Updating Least Squares

The key to sequential updating is the idea of *updating*. New observations come in, and they change our best least squares estimate of the parameters $\boldsymbol{x}$. We want to compute that change—to update the estimate $\hat{\boldsymbol{x}}$. The process will be efficient if we can express the new estimate as *a linear combination of the old estimate* $\hat{\boldsymbol{x}}_{\text{old}}$ *and the new observation* $\boldsymbol{b}_{\text{new}}$:

$$\hat{\boldsymbol{x}}_{\text{new}} = L\hat{\boldsymbol{x}}_{\text{old}} + K\boldsymbol{b}_{\text{new}}. \tag{20.1}$$

This section will make three key points.

The first point is that the process is *recursive*. We do not store the old observations $\boldsymbol{b}_{\text{old}}$! Those measurements were already used in the estimate $\hat{\boldsymbol{x}}_{\text{old}}$. Built into equation (20.1) is the expectation (or hope) that all the information from $\boldsymbol{b}_{\text{old}}$ that we need for $\hat{\boldsymbol{x}}_{\text{new}}$ is available in $\hat{\boldsymbol{x}}_{\text{old}}$. Since $\boldsymbol{x}$ is a much shorter vector than $\boldsymbol{b}$ (which is growing in length with each new measurement) the filter is efficient.

The second point is that the update formula (20.1) can be written (and derived) in many equivalent ways. This makes a lot of expositions of the Kalman filter difficult to follow. The frustrated reader finally just asks for the damn formula. But some of the variations help the intuition, for example by separating the "prediction" from the "correction":

$$\hat{\boldsymbol{x}}_{\text{new}} = \hat{\boldsymbol{x}}_{\text{old}} + K(\boldsymbol{b}_{\text{new}} - A_{\text{new}}\hat{\boldsymbol{x}}_{\text{old}}). \tag{20.2}$$

Direct comparison with (20.1) gives the relation $L = I - KA_{\text{new}}$. The *Kalman gain matrix* K becomes the crucial matrix to identify. It multiplies the mismatch $\boldsymbol{b}_{\text{new}} - A_{\text{new}}\hat{\boldsymbol{x}}_{\text{old}}$ (the *innovation*) between old estimates and new measurements. From the estimate $\hat{\boldsymbol{x}}_{\text{old}}$ we would have predicted measurements $A_{\text{new}}\hat{\boldsymbol{x}}_{\text{old}}$. The difference between this prediction and the actual $\boldsymbol{b}_{\text{new}}$ is multiplied by K to give the correction in $\hat{\boldsymbol{x}}$. That is (20.2).

This gain matrix K involves the statistics of $\boldsymbol{b}_{\text{new}}$ and $\hat{\boldsymbol{x}}_{\text{old}}$, because they tell how much weight to give to this mismatch. (Thus the crucial covariance matrix P must also be updated! This will be our third point.) Variations of the updating formula are introduced for the sake of numerical stability, when this formula is to be applied many times. All the equivalent forms must be related by matrix identities and we will try to make those clear.

The third key point is that the reliability of $\hat{\boldsymbol{x}}$ is a crucial part of the output. This is the *error covariance matrix* $P = \Sigma_{\hat{\boldsymbol{x}}}$ for the estimate. P tells us the statistical properties of $\hat{\boldsymbol{x}}$ based on the statistical properties of $\boldsymbol{b}$. This matrix does not depend on the particular measurements $\boldsymbol{b}$ or measurement errors $\boldsymbol{e}$. Those are only random samples from a whole population of possible measurements and errors. We assume a Gaussian (normal) distribution of this error population, with a known covariance matrix Σ_e. (Well, barely known. We may struggle to find a realistic Σ_e.) The least-squares solution $\hat{\boldsymbol{x}}$ to a static equation $A\boldsymbol{x} = \boldsymbol{b}$ is weighted by Σ_e^{-1}, and $P = (A^{\mathrm{T}}\Sigma_e^{-1}A)^{-1}$ is the error covariance matrix for $\hat{\boldsymbol{x}}$. It is this covariance that we update when measurements $\boldsymbol{b}_{\text{new}}$ arrive with $\Sigma_{e,\text{new}}$:

$$\begin{aligned} P_{\text{new}} &= (I - KA)P_{\text{old}}(I - KA)^{\mathrm{T}} + K\Sigma_{e,\text{new}}K^{\mathrm{T}} \\ &= K(AP_{\text{old}}A^{\mathrm{T}} + \Sigma_{e,\text{new}})K^{\mathrm{T}} - KAP_{\text{old}} - P_{\text{old}}A^{\mathrm{T}}K^{\mathrm{T}} + P_{\text{old}}. \end{aligned} \tag{20.3}$$

To obtain this formula from (20.1) with $L = I - KA$ *we assume that the errors* $\boldsymbol{e}_{\text{new}}$ *in* $\boldsymbol{b}_{\text{new}}$ *are statistically independent from the errors* $\boldsymbol{e}_{\text{old}}$:

$$\text{The covariance matrix for} \quad \boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_{\text{old}} \\ \boldsymbol{b}_{\text{new}} \end{bmatrix} \quad \text{is} \quad \Sigma_e = \begin{bmatrix} \Sigma_{e,\text{old}} & 0 \\ 0 & \Sigma_{e,\text{new}} \end{bmatrix}.$$

Now we want to determine K so that P_{new} becomes as small as possible. A necessary condition is that $dP_{\text{new}}/dK = 0$:

$$2K(AP_{\text{old}}A^{\text{T}} + \Sigma_{e,\text{new}}) - 2P_{\text{old}}A^{\text{T}} = 0$$

or

$$K = P_{\text{old}}A^{\text{T}}(AP_{\text{old}}A^{\text{T}} + \Sigma_{e,\text{new}})^{-1}. \tag{20.4}$$

We introduce the notation $Q = AP_{\text{old}}A^{\text{T}} + \Sigma_{e,\text{new}}$. From (20.4) we get $KQ = P_{\text{old}}A^{\text{T}}$ or $AP_{\text{old}} = QK^{\text{T}}$. We insert into (20.3):

$$P_{\text{new}} = KQK^{\text{T}} - KQK^{\text{T}} - KAP_{\text{old}} + P_{\text{old}} = (I - KA)P_{\text{old}}. \tag{20.5}$$

By this we have found all formulas for the recursive least-squares update.

In order to start the recursive least squares procdure we need estimates for the initial weight matrix and an initial solution vector. The M-file rls shows how to obtain these estimates. We restrict ourselves to quote the core code:

```
for i = 1:size(b,1)
  K = P * A(i,:)' * inv(A(i,:) * P * A(i,:)' + Sigma(i,i));
  P = (eye(size(A,2)) - K * A(i,:)) * P;
  x = x + K * (b(i) - A(i,:) * x);
end
```