Overview

Last time we introduced the Gram Schmidt process as an algorithm for turning a basis for a subspace into an orthogonal basis for the same subspace. Having an orthogonal basis (or even better, an orthonormal basis!) is helpful for many problems associated to orthogonal projection.

Today we'll discuss the "Least Squares Problem", which asks for the best approximation of a solution to a system of linear equations in the case when an exact solution doesn't exist.

From Lay, §6.5

1. Introduction

Problem: What do we do when the matrix equation $A\mathbf{x} = \mathbf{b}$ has no solution $\mathbf{x}$?
Such inconsistent systems $A\mathbf{x} = \mathbf{b}$ often arise in applications, sometimes with large coefficient matrices.

Answer: Find $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is as close as possible to $\mathbf{b}$.

In this situation $A\hat{\mathbf{x}}$ is an *approximation* to $\mathbf{b}$. The **general least squares problem** is to find an $\hat{\mathbf{x}}$ that makes $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ as small as possible.

Definition

For an $m \times n$ matrix A, a *least squares solution* to $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n.$$

The name "least squares" comes from $\|\cdot\|^2$ being the sum of the squares of the coordinates.

It is now natural to ask ourselves two questions:

(1) Do least square solutions always exist?
The answer to this question is YES. We will see that we can use the Orthogonal Decomposition Theorem and the Best Approximation Theorem to show that least square solutions always exist.

(2) How can we find least square solutions?
The Orthogonal Decomposition Theorem —and in particular, the uniqueness of the orthogonal decomposition— gives a method to find all least squares solutions.

Solution of the general least squares problem

Consider an $m \times n$ matrix $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix}$.

- If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is a vector in $\mathbb{R}^n$, then the definition of matrix-vector multiplication implies that

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n\,.$$

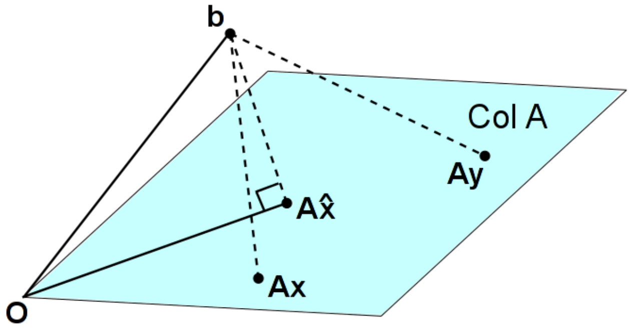
  So, the vector $A\mathbf{x}$ is the linear combination of the columns of A with weights given by the entries of $\mathbf{x}$.

- For any vector $\mathbf{x}$ in $\mathbb{R}^n$ that we select, the vector $A\mathbf{x}$ is in Col A. We can solve $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{b}$ is in Col A.

- If the system $A\mathbf{x} = \mathbf{b}$ is inconsistent it means that $\mathbf{b}$ is NOT in Col A.
- So we seek $\hat{\mathbf{x}}$ that makes $A\hat{\mathbf{x}}$ the closest point in Col A to $\mathbf{b}$.
- The Best Approximation Theorem tells us that the closest point in Col A to $\mathbf{b}$ is $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$.
- So we seek $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. In other words, the least squares solutions of $A\mathbf{x} = \mathbf{b}$ are exactly the solutions of the system

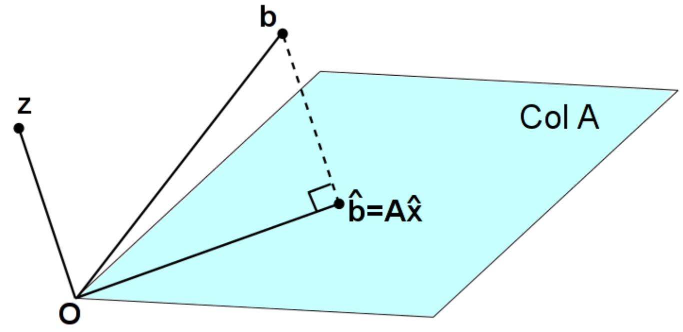
$$A\hat{\mathbf{x}} = \hat{\mathbf{b}}\,.$$

  By construction, the system $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ is always consistent.



