(4.1) Least Squares and the Normal Equations

Saturday, December 9, 2017 1:33 PM

References:

- Numerical Analysis 2nd Edition
  - Timothy Sauer

Notes composed by: Christ-Brian Amedjonekou

Personal Statement: These notes are based off the above reference and is part of a series of notes that will cover topics in Numerical Analysis. My goal is to be as straight to the point as I can be with these notes and hopefully set them up in such a way that they act as a supplement of the above references as well as the condensed version of said references. It will include a brief overview to the topic followed by the rules of said topic.

Topics:

- Inconsistent systems of equations
- Fitting models for Data
- Conditioning of Least Squares

Overview: In this section, we'll cover Least Squares and The Normal Equations. In the previous section we solve the linear system when it's consistent but when the system is inconsistent we won’t be able to come up with a solution so the next best thing is finding the least squares approximation.

1. **Inconsistent systems of equations**

- The **least square** method is a product of two problems.

  1. An inconsistent system. This system of equations has no solutions. The cause of this is most often having more equations than variables. As a result we cannot find a solution that will satisfy all equations within the system.

  2. Fitting data with a high degree polynomial (covered in Chapter 3) becomes undesirable due to the high volume of data points or the data point themselves have some sort of margin of error.

  - The **least square** approach is more ideal in these cases because the goal then becomes using a simpler model that can approximated the data points.

- Say you have an **inconsistent system**, for example the following:

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & -1 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

$$Ax = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = b$$

  - The above can be represented as a vector equation which is shown below

$$Ax = [x_1] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + [x_2] \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = b$$

  - Which can represent "b" as a linear combination of the other two vectors.

$$x_1 v_1 + x_2 v_2 + \ldots + x_k v_k = b$$

- But we know that there are no weights that satisfies the above .

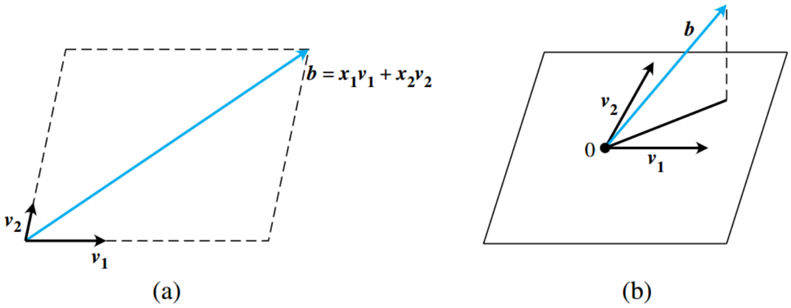

Figure 4.1 Geometric solution of a system of three equations in two unknowns. (a) Equation (4.3) requires that the vector b, the right-hand side of the equation, is a linear combination of the columns vectors v_1 and v_2. (b) If b lies outside of the plane defined by v_1 and v_2, there will be no solution. The least squares solution $\overline{x}$ makes the combination vector $A\overline{x}$ the one in the plane Ax that is nearest to b in the sense of Euclidean distance.

- As "b" does not lie in the plane defined by the vectors "v_1 " and "v_2 "

- You won't be able to find the exact answer since there isn't one. But we still don't want to come up empty handed. The next best thing is to get an approximation, or as close as we could get to the vector which would have satisfied (acted as the solution) the matrix equation above if there was one.
  - This approximated vector is denoted by the following:

$$\overline{x} = x_{bar} = approximate\ solution\ vector = weights$$

  - The idea for the **least squares method**, as aforementioned, is to find an **approximate vector, $\overline{x}$**, that will get us **close to the the vector b**. In other words we want to **minimize the vector b**. The following demostrates this conscept:

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

$$b - A\overline{x}$$

- The above is the **residual vector** which is perpendicular to the **plane of Ax** denoted by $\{\boldsymbol{Ax}|\boldsymbol{x}\epsilon\boldsymbol{R^n}\}$, which is denoted by the following:

$$b - A\overline{x} \perp \{\boldsymbol{Ax} \mid \boldsymbol{x}\ \epsilon\ \boldsymbol{R^n}\}$$

