Derivation of the Kalman Filter

Kai Borre

Danish GPS Center, Denmark

Block Matrix Identities

The key formulas give the inverse of a 2 by 2 block matrix, assuming T is invertible:

$$\begin{bmatrix} T & U \\ V & W \end{bmatrix}^{-1} = \begin{bmatrix} L & M \\ N & P \end{bmatrix}. \tag{1}$$

Our applications have symmetric matrices. All blocks on the diagonal are symmetric; the blocks off the diagonal are $U = V^{\mathrm{T}}$ and $M = N^{\mathrm{T}}$. The key to Kalman's success is that the matrix to be inverted is *block tridiagonal*. Thus U and V are nonzero only in their last block entries, and T itself is block tridiagonal. But the general formula gives the inverse without using any special properties of T, U, V, W:

$$\begin{aligned} L &= T^{-1} + T^{-1}UPVT^{-1} \\ M &= -T^{-1}UP \\ N &= -PVT^{-1} \\ P &= (W - VT^{-1}U)^{-1}. \end{aligned} \tag{2}$$

The simplest proof is to multiply matrices and obtain I. The actual derivation of (2) is by block elimination. Multiply the row $[\,T \quad U\,]$ by VT^{-1} and subtract from $[\,V \quad W\,]$:

$$\begin{bmatrix} I & 0 \\ -VT^{-1} & I \end{bmatrix} \begin{bmatrix} T & U \\ V & W \end{bmatrix} = \begin{bmatrix} T & U \\ 0 & W - VT^{-1}U \end{bmatrix}. \tag{3}$$

The two triangular matrices are easily inverted, and then block multiplication produces (2). This is only Gaussian elimination with blocks. In the scalar case the last corner is $W - VU/T$. In the matrix case we keep the blocks in the right order! The inverse of that last entry is P.

Now make a trivial but valuable observation. We could eliminate in the opposite order. This means that we subtract UW^{-1} times the *second* row $[\, V \quad W \,]$ from the *first* row $[\, T \quad U \,]$:

$$\begin{bmatrix} I & -UW^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} T & U \\ V & W \end{bmatrix} = \begin{bmatrix} T - UW^{-1}V & 0 \\ V & W \end{bmatrix}.$$

Inverting this new right side yields different (but still correct) formulas for the blocks L, M, N, P in the inverse matrix.

We pay particular attention to the (1, 1) block. It becomes $L = (T - UW^{-1}V)^{-1}$. That is completely parallel to $P = (W - VT^{-1}U)^{-1}$, just changing letters.

Now compare the new form of L with the form in (2). Their equality is the most important formula in matrix update theory. We only mention four of its originators: **Sherman–Morrison–Woodbury–Schur**:

$$(T - UW^{-1}V)^{-1} = T^{-1} + T^{-1}U(W - VT^{-1}U)^{-1}VT^{-1}. \qquad (4)$$

We are looking at this as an update formula, when the matrix T is *perturbed* by $UW^{-1}V$. Often this is a perturbation of low rank. Previously we looked at (1) as a bordering formula, when the matrix T was *bordered* by U and V and W. Well, matrix theory is beautiful.

Derivation of the Kalman Filter

We will base all steps on the two previous matrix identities. Those identities come from the inverse of a 2 by 2 *block matrix*. The problem is to update the last entries of $(\mathcal{A}^{\mathrm{T}}\Sigma^{-1}\mathcal{A})^{-1}$, when new rows are added to the big matrix $\mathcal{A}$. Those new rows will be of two kinds, coming from the *state equation* and the *observation equation*. A typical filtering step has two updates, from $\mathcal{A}_{k-1}$ to $\mathcal{S}_k$ (by including the state equation) and then to $\mathcal{A}_k$ (by including the observation equation):

$$\mathcal{A}_1 = \begin{bmatrix} A_0 & \\ -F_0 & I \\ & A_1 \end{bmatrix} \rightarrow \quad \mathcal{S}_2 = \begin{bmatrix} A_0 & & \\ -F_0 & I & \\ & A_1 & \\ & -F_1 & I \end{bmatrix}$$

$$\rightarrow \quad \mathcal{A}_2 = \begin{bmatrix} A_0 & & \\ -F_0 & I & \\ & A_1 & \\ & -F_1 & I \\ & & A_2 \end{bmatrix}.$$

The step to $\mathcal{S}_k$ adds a new row and column (this is *bordering*). The second step only adds a new row (this is *updating*). The least-squares solution $\hat{\boldsymbol{x}}_{k-1|k-1}$ for the old state leads to $\hat{\boldsymbol{x}}_{k|k-1}$ (using the state equation) and then to $\hat{\boldsymbol{x}}_{k|k}$ (using the new observation $\boldsymbol{b}_k$).

