MAT188 WEEK 10 Orthogonal Transformations and Least Square

Vocabulary: orthogonal projection, orthogonal linear transformation, orthogonal matrix, least square solution.

Reading from the textbook: Sec 5.3, Sec 5.4

Introduce

This week, we will discuss the concept of orthogonal linear transformations. When we first introduce the definition of an orthogonal transformation. We use the word *orthogonal* in a different sense than we did last week. It may seem these two usage are somewhat disconnected, however, you will see there is indeed a connection.

A linear transformation T mapping $\mathbb{R}^n$ to $\mathbb{R}^n$ is called an *orthogonal transformation* if the magnitude of $\vec{x}$ equals the magnitude of $T(\vec{x})$, for all $\vec{x}$ in $\mathbb{R}^n$. Thus, we say orthogonal transformations T preserve the magnitude (or the norm) of vectors. Formally, we have the definition[1]:

Definition (Orthogonal Linear Transformation - Orthogonal Matrix)

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is called orthogonal if $||T(\vec{x})|| = ||\vec{x}||$ for all $\vec{x} \in \mathbb{R}^n$. If $T(\vec{x}) = A\vec{x}$ is an orthogonal transformation, we say A is an orthogonal matrix.

We have already seen examples of orthogonal linear transformations in this course. For instance, rotations and reflections are both orthogonal transformations.

Example The rotation in $\mathbb{R}^2$ given by

$$T(\vec{x}) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \vec{x}$$

is an orthogonal linear map since when we rotate vectors, their magnitude does not change. Hence the matrix $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ is an orthogonal matrix.

Example The reflection with respect to a line ℓ through the origin in $\mathbb{R}^2$ given by $\text{ref}(\vec{x}) = 2\,\text{Proj}_\ell\,\vec{x} - \vec{x}$ is an orthogonal linear transformation, since the length of $\vec{x}$, and $\text{ref}(\vec{x})$ are the same.

Are there other examples that we have not seen? Given a linear transformation and its standard matrix, how can we tell if it preserves the norm of all vectors? For that we need a different characterization of orthogonal maps. First let's see that orthogonal transformations preserve right angles.

Theorem (Orthogonal transformations preserve orthogonality)**.** *Let T be an orthogonal linear map from $\mathbb{R}^n$ to $\mathbb{R}^n$. If $\vec{v}$ and $\vec{w}$ are orthogonal, then so are $T(\vec{v})$ and $T(\vec{w})$.*

In fact, orthogonal transformations preserve all angles because they preserve the dot product between vectors. We can give an alternative characteristic of orthogonal linear transformations in terms of dot product.

Theorem (Orthogonal linear transformations preserves dot product)**.** *A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transformation if and only if $T(\vec{v}) \cdot T(\vec{w}) = \vec{v} \cdot \vec{w}$ for all $\vec{v}, \vec{w} \in \mathbb{R}^n$.*

Notice that this in particular implies if $\vec{u}$ and $\vec{v}$ are orthogonal, then $T(\vec{u})$ and $T(\vec{v})$ are also orthogonal!

Since an orthogonal map preserves orthogonality and length, it preserves orthonormal sets. That is, an orthogonal linear map takes orthonormal sets to orthonormal sets. The converse of this statement is also true[2]: if a linear map preserves orthonormality, it should preserve length and hence is an orthogonal map. This observation is particularly useful because it tells us that the columns of the standard matrix of an orthogonal linear transformation, which are the output of an orthonormal basis (the standard basis),

[1] Note that by now we can use the term "orthogonal" in three different senses. We can say "orthogonal vectors", "an orthogonal linear transformation" and "an orthogonal matrix".

