Plane Transformations & Least Squares

Outcomes In today's lecture we will learn how to:

- Induce transformations through matrix multiplication.
- Calculate least squares solutions for inconsistent systems of linear equations.
- Determine best fit linear and polynomial functions for given data points.

Contents

- Scaling, reflections and rotations of vectors in the plane.
- Least squares solutions for approximating overdetermined systems.
- Pseudoinverse.
- Normal system of equations.
- Least squares line.
- Least squares polynomials.

Exercises

1. Find the 2×2 matrix which will rotate vectors in $\mathbb{R}^2$ by an angle θ in the *clockwise* direction.

2. By using the pseudoinverse, find a least squares solution for the following systems $A\boldsymbol{x} = \boldsymbol{b}$.

(a) $A = \begin{bmatrix} 3 & 1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ (b) $A = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 0 & -1 \\ -1 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 1 \\ 2 \\ -1 \end{bmatrix}$

3. Find a least squares solution for the following systems by constructing and then solving the normal equations using Gaussian Elimination.

(a) $\begin{aligned} x_1 - 2x_2 &= 4 \\ -3x_2 &= 1 \\ 2x_1 + 5x_2 &= -2 \\ 3x_1 &= 4 \end{aligned}$ (b) $\begin{aligned} x_1 + x_2 - x_3 &= 2 \\ -x_2 + 2x_3 &= 6 \\ 3x_1 + 2x_2 - x_3 &= 11 \\ -x_1 + x_3 &= 0 \end{aligned}$

4. Find the least squares approximating line for the given data points.

   (a) $(2,1),(3,2),(4,3),(5,2)$ (b) $(1,1),(2,3),(3,4),(4,5),(5,7)$

5. Find the least squares approximating quadratic for the data points $(-2,4)$, $(-1,7)$, $(0,3)$, $(1,0)$, $(2,-1)$.

These exercises should take around 90 minutes to complete.

(Answers: 1. $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$; 2.(a) $\hat{\boldsymbol{x}} = \begin{bmatrix} \frac{1}{5} \\ \frac{7}{15} \end{bmatrix}$ (b) $\hat{\boldsymbol{x}} = \begin{bmatrix} \frac{24}{17} \\ -\frac{8}{17} \end{bmatrix}$; 3.(a) $\hat{\boldsymbol{x}} = \begin{bmatrix} \frac{4}{3} \\ -\frac{5}{6} \end{bmatrix}$ (b) $\hat{\boldsymbol{x}} = \begin{bmatrix} \frac{42}{11} \\ \frac{19}{11} \\ \frac{42}{11} \end{bmatrix}$; 4.(a) $y = 0.6 + 0.4x$ (b) $y = -0.2 + 1.4x$; 5. $y = \frac{18}{5} - \frac{17}{10}x - \frac{1}{2}x^2$)

Plane Transformations

We'll now explore the geometrical interpretation of multiplication by a square matrix by looking at transformations in the Cartesian plane.

Consider a general 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Clearly, A can only multiply vectors in $\mathbb{R}^2$ and the result of such a multiplication is another vector in $\mathbb{R}^2$. Note that we can define a *function* $T(\boldsymbol{x}) = A\boldsymbol{x}$ in this manner, *i.e.*

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix}.$$

We say that T maps from $\mathbb{R}^2$ to $\mathbb{R}^2$ and it is actually an example of a *linear transformation.* Let us consider some simple linear transformations in the plane.

Notice that $A = I$ is not particularly interesting, as $A\boldsymbol{x} = I\boldsymbol{x} = \boldsymbol{x}$, *i.e.* the vector $\boldsymbol{x}$ remains untouched. However, consider what happens if we change the diagonal entries. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. Then

$$A\boldsymbol{x} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = 2\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

i.e. multiplication by A results in a *stretch* of the vector $\boldsymbol{x}$ by a factor of 2. Similarly, a matrix of the form $A = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$ would result in stretch by a factor k, assuming $k > 1$. (The effect in the case of $0 < k < 1$ is called a *compression.*)

Reflections can also be easily simulated with matrix multiplication. Consider $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Then,

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix},$$

i.e. the resulting vector (or point, if you like) is a *reflection* of the original vector in the x_1 axis. Similarly, $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ yields

$$A\boldsymbol{x} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix},$$

i.e. a reflection in the line $x_2 = x_1$!