- Remember that if two vectors are orthogonal then their dot product = 0.
- With that be said, the following represents the fact above:

$$(Ax)^T * (b - A\overline{x}) = 0,\ \ for\ all\ x\ \ \epsilon\ R^n$$

$$x^T * A^T * (b - A\overline{x}) = 0,\ \ for\ all\ x\ \ \epsilon\ R^n$$

- The result is that $A^T * (b - A\overline{x})$ is perpendicular to all $x\ \ \epsilon\ R^n$, therefore the
following can be concluded:

$$A^T * (b - A\overline{x}) = 0$$

- Which gives us the system of equations that defines the least squares solution (approximation) denoted by the following:

$$A^T * (b - A\overline{x}) = 0$$

$$A^T b - A^T A\bar{x} = 0$$

$$A^T A\bar{x} = A^T b$$

- The following is an example of solving for $\bar{x}$ using the previously shown inconsistent system.

$$Ax = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = b$$

$$A^T A\bar{x} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

$$A^T A\bar{x} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix} = A^T b$$

$$\begin{bmatrix} 3 & 1 & 6 \\ 1 & 3 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & \frac{7}{4} \\ 0 & 1 & \frac{3}{4} \end{bmatrix}$$

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{7}{4} \\ \frac{3}{4} \end{bmatrix}$$

- Now that we found the least square solution, we want to find the Euclidean Length of the residual vector ($r = b - A\bar{x}$) that our solution minimizes. The following demostrates that process:

$$r = b - A\bar{x}, \quad where\ r = residual\ vector$$

$$r = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{7}{4} \\ \frac{3}{4} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}$$

$$r = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}$$

- Now that we have the residual vector, we'll find the **Euclidean Length**.
- The **Euclidean Length** is one way to express the size of the residual vector.
- The **Euclidean Length** is also called the **2-norm** demonstrate by the following:

$$||r||_2 = \sqrt{r_1^2 + r_2^2 + \ldots + r_m^2}$$

- For the above example the following is the Euclidean Length.

$$||r||_2 = \sqrt{r_1^2 + r_2^2 + r_3^2}$$

$$||r||_2 = \sqrt{\left(-\frac{1}{2}\right)^2 + 0^2 + \left(\frac{1}{2}\right)^2} = .7071$$

- The **square error** is another way to express the size of the residual and is demonstrated by the following:

$$SE = r_1^2 + r_2^2 + \ldots + r_m^2$$

  - For the above example the following is the Euclidean Length.

$$SE = r_1^2 + r_2^2 + \ldots + r_m^2$$

$$SE = \left(-\frac{1}{2}\right)^2 + 0^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2} = .5$$

- The **root mean square error** is the final way to the size of the residual and is
demonstrated by the following:

$$RMSE = \sqrt{\frac{SE}{m}} = \frac{||r||_2}{\sqrt{m}} = r_1^2 + r_2^2 + \ldots + r_m^2$$

$$||r||_2 = \sqrt{\left(-\frac{1}{2}\right)^2 + 0^2 + \left(\frac{1}{2}\right)^2} = .7071$$

$$RMSE = \sqrt{\frac{SE}{m}} = \frac{||r||_2}{\sqrt{m}} = \frac{.7071}{\sqrt{3}} = .4082$$

2. **Fitting models for Data**

- Let's say we are given a set of points $(t_1, y_1) \ldots (t_m, y_m)$ and a model, represented by the lines, $y = c_1 + c_2 t$. **Least Squares Approximation** will allow us to draw the ***best fit*** line where the **square error** is the smallest possible amount it can be for all lines, $y = c_1 + c_2 t$.
  - All we have to do here is place all points in the model represented by the lines, $y = c_1 + c_2 t$,
shown by the following:

$$Given\ Points\ (t,\ y) : (1,2),\ (-1,\ 1),\ (1,\ 3)$$

$$\begin{aligned} Lines: \quad & c_1 + c_2(1) = 2 \\ & c_1 + c_2(-1) = 1 \\ & c_1 + c_2(1) = 3 \end{aligned}$$