[2] although it takes some work to prove!

are orthonormal. Conversely[3], if the standard matrix of a linear map has orthonormal columns, then the linear map should be orthogonal. For a linear map $T(\vec{x}) = A\vec{x}$, recall that $T(\vec{e}_1), T(\vec{e}_2), \dots, T(\vec{e}_n)$ are precisely the columns of A, the standard matrix of T. We thus say that $T(\vec{x}) = A\vec{x}$ is orthogonal, or equivalently A is orthogonal, exactly when the set $\{T(\vec{e}_1), T(\vec{e}_2), \dots, T(\vec{e}_n)\}$ forms an orthonormal basis for $\mathbb{R}^n$.

Theorem (Orthogonal transformations and orthonormal bases)**.** *A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transformation if and only if the vectors $T(\vec{e}_1), T(\vec{e}_2), \dots, T(\vec{e}_n)$ form an orthonormal basis for $\mathbb{R}^n$.*

Exercise Let $P : \mathbb{R}^2 \to \mathbb{R}^2$ be the orthogonal projection onto the x axis, and let $R : \mathbb{R}^2 \to \mathbb{R}^2$ be the 90 degree counterclockwise rotation. Let A and B be their standard matrices.

(1) Explain geometrically, do P and R preserve the norm of vectors. Which one does and which one doesn't?
(2) Verify the theorem above: Find A, and B. Do they have orthonormal columns? which one does and which one doesn't?
(3) Compute $A^T A$ and $B^T B$. What do you notice? Is there a connection between your answer and the previous part?
(4) Pick two vectors $\vec{v}, \vec{w}$ in $\mathbb{R}^2$. Draw $\vec{v}$, $\vec{w}$, $P(\vec{v})$, $P(\vec{w})$, $R(\vec{v})$, $R(\vec{w})$. Compare the angle between $\vec{v}$ and $\vec{w}$ and their outputs after applying P and R. Do P and R preserve the angle between $\vec{v}$ and $\vec{w}$?

Your observations in the previous exercise about A and B generalize. Suppose $A = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_n \\ | & & | \end{bmatrix}$ is an orthogonal matrix (that is the standard matrix of an orthogonal linear transformation). Then

$$A^T A = \begin{bmatrix} - & \vec{u}_1^T & - \\ & \vdots & \\ - & \vec{u}_m^T & - \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{bmatrix}$$

The off diagonal entrees of this matrix are zero and the main diagonal entries are 1. In other words this is identity. Since both A^T and A are square matrices that implies that AA^T is also identity. So the inverse of an orthogonal matrix A is given by A^T, its transpose.

Theorem (Orthogonal matrices)**.** *Consider an $n \times n$ matrix A. The matrix A is orthogonal if and only if $A^T A = I_n$ or, equivalently, if $A^{-1} = A^T$.*

Let's now look back at a linear transformation we have studied before: orthogonal projection. Recalling from readig 8, if we have an orthonormal basis $\mathcal{U} = \{\vec{u}_1, \dots, \vec{u}_m\}$ for a subspace $V \subseteq \mathbb{R}^n$, the projection of a vector $\vec{x} \in \mathbb{R}^n$ onto V is given by

$$\mathrm{Proj}_V(\vec{x}) = \vec{u}_1(\vec{u}_1 \cdot \vec{x}) + \cdots + \vec{u}_m(\vec{u}_m \cdot \vec{x}).$$

First, we notice that this is a linear combination of the column vectors $\vec{u}_1, \dots, \vec{u}_m$, with coefficients $\vec{u}_1 \cdot \vec{x}, \dots, \vec{u}_m \cdot \vec{x}$, so, it can be written as the matrix-vector product

$$\mathrm{Proj}_V(\vec{x}) = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{bmatrix} \begin{bmatrix} \vec{u}_1 \cdot \vec{x} \\ \vdots \\ \vec{u}_m \cdot \vec{x} \end{bmatrix}.$$

[3] the proof of the converse needs more work that we did not include in this reading.