Finally, rotations can be induced by the general *rotation matrix*,

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Let $\theta = \frac{\pi}{4}$ and $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then

$$A\boldsymbol{x} = \begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos\frac{\pi}{4} \\ \sin\frac{\pi}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Note that the resulting vector still has length 1, but it is has been rotated by $\theta = 45°$ in the *anti-clockwise* direction. Similarly, for a general θ, this matrix will induce an anti-clockwise rotation of angle θ.

The other type of operation we can achieve with a matrix multiplication is a *shear* in the x_1 and x_2 directions. Anyhow, it be can be easily shown that *multiplication by an invertible* 2×2 *matrix is equivalent to a sequence of shears, compressions, stretches and reflections*. A similar statement can be made for multiplication by an invertible $n \times n$ matrix, except that the geometry is a little more complicated since we are now dealing with vectors in $\mathbb{R}^n$.

Least Squares Approximation

The least squares technique is a method that can be used to approximate solutions to inconsistent systems of linear equations as well to approximate mathematical relationships between variables.

Overdetermined & Inconsistent Systems of Linear Equations

As we have seen from earlier lectures, systems of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ can have a unique solution, infinite solutions or no solutions. When a system of equations does not have a solution we refer to this as being inconsistent, this arises when $r(A) \neq r([A|\boldsymbol{b}])$. When modeling real world problems using a system of linear equations inaccuracies are often introduced from measurements and this in turns leads to an inconsistent system. In situations such as this where no solution exists but a solution is still required, the idea is to then find an $\boldsymbol{x}$ that makes $A\boldsymbol{x}$ as close as possible to $\boldsymbol{b}$, that is we search for the best possible approximate solution which is called a least squares solution.

Consider the system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ where A is an $m \times n$ matrix, where $m > n$ (*i.e.* we have more equations than unknown variables) and $r(A) = n$. A system of linear equations where there are more equations than variables is referred to as an overdetermined system of linear equations. In overdetermined systems we typically expect to find the system to be inconsistent.

Ex: Show that the following overdetermined system of linear equations is inconsistent.

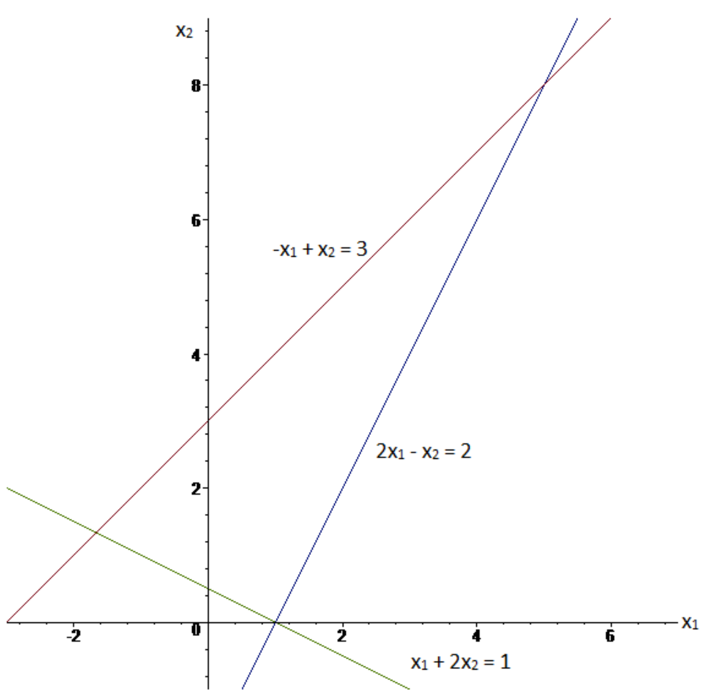
$$\begin{aligned} x_1 + 2x_2 &= 1 \\ 2x_1 - x_2 &= 2 \\ -x_1 + x_2 &= 3 \end{aligned}$$

