. The prediction errors are also called the residuals.

How is this related to linear algebra? Well, if we assume a linear model that relates the input h and the firing rate $\dot{r}$, we are actually looking for two parameters x_1, x_2 such that:

$$x_1 h + x_2 = \dot{r}$$

For one measurement we can write the following equation:

$$(h\ 1)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \dot{r}$$

For all m measurements we get a system of equations:

$$\underbrace{\begin{pmatrix} h_1 & 1 \\ h_2 & 1 \\ \dots & \dots \\ h_m & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{\bar{x}} = \underbrace{\begin{pmatrix} \dot{r}_1 \\ \dot{r}_2 \\ \dots \\ \dot{r}_m \end{pmatrix}}_{\bar{b}}$$

Or:

$$A\bar{x} = \bar{b}$$

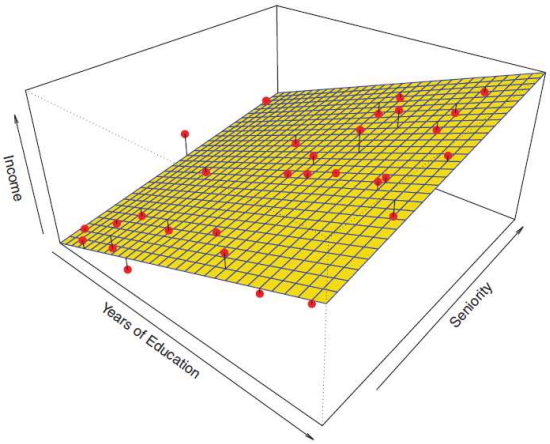
As we already said, this equation has no solution[2]. We will look for a solution – a parameters vector $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ – that minimizes the residuals. Finding this best linear fit is also called linear regression.

Note – 2D problems

[1] You will study this model analytically in the course Theoretical and Computational Neuroscience A, when you talk about dynamics of rate networks.

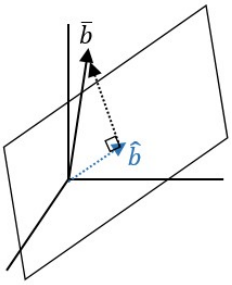
[2] When we first learned of such inconsistent systems, we said we could use Gaussian elimination to figure out that indeed the system has no solution. Now, we will see how we can still proceed and find some kind of "best" solution.

While we are dealing with the simplest cast of 1D problems, we often encounter linear regression in higher dimensions. For example, one could try and predict a person's income as a linear combination of her seniority and her years of education[3] (the observations are shown in red, and the yellow plane indicates the linear regression fit to the data):



3.2 The solution – Projecting $\bar{b}$ onto $Im(A)$

The equation $A\bar{x} = \bar{b}$ has no solution because $\bar{b} \neq Im(A)$. We could solve it if $\bar{b} \in Im(A)$. So instead of solving the original equation, we will project $\bar{b}$ onto $Im(A)$, and solve this new equation. We will denote the projected vector by $\hat{b}$ and the solution that we find by $\hat{x}$ (just to remind ourselves that we moved to solving a different equation):



$$A\hat{x} = \hat{b}$$

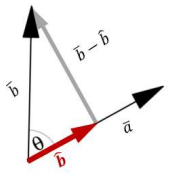
Why is this a good solution? Because it has the property that of all vectors in $Im(A)$, it gives us the one that is closest to the original $\bar{b}$.

3.3 Projection matrix

To project $\bar{b}$ onto $Im(A)$ means to map it to another vector. We should be able to find a matrix that does this. This will be the projection matrix onto $Im(A)$.

3.3.1 Projection matrix in 1D

To see how we can construct the projection matrix, let's start with a 1D example. We want to project $\bar{b}$ onto $\bar{a}$:



The projection $\hat{b}$ is going to be some multiple of $\bar{a}$, so: $\hat{b} = x\bar{a}$.