We seek $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is the closest point to $\mathbf{b}$ in Col A.

Equivalently, we need to find $\hat{\mathbf{x}}$ with the property that $A\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{b}$ onto Col(A).



Since $\hat{\mathbf{b}}$ is the closest point to $\mathbf{b}$ in Col A, we need $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.

The normal equations

- By the Orthogonal Decomposition Theorem, the projection $\hat{\mathbf{b}}$ is the unique vector in Col A with the property that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col A.
- Since for every $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ the vector $A\hat{\mathbf{x}}$ is automatically in Col A, requiring that $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ is the same as requiring that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to Col A.
- This is equivalent to requiring that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of A. This means

$$\mathbf{a}_1^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0, \ \mathbf{a}_2^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0, \ \cdots, \ \mathbf{a}_n^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0.$$

- This gives

$$\begin{aligned} \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix} (\mathbf{b} - A\hat{\mathbf{x}}) &= \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ A^T(\mathbf{b} - A\hat{\mathbf{x}}) &= \mathbf{0} \\ A^T\mathbf{b} - A^TA\hat{\mathbf{x}} &= \mathbf{0} \end{aligned}$$

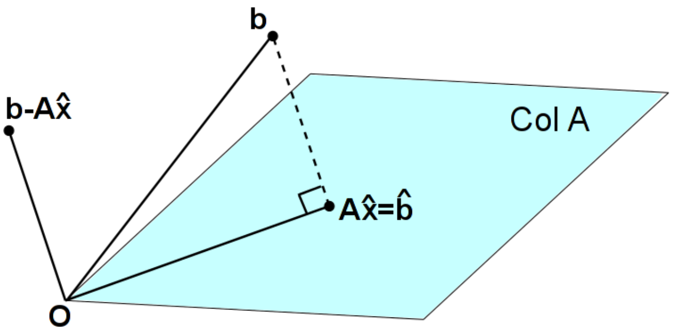
$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$$

These are the normal equations for $\hat{\mathbf{x}}$.

Theorem

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations

$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}.$$



Since $A\hat{\mathbf{x}}$ is automatically in Col A and $\hat{\mathbf{b}}$ is the unique vector in Col A such that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col A, requiring that $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ is the same as requiring that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to Col A.

Examples

Example 1

Find a least squares solution to the inconsistent system $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}.$$

To solve the normal equations $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$, we first compute the relevant matrices:

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 11 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}.$$

So we need to solve $\begin{bmatrix} 3 & 3 \\ 3 & 11 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$. The augmented matrix is

$$\left[\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 11 & 14 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 1 & 2 \\ 3 & 11 & 14 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 8 & 8 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & 1 \end{array}\right] \to \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right.$$

This gives $\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Note that $A\hat{\mathbf{x}} = \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix}$ and this is the closest point in Col A to $\mathbf{b} = \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}$.

We could note in this example that $A^TA = \begin{bmatrix} 3 & 3 \\ 3 & 11 \end{bmatrix}$ is invertible with inverse $\frac{1}{24}\begin{bmatrix} 11 & -3 \\ -3 & 3 \end{bmatrix}$. In this case the normal equations give

$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b} \iff \hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}.$$

So we can calculate

$$\begin{aligned} \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \\ &= \frac{1}{24}\begin{bmatrix} 11 & -3 \\ -3 & 3 \end{bmatrix}\begin{bmatrix} 6 \\ 14 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \end{aligned}$$

Example 2

Find a least squares solution to the inconsistent system $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 3 & -1 \\ 1 & -2 \\ 2 & 3 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}.$$

Notice that
$A^TA = \begin{bmatrix} 3 & 1 & 2 \\ -1 & -2 & 3 \end{bmatrix}\begin{bmatrix} 3 & -1 \\ 1 & -2 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 14 & 1 \\ 1 & 14 \end{bmatrix}$ is invertible. Thus the normal equations become

$$\begin{aligned} A^TA\hat{\mathbf{x}} &= A^T\mathbf{b} \\ \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \end{aligned}$$