Soln: This system is overdetermined as the there are $m = 3$ equations but only $n = 2$ unknowns (*i.e.* $m > n$). By using the Gaussian Elimination method in attempting to solve the system we get,

$$[A|\,\boldsymbol{b}] = \left[\begin{array}{cc|c} 1 & 2 & 1 \\ 2 & -1 & 2 \\ -1 & 1 & 3 \end{array}\right] \begin{array}{l} \\ \mathrm{R}_2 \to \mathrm{R}_2 - 2\mathrm{R}_1 \\ \mathrm{R}_3 \to \mathrm{R}_3 + \mathrm{R}_1 \end{array} \sim \left[\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & -5 & 0 \\ 0 & 3 & 4 \end{array}\right] \; \mathrm{R}_3 \to 5\mathrm{R}_3 + 3\mathrm{R}_2$$

$$\sim \left[\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & -5 & 0 \\ 0 & 0 & 20 \end{array}\right].$$

Clearly, $\mathrm{r}(A) = 2$ and $\mathrm{r}([A|\,\boldsymbol{b}]) = 3 \neq \mathrm{r}(A)$. Hence, the equations are inconsistent and there is no solution. This lack of solution can also be seen when the three equations are plotted in the Cartesian plane.



As can be seen in the plot above the three lines do not have a common point (x_1, x_2) of intersection, hence the lack of solution for the system.

To approximate the solution to an inconsistent system $A\boldsymbol{x} = \boldsymbol{b}$ we make use of the pseudoinverse of A (also called the generalized inverse or Moore-Penrose inverse). The pseudoinverse of A is denoted by $\text{pinv}(A)$, and defined as

$$\text{pinv}(A) = (A^T A)^{-1} A^T$$

Again let $A\boldsymbol{x} = \boldsymbol{b}$ be a system of m linear equations in n unknown variables. Since A is of order $m \times n$, consequently A^T will be an $n \times m$ matrix and $A^T A$ is an $n \times n$ square matrix. Multiplying each side of $A\boldsymbol{x} = \boldsymbol{b}$ by A^T leads to the normal system of equations given by

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

The solution to these normal equations is what is referred to as the least squares solutions of the system $A\boldsymbol{x} = \boldsymbol{b}$. If the system $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent, then these least squares solutions are taken as the approximate solution to $A\boldsymbol{x} = \boldsymbol{b}$. The normal system of equations can be regarded as being a system of linear equations with an $n \times n$ coefficient matrix $(A^T A)$. If $r(A) = n$ then the square matrix $(A^T A)$ is invertible, which leads to a unique least squares solution, denoted by $\hat{\boldsymbol{x}}$, to the normal equations. That is if we multiple both sides of the normal equations by the inverse $(A^T A)^{-1}$ we get

$$\begin{aligned}
(A^T A)^{-1}(A^T A)\hat{\boldsymbol{x}} &= (A^T A)^{-1} A^T\boldsymbol{b} \\
I\hat{\boldsymbol{x}} &= (A^T A)^{-1} A^T\boldsymbol{b} \\
\hat{\boldsymbol{x}} &= (A^T A)^{-1} A^T\boldsymbol{b}
\end{aligned}$$

Which leads to,

$$\hat{\boldsymbol{x}} = \text{pinv}(A)\boldsymbol{b}$$

This unique least squares solution $\hat{\boldsymbol{x}} = \text{pinv}(A)\boldsymbol{b}$ is generally taken to be the closet we can get to a true solution, that is it's the best approximate solution to the inconsistent system of equations. In the situation where the system of equations $A\boldsymbol{x} = \boldsymbol{b}$ actually has a unique solution, then the least squares solution will be that unique solution (*i.e.* $\hat{\boldsymbol{x}} = \boldsymbol{x}$).

Ex: Determine the least squares solution for the inconsistent system of linear equations given in the previous example.

