Solutions for Challenge Problem Set 4, Math 291 Fall 2017

Eric A. Carlen[1]
Rutgers University

December 20, 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 $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ 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,

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

SOLUTION Clearly $\tilde{\mathbf{b}} \in \mathrm{Span}(\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$, and by the properties of the Gram-Schmidt algorithm, the span of the columns of A equals $\mathrm{Span}(\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$. Next, note that $\mathbf{b} - \tilde{\mathbf{b}}$ is orthogonal to each $\mathbf{u}_j$, and hence to each vector in the span of the columns of A. Then for $\mathbf{y}$ in the span of the columns of A, $\mathbf{y} - \tilde{\mathbf{b}}$ belongs to the span of the columns of A, and $\tilde{\mathbf{b}} - \mathbf{b}$ is orthogonal to the span of the columns of A.

Then since $\mathbf{y} - \mathbf{b} = (\mathbf{y} - \tilde{\mathbf{b}}) + (\tilde{\mathbf{b}} - \mathbf{b})$, the Pythagorean Theorem gives

$$\|\mathbf{y} - \mathbf{b}\|^2 = \|\mathbf{y} - \tilde{\mathbf{b}}\|^2 + \|\tilde{\mathbf{b}} - \mathbf{b}\|^2 \geq \|\tilde{\mathbf{b}} - \mathbf{b}\|^2$$

and there is equality if and only if $\mathbf{y} = \tilde{\mathbf{b}}$.

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.)

SOLUTION By definition $A\mathbf{x}_0 = \tilde{\mathbf{b}}$ and we can define $\mathbf{y} := A\mathbf{x}$. Then $\mathbf{y}$ is in the span of the columns of A, and so by the first exercise, $\|\tilde{\mathbf{b}} - \mathbf{b}\|^2 \le \|\mathbf{y} - \mathbf{b}\|^2$, which is then the same as $\|A\mathbf{x}_0 - \mathbf{b}\|^2 \le \|A\mathbf{x} - \mathbf{b}\|^2$.

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

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

SOLUTION By linearity,

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

Also

$$\|\mathbf{x}_1\| = \sum_{j=1}^{s} (\mathbf{x}_0 \cdot \mathbf{w}_j)^2 \le \sum_{j=1}^{n} (\mathbf{x}_0 \cdot \mathbf{w}_j)^2$$

with equality only in case $\mathbf{x}_0 \cdot \mathbf{w}_j = 0$ for each $j = s+1, \dots, n$, and in this case $\mathbf{x}_1 = \mathbf{x}_0$.

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 find the unique least length, least squares solution of $A\mathbf{x} = \mathbf{b}$ for

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

SOLUTION We first compute the QR factorization of A:

$$Q = \begin{bmatrix} 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} 2 & 2 & 0 & 2 \\ 0 & 2 & 4 & 6 \\ 0 & 0 & 0 & 2 \end{bmatrix} \ .$$

Then $\tilde{bb}$ is given by $\tilde{\mathbf{b}} = QQ^T\mathbf{b}$, and then we get a least squares solution by solving $A\mathbf{x} = \tilde{\mathbf{b}}$, which since $A = QR$ and $Q^TQ = I_{3\times 3}$, is equivalent to solving $R\mathbf{x} = Q^t\mathbf{b}$. Since the third column of R is non-pivotal, we can always find a solution with $x_3 = 0$ by back substitution.

Proceeding to actual computations, take $\mathbf{b} = (1,2,3,4)$. Then $Q^t\mathbf{b} = (2,4,3)$ and $R\mathbf{x} = (2,4,3)$ has the solution $(2,-5/2,0,3/2)$. (This is the unique solution with $x_3 = 0$.

Next, taking $bb = (1,1,1,1)$, $Q^t\mathbf{b} = (2,4,3)$ and $R\mathbf{x} = (1,1,1)$ has the solution $(1,-1,0,1/2)$.

Next, taking $bb = (1,1,-1,1)$, $Q^t\mathbf{b} = (2,4,3)$ and $R\mathbf{x} = (2,0,0)$ has the solution $(1,0,0,0)$.

To convert our least squares solutions into least-length least squares solutions, apply the Gram-Schmidt algorithm to the rows of A. Let $W = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3]$ be the 3×4 matrix that has these vectors as its columns. Then define $P = WW^T$, and note that for any $\mathbf{x} \in \mathbb{R}^4$,

$$P\mathbf{x} = \sum_{j=1}^{3}(\mathbf{x}\cdot\mathbf{w}_j)\mathbf{w}_j$$

and hence applying P to our least squares solutions, converts them to least-length least squares solutions. Doing the computations, we find

$$P = \begin{bmatrix} 5/9 & 4/9 & -2/9 & 0 \\ 4/9 & 5/9 & 2/9 & 0 \\ -2/9 & 2/9 & 8/9 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Then $P(2,-5/2,0,3/2) = (0,-1/2,-1,3/2)$ is the least-length least squares solution for the first choice of $\mathbf{b}$, $P(1,-1,0,1/2) = (1/9,-1/9,-4/9,1/2)$ is the least-length least squares solution for the second choice of $\mathbf{b}$, and $P(1,0,0,0) = (5/9,4/9,-4/9,0)$ is the least-length least squares solution for the third choice of $\mathbf{b}$.

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

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

And the vector $\mathbf{y} := (y_1, y_2, \ldots, y_n)$. 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, $\mathrm{rank}(A) = 2$ and then that the 2×2 matrix $A^T A$ is invertible. Note that if $\mathbf{x}$ solves $A\mathbf{x} = \mathbf{y}$, then $A^T A\mathbf{x} = A^T\mathbf{y}$, 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.

SOLUTION If $A^T A\mathbf{x} = \mathbf{0}$B dor any bx, then $0 = \mathbf{x} \cdot A^T A\mathbf{x} = A\mathbf{x} \cdot A\mathbf{x} = \|A\mathbf{x}\|^2$. Since the columns of A are linearly independent under the rank assumption, this means $\mathbf{x} = 0$. Hence the null space of A is trivial, and since it is a square matrix, the fundamental Theorem of Linear Algebra says it is invertible.

Now let $\tilde{\mathbf{y}}$ be related to $\mathbf{y}$ as in Exercise 1, so that $A\mathbf{x} = \tilde{\mathbf{y}}$ has a unique solution $\mathbf{x}_0$ which is the least squares solutions to $A\mathbf{x} = \mathbf{y}$. Then $\mathbf{y} = (\mathbf{y} - \tilde{\mathbf{y}}) + \tilde{\mathbf{y}}$, and since $(\mathbf{y} - \tilde{\mathbf{y}})$ is orthogonal to the columns of A, $A^T(\mathbf{y} - \tilde{\mathbf{y}}) = \mathbf{0}$. Therefore, $A^T\mathbf{y} = A^T\tilde{\mathbf{y}}$. So

$$(A^T A)^{-1} A^T \mathbf{y} = (A^T A)^{-1} A^T \tilde{\mathbf{y}} = (A^T A)^{-1} A^T A\mathbf{x}_0 = \mathbf{x}_0 \ .$$

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

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

SOLUTION Forming A and $\mathbf{y}$ as above, we compute $A^t A = \begin{bmatrix} 71 & 22 \\ 22 & 8 \end{bmatrix}$ and $A^t\mathbf{y} = (68.46, 19.74)$. Then the least-squares solution (also least length since it is unique) is

$$(a, b) = (A^T A)^{-1} A^T \mathbf{y} = (1.350. - 1.245) \ .$$

So the best fit line is $y = (1.350)x - 1.245$.