Remember that least squares works with the symmetric block tridiagonal matrices $\mathcal{T}_k = \mathcal{A}_k^{\mathrm{T}} \Sigma_k^{-1} \mathcal{A}_k$. So we need the change in $\mathcal{T}_{k-1}$ when $\mathcal{A}_{k-1}$ is bordered and then updated:

$$\begin{array}{llll} \text{Bordering by the row} & R = \begin{bmatrix} 0 & \ldots & -F_{k-1} & I \end{bmatrix} & \text{adds} & R^{\mathrm{T}} \Sigma_{\epsilon,k}^{-1} R \\ \text{Updating by the row} & W = \begin{bmatrix} 0 & \ldots & 0 & A_k \end{bmatrix} & \text{adds} & W^{\mathrm{T}} \Sigma_{e,k}^{-1} W. \end{array}$$

We are multiplying matrices using "columns times rows". Every matrix product BA is the sum of (column j of B) times (row j of A). In our case B is the transpose of A, and these are *block* rows and columns. The matrices $\Sigma_{\epsilon,k}^{-1}$ and $\Sigma_{e,k}^{-1}$ appear in the middle, because this is *weighted* least squares.

The complete system, including all errors up to ϵ_k and e_k in the error vector $\mathcal{e}_k$, is $\mathcal{A}_k \mathcal{x}_k = \mathcal{b}_k - \mathcal{e}_k$:

$$\mathcal{A}_2 \mathcal{x}_2 = \begin{bmatrix} A_0 & & \\ -F_0 & I & \\ & A_1 & \\ & -F_1 & I \\ & & A_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_0 \\ \mathbf{0} \\ \boldsymbol{b}_1 \\ \mathbf{0} \\ \boldsymbol{b}_2 \end{bmatrix} - \begin{bmatrix} \boldsymbol{e}_0 \\ \boldsymbol{\epsilon}_1 \\ \boldsymbol{e}_1 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{e}_2 \end{bmatrix} = \mathcal{b}_2 - \mathcal{e}_2. \quad (5)$$

The weight matrix Σ_k^{-1} for the least-squares solution is block diagonal:

$$\Sigma_2^{-1} = \begin{bmatrix} \Sigma_{e,0}^{-1} & & & \\ & \ddots & & \\ & & \Sigma_{\epsilon,2}^{-1} & \\ & & & \Sigma_{e,2}^{-1} \end{bmatrix}.$$

Our task is to compute the ***last corner block*** $P_{k|k}$ *of the matrix* $\left(\mathcal{A}_k^{\mathrm{T}} \Sigma_k^{-1} \mathcal{A}_k\right)^{-1}$. We do that in two steps. The "prediction" finds $P_{k|k-1}$ from the bordering step, when the state row is added to $\mathcal{A}_{k-1}$ (producing $\mathcal{S}_k$). Then the "correction" finds $P_{k|k}$ when the observation row is included too.

The second task is to compute the prediction $\hat{\boldsymbol{x}}_{k|k-1}$ and correction $\hat{\boldsymbol{x}}_{k|k}$ for the new state vector $\boldsymbol{x}_k$. This is the last component of $\hat{x}_k$. If we compute *all* components of $\hat{x}_k$, then we are smoothing old estimates as well as filtering to find the new $\hat{\boldsymbol{x}}_{k|k}$.

Naturally Σ affects all the update formulas. The derivation of these formulas will be simpler if we begin with the case $\Sigma = I$, in which all noise is "white". Then we adjust the formulas to account for the weight matrices.

We compute the last entries in $(\mathcal{A}^{\mathrm{T}}\Sigma^{-1}\mathcal{A})^{-1}$ and then in $(\mathcal{A}^{\mathrm{T}}\Sigma^{-1}\mathcal{A})^{-1}\mathcal{A}^{\mathrm{T}}\Sigma^{-1}\mathcal{b}$. These entries are $P_{k|k}$ and $\hat{\boldsymbol{x}}_{k|k}$. Block matrices will be everywhere.

There is another approach to the same result. Instead of working with $\mathcal{A}^{\mathrm{T}}\mathcal{A}$ (which here is block tridiagonal) we could ***orthogonalize*** the columns of $\mathcal{A}$. This Gram-Schmidt process factors $\mathcal{A}$ into a block orthogonal $\mathcal{Q}$ times a block bidiagonal $\mathcal{R}$. Those factors are updated at each step of the ***square root information filter***. This $\mathcal{Q}\mathcal{R}$ approach adds new insight about "orthogonalizing the innovations".

