25. Least Squares

Consider the linear system $L(x) = v$, where $L : U \stackrel{\text{linear}}{\Longrightarrow} W$, and $v \in W$ is given. As we have seen, this system may have no solutions, a unique solution, or a space of solutions. But if v is not in the range of L then there will never be any solutions for $L(x) = v$.

However, for many applications we do not need a exact solution of the system; instead, we try to find the best approximation possible. To do this, we try to find x that minimizes $||L(x) - v||$.

"My work always tried to unite the Truth with the Beautiful, but when I had to choose one or the other, I usually chose the Beautiful." – Herman Weyl.

This method has many applications, such as when trying to fit a (perhaps linear) function to a 'noisy' set of observations. For example, suppose we measured the position of a bicycle on a racetrack once every five seconds. Our observations won't be exact, but so long as the observations are right on average, we can figure out a best-possible linear function of position of the bicycle in terms of time.

Suppose M is the matrix for L in some bases for U and W, and v and x are given by column vectors V and X in these bases. Then we need to approximate

$$MX - V \sim 0$$

Note that if $\dim U = n$ and $\dim W = m$ then M can be represented by an $m \times n$ matrix. We can write $W = L(U) \oplus L(U)^{\perp}$. Then we can uniquely write $V = V^{\parallel} + V^{\perp}$, with $V^{\parallel} \in L(U)$ and $V^{\perp} \in L(U)^{\perp}$.

Then we should solve $L(u) = V^{\parallel}$. In components, $V^{\perp}$ is just $V - MX$, and is the part we will eventually wish to minimize.

In terms of M, recall that $L(V)$ is spanned by the columns of M. (In the natural basis, the columns of M are $Me_1, \ldots, Me_n$.) Then $v^{\perp}$ must be perpendicular to the columns of M. *i.e.*, $M^T(V - MX) = 0$, or

$$M^T MX = M^T V.$$

Solutions X to $M^T MX = M^T V$ are called *least squares* solutions to $MX = V$.

Notice that any solution X to $MX = V$ is a least squares solution. However, the converse is often false. In fact, the equation $MX = V$ may have no solutions at all, but still have least squares solutions to $M^T MX = M^T V$.

Observe that since M is an $m \times n$ matrix, then M^T is an $n \times m$ matrix. Then $M^T M$ is an $n \times n$ matrix, and is symmetric, since $(M^T M)^T = M^T M$. Then, for any vector X, we can evaluate $X^T M^T M X$ to obtain a number. This is a very nice number, though! It is just the length $|MX|^2 = (MX)^T (MX) = X^T M^T M X$.

Now suppose that $\ker L = \{0\}$, so that the only solution to $MX = 0$ is $X = 0$; in particular, M is invertible. But if M is invertible, then so is $M^T M$, since $(M^T M)^{-1} = M^{-1} M^{-T}$. Then the only solution to $M^T M X = 0$ is $X = 0$.

In this case, the least squares solution (the X that solves $M^T M X = MV$ is unique, and is equal to

$$X = (M^T M)^{-1} M^T V.$$

In a nutshell, this is the least square method.

- Compute $M^T M$ and $M^T V$.
- Solve $(M^T M)X = M^T V$ by Gaussian elimination.

Example Captain Conundrum falls off of the leaning tower of Pisa and makes three rather shaky measurements of his velocity at three different times.

t	$\frac{m}{s}$
1	11
2	19
3	31

Having taken some calculus, he believes that his data are best approximated by a straight line

$$v = at + b.$$

Then we should find a and b to best fit the data.

$$\begin{aligned} 11 &= a \cdot 1 + b \\ 19 &= a \cdot 2 + b \\ 31 &= a \cdot 3 + b. \end{aligned}$$

As a system of linear equations, this becomes:

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 11 \\ 19 \\ 31 \end{pmatrix}.$$

There is likely no actual straightline solution, so instead solve $M^T M X = MV$.

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 11 \\ 19 \\ 31 \end{pmatrix}.$$

This simplifies to the system:

$$\left(\begin{array}{cc|c} 14 & 6 & 142 \\ 6 & 3 & 61 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 10 \\ 0 & 1 & \frac{1}{3} \end{array}\right).$$

Then the least-squares fit is the line

$$v = 10t + \frac{1}{3}.$$

References

- Hefferon, Chapter Three, Section VI.2: Gram-Schmidt Orthogonalization

Wikipedia:

- Linear Least Squares
- Least Squares

Review Questions

1. Let $L : U \to V$ be a linear transformation. Suppose $v \in L(U)$ and you have found a vector u_{ps} that obeys $L(u_{ps}) = v$.

   Explain why you need to compute $\ker L$ to describe the solution space of the linear system $L(u) = v$.

2. Suppose that M is an $m \times n$ matrix with trivial kernel. Show that for any vectors u and v in $\mathbb{R}^m$:

   - $u^T M^T M v = v^T M^T M u$
   - $v^T M^T M v \geq 0$.
   - If $v^T M^T M v = 0$, then $v = 0$.

   (Hint: Think about the dot product in $\mathbb{R}^n$.)