Soln: Here we have

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

As previously determined the rank of A is 2, thus the system has a unique least squares solution. In computing $\text{pinv}(A)$ we get

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

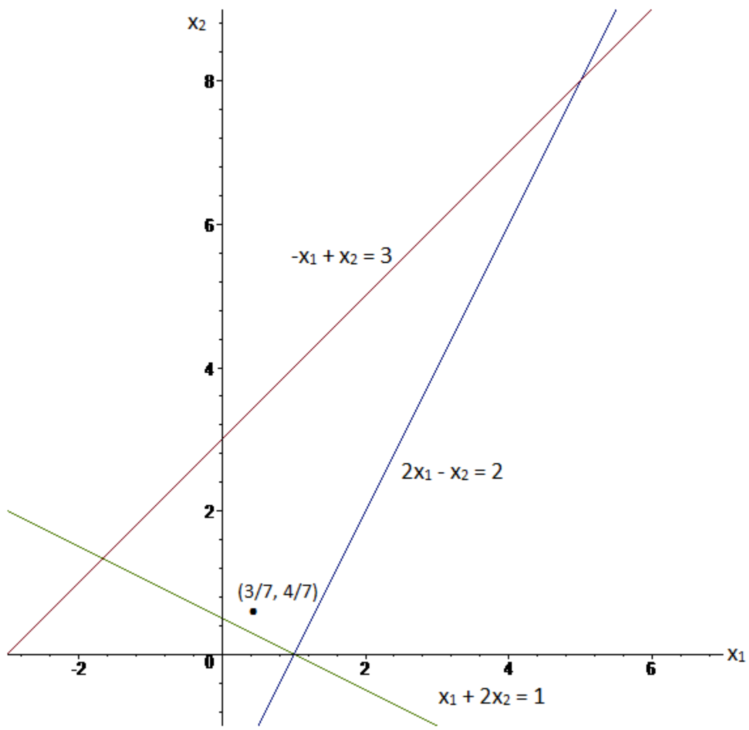
$$(A^TA)^{-1} = \frac{1}{(6)(6)-(-1)(-1)}\begin{bmatrix} 6 & -(-1) \\ -(-1) & 6 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix}$$

$$\text{pinv}(A) = (A^TA)^{-1}A^T = \frac{1}{35}\begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix}\begin{bmatrix} 1 & 2 & -1 \\ 2 & -1 & 1 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 8 & 11 & -5 \\ 13 & -4 & 5 \end{bmatrix}$$

The least squares solution is

$$\hat{\boldsymbol{x}} = \text{pinv}(A)\boldsymbol{b} = \frac{1}{35}\begin{bmatrix} 8 & 11 & -5 \\ 13 & -4 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 15 \\ 20 \end{bmatrix} = \begin{bmatrix} \frac{3}{7} \\ \frac{4}{7} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{x}}_1 \\ \hat{\boldsymbol{x}}_2 \end{bmatrix}$$

Hence the best approximate solution is for $x_1 = \frac{3}{7}$ and $x_2 = \frac{4}{7}$, as plotted below.



Finally, it should be noted that when solving the normal equations $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$ to determine the least squares solution $\hat{\boldsymbol{x}}$, it is computationally more efficient (*i.e.* quicker) to use Gaussian Elimination rather than to determine $(A^TA)^{-1}$ and then take the product $(A^TA)^{-1}A^T\boldsymbol{b}$.

Ex: For the following inconsistent system of linear equations, use Gaussian Elimination to find the least squares solution to the normal equations.

$$\begin{aligned} x_1 + 2x_2 &= 3 \\ x_1 + x_2 &= 1 \\ 2x_1 + 3x_2 &= 3 \end{aligned}$$

<u>Soln:</u> Begin by forming A^TA and $A^T\boldsymbol{b}$ where

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 2 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix}$$

Hence

$$A^TA = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 9 \\ 9 & 14 \end{bmatrix}$$

$$A^T\boldsymbol{b} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 16 \end{bmatrix}$$