Updates of the Covariance Matrices

We now compute the blocks $P_{k|k-1}$ and $P_{k|k}$ in the lower right corners of $(\mathcal{S}_k^{\mathrm{T}}\mathcal{S}_k)^{-1}$ and $(\mathcal{A}_k^{\mathrm{T}}\mathcal{A}_k)^{-1}$. Then we operate on the right side, to find the new state $\hat{\boldsymbol{x}}_{k|k}$.

Remember that $\mathcal{S}_k$ comes by adding the new row $[\,V \quad I\,]$ $= [\,0 \quad \dots \quad -F_{k-1} \quad I\,]$. Then $\mathcal{A}_{k-1}^{\mathrm{T}}\mathcal{A}_{k-1}$ grows to $\mathcal{S}_k^{\mathrm{T}}\mathcal{S}_k$ by adding $[\,V \quad I\,]^{\mathrm{T}}[\,V \quad I\,]$:

$$\mathcal{S}_k^{\mathrm{T}}\mathcal{S}_k = \begin{bmatrix} \mathcal{A}_{k-1}^{\mathrm{T}}\mathcal{A}_{k-1} + V^{\mathrm{T}}V & V^{\mathrm{T}} \\ V & I \end{bmatrix}. \tag{6}$$

Because V is zero until its last block, equation (6) is really the addition of a 2 by 2 block in the lower right corner. This has two effects at once. It *perturbs* the existing matrix $T = \mathcal{A}_{k-1}^{\mathrm{T}}\mathcal{A}_{k-1}$ and it *borders* the result T_{new}. The perturbation is by $V^{\mathrm{T}}V$. The bordering is by the row V and the column V^{T} and the corner block I. Therefore the update formula for T_{new}^{-1} goes inside the bordering formula:

$$\begin{array}{lll} \text{Update to} & T + V^{\mathrm{T}}V: & T_{\text{new}}^{-1} = T^{-1} - T^{-1}V^{\mathrm{T}}(I + VT^{-1}V^{\mathrm{T}})^{-1}VT^{-1} \\ \text{Border by} & V, V^{\mathrm{T}}, I: & \qquad P = (I - VT_{\text{new}}^{-1}V^{\mathrm{T}})^{-1}. \end{array}$$

Substitute the update into the bordering formula and write Z for the block

$VT^{-1}V^{\mathrm{T}}$. This gives the great simplification $P = I + Z$:

$$\begin{aligned} P &= \left(I - V\left(T^{-1} - T^{-1}V^{\mathrm{T}}(I+Z)^{-1}VT^{-1}\right)V^{\mathrm{T}}\right)^{-1} \\ &= \left(I - Z + Z(I+Z)^{-1}Z\right)^{-1} = I + Z. \qquad (7) \end{aligned}$$

This block P is $P_{k|k-1}$. It is the corner block in $(\mathcal{S}_k^{\mathrm{T}}\mathcal{S}_k)^{-1}$. The row $[\,V \quad I\,]$ has been included, and the matrix Z is

$$
\begin{aligned}
VT^{-1}V^{\mathrm{T}} = \begin{bmatrix} 0 & \dots & -F_{k-1} \end{bmatrix} & \begin{bmatrix} \cdot & \cdots & \cdot \\ \vdots & \ddots & \vdots \\ \cdot & \cdots & P_{k-1|k-1} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ -F_{k-1}^{\mathrm{T}} \end{bmatrix} \\
&= F_{k-1}P_{k-1|k-1}F_{k-1}^{\mathrm{T}}.
\end{aligned}
$$

Therefore $P = I + Z$ in equation (7) is exactly the Kalman update formula

$$P_{k|k-1} = I + F_{k-1}P_{k-1|k-1}F_{k-1}^{\mathrm{T}}. \tag{8}$$

The identity matrix I will change to $\Sigma_{\epsilon,k}$ when the covariance of the state equation error $\boldsymbol{\epsilon}_k$ is accounted for.

Now comes the second half of the update. The new row $W = [\,0 \quad \ldots \quad 0 \quad A_k\,]$ enters from the observation equations. This row is placed below $\mathcal{S}_k$ to give $\mathcal{A}_k$. Therefore $W^{\mathrm{T}} W$ is added to $\mathcal{S}_k^{\mathrm{T}} \mathcal{S}_k$ to give $\mathcal{A}_k^{\mathrm{T}} \mathcal{A}_k$. We write $\mathcal{Y}$ for the big tridiagonal matrix $\mathcal{S}_k^{\mathrm{T}} \mathcal{S}_k$ before this update, and we use the update formula (4) for the new inverse:

$$(\mathcal{A}_k^{\mathrm{T}} \mathcal{A}_k)^{-1} = (\mathcal{Y} + W^{\mathrm{T}} I W)^{-1} = \mathcal{Y}^{-1} - \mathcal{Y}^{-1} W^{\mathrm{T}} (I + W \mathcal{Y}^{-1} W