Furthermore,

$$A^T\mathbf{b} = \begin{bmatrix} 3 & 1 & 2 \\ -1 & -2 & 3 \end{bmatrix}\begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 19 \\ -4 \end{bmatrix}$$

So in this case

$$\begin{aligned} \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \\ &= \begin{bmatrix} 14 & 1 \\ 1 & 14 \end{bmatrix}^{-1}\begin{bmatrix} 19 \\ -4 \end{bmatrix} \\ &= \frac{1}{195}\begin{bmatrix} 14 & -1 \\ -1 & 14 \end{bmatrix}\begin{bmatrix} 19 \\ -4 \end{bmatrix} \\ &= \frac{1}{13}\begin{bmatrix} 18 \\ -5 \end{bmatrix}. \end{aligned}$$

With these values, we have

$$A\hat{\mathbf{x}} = \frac{1}{13}\begin{bmatrix}59\\28\\21\end{bmatrix} \sim \begin{bmatrix}5.54\\2.15\\1.62\end{bmatrix}$$

which is as close as possible to $\begin{bmatrix}4\\3\\2\end{bmatrix}$. □

Example 3

For $A = \begin{bmatrix}1 & 0 & 2\\2 & 1 & 5\\-1 & 1 & -1\\0 & 1 & 1\end{bmatrix}$, what are the least squares solutions to

$A\mathbf{x} = \mathbf{b} = \begin{bmatrix}1\\-1\\-1\\2\end{bmatrix}$?

$$A^TA = \begin{bmatrix}6 & 1 & 13\\1 & 3 & 5\\13 & 5 & 31\end{bmatrix}, \quad A^T\mathbf{b} = \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

For this example, solving $A^TA\hat{x} = A^T\mathbf{b}$ is equivalent to finding the null space of A^TA

$$\begin{bmatrix}6 & 1 & 13\\1 & 3 & 5\\13 & 5 & 31\end{bmatrix} \xrightarrow{rref} \begin{bmatrix}1 & 0 & 2\\0 & 1 & 1\\0 & 0 & 0\end{bmatrix}$$

Here, x_3 is free and $x_2 = -x_3, x_1 = -2x_3$.

So $\mathrm{Nul}\ A^TA = \mathbb{R}\begin{bmatrix}2\\1\\-1\end{bmatrix}$.

Here $A\hat{\mathbf{x}} = \mathbf{0}$ –not a very good approximation!
Remember that we are looking for the vectors that map to the closest point to $\mathbf{b}$ in $\mathrm{Col}\ A$.

The question of a "best approximation" to a solution has been reduced to solving the normal equations.

An immediate consequence is that there is going to be a unique least squares solution if and only if $A^T A$ is invertible (note that $A^T A$ is always a square matrix).

Theorem

The matrix $A^T A$ is invertible if and only if the columns of A are linearly independent. In this case the equation $A\mathbf{x} = \mathbf{b}$ has only one least squares solution $\hat{\mathbf{x}}$, and it is given by

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

For the proof of this theorem see Lay 6.5 Exercises 19 - 21.

Formula (1) for $\hat{\mathbf{x}}$ is useful mainly for theoretical calculations and for hand calculations when $A^T A$ is a 2×2 invertible matrix.

When a least squares solution $\hat{\mathbf{x}}$ is used to produce $A\hat{\mathbf{x}}$ as an approximation to $\mathbf{b}$, the distance from $\mathbf{b}$ to $A\hat{\mathbf{x}}$ is called the **least squares error** of this approximation.

