Kalman Filters and Least Squares

Dávid Iván

December 1, 2021

Contents

1 Weighted Least Squares

We have a linear system to solve:

$$Ax = b \tag{1}$$

Where $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. When A has more rows than columns ($m > n$), we don't (necessarily) have an exact solution. This is where we use least squares to get an approximate solution. x_E is the solution to the Euclidean distance minimization:

$$\begin{aligned} x_E &= \text{argmin}_x ||Ax - b||^2 \\ &= \text{argmin}_x \left[(Ax - b)^T (Ax - b)\right] \\ &= (A^T A)^{-1} A^T b \end{aligned} \tag{2}$$

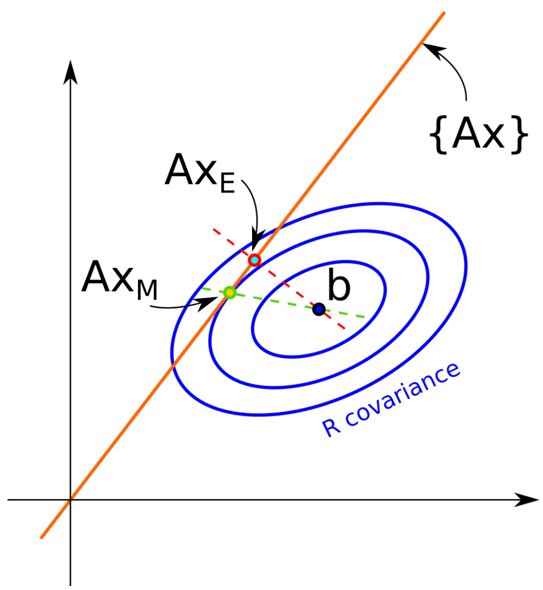


Figure 1: difference between ordinary least squares and weighted least squares solution. We want to be close to b either in terms of Euclidean distance or Mahalanobis distance.

x_M is the solution to the Mahalanobis distance minimization (the weighted least squares problem):

$$
\begin{aligned}
x_M &= \text{argmin}_x ||Ax - b||_R^2 \\
&= \text{argmin}_x \left[(Ax-b)^T R^{-1}(Ax-b)\right] \\
&= (A^T R^{-1} A)^{-1} A^T R^{-1} b
\end{aligned} \tag{3}
$$

This can be justified when the residual is thought of as a Gaussian random variable with zero mean and covariance R:

$$
Ax - b = r \propto N(0, R) \tag{4}
$$

Of course, when R covariance matrix is the identity, then the two solutions are the same: $x_E = x_M$. Figure 1 shows the difference between x_E and x_M.

1.1 Block Least Squares

Solving

$$
\text{argmin}_x \left[||A_1 x - b||_{R_1}^2 + ||A_2 x - b||_{R_2}^2\right] \tag{5}
$$

is equivalent to solving the following "normal" weighted least squares problem.

Let's divide up the matrices A, b and R as shown on Figure 2.

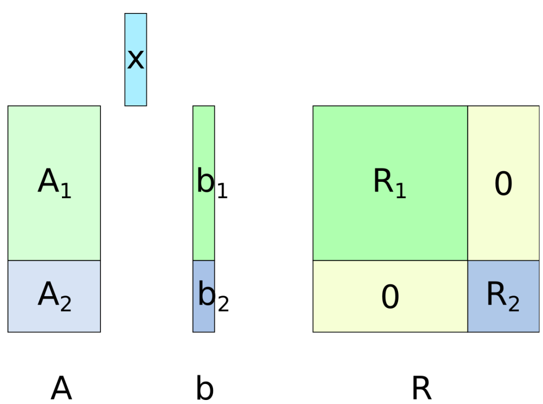


Figure 2: weighted least squares with block matrices

$A_1 \in \mathbb{R}^{m_1 \times n}$, $A_2 \in \mathbb{R}^{m_2 \times n}$, $m_1 + m_2 = m$. $b_1 \in \mathbb{R}^{m_1}$, $b_2 \in \mathbb{R}^{m_2}$. $R_1 \in \mathbb{R}^{m_1 \times m_1}$, $Q_2 \in \mathbb{R}^{m_2 \times m_2}$. R_1 and R_2 are both positive (semi)definite matrices.

