Lecture notes by Steven Veshkini (03/19/2015)

Least Squares

In this lecture, we focused on finding a way to approximate the intersection and general curve of data. This is possible using the least squares algorithm, which makes it easy to approximate the solution to overdetermined systems (i.e more equations than unknowns). The main example was with our GPS system; given some number of beacons and the readings of distances between a user and the 3 beacons, how is it possible to determine which distance reading is correct? It turns out that if we add more beacons to get more data, we can make more robust measurements and use the least squares approach to find out the answer.

Let's think about how we would setup a system in which we have more equations (n) than unknowns (m).

$a_{11}x_1 + a_{12}x_2 + ... + a_{1m}x_m = b_1 + e_1$

$a_{21}x_1 + ... + ... + a_{2m}x_m = b_2 + e_2$

$a_{n1}x_1 + ... + ... + a_{nm}x_m = b_n + e_n$

With the above equations, we are trying to approximate a solution to the system with a range for the error. Essentially, we know b (in the above example it would be the distance readings), but we already know that it's going to be a little bit off, so we capture the error with the e term.

We can put the above equations into a matrix and attempt to generate an x such that e is as small as possible. Let's try to do so with in a simple case.

$a_{11}x = b_1 + e_1$
$a_{21}x = b_2 + e_2$

In this case, we know that the b's are noisy, so we try to separate the b's out from the noise.

$\vec{e} = A\vec{x}$ - $\vec{b}$

We want to choose an $\vec{x}$ that minimizes $||\vec{e}||^2$.

In two dimensions, we know that the orthogonal projection is the shortest distance from a point onto a line. This principles generalizes to higher dimensions. (So, in 3D the shortest distance from a point to a plane is the orthogonal projection of the point onto the plane (as you proved in the HW.))

$$\min_x \|\vec{e}\|^2 = \min_x \left(\left\| A\vec{x} - \vec{b} \right\|^2 \right) \tag{1}$$

which is essentially

$$\min_x [(a_{11}x - b_1)^2 + (a_{21}x - b_2)^2]. \tag{2}$$

To calculate the minimum of a function we use the idea of finding critical points, i.e. points where the slope of the function is zero. For this, we take the derivative of a function and set it equal to zero to find the critical point. Then, we can check whether the point corresponds to a maximum or minimum by using the second derivative. A negative second derivative means that it is a minimum.

If we do this, we get the following for the minimum of $\vec{x}$ in the two dimensional case :

$$x = \frac{b_1 a_{11} + b_2 a_{21}}{(a_{11})^2 + (a_{21})^2} = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|^2}$$

Based on this strategy, we can try to generalize this to n dimensions.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{n1} & \dots & \dots & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix} + \vec{e}$$

In order to achieve the same result as with 2 dimensions, we need $\vec{e}$ to be orthogonal to all the columns in the matrix (in order to minimize e).

$$\begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix} = 0$$

This can essentially be simplified to:

$$(A^T)(\vec{e}) = 0$$

Since $\vec{e}$ is just $A\vec{x}$ we can simplify.

$$A^T(A\vec{x}) = 0$$

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

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

$$\vec{x} = (A^T A)^{-1} A^T \vec{b}$$

We have just derived the least squares algorithm.

Application of Least Squares

It turns out Gauss used this technique to predict where certain planets would be in their orbit. A scientist named Piazzi made 19 observations over the period of a month in regards to the orbit of Ceres (can be viewed as equations). Gauss used some of these observations. He also knew the general shape of the orbit of planets due to Kepler's laws of planetary motion. Gauss set up equations like so:

$$\alpha x^2 + \beta xy + \gamma y^2 + \delta x + \varepsilon y = \phi$$

If one divides the whole equation by ϕ, nothing significant happens so we can ignore the denominator and treat the right side of the equation as 1.

$$\alpha x^2 + \beta xy + \gamma y^2 + \delta x + \varepsilon y = 1$$

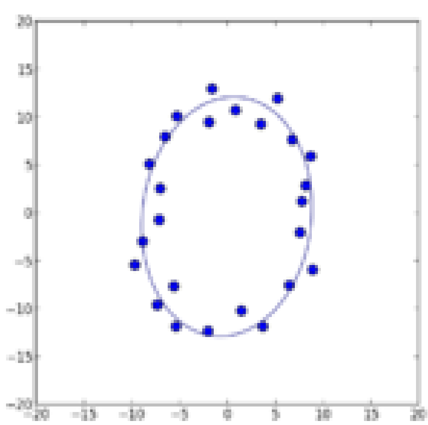
We can set up a matrix like so:

$$\begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_n^2 & \cdots & \cdots & \cdots & y_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 1 \\ \cdots \\ \cdots \\ 1 \end{bmatrix}$$

Using this matrix, we can use the least squares formula.

$$\vec{x} = (A^T A)^{-1} A^T \vec{b}$$

In doing so, a possible result would be the following:



As we can see, the least squares method is useful for fitting known equations to a curve provided that we know the general shape of the curve.