Challenge Problem Set 4, Math 291 Fall 2017

Eric A. Carlen[1]
Rutgers University

November 14, 2017

This challenge problem set concerns least squares solutions of linear equations.

0.1 Least squares solutions

Consider the equation $A\mathbf{x} = \mathbf{b}$ where A is an $m \times n$ matrix. We know that this equation is solvable if and only if $\mathbf{b}$ lies in the span of the columns of A. And we know that in this case the solution is unique if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the solution $\mathbf{x} = \mathbf{0}$.

Suppose that the span of the columns of A is not all of $\mathbb{R}^m$, so that there exist vectors $\mathbf{b} \in \mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{b}$ has no solution. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ be the orthonormal set obtained by applying the Gram-Schmidt Algorithm to the columns of A.

Exercise 1: For any $\mathbf{b} \in \mathbb{R}^m$, define

$$\tilde{\mathbf{b}} = \sum_{j=1}^{r} (\mathbf{b} \cdot \mathbf{u}_j)\mathbf{u}_j \ .$$

Show that $\tilde{\mathbf{b}}$ belongs to the span of the columns of A, and that for all $\mathbf{y}$ in the span of the columns of A,

$$\|\tilde{\mathbf{b}} - \mathbf{b}\|^2 \leq \|\mathbf{y} - \mathbf{b}\|^2 \ .$$

Exercise 2. Let $\tilde{\mathbf{b}}$ be defined as in Exercise 1. Since $\tilde{\mathbf{b}}$ belongs to the span of the columns of A, there is at least one $\mathbf{x}_0\ in \mathbb{R}^n$ such that $A\mathbf{x}_0 = \tilde{\mathbf{b}}$. Show that for all $\mathbf{x} in \mathbb{R}^n$,

$$\|A\mathbf{x}_0 - \mathbf{b}\|^2 \leq \|A\mathbf{x} - \mathbf{b}\|^2 \ .$$

(Use the result of Exercise 1.)

The vector $\mathbf{x}_0$ considered in Exercise 2 may not solve $a\mathbf{x} = \mathbf{b}$, but it comes as close as possible: It makes $\|A\mathbf{x} - \mathbf{b}\|^2$ as small as possible. It is what is commonly called a *least squares solution* of the equation $A\mathbf{x} = \mathbf{b}$.

However, if $A\mathbf{x} = \mathbf{0}$ has solutions other than $\mathbf{x} = \mathbf{0}$, there will be infinitely many such least squares solutions. Among these, there is one that is special: It has a length that is less than that of any other least squares solution.

To see this, let $\{\mathbf{w}_1, \ldots, \mathbf{w}_s\}$ be the orthonormal basis that one gets by applying the Gram-Schmidt algorithm to the rows of A. Let $\{\mathbf{w}_1, \ldots, \mathbf{w}_s, \mathbf{w}_{s+1}, \ldots, \mathbf{w}_n\}$ be an orthonormal basis of $\mathbb{R}^n$ that one gets by extending $\{\mathbf{w}_1, \ldots, \mathbf{w}_s\}$.

Exercise 3: Show that for $s+1 \le j \le n$, $A\mathbf{w}_j = 0$. Then show that if $\mathbf{x}_0$ satisfies $A\mathbf{x}_0 = \tilde{\mathbf{b}}$, then so does

$$\mathbf{x}_1 := \mathbf{x}_0 - \sum_{j=s+1}^{n} (\mathbf{x}_0 \cdot \mathbf{w}_j)\mathbf{w}_j \ .$$

Moreover, show that $\|\mathbf{x}_1\| \le \|\mathbf{x}_0\|$, and there is equality if and only if $\mathbf{x}_0$ is in the span of the rows of A. This least length solution is called the least length, least squares solution of $A\mathbf{x} = \mathbf{b}$. Thus, *every* matrix equation $A\mathbf{x} = \mathbf{b}$ has a unique least length, least squares solution

Exercise 4. Let $A = \begin{bmatrix} 1 & 0 & -2 & -1 \\ 1 & 2 & 2 & 3 \\ -1 & 0 & 2 & 3 \\ 1 & 2 & 2 & 5 \end{bmatrix}$.

Apply the Gram-Schmidt Algorithm to the columns and rows of A, and fins the unique least length, least squares solution of $A\mathbf{x} = \mathbf{b}$ for

$\mathbf{b} = (1,2,3,4)$ and then $\mathbf{b} = (1,1,1,-1)$ and then $\mathbf{b} = (1,1,-1,1)$ and then .

Here is a typical application of the least squares method. Suppose you know, for theoretical reasons that some variable y must depend on another variable x in a "linear way";

$$y = ax + b \ ,$$

for some constants a and b. To determine a and b from measurements, you could measure the output variable y for two values of the input variable x, say x_1 and x_2 and you would get a pair of equations that you could then solve for a and b:

$$\begin{aligned} ax_1 + b &= y_1 \\ ax_2 + b &= y_2 \end{aligned}$$

The problem is that all laboratory measurements have some error in them, and to bet better values of a and b, you might make more measurements, say 6:

$$\begin{aligned} ax_1 + b &= y_1 \\ ax_2 + b &= y_2 \\ ax_3 + b &= y_3 \\ ax_4 + b &= y_4 \\ ax_5 + b &= y_5 \\ ax_6 + b &= y_6 \end{aligned}$$

Now the problem is, in general, over-determined. There are 6 equations and only 2 variables, and in general no solution – there will be a solution if and only if all of the point (x_j, y_j) lie exactly on some line. Usually there will be some scatter. One can find the "least squares fit" of a line to this scattered data using least squares solutions.

Introduce the matrix

$$A := \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \vdots \\ x_n & 1 \end{bmatrix}$$

And the vector $\mathbf{y} := (y_1, y_2, \dots, y_n)$. Then with $\mathbf{x} = (a, b)$, we seek a least squares solution to $A\mathbf{x} = \mathbf{y}$. Using these values of a and b, you can plot a line that is the best least-squares fir of the data.

Exercise 5. Show that as long as all of the x_j are not the same, $\text{rank}(A) = 2$ and then that the 2×2 matrix $A^T Aisinvertible.Notethatif\mathbf{x}solvesAx= \mathbf{y}, thenA^T A\mathbf{x} = A^T\mathbf{y}$, and hence $\mathbf{x} = (A^T A)^{-1} A^T\mathbf{y}$. Show that for all $\mathbf{y} \in \mathbb{R}^n$, without any other assumptions, $(A^T A)^{-1} A^T\mathbf{y}$ is the unique least squares solution of $A\mathbf{x} = \mathbf{y}$ in this case. Notice that $A^T A$ is a 2×2 matrix no matter how large n is, so that it is always easy to invert.

Exercise 6. Find the line $y = ax + b$ that is the best least squares fit to the points

$$(1, 0.12)\ ,\ (1.5, 0.77)\ ,\ (2, 1.48)\ ,\ (2.5, 2.11)\ ,\ (3, 0.2.75)\ ,\ (3.5, 3.49)\ ,\ (4, 4.21)\ ,\ (4.5, 4.81)\ ,\ .$$