Recalling that the vector $\begin{bmatrix} \vec{u}_1 \cdot \vec{x} \\ \vdots \\ \vec{u}_m \cdot \vec{x} \end{bmatrix}$ can also be written as a matrix-vector product where the rows of the corresponding matrix are given by $\vec{u}_1^T$ to $\vec{u}_m^T$. Thus,

$$\mathrm{Proj}_V(\vec{x}) = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{bmatrix} \begin{bmatrix} - & \vec{u}_1^T & - \\ & \vdots & \\ - & \vec{u}_m^T & - \end{bmatrix} \vec{x}.$$

We found the standard matrix for the orthogonal projection onto V:

$$\mathrm{Proj}_V(\vec{x}) = QQ^T\vec{x}$$

where the matrix Q has orthonormal columns which form an orthonormal basis for the subspace V, and Q^T is the transpose of Q. Note that Q is not an orthogonal matrix since it is not square, unless $V = \mathbb{R}^n$.

Let's see how this looks in an example.

Example Consider the subspace V of $\mathbb{R}^3$ with orthonormal basis $\mathcal{U} = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$. Using our previous formula, the projection of the vector $\vec{x} = \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \end{bmatrix}$ onto V is given by

$$\begin{aligned} \mathrm{Proj}_V(\vec{x}) &= \vec{u}_1(\vec{u}_1 \cdot \vec{x}) + \vec{u}_2(\vec{u}_2 \cdot \vec{x}) \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \left(\begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \end{bmatrix} \right) + \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \left(\begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \end{bmatrix} \right) \\ &= 2 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 0 \\ \sqrt{2} \end{bmatrix}. \end{aligned}$$

Now, using the matrix multiplication approach, we obtain

$$\mathrm{Proj}_V(\vec{x}) = QQ^T\vec{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 0 \\ \sqrt{2} \end{bmatrix}.$$

Watch a video::this video to recap on orthogonal projection and its standard matrix.

We now have two ways to compute the projection of a vector onto a subspace, but both approaches require that we have access to an orthonormal basis for that subspace, which can be rather time-consuming to obtain! Can we compute the projection without having access to an orthonormal basis? The answer turns out to be yes, and this is our main objective in the next section.

Solidify

As mentioned at in the conclusion of the previous section, our main objective here is to find a formula for the projection onto a subspace without first needing to compute an orthonormal basis for that subspace.

Suppose we have a basis $\mathcal{B} = \{\vec{u}_1, \ldots, \vec{u}_m\}$ for a subspace V of $\mathbb{R}^n$. We can then define a matrix A which has columns $\vec{u}_1, \ldots, \vec{u}_m$. Thus, A is an $n \times m$ matrix with $Col(A) = V$. The matrix A will have a crucial role. In fact, we will be using the product $A^T A$ and its inverse. Let's first convince ourselves that $A^T A$ is always invertible.

Since the columns of A are linearly independent, A has a QR factorization, where Q is an $n \times m$ matrix with orthonormal columns and R is an $m \times m$ upper triangular invertible matrix. See the end of post class reading 8 if you need a refresher on Q and R. We use this factorization to prove that $A^T A$ is an $m \times m$ invertible matrix. Using $A = QR$, we obtain

$$\begin{aligned} A^T A &= (QR)^T QR \\ &= R^T Q^T QR. \end{aligned}$$

Notice here that the order of matrix product switches when we take the transpose. One can prove this by working directly with the definition of matrix multiplication introduced in PCE 4. So we leave the proof of this as a exercise. Since Q is an $n \times m$ matrix with orthonormal columns, we have $Q^T Q = I_m$, the $m \times m$ identity matrix. Thus, $A^T A = R^T R$, which is the product of invertible matrices, implying that $A^T A$ is an invertible matrix. This result is summarized in the following theorem.

Theorem. *If A is an $n \times m$ matrix with linearly independent columns, then $A^T A$ is an $m \times m$ invertible matrix.*

For reference, the following theorem provides a useful list of properties associated with the transpose of a matrix.

