1 Estimation of a Constant Scalar

In this section we will look at how to estimate a constant from a series of nosy measurements of that constant. For example, we might have a temperature sensor attached to an engine and we would want to estimate the temperature of the engine. Each temperature measurement is very noisy since we have only a cheap sensor. So we will take multiple measurements with the sensor to try to get a better estimate of the true temperature.

So lets put this in mathematical terms, Suppose we have a number of measurements y_i where $i \in \{1, \ldots, k\}$ of the unknown quantity x (in this case the engine temperature). Each of the measurements are corrupted by some amount of random, white, uncorrelated noise of zero mean v_i, such that $E(V) = 0$ and $E(v_i v_j^T) = 0$ where $v_i \in V$. Each measurement of the quantity can be expressed as:

$$y_i = x + v_i \tag{1}$$

Now to estimate the true temperature from all the measurements, a reasonable idea would just to average together all the measurements that have been made:

$$\bar{y} = \frac{1}{k} \sum_{i=1}^{k} y_i \tag{2}$$

$$\begin{aligned} \bar{y} &= \frac{1}{k} \sum_{i=1}^{k} (x + v_i) && (3) \\ &= \frac{1}{k}(kx) + \frac{1}{k} \sum_{i=1}^{k} v_i && (4) \\ &= x + \bar{v} && (5) \end{aligned}$$

We know that the mean of all the noise (or the expected value) is equal to zero from the probability distribution:

$$\bar{v} = E(V) = 0 \tag{6}$$

So the best estimate of x which we will denote as $\hat{x}$ is calculated from the mean of all the measurements:

$$\hat{x} = \bar{y} \tag{7}$$

This relies on the fact that if we take enough measurements, the average value of all the noise should be zero mean, so it should all cancel out. The number of measurements that need to be made will depend on the required accuracy of the estimate required and the size of the noise distribution.

2 Linear Least Squares

Lets extend the estimation of a constant scalar in the previous section to a constant vector. Suppose we that we have a measurement noisy measurement y_i and this measurement is now a linear combination of the elements of the vector x to be estimated, such that:

$$\begin{aligned} y_1 &= H_{11}x_1 + \cdots + H_{1n}x_n + v_1 && (8)\\ &\vdots && (9)\\ y_k &= H_{k1}x_1 + \cdots + H_{kn}x_n + v_k && (10)\\ & && (11) \end{aligned}$$

The quantity that we want to estimate is now a n-dimensional vector and we can write the above system of equations in a matrix form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1n} \\ H_{21} & H_{22} & \dots & H_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ H_{k1} & H_{k2} & \dots & H_{kn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix} \tag{12}$$

This equation can be be written as:

$$y = Hx + v \tag{13}$$

We want to calculate the value of x, but we don't know the random values of v, so we want to come up with the best estimate of x that we can with all the information that we have available. Lets denote this best estimate again as $\hat{x}$. The error residual ϵ or difference between the measurements y and the best estimate vector $H\hat{x}$ is:

$$\epsilon = y - H\hat{x} \tag{14}$$

If we can make the vector ϵ as small as possible (in magnitude), then the estimate $\hat{x}$ should be as close to the true value of x that we can get. Lets define a cost function J, this function is a single scalar value that is related to the error residual, if we make this value of J small, then the whole error residual vector ϵ should also be small. Let the cost function J be:

$$\begin{aligned} J &= \epsilon_1^2 + \epsilon_2^2 + \cdots + \epsilon_k^2 && (15)\\ &= \epsilon^T \epsilon && (16) \end{aligned}$$

The cost function J is simply the sum of the squared errors. We can expand the cost function to be:

$$\begin{aligned} J &= \epsilon^T \epsilon && (17)\\ &= (y - H\hat{x})^T (y - H\hat{x}) && (18)\\ &= y^T y - \hat{x}^T H^T y - y^T H\hat{x} + x^T H^T H\hat{x} && (19) \end{aligned}$$

To find the minimum of this equation, we can differentiate it with respect to $\hat{x}$ and set the derivate to zero. This calculates the stationary points of the equation, and we know that the stationary points can be a maximum or minimum points on a curve. Solving the derivative for $\hat{x}$ should then calculate the location or value of $\hat{x}$ that minimizes the function J.

$$\frac{\partial J}{\partial \hat{x}} = -y^T H - y^T H + 2\hat{x}^T H^T H \tag{20}$$

$$= 0 \tag{21}$$

Solving for $\hat{x}$ gives:

$$H^T y = H^T H \hat{x} \tag{22}$$

$$\hat{x} = \left(H^T H\right)^{-1} H^T y \tag{23}$$

This is then the equation for the least squares solution (the equation that minimizes the sum of the squared error). For this equation to be tractable, the matrix H must be full rank and the inverse of $(H^T H)$ must exist. The number of measurements k must be greater than the number of elements n in the vector $\hat{x}$ to estimate.

Lets take another look at our engine temperature problem. This time we would like to calculate how the engine temperature changes with the speed of the engine, i.e. RPMs (revolutions per minute). We take measurements of the engine temperature y_i at different RPMs speeds r_i and we want to find a linear line (of line form $y = ax + b$) of best fit between the temperature and RPM, such that:

$$y_i = x_1 r_i + x_2 + v_i \tag{24}$$

We can do this via the least squares solution of writing the measurements as a linear combination of the vector we want to estimate:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} r_1 & 1 \\ r_2 & 1 \\ \vdots & \vdots \\ r_k & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix} \tag{25}$$

so that the estimate vector $\hat{x}$ is just the parameters of the line of best fit that we want to estimate.