In order to solve the problem, we must create a series of experiments. We are going to assume that we are exploring in one of the experiments where we obtain N observables values y dependent on n, *a priori*, linearly independent variables, $\{x_{jk}\}$, and need to find the most likely linear fit for it. Assuming,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & x_{2n} \\ x_{N1} & x_{N2} & x_{N3} & \dots & x_{Nn} \end{bmatrix}, \tag{1}$$

$$\mathbf{Y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \tag{2}$$

with,

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix}, \tag{3}$$

The system may be expressed easily by $\mathbf{Y} = \mathbf{X} \cdot \mathbf{a}$, where $\mathbf{a}$ is unknown.

How do we find the best fit? The answer is, we need to find the minimum of the Likelihood function $L(\mathbf{a})$ defined as follows,

$$L(\mathbf{a}) := \frac{1}{2N} \sum_{j=1}^{N} \left[y_j - (a_1 x_{j1} + a_2 x_{j2} + \dots + a_n x_{jn})\right]^2 = \frac{1}{2N} \left(\mathbf{Y} - \mathbf{X} \cdot \mathbf{a}\right)^{\mathrm{T}} \cdot \left(\mathbf{Y} - \mathbf{X} \cdot \mathbf{a}\right). \tag{4}$$

After some algebra, the minimization with respect to the fitting parameters a_1, a_2, etc, is the optimal $\mathbf{a}$ as follows,

$$\mathbf{a}_{OPT} = \left(\mathbf{X} \cdot \mathbf{X}^{\mathrm{T}}\right)^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{Y}. \tag{5}$$

Now we need to find the standard deviation of the y's.

$$\sigma^2 = 2L(\mathbf{a}_{OPT}). \tag{6}$$

Work up to this point, we will continue to upload the second part.