Example 4

Given $A = \begin{bmatrix} 3 & -1 \\ 1 & -2 \\ 2 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}$ as in Example 2, we found

$$A\hat{\mathbf{x}} = \frac{1}{13}\begin{bmatrix} 59 \\ 28 \\ 21 \end{bmatrix} \sim \begin{bmatrix} 5.54 \\ 2.15 \\ 1.62 \end{bmatrix}$$

Then the **least squares error** is given by $||\mathbf{b} - A\hat{\mathbf{x}}||$, and since

$$\mathbf{b} - A\hat{\mathbf{x}} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix} - \begin{bmatrix} 5.54 \\ 2.15 \\ 1.62 \end{bmatrix} = \begin{bmatrix} -1.54 \\ 0.85 \\ 0.38 \end{bmatrix},$$

we have

$$||\mathbf{b} - A\hat{\mathbf{x}}|| = \sqrt{(-1.54)^2 + .85^2 + .38^2} \approx \sqrt{3.24}.$$

Alternative calculations

Note: we didn't cover the QR decomposition in class; these slides are just provided as a reference for your own interest.
In some cases the normal equations for a least squares problem can be *ill conditioned*; that is, small errors in the calculations of the entries of A^TA can sometimes cause relatively large errors in the solution $\hat{\mathbf{x}}$. If the columns of A are linearly independent, the least squares solution can be computed more reliably through a QR factorisation of A.

Theorem

Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorisation of A. Then for each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution, given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}. \tag{2}$$

Proof. Let $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$. Then

$$A\hat{\mathbf{x}} = QR\hat{\mathbf{x}} = QRR^{-1}Q^T\mathbf{b} = QQ^T\mathbf{b}.$$

The columns of Q form an orthonormal basis for Col A. Hence $QQ^T\mathbf{b}$ is the orthogonal projection of $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto Col A.
Thus $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, which shows that $\hat{\mathbf{x}}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$.
The uniqueness of $\hat{\mathbf{x}}$ follows from the previous theorem. □

Note that $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$ is equivalent to

$$R\hat{\mathbf{x}} = Q^T\mathbf{b} \tag{3}$$

Because R is upper triangular it is faster to solve (3) by back-substitution or row operations than to compute R^{-1} and use (2).

3.1 Examples

Example 5

We are given

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 4 \\ 1 & -1 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \\ 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} -1 \\ 6 \\ 5 \\ 7 \end{bmatrix}$$

Using this QR factoristaion of A we want to find the least squares solution of $A\mathbf{x} = \mathbf{b}$.

We will use the equation $R\hat{\mathbf{x}} = Q^T\mathbf{b}$ to solve this problem.

We calculate

$$Q^T\mathbf{b} = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -1/2 & 1/2 & -1/2 & 1/2 \end{bmatrix} \begin{bmatrix} -1 \\ 6 \\ 5 \\ 7 \end{bmatrix}$$

$$= \begin{bmatrix} 17/2 \\ 9/2 \end{bmatrix}$$

The least squares solution $\hat{\mathbf{x}}$ satisfies $R\hat{\mathbf{x}} = Q^T\mathbf{b}$; that is

$$\begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 17/2 \\ 9/2 \end{bmatrix}.$$

This is easily solved to give

$$\hat{\mathbf{x}} = \begin{bmatrix} 29/10 \\ 9/10 \end{bmatrix},$$

and

$$A\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 13/2 \\ 2 \\ 13/2 \end{bmatrix}.$$

Example 6

We want to find the least squares solution for $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Gram-Schmidt on the columns of A yields

$$Q = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{2} & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix}.$$

Now we know that $R = Q^TA$.

Thus

$$R = \begin{bmatrix} \sqrt{6} & \sqrt{6}/2 & 11/\sqrt{6} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{3} \end{bmatrix}, Q^T\mathbf{b} = \begin{bmatrix} \sqrt{6}/3 \\ 0 \\ -2/\sqrt{3} \end{bmatrix}.$$

So we need to solve

$$\begin{bmatrix} \sqrt{6} & \sqrt{6}/2 & 11/\sqrt{6} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{3} \end{bmatrix} \widehat{\mathbf{x}} = \begin{bmatrix} \sqrt{6}/3 \\ 0 \\ -2/\sqrt{3} \end{bmatrix}$$

Thus $\widehat{\mathbf{x}} = \begin{bmatrix} 5 \\ -2 \\ -2 \end{bmatrix}$ almost immediately. Then $A\widehat{\mathbf{x}} = \mathbf{b}$, an exact solution this time. □