  - Then we create a matrix, shown by the following:

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & -1 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

$$Ax = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = b$$

  - We know this matrix has no solution, so we use the least squares method shown by the following:

$$Ax = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = b$$

$$A^T A\overline{x} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

$$A^T A\overline{x} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix} = A^T b$$

$$\begin{bmatrix} 3 & 1 & 6 \\ 1 & 3 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & \frac{7}{4} \\ 0 & 1 & \frac{3}{4} \end{bmatrix}$$

$$\overline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{7}{4} \\ \frac{3}{4} \end{bmatrix}$$

- Now that we found the least square solution, we want to find the Euclidean Length of the residual vector $(r = b - A\overline{x})$ that our solution minimizes. The following demostrates that process:

$$r = b - A\overline{x}, \quad where\ r = residual\ vector$$

$$r = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{7}{4} \\ \frac{3}{4} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}$$

$$r = \begin{bmatrix} -\frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}$$

  - Now that we have the residual vector, we'll find the **Euclidean Length**.
  - The **Euclidean Length** is one way to express the size of the residual vector.
  - The **Euclidean Length** is also called the **2-norm** demonstrate by the following:

$$||r||_2 = \sqrt{r_1^2 + r_2^2 + \ldots + r_m^2}$$

    - For the above example the following is the Euclidean Length.

$$||r||_2 = \sqrt{r_1^2 + r_2^2 + r_3^2}$$

$$||r||_2 = \sqrt{\left(-\frac{1}{2}\right)^2 + 0^2 + \left(\frac{1}{2}\right)^2} = .7071$$

  - The **square error** is another way to express the size of the residual and is demonstrated by the following:

$$SE = r_1^2 + r_2^2 + \ldots + r_m^2$$

    - For the above example the following is the Euclidean Length.

$$SE = r_1^2 + r_2^2 + \ldots + r_m^2$$

$$SE = \left(-\frac{1}{2}\right)^2 + 0^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2} = .5$$

- The **root mean square error** is the final way to the size of the residual and is
demonstrated by the following:

$$RMSE = \sqrt{\frac{SE}{m}} = \frac{||r||_2}{\sqrt{m}} = r_1^2 + r_2^2 + \ldots + r_m^2$$

$$||r||_2 = \sqrt{\left(-\frac{1}{2}\right)^2 + 0^2 + \left(\frac{1}{2}\right)^2} = .7071$$

$$RMSE = \sqrt{\frac{SE}{m}} = \frac{||r||_2}{\sqrt{m}} = \frac{.7071}{\sqrt{3}} = .4082$$

- After finding the least squares approximation, the square errors, $2 - norm$, and root mean square errors we can we have the following:

t	y	line	error
1	2	2.5	-1/2
-1	1	1.0	0
1	3	2.5	1/2

- The values under "line" actually is the vector "$A\overline{x} = \overline{b}$"

- The above was an example of fitting models for data. The General Case is the following:

Fitting data by least squares

Given a set of m data points $(t_1, y_1), \ldots, (t_m, y_m)$.

STEP 1. Choose a model. Identify a parameterized model, such as $y = c_1 + c_2 t$, which will be used to fit the data.

STEP 2. Force the model to fit the data. Substitute the data points into the model. Each data point creates an equation whose unknowns are the parameters, such as c_1 and c_2 in the line model. This results in a system $Ax = b$, where the unknown x represents the unknown parameters.

STEP 3. Solve the normal equations. The least squares solution for the parameters will be found as the solution to the system of normal equations $A^T Ax = A^T b$.

3. Conditioning of Least Squares

Solving the normal equations in double precision cannot deliver an accurate value for the least squares solution. The condition number of $A^T A$ is too large to deal with in double precision arithmetic, and the normal equations are ill-conditioned, even though the original least squares problem is moderately conditioned. There is clearly room for improvement in the normal equations approach to least squares. In Example 4.15, we revisit this problem after developing an alternative that avoids forming $A^T A$. ◄