To find the right scalar x, we demand that the gray residual vector $(\bar{b} - \hat{b})$ is orthogonal to $\bar{a}$:

$$\bar{a}^T(\bar{b} - \hat{b}) = 0$$
$$\bar{a}^T\hat{b} = \bar{a}^T\bar{b}$$

[3] This example is taken from the ISLR book, sixth printing.

$$\bar{a}^T(x\bar{a}) = \bar{a}^T\bar{b}$$
$$x = \frac{\bar{a}^T\bar{b}}{\bar{a}^T\bar{a}}$$

(This is actually a proof for the formula we had above for $proj_{\bar{w}}\bar{v}$)
Overall, the projected vector is:

$$\hat{b} = \bar{a}x = \bar{a}\frac{\bar{a}^T\bar{b}}{\bar{a}^T\bar{a}}$$

But notice that we can think of it not as a vector ($\bar{a}$) times a scalar, but as a matrix times a vector:

$$\hat{b} = \underbrace{\left(\frac{\bar{a}\bar{a}^T}{\bar{a}^T\bar{a}}\right)}_{P}\bar{b}$$

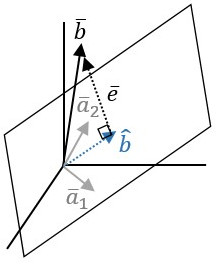
This matrix P is the projection matrix. P takes a vector and projects it onto $\bar{a}$.

Properties of any projection matrix

1. P is symmetric: $P^T = P$
2. $P^2 = P$ (applying the projection twice is the same as applying it once)

3.3.2 Projection matrix in higher dimensions

Now we can go back to projecting $\bar{b}$ onto the plane. We will use an example in 2D to learn the general case of n-dimensions.
To represent the plane, we can take two linearly independent vectors that span the plane: $\bar{a}_1, \bar{a}_2$.
How can we find such vectors?
Well, the plane we're looking at is just the column space of A. So $\bar{a}_1, \bar{a}_2$ are just the columns of A:



$$\hat{b} = x_1\bar{a}_1 + x_2\bar{a}_2 = A\hat{x}$$

We are looking for the coefficients x_1, x_2 that will make the residual perpendicular to the plane:

$$\bar{e} = \bar{b} - \hat{b} \text{ should be perpendicular to the plane}$$

So we demand that it is perpendicular to each of the vectors that span the plane:

$$\bar{a}_1^T(\bar{b} - A\hat{x}) = 0$$
$$\bar{a}_2^T(\bar{b} - A\hat{x}) = 0$$

Notice that $\bar{a}_1^T$ is a row vector, and $(\bar{b} - A\hat{x})$ is a column vector. We can rewrite this using matrices:

$$\begin{pmatrix} -\bar{a}_1^T - \\ -\bar{a}_2^T - \end{pmatrix}(\bar{b} - A\hat{x}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
$$A^T\underbrace{(\bar{b} - A\hat{x})}_{\bar{e}} = 0$$

Now, look at the equation. We demanded that the error $\bar{e}$ is orthogonal to the plane (which is $Im(A) = colsp(A)$), and we got here that it is in the kernel of A^T: $\bar{e} \in \ker(A^T)$. This is exactly what we expected[4].

[4] Go back a few pages and see that any vector in $\ker(A^T)$ is orthogonal to any vector $Im(A)$.

This last equation gives us:

$$A^T A\hat{x} = A^T \bar{b}$$

The "formal" solution for $\hat{x}$ is[5]:

$$\hat{x} = (A^T A)^{-1} A^T \bar{b}$$

This is the solution we were looking for! This is the vector of unknowns in the "new" equation $A\hat{x} = \hat{b}$, which does have a solution.

To complete the picture – what is the projected vector $\hat{b}$?

$$\hat{b} = A\hat{x} = \underbrace{A(A^T A)^{-1} A^T}_{\substack{the\ projection \\ matrix}} \bar{b}$$

Compare the resulting projection matrix to what we found in the 1D case[6]:

$$P_{1D} = \frac{\bar{a}\bar{a}^T}{\bar{a}^T \bar{a}} \qquad P_{2D} = A(A^T A)^{-1} A^T$$

To summarize, $P = A(A^T A)^{-1} A^T$ is the projection matrix that projects any vector $\bar{b}$ onto the column space of A. Of all the vectors in $Im(A)$, this projection ($\hat{b}$) is the closest vector to $\bar{b}$.