Using Gaussian Elimination we set up the augmented matrix $[A^TA|A^T\boldsymbol{b}]$ then reduce it into row echelon form

$$\left[\begin{array}{cc|c} 6 & 9 & 10 \\ 9 & 14 & 16 \end{array}\right] R_2 = 6R_2 - 9R_1 \quad \sim \quad \left[\begin{array}{cc|c} 6 & 9 & 10 \\ 0 & 3 & 6 \end{array}\right]$$

Row 2: $3x_2 = 6 \;\Rightarrow\; x_2 = 2$

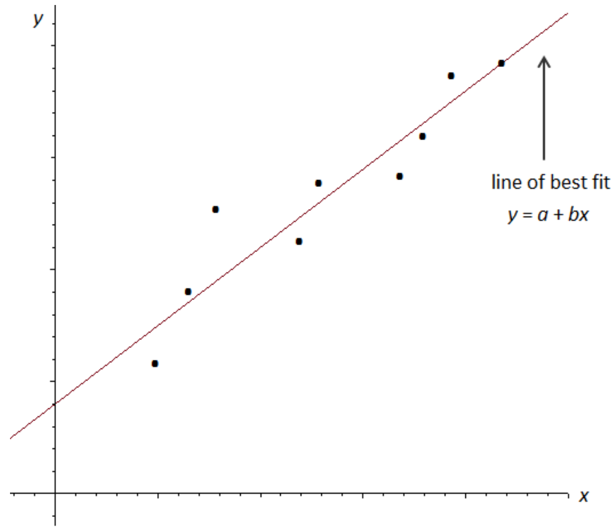
Row 1: $6x_1 + 9x_2 = 10 \;\Rightarrow\; 6x_1 + 9(2) = 10 \;\Rightarrow\; 6x_1 = -8 \;\Rightarrow\; x_1 = -\dfrac{4}{3}$

Hence the best approximate solution is for $x_1 = -\frac{4}{3}$ and $x_2 = 2$.

<u>Least Squares Lines & Curves</u>

A common task in engineering is the gathering of experimental data which is then used to build or verify a mathematical relationship between the variables being measured. This mathematical relationship in turn can be used to predict the value of one variable as a function of other variables. For example, an electrical engineer might want to establish a relationship between the resistance R of a wire at temperature t. In trying to establish this relationship they set the temperature t then measure the corresponding resistance R of the wire leading to a set of data points (t_i, R_i). Or, an mechanical engineer might be interested in the slip S (%) of a vehicle and how it relates to the radius r of a tyre, hence the engineer generates data by varying the radius r_1, r_2,..., r_m and measuring the corresponding slip S_1, S_2,...,S_m, leading to the data points (r_i, S_i). The least squares problem plays an important role in such mathematical modelling of real world phenomena.

In each of these examples, the data points (x_i, y_i) come from the measurement of an independent variable x_i and a dependent variable y_i. Using these data points we then try to develop a relationship between the variables x and y that can be used to predict new values of y for given values of x. Often when these experimental data points (x_i, y_i) are plotted they seem to lie close to a line. What we therefore what to determine is the equation of a line $y = a + bx$ that best fits these data points.



The problem is in determining the values of the constants a and b such that the line $y = a + bx$ gives the best approximation of the data. In most situations there is too much data to lead to an exact fit (as seen in the example that follows), or there is experimental errors in the measurements, hence we can't find a value for a and b such that $y_i = a + bx_i$ is true for all the data points (x_i, y_i).

Ex: An experiment is conducted leading to the gathering of the data points $(1, 1)$, $(2, 3)$, $(5, 4)$. Find, if possible, the line $y = a + bx$ that fits the given data points.