Theorem (Properties of the Transpose)**.**

(1) $(A+B)^T = A^T + B^T$ *for all $m \times n$ matrices A and B.*
(2) $(kA)^T = kA^T$ *for all $m \times n$ matrices A and scalars k.*
(3) $(AB)^T = B^T A^T$ *for all $m \times p$ matrices A and $p \times n$ matrices B.*
(4) $rank(A^T) = rank(A)$ *for all matrices A.*
(5) $(A^T)^{-1} = (A^{-1})^T$ *for all invertible matrices A.*

With these results established, we present a formula for the projection onto a subspace V without needing to compute an orthonormal basis.

Theorem (Standard matrix of orthogonal projection)**.** *Consider a subspace $V \in \mathbb{R}^n$ with basis $B = \{v_1, \ldots, v_m\}$. Then if we take*

$$A = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_m \\ | & & | \end{bmatrix},$$

then the projection of $\vec{b}$ onto V is given by the following formula

$$\mathrm{Proj}_V(\vec{b}) = A(A^T A)^{-1} A^T \vec{b}.$$

The proof of this theorem is left for the post-class reading.
Although this formula looks complicated, it is a far easier task to find a basis for a subspace than it is to find an orthonormal basis for a subspace, thus, this formula comes in quite handy. We also note that if the columns of A happen to be orthonormal, then $A^T A = I_m$, and we arrive at the same projection equation discussed earlier in this reading, $\mathrm{Proj}_V(\vec{b}) = AA^T\vec{b}$.

Before discussing a thrilling application of this formula, we first need to discuss an important geometric property of orthogonal projection. Namely, that $\operatorname{Proj}_V(\vec{x})$ yields the vector in V which is "closest" to $\vec{x}$, meaning, for any $\vec{x} \in \mathbb{R}^n$, $||\operatorname{Proj}_V(\vec{x}) - \vec{x}|| \leq ||\vec{v} - \vec{x}||$ for all $\vec{v} \in V$. This is quite intuitive in $\mathbb{R}^2$ and $\mathbb{R}^3$ and can be proved in $\mathbb{R}^n$. The following theorem provides a precise statement for this idea.

Theorem. *Suppose V is a subspace of $\mathbb{R}^n$ and $\vec{x} \in \mathbb{R}^n$. The closest vector in V to $\vec{x}$ is given by $\operatorname{Proj}_V(\vec{x})$. In other words, $||\operatorname{Proj}_V(\vec{x}) - \vec{x}|| \leq ||\vec{v} - \vec{x}||$ for any $\vec{v} \in V$.*

For a proof of this theorem, see the post-class notes.
With all of these tools in our tool belt, we proceed to our main application of this section, **least-squares solutions**.
Let's suppose we have 3 points which are co-linear, say $(1, 2)$, $(2, 4)$, and $(3, 6)$. How would we go about finding an equation for the line that passes through these points? Using $y = c_0 + c_1 x$, we could set up a system of equations to solve for c_0 and c_1, given by

$$\begin{aligned} c_0 + c_1 &= 2 \\ c_0 + 2c_1 &= 4 \\ c_0 + 3c_1 &= 6, \end{aligned}$$

or equivalently, we could set up the matrix equation $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$, which of course turns out to be consistent, with $c_0 = 0$ and $c_1 = 2$.

But what if we change one of the numbers slightly so that the matrix equation is no longer consistent, for example, suppose the matrix equation is $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$. How can we proceed?
Our goal is still to find the line which best matches the data. There are several ways one could quantify what the best fit is, but one popular choice is to minimize the squared error. Namely, we could search for a solution $\begin{bmatrix} c_0^* \\ c_1^* \end{bmatrix}$ which minimizes the error $\left|\left| \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} - \begin{bmatrix} c_0 + c_1 \\ c_0 + 2c_1 \\ c_0 + 3c_1 \end{bmatrix} \right|\right|^2$. This minimizer is called the least-squares solution. For example, our previous choice of $c_0 = 0$ and $c_1 = 2$ yields a squared error of 1, but is there a choice of coefficients which leads to a smaller squared error?

