1 Weighted Least Squares Estimation

In the previous solutions, we did not know or include any information about the noise values v_i in the solutions. What happens if we do know some information? What happens if we know some information about how accurate each measurement is or that some measurements may be more accurate than others? How we include that information into the solution, and can we use that information to calculate how accurate our solution is?

Lets extend the example, suppose the vector x to be estimated is a constant n-dimensional vector, and y is a k-dimensional nosy measurement vector that is a linear combination of x via the model matrix H. Each element in the measurement matrix has some additive measurement noise components v_i and that noise has a variance of σ_i^2. The problem can be expressed mathematically as:

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

$$E(v_i^2) = \sigma_i^2 \quad (i = 1, \dots, k) \tag{2}$$

$$E(vv^T) = R = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k^2 \end{bmatrix} \tag{3}$$

Where we want to minimize the cost function J with respect to $\hat{x}$ such that:

$$J = \frac{\epsilon_1^2}{\sigma_1^2} + \frac{\epsilon_2^2}{\sigma_2^2} + \dots + \frac{\epsilon_1^k}{\sigma_k^2} \tag{4}$$

$$= \epsilon^T R^{-1} \epsilon \tag{5}$$

So this new cost function minimizes the sum of the squares of the weighted error residuals (hence the name weighted least squares). Each weighting is based on the expected noise variance, the larger the variance, the small the weight is given to that measurement. It is only the relative weightings between the measurements that matter, if all the weights were the same then the solution would be the same as the least squares. The overall magnitude of the wights do not matter, they don't change the minimum solution $\hat{x}$, only the relative size of J at the minimum point.

We can again calculate the least squares solution by differentiating the cost function and finding the minimum point by calculating the solution that sets the derivative to zero:

$$
\begin{aligned}
J &= \epsilon^T R^{-1} \epsilon && (6)\\
&= (y - H\hat{x})^T R^{-1} (y - H\hat{x}) && (7)\\
&= y^T R^{-1} y - \hat{x}^T H^T R^{-1} y - y^T R^{-1} H\hat{x} + x^T H^T R^{-1} H\hat{x} && (8)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial J}{\partial \hat{x}} &= -y^T R^{-1} H + \hat{x}^T H^T R^{-1} H && (9)\\
&= 0 && (10)\\
H^T R^{-1} y &= H^T R^{-1} H\hat{x} && (11)\\
\hat{x} &= (H^T R^{-1} H)^{-1} H^T R^{-1} y && (12)
\end{aligned}
$$

We can also estimate the uncertainty on the estimates using the transformation of uncertainty error propagation. We know that if we have a linear relationship $f = Ax$, then we can transform the x uncertainty covariance matrix Σ_x into the f uncertainty covariance matrix Σ_f via the relationship $\Sigma_f = A\Sigma_x A^T$, so extending that to the weighted least squares solution:

$$\Sigma_{\hat{x}} = \left[(H^T R^{-1} H)^{-1} H^T R^{-1}\right] \Sigma_y \left[(H^T R^{-1} H)^{-1} H^T R^{-1}\right]^T \tag{13}$$

If assume that we have modelled the uncertainty correctly such that $\Sigma_y = R$ then the equation simplifies to:

$$\Sigma_{\hat{x}} = (H^T R^{-1} H)^{-1} \tag{14}$$