The WLS solution is

$$
x^* = (A^T R^{-1} A)^{-1} A^T R^{-1} b \tag{6}
$$

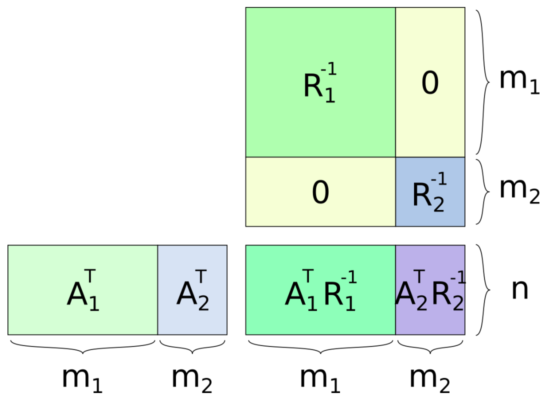


Figure 3: Calculating the matrix $A^T R^{-1}$

Figure 3 shows how to calculate $A^T R^{-1}$

Now the WLS solution can be calculated as follows:

$$x^* = (A_1^T R_1^{-1} A_1 + A_2^T R_2^{-2} A_2)^{-1} (A_1^T R_1^{-1} b_1 + A_2^T R_2^{-2} b_2) \tag{7}$$

Now let's solve the following minimization problem:

$$x_2^* = \operatorname{argmin}_x \left[||A_1 x - b||_{R_1}^2 + ||A_2 x - b||_{R_2}^2 \right] = \operatorname{argmin} \left[L_1(x) + L_2(x) \right] \tag{8}$$

First calculate the derivative:

$$\frac{d}{dx}(L_1(x) + L_2(x)) = 2(A_1 x - b_1)^T R_1^{-1} A_1 + 2(A_2 x - b_2)^T R_2^{-1} A_2 \tag{9}$$

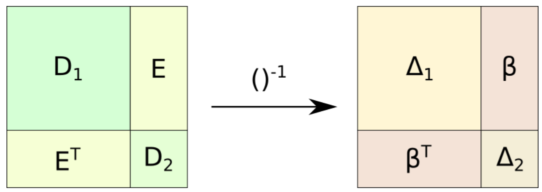
Setting it to zero and solving it:

$$\begin{aligned}
2(A_1 x - b_1)^T R_1^{-1} A_1 + 2(A_2 x - b_2)^T R_2^{-1} A_2 = 0 \\
A_1^T R_1^{-1} (A_1 x - b_1) + A_2^T R_2^{-1} (A_2 x - b_2) = 0 \\
A_1^T R_1^{-1} A_1 x - A_1^T R_1^{-1} b_1 + A_2^T R_2^{-1} A_2 x - A_2^T R_2^{-1} b_2 = 0 \\
A_1^T R_1^{-1} A_1 x + A_2^T R_2^{-1} A_2 x = A_1^T R_1^{-1} b_1 + A_2^T R_2^{-1} b_2 \\
(A_1^T R_1^{-1} A_1 + A_2^T R_2^{-1} A_2) x = A_1^T R_1^{-1} b_1 + A_2^T R_2^{-1} b_2 \\
x_2^* = (A_1^T R_1^{-1} A_1 + A_2^T R_2^{-1} A_2)^{-1} (A_1^T R_1^{-1} b_1 + A_2^T R_2^{-1} b_2)
\end{aligned} \tag{10}$$

Now we see that indeed, $x_2^* = x^*$.

2 Matrix inversion

Let's calculate the following matrix inversion.



Where $D_1 \in \mathbb{R}^{N \times N}$, symmetric, invertible. $D_2 \in \mathbb{R}^{n \times n}$, symmetric, invertible. $E \in \mathbb{R}^{N \times n}$.

It is easy to derive the following expressions:

$$\Delta_2 = (D_2 - E^T D_1^{-1} E)^{-1} \tag{11}$$

$$\begin{aligned} \Delta_1 &= D_1^{-1} + D_1^{-1} E \Delta_2 E^T D_1^{-1} \\ &= (D_1 - E D_2^{-1} E^T)^{-1} \end{aligned} \tag{12}$$

$$\beta = -D_1^{-1} E \Delta_2 \tag{13}$$

Note that if $N >> n$ and we already know D_1^{-1} (calculated earlier), then we might use the first formulae for Δ_1, instead of the second, where we would have to invert a big $(N \times N)$ matrix.