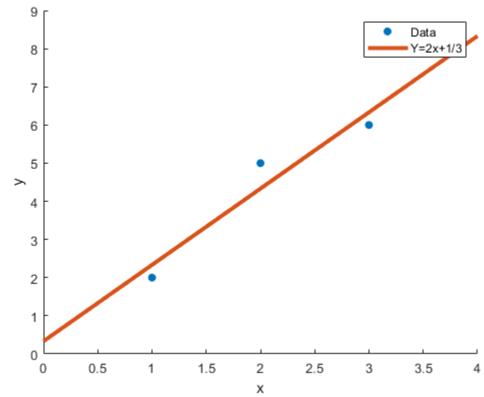
Every vector of the form $\begin{bmatrix} c_0 + c_1 \\ c_0 + 2c_1 \\ c_0 + 3c_1 \end{bmatrix}$ is in the $Col(A)$, where $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$. Therefore, we seek a vector $\vec{v}$ in the column space of A which minimizes $\left|\left| \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} - \vec{v} \right|\right|^2$. This may look awfully familiar. Recalling from the previous theorem, the vector in a subspace V which is closest to $\vec{x}$ is $\operatorname{Proj}_V(\vec{x})$. Thus, the coefficients c_0^* and c_1^* which minimize the squared error must satisfy

$$A \begin{bmatrix} c_0^* \\ c_1^* \end{bmatrix} = A(A^T A)^{-1} A^T \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}.$$

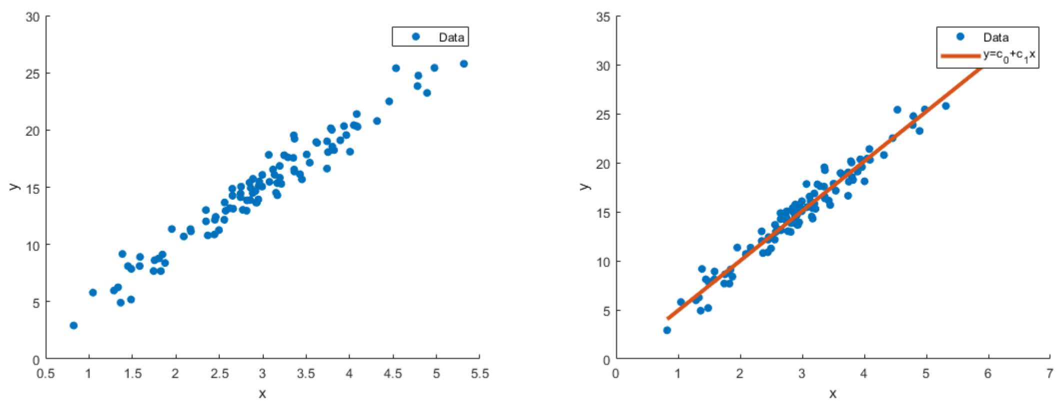
Substituting $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, yields

$$\begin{bmatrix} c_0^* \\ c_1^* \end{bmatrix} = \begin{bmatrix} \frac{4}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ 2 \end{bmatrix}$$

A plot showing this result is included below.



We can apply this to a much larger data set. Suppose we have a set of points (x_i, y_i) for $i = 1, \ldots, n$ which live in $\mathbb{R}^2$. We can apply the above approach to find the least-squares solution for the line $y = c_0 + c_1 x$. A plot showing the results is provided below.



Watch a video::this video for a recap on the connection between orthogonal projection and data fitting. Finding least square solutions to linear systems is a common technique used in many real-world applications.
Watch a video::This video walks you through one such application. We won't test you on the example covered in this video, nor expect you to be able to regenerate it. It is here for you to appreciate the theory you have learned so far in this course.