3.4 Linear regression (the least squares solution)

Let's look at a concrete example of linear regression (fitting a line) with three measurements.
Let the measurements (y) as a function of time (t) be $(1,1), (2,2), (3,2)$. No line passes through all these points (draw it to convince yourselves). We are looking for the best fitting line, so we model the measurements as some linear function:

$$y = x_1 + x_2 t$$

From each point we get one equation:

$$x_1 + x_2 = 1$$
$$x_1 + 2x_2 = 2$$
$$x_1 + 3x_2 = 2$$

If we write this in matrix form:

$$A\bar{x} = \bar{b}$$
$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

As we saw, the "best" solution is the one that we get by projection $\bar{b}$ onto $Im(A)$ and solving the new system of equations, which gives us:

$$A^T A\hat{x} = A^T \bar{b}$$

If we calculate explicitly we get:

$$A^T A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix}$$

[5] Remember that in real life, you can simply solve the system of equations, without having to calculate the inverse explicitly

[6] It seems like we could simplify it like that: $A(A^T A)^{-1} A^T = A(A^{-1}(A^T)^{-1})A^T = (AA^{-1})((A^T)^{-1}A^T) = I$, which really doesn't make any sense. What is the catch then? We can't use the formula for $(AB)^{-1}$ in this case, because A isn't square.

$$A^T\bar{b} = \begin{pmatrix}1 & 1 & 1\\ 1 & 2 & 3\end{pmatrix}\begin{pmatrix}1\\2\\2\end{pmatrix} = \begin{pmatrix}5\\11\end{pmatrix}$$

So we get:

$$\begin{pmatrix}3 & 6\\ 6 & 14\end{pmatrix}\hat{x} = \begin{pmatrix}5\\11\end{pmatrix}$$

And these two equations we can solve:

$$x_1 = \frac{2}{3}, x_2 = \frac{1}{2}$$

So the best fitting line is:

$$y = \frac{2}{3} + \frac{1}{2}t$$

Why is solving these equations equivalent to the least-squares solution?

The least squares solution is the solution we get when we demand that the sum of square errors is minimized. The errors are just the difference between the modeled data points ($A\hat{x}$) and the measured points ($\bar{b}$). The sum of squared errors is:

$$|\bar{e}|^2 = e_1^2 + e_2^2 + e_3^2 = (x_1 + x_2 - 1)^2 + (x_1 + 2x_2 - 2)^2 + (x_1 + 3x_2 - 2)^2$$

If we didn't know linear algebra, we could still find the minimum of this squared error. We would take its derivative with respect to x_1 and demand that it is 0, then take its derivative with respect to x_2 and demand that it is also 0. For example:

$$\frac{\partial(|\bar{e}|^2)}{\partial x_1} = 2(x_1 + x_2 - 1) + 2(x_1 + 2x_2 - 2) + 2(x_1 + 3x_2 - 2) = 0$$

$$6x_1 + 12x_2 - 10 = 0$$

$$3x_1 + 6x_2 - 5 = 0$$

Which is just the first equation in the system of equations we got the linear-algebra way.

5. Orthonormal basis and orthogonal matrices

We already talked about an orthogonal basis, for which $\bar{u}_1, \bar{u}_2, \dots, \bar{u}_n \in \mathbb{R}^n$ are nonzero pairwise orthogonal vectors:

$$\bar{u}_i^T\bar{u}_j \begin{cases} = 0 \; for \; i \neq j \\ > 0 \; for \; i = j \end{cases}$$

If the norm of each vector is 1, the basis is called an orthonormal basis. To make it easier to identify, we will denote the orthonormal vectors by $\{q_i\}_{i=1}^n$:

$$\bar{q}_i^T\bar{q}_j \begin{cases} = 0 \; for \; i \neq j \\ 1 \; for \; i = j \end{cases}$$

