LEAST SQUARES SOLUTIONS

1. Orthogonal projection as closest point

The following minimizing property of orthogonal projection is very important:

Theorem 1.1. *Fix a subspace $V \subset \mathbb{R}^n$ and a vector $\vec{x} \in \mathbb{R}^n$. The orthogonal projection* $\mathrm{proj}_V(\vec{x})$ *onto V is the vector in V closest to $\vec{x}$. That is,*

$$||\vec{x} - \mathrm{proj}_V(\vec{x})|| < ||\vec{x} - \vec{v}||$$

for all $\vec{v} \in V$ with $\vec{v} \neq \mathrm{proj}_V(\vec{x})$.

Proof. Algebraically, have

$$\vec{x} = \mathrm{proj}_V(\vec{x}) + \vec{x}^{\perp}$$

Notice, for any $\vec{v} \in V$, $\mathrm{proj}_V(\vec{x}) - \vec{v} \in V$ and so is orthogonal to $\vec{x}^{\perp}$. Hence,

$$||\vec{x} - \vec{v}||^2 = ||(\mathrm{proj}_V\vec{x} - \vec{v}) + \vec{x}^{\perp}||^2 = ||\mathrm{proj}_V\vec{x} - \vec{v}||^2 + ||\vec{x}^{\perp}||^2.$$

Where the last equality used the Pythagorean theorem. Hence,

$$||\vec{x} - \vec{v}||^2 = ||\vec{x} - \mathrm{proj}_V(\vec{x})||^2 + ||\vec{x} - \mathrm{proj}_V(\vec{x})||^2 > ||\vec{x} - \mathrm{proj}_V(\vec{x})||^2$$

where we used that $\vec{x} \neq \mathrm{proj}_V(\vec{x}) \Rightarrow ||\mathrm{proj}_V\vec{x} - \vec{v}|| > 0$. □

2. Orthogonal Complements and the Transpose

Recall, the fundamental identity

$$A \in \mathbb{R}^{n\times m}, \vec{x} \in \mathbb{R}^m, \vec{y} \in \mathbb{R}^n$$

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

This implies, that if $\vec{y} \in \ker(A^{\top})$, then $\vec{y}$ is orthogonal to $\mathrm{Im}\,(A)$. That is,

$$\ker(A^{\top}) \subset \mathrm{Im}\,(A)^{\perp}.$$

Conversely, if $\vec{y} \in (im(A))^{\perp}$, then $A^{\top}\vec{y}$ is orthogonal to every vector $\vec{x}$. That is, $A^{\perp}\vec{y} \in (\mathbb{R}^m)^{\perp} = \left\{\vec{0}\right\}$

$$\mathrm{Im}\,(A)^{\perp} \subset \ker(A^{\top})$$

Hence, $\ker(A^{\top}) = im(A)^{\perp}$. Note that,

$$n - \dim(\mathrm{Im}\,(A)) = \dim(\mathrm{Im}\,(A)^{\perp}) = \dim(\ker(A^{\top})) = n - \dim(\mathrm{Im}\,(A^{\top})).$$

Hence,

$$\mathrm{rank}(A) = \dim(\mathrm{Im}\,(A)) = \dim(\mathrm{Im}\,(A^{\perp})) = \mathrm{rank}(A^{\top})$$

Moreover,

Theorem 2.1.
(1) *If $A \in \mathbb{R}^{n\times m}$, then* $\ker(A) = \ker(A^{\top}A)$
(2) *If $A \in \mathbb{R}^{n\times m}$ has* $\ker(A) = \left\{\vec{0}\right\}$*, then $A^{\top}A$ is invertible.*

Proof. To see Claim (1) observe that, as we have seen before, $\ker(A) \subset \ker(A^\top A)$. If $\vec{x} \in \ker(A^\top A)$, then $A\vec{x} \in \ker(A^\top) = (\mathrm{Im}(A))^\top$. This means $A\vec{x} \in \mathrm{Im}(A) \cap \mathrm{Im}(A)^\top$ so $A\vec{x} = \vec{0}$. Hence, $\vec{x} \in \ker(A)$ and so $\ker(A^\top A) \subset \ker(A)$. Together, this proves the claim. Claim (2) follows from the fact that $A^\top A \in \mathbb{R}^{m \times m}$ is square. By by the first claim, $\ker(A^\top A) = \ker(A) = \left\{\vec{0}\right\}$. This means $A^\top A$ is invertible by previous work. $\square$

3. Least Squares Solution

Suppose we have an inconsistent system

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

Here $A \in \mathbb{R}^{n \times m}$ and $\vec{b} \in \mathbb{R}^n$. Recall, this means that $\vec{b} \notin \mathrm{Im}\,(A)$. We say $\vec{x}^* \in \mathbb{R}^m$ is a *least squares solution* if

$$||\vec{b} - A\vec{x}^*|| \leq ||\vec{b} - A\vec{x}||$$

for all $\vec{x} \in \mathbb{R}^m$. In other words, $\vec{y}^* = A\vec{x}^*$ is the vector in $\mathrm{Im}\,(A)$ that is closest to $\vec{b}$, that is is closest to being a true solution. We call the distance,

$$E = ||A\vec{x}^* - \vec{b}||$$

the *error*. This measures how far a least squares solution is from being a true solution.

How to compute a least squares solution? From our first observation in this handout, the point $\vec{y}^*$ closest to $\vec{b}$ is given by

$$\vec{y}^* = \mathrm{proj}_V(\vec{b})$$

where $V = \mathrm{Im}\,(A)$. As $\vec{y}^* \in \mathrm{Im}\,(A)$, then can find the least square solutions by solving the (now consistent) system

$$A\vec{x} = \vec{y}^* = proj_V(\vec{b})$$

Notice, this approach requires us to find $\mathrm{proj}_V(\vec{b})$. In general, this is non-trivial step as we need an orthonormal basis of $\mathrm{Im}\,(A)$.