Soln: We begin by substituting the given points (x, y) into the equation $y = a+bx$, which leads to the system of linear equations,

$$\begin{aligned} a + b &= 1 \\ a + 2b &= 3 \\ a + 5b &= 4 \end{aligned}$$

Note that this system is overdetermined as the there are $m = 3$ equations but only $n = 2$ unknowns (*i.e.* a and b). By using the Gaussian Elimination method in attempting to solve the system we get,

$$[A|\,\boldsymbol{b}] = \begin{bmatrix} 1 & 1 & | & 1 \\ 1 & 2 & | & 3 \\ 1 & 5 & | & 4 \end{bmatrix} \begin{matrix} \\ \mathrm{R}_2 \to \mathrm{R}_2 - \mathrm{R}_1 \\ \mathrm{R}_3 \to \mathrm{R}_3 - \mathrm{R}_1 \end{matrix} \sim \begin{bmatrix} 1 & 1 & | & 1 \\ 0 & 1 & | & 2 \\ 0 & 4 & | & 3 \end{bmatrix} \mathrm{R}_3 \to \mathrm{R}_3 - 4\mathrm{R}_2$$

$$\sim \begin{bmatrix} 1 & 1 & | & 1 \\ 0 & 1 & | & 2 \\ 0 & 0 & | & -5 \end{bmatrix}.$$

Clearly, $\mathrm{r}(A) = 2 \neq \mathrm{r}([A|\,\boldsymbol{b}]) = 3$. Hence, the equations are inconsistent and there is no solution for a and b. Therefore we can not find an equation $y = a + bx$ that exactly fits the data points.

In practice when carrying out an engineering investigation typically more than two data points will be collected. By making many measurements errors in the physical measurements should roughly cancel each other out. However having multiple (*i.e.* more than two) data points leads to an overdetermined system when trying to fit a linear equation $y = a + bx$, and being overdetermined we would expect the system to be inconsistent, having no actual solution for a and b. It is therefore our task to determine values for a and b that comes as close as possible to satisfying all the data points.

Suppose we are given m data points (x_1, y_1), (x_2, y_2),...,(x_m, y_m), where $m > 2$ and at least two of the x_i are distinct, our objective is to find a linear function $y = a_0 + a_1 x$ (*i.e.* line) that best fits the data points (x_i, y_i). If there didn't exist any errors in the measurements and the data was truly linear, then for some value of a_0 and a_1 we would have,

$$y_i = a_0 + a_1 x_i \quad \text{for} \quad i = 1, 2, ..., m.$$

That is,

$$\begin{aligned} a_0 + a_1 x_1 &= y_1 \\ a_0 + a_1 x_2 &= y_2 \\ a_0 + a_1 x_3 &= y_3 \\ &\vdots \\ a_0 + a_1 x_m &= y_m \end{aligned}$$

which is a system of linear equations and can be written in matrix form as

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix},$$

i.e. $A\boldsymbol{x} = \boldsymbol{b}$, where $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$. As these m linear equations in two unknown variables a_0 and a_1 form an overdetermined system, we expect the system to be inconsistent, that is there will be no solution for $\boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$. Since this linear system $A\boldsymbol{x} = \boldsymbol{b}$ is typically inconsistent we therefore instead find a least squares solution (*i.e.* best approximate solution) to $A\boldsymbol{x} = \boldsymbol{b}$. This can be done using the exact same method as in the last section. Since at least two of the x_i are distinct it means $r(A) = 2$, and as a result $A\boldsymbol{x} = \boldsymbol{b}$ has a unique least square solution $\hat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b} =$

pinv$(A)\boldsymbol{b}$. The least squares solution $\hat{\boldsymbol{x}}$ to the system $A\boldsymbol{x} = \boldsymbol{b}$ gives the coefficients a_0 and a_1 of the linear function $y = a_0 + a_1 x$, we refer to this line as the least squares line.

Ex: For the previous example, find an equation of the least squares line for the given data.