This is often written using Kronecker's delta, δ_{ij}, which is 1 if $i = j$ and 0 otherwise:

$$\bar{q}_i^T\bar{q}_j = \delta_{ij}$$

The change-of-basis matrix to this orthonomal basis is simply the basis vectors in the columns:

$$Q = \begin{pmatrix} | & | & | & | \\ \bar{q}_1 & \bar{q}_2 & \dots & \bar{q}_n \\ | & | & | & | \end{pmatrix}$$

And it has the special property that:

$$Q^T Q = I$$

Examples

Let $Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{pmatrix}$. Then: $Q^T Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Notice that from this example we see that the matrix Q is not square, but we still get the identity[7].

5.1 Orthogonal matrices

Definition: Let Q be a real **square** matrix whose columns are pairwise orthogonal unit vectors. Then Q is called an orthogonal matrix.

This definition is equivalent to saying that:

- The rows of Q are pairwise orthogonal unit vectors.
- Q is invertible, and $Q^T Q = QQ^T = I$.

For example, any permutation matrix and any rotation matrix are orthogonal:

$$Q = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad Q = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

Properties of an orthogonal matrix:

1. $\boldsymbol{Q^{-1} = Q^T}$ (because $Q^T Q = QQ^T = I$)
2. **The columns of $\boldsymbol{Q}$ are orthogonal.**
3. **The rows of $\boldsymbol{Q}$ are also orthogonal.**
   Because $QQ^T = I$ also tells us that the inner product of every row with every other row is 0
4. **Its determinant is $\boldsymbol{\pm 1}$**
5. **If preserves the inner product**: $\langle A\bar{v}, A\bar{w} \rangle = \langle \bar{v}, \bar{w} \rangle$
   Proof: $\langle A\bar{v}, A\bar{w} \rangle = (A\bar{v})^T (A\bar{w}) = \bar{v}^T A^T A \bar{w} = \bar{v}^T \bar{w} = \langle \bar{v}, \bar{w} \rangle$
6. **It preserves the norm** (because it preserves the inner product)

Transforms that preserve the inner product

The 5th property of an orthogonal matrix is not trivial. Other transformations do not preserve the inner product.

Example 1

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \bar{v} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \bar{w} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

$$\bar{v}^T \bar{w} = 6 \qquad \text{but} \qquad (A\bar{v})^T (A\bar{w}) = \begin{pmatrix} 4 & 0 \end{pmatrix} \begin{pmatrix} 6 \\ 2 \end{pmatrix} = 24$$

[7] What about QQ^T? Is it also equal to the identity matrix? Are you sure?

Here A changed the norm of each vector, and therefore changed the inner product as well.

Example 2

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \bar{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \bar{w} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\bar{v}^T \bar{w} = 0 \quad \text{but} \quad (A\bar{v})^T (A\bar{w}) = (1 \quad 0) \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2$$

Here A changed both the norm of a vector and the angle between different vectors.

Theorem: The change of basis matrix between two orthonormal bases is an orthogonal matrix.

5. Gram-Schmidt orthogonalization process

Another case where projecting vectors is useful is for constructing an orthonormal basis.

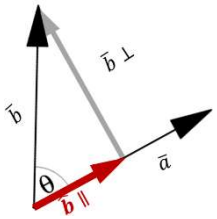
Let $\bar{a}, \bar{b}, \bar{c}$ be linearly independent vectors (of dimension n). These vectors span some subspace U, and in fact they are a basis to this subspace. We can use this given basis $\{\bar{a}, \bar{b}, \bar{c}\}$ to construct an orthonormal basis. This is done using the Gram-Schmidt orthogonalization process.

We will start by creating three orthogonal vectors, let's call them $\{\tilde{a}, \tilde{b}, \tilde{c}\}$ and later we will normalize them (just divide each vector by its norm, so that its new norm is 1).

<u>Step 1 – orthogonalize</u>

$\bar{a}$ is fine as it is, so $\tilde{a} = \bar{a}$.