4. The normal equation

The following theorem gives a more direct method for finding least squares solutions.

Theorem 4.1. *The least square solutions of*

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

are the exact solutions of the (necessarily consistent) system

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

This system is called the normal equation *of* $A\vec{x} = \vec{b}$.

Proof. We have the following equivalent statements: $\vec{x}^*$ is a least squares solution

$$\iff \vec{b} - A\vec{x}^* = \vec{b} - \mathrm{proj}_V(\vec{b}) \in V^\perp = (\mathrm{Im}\,(A))^\perp = \ker(A^\top)$$

$$\iff A^\top(A\vec{x}^* - \vec{b}) = \vec{0} \iff A^\top A\vec{x}^* = A^\top \vec{b}.$$

$\square$

Notice, that $\ker(A^\top A) = \ker(A)$ so the least square solution need not be unique. However, the lack of uniqueness is encoded in $\ker(A)$.

Theorem 4.2. *If* $\ker(A) = \left\{\vec{0}\right\}$, *then the linear system*

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

has the unique least squares solution

$$\vec{x}^* = (A^\top A)^{-1} A^\top \vec{b}$$

Usually, it is more computationally efficient to apply Gaussian elimination to the normal equation.

EXAMPLE: Find a least squares solution to

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 1 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

The normal equation of this system is

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 1 \end{bmatrix}^\top \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 1 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 1 \end{bmatrix}^\top \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 5 & 0 \\ 0 & 6 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

We can solve this by inspection and see that the unique least squares solution is

$$\vec{x}^* = \begin{bmatrix} 1/5 \\ 1/3 \end{bmatrix}.$$

Notice,

$$\vec{b} - A\vec{x}^* = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 13/15 \\ 1/3 \\ -1/15 \end{bmatrix} = \frac{1}{15} \begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix}$$

So $\vec{x}^*$ is not a true solution. The (square of) the error is

$$||\vec{b} - A\vec{x}^*||^2 = \frac{30}{15^2} = \frac{2}{15}.$$

5. Data fitting

An important use of least square methods is in finding functions that fit some data (this is also called regression analysis). While there are many different methods to do this, the first was to use least squares. To understand the idea, recall that, in general, it is not possible to find a quadratic polynomial whose graph goes through more than three specified points. However, using least squares, one can try and find the quadratic polynomial that is as close as possible (in a certain sense) to going through the points.

EXAMPLE: Fit a quadratic polynomial $p \in P_2$ to

$$(1, 2), (0, 0), (2, 0), (-1, -2).$$

Write $p(x) = a_0 + a_1 x + a_2 x^2$. Our goal is to find a p that is as close as possible to satisfying:

$$\left\{ \begin{array}{c} p(1) = 2 \\ p(0) = 0 \\ p(2) = 0 \\ p(-1) = -2 \end{array} \right. \iff \left\{ \begin{array}{c} a_0 + a_1 + a_2 = 2 \\ a_0 = 0 \\ a_0 + 2a_1 + 4a_2 = 0 \\ a_0 - a_1 + a_2 = -2 \end{array} \right.$$

We can rewrite this system in matrix form as:

$$\begin{bmatrix}1 & 1 & 1\\ 1 & 0 & 0\\ 1 & 2 & 4\\ 1 & -1 & 1\end{bmatrix}\begin{bmatrix}a_0\\ a_1\\ a_2\end{bmatrix} = \begin{bmatrix}2\\ 0\\ 0\\ -2\end{bmatrix}.$$

The normal equation of this system is

$$\begin{bmatrix}1 & 1 & 1\\ 1 & 0 & 0\\ 1 & 2 & 4\\ 1 & -1 & 1\end{bmatrix}^\top \begin{bmatrix}1 & 1 & 1\\ 1 & 0 & 0\\ 1 & 2 & 4\\ 1 & -1 & 1\end{bmatrix}\vec{a} = \begin{bmatrix}1 & 1 & 1\\ 1 & 0 & 0\\ 1 & 2 & 4\\ 1 & -1 & 1\end{bmatrix}^\top \begin{bmatrix}2\\ 0\\ 0\\ -2\end{bmatrix}$$

$$\begin{bmatrix}4 & 2 & 6\\ 2 & 6 & 8\\ 6 & 8 & 18\end{bmatrix}\vec{a} = \begin{bmatrix}0\\ 4\\ 0\end{bmatrix}$$

$$\operatorname{rref}\begin{bmatrix}4 & 2 & 6 & | & 0\\ 2 & 6 & 8 & | & 4\\ 6 & 8 & 18 & | & 0\end{bmatrix} = \begin{bmatrix}1 & 0 & 0 & | & 3/5\\ 0 & 1 & 0 & | & 9/5\\ 0 & 0 & 1 & | & -1\end{bmatrix}$$

So

$$\vec{a}^* = \begin{bmatrix}3/5\\ 9/5\\ -1\end{bmatrix}$$

and

$$p^*(x) = \frac{3}{5} + \frac{9}{5}x - x^2$$

is least squares solution in sense that of all quadratic polynomials p^* minimizes

$$E(p)^2 = (p(1) - 2)^2 + (p(0) - 0)^2 + (p(2) - 0)^2 + (p(-1) + 2)^2.$$

That is,

$$E(p)^2 \geq E(p^*)^2 = \left(\frac{-3}{5}\right)^2 + \left(\frac{3}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{-1}{5}\right)^2 = \frac{20}{25} = \frac{4}{5}.$$