Soln: We begin by setting up the matrix A and $\boldsymbol{b}$,

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$$

In computing pinv(A) we get

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & 30 \end{bmatrix}$$

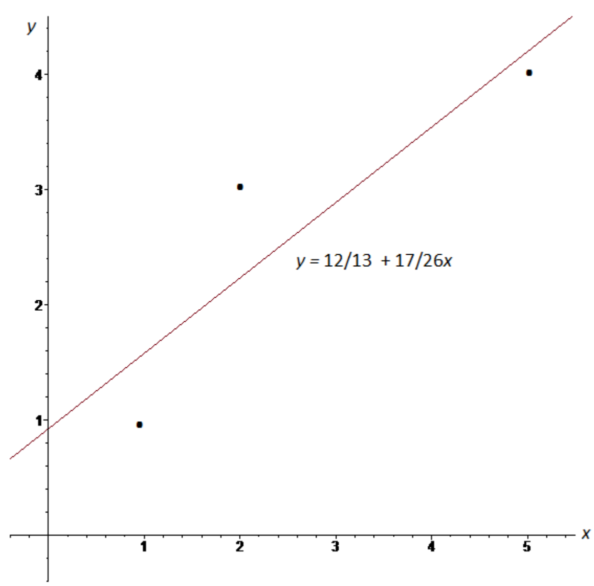
$$(A^T A)^{-1} = \frac{1}{(3)(30) - (8)(8)} \begin{bmatrix} 30 & -8 \\ -8 & 3 \end{bmatrix} = \frac{1}{26} \begin{bmatrix} 30 & -8 \\ -8 & 3 \end{bmatrix}$$

$$\text{pinv}(A) = (A^T A)^{-1} A^T = \frac{1}{26} \begin{bmatrix} 30 & -8 \\ -8 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 5 \end{bmatrix} = \frac{1}{26} \begin{bmatrix} 22 & 14 & -10 \\ -5 & -2 & 7 \end{bmatrix}$$

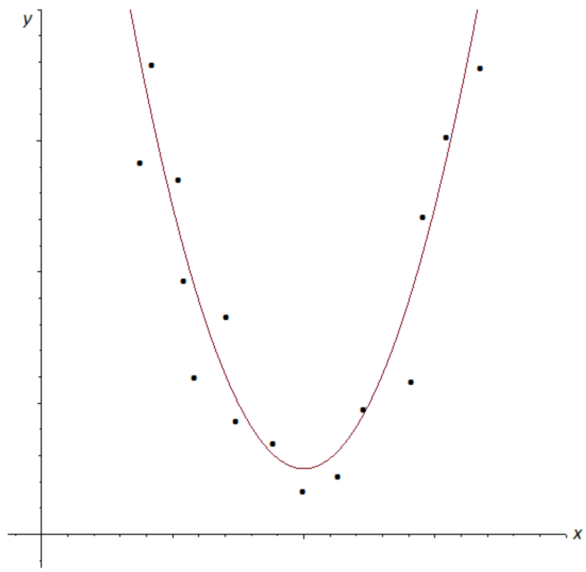
The least squares solution is

$$\hat{\boldsymbol{x}} = \text{pinv}(A)\boldsymbol{b} = \frac{1}{26} \begin{bmatrix} 22 & 14 & -10 \\ -5 & -2 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} = \frac{1}{26} \begin{bmatrix} 24 \\ 17 \end{bmatrix} = \begin{bmatrix} \frac{12}{13} \\ \frac{17}{26} \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

Hence the line that best fits the data is $y = \frac{12}{13} + \frac{17}{26}x$. This line and the data points are shown below.



In some situations a straight line does a poor job in representing engineering data. That is when the data points (x_1, y_1), (x_2, y_2),...,(x_m, y_m) are plotted the points do not lie close to any line. So for example in the plot below it would make more sense to find a parabola rather than a least squares line.



When a straight line will not suffice it will therefore be more appropriate to propose some other functional relationship between x and y, such as a polynomial function of degree n,

$$y = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$$

The previous technique for finding the least squares line for fitting a given set of points can be easily extended to the problem of finding a polynomial of degree n. That is suppose we are given m data points (x_1, y_1), (x_2, y_2),...,(x_m, y_m), where at least $n+1$ of the x_i are distinct (in particular $m \geq n+1$) and we are interested in finding the polynomial of degree n that best fits the given data. By analogy with the preceding least squares line analysis, we define

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ 1 & x_3 & x_3^2 & \cdots & x_3^n \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$$