We move to the next vector. We want to take only that part of $\bar{b}$ which is orthogonal to $\tilde{a}$. Earlier in the lesson we were interested in the projection of $\bar{b}$ onto $\bar{a}$, but now we are interested in the other component. To find it, we just have to subtract from $\bar{b}$ the part that is along $\bar{a}$:



$$\tilde{b} = \bar{b}_{\perp} = \bar{b} - \bar{b}_{\parallel}$$

Convince yourself geometrically this this is the case. So we get:

$$\tilde{b} = \bar{b} - \frac{\tilde{a}^T \bar{b}}{\tilde{a}^T \tilde{a}} \tilde{a}$$

Similarly, to get $\tilde{c}$ that is orthogonal to both $\tilde{a}$ and $\tilde{b}$, we have to subtract from $\bar{c}$ its projection on both previous vectors we found:

$$\tilde{c} = \bar{c} - \frac{\tilde{a}^T \bar{c}}{\tilde{a}^T \tilde{a}} \tilde{a} - \frac{\tilde{b}^T \bar{c}}{\tilde{b}^T \tilde{b}} \tilde{b}$$

<u>Step 2 – normalize</u>

$$\bar{q}_1 = \frac{\tilde{a}}{|\tilde{a}|}, \bar{q}_2 = \frac{\tilde{b}}{|\tilde{b}|}, \bar{q}_3 = \frac{\tilde{c}}{|\tilde{c}|}$$

And generally, if $\{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n\}$ is a basis of some vector space V, we can use it to construct an orthogonal basis $\{\bar{w}_1, \bar{w}_2, \dots, \bar{w}_n\}$ for V:

$$w_1 = v_1$$
$$w_2 = v_2 - \frac{\langle v_2, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1$$
$$w_3 = v_3 - \frac{\langle v_3, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle v_3, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$w_n = v_n - \frac{\langle v_n, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle v_n, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2 - \cdots - \frac{\langle v_n, w_{n-1} \rangle}{\langle w_{n-1}, w_{n-1} \rangle} w_{n-1}$$

And all that is left is to normalize them to get $\{\bar{q}_1, \bar{q}_2, \dots, \bar{q}_n\}$.
Note: For $\{\bar{q}_1, \bar{q}_2, \dots, \bar{q}_n\}$ to be a basis for V, they must be (1) linearly independent, and also (2) they must span V. They are linearly independent because they are orthogonal, and you can show that their span is the same as the span of $\{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n\}$, which completes the proof.

Example – Gram-Schmidt orthogonalization
The following vectors span some subspace (a plane) in $\mathbb{R}^3$:
$$\bar{a} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \bar{b} = \begin{pmatrix} 1 \\ 0 \\ 5 \end{pmatrix}$$
We will use them to construct an orthonormal basis for this subspace.
$\tilde{a} = \bar{a}$
$\tilde{b} = \bar{b} - \frac{\tilde{a}^T \bar{b}}{\tilde{a}^T \tilde{a}} \tilde{a} = \begin{pmatrix} 1 \\ 0 \\ 5 \end{pmatrix} - \frac{6}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ 3 \end{pmatrix}$
You can check and see that $\tilde{a}^T \tilde{b} = 0$.
Now all we have to do it normalize each vector:
$$\bar{q}_1 = \frac{\tilde{a}}{|\tilde{a}|} = \frac{1}{\sqrt{\tilde{a}^T \tilde{a}}} \tilde{a} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \bar{q}_2 = \frac{\tilde{b}}{|\tilde{b}|} = \frac{1}{\sqrt{\tilde{b}^T \tilde{b}}} \tilde{b} = \frac{1}{\sqrt{14}} \begin{pmatrix} -1 \\ -2 \\ 3 \end{pmatrix}$$

Question: What is the relationship between the span of $\bar{a}, \bar{b}$ and the span of $\bar{q}_1, \bar{q}_2$?
Answer: It is the same span. Both pairs of vectors are bases for the same space.