As the linear system of equations $A\boldsymbol{x} = \boldsymbol{b}$ is overdetermined, since there are more equations m than n variables, we expect the system to be inconsistent. Therefore given that we can't determine a polynomial such that all the data points lie exactly on the graph we instead find the best fitting approximating polynomial. The condition that at least

$n+1$ of the x_i are distinct ensures that $A\boldsymbol{x} = \boldsymbol{b}$ has a unique least square solution $\hat{\boldsymbol{x}} = (A^TA)^{-1}A^T\boldsymbol{b} = \text{pinv}(A)\boldsymbol{b}$. The least squares solution $\hat{\boldsymbol{x}}$ to the system $A\boldsymbol{x} = \boldsymbol{b}$ gives the coefficients a_0, a_1, a_2,..., a_n of the polynomial function $y = a_0 + a_1x + a_2x^2 + ... + a_nx^n$, this polynomial is called the least squares polnomial of degree n.

Ex: Find a quadratic least squares polynomial $y = a_0 + a_1x + a_2x^2$ for the data points $(-2,4)$, $(-1,2)$, $(0,0)$, $(1,1)$, $(2,3)$.

Soln: Instead of solving the overdetermined system $A\boldsymbol{x} = \boldsymbol{b}$ by determining $(A^TA)^{-1}$ and then taking the product $(A^TA)^{-1}A^T\boldsymbol{b}$, we will use the more efficient method of using Gaussian Elimination to solve the normal equations $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$ to determine the least squares solution $\hat{\boldsymbol{x}}$.

Begin by forming A^TA and $A^T\boldsymbol{b}$ where

$$A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 1 \\ 3 \end{bmatrix}$$

Hence

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix}$$

$$A^T\boldsymbol{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ -3 \\ 31 \end{bmatrix}$$

We now set up the augmented matrix $[A^TA|A^T\boldsymbol{b}]$ then reduce it into row echelon form

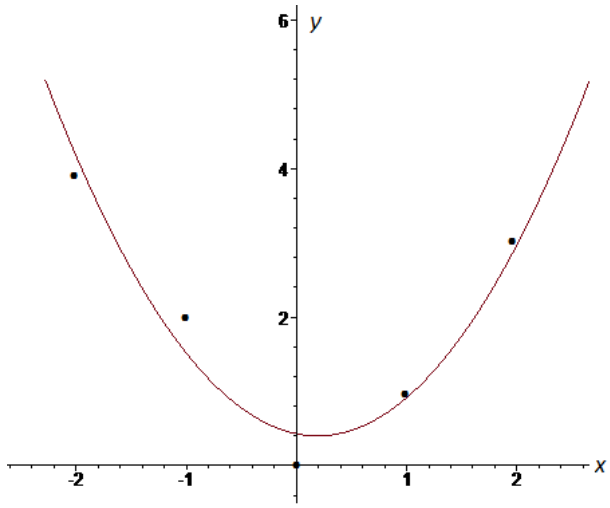
$$\left[\begin{array}{ccc|c} 5 & 0 & 10 & 10 \\ 0 & 10 & 0 & -3 \\ 10 & 0 & 34 & 31 \end{array}\right] \begin{array}{c} \\ \\ R_3 = R_3 - 2R_1 \end{array} \sim \left[\begin{array}{ccc|c} 5 & 0 & 10 & 10 \\ 0 & 10 & 0 & -3 \\ 0 & 0 & 14 & 11 \end{array}\right]$$

Row 3: $14a_2 = 11 \;\Rightarrow\; a_2 = \frac{11}{14}$

Row 2: $10a_1 = -3 \;\Rightarrow\; a_1 = \frac{-3}{10}$

Row 1: $5a_0 + 10a_2 = 10 \;\Rightarrow\; 5a_0 + 10(\frac{11}{14}) = 10 \;\Rightarrow\; 5a_0 = \frac{30}{14} \;\Rightarrow\; a_0 = \frac{3}{7}$

This means that the least squares approximating quadratic for the data points is $y = \frac{3}{7} - \frac{3}{10}x + \frac{11}{14}x^2$. The graph along with the data points are shown below.



6
4
2
y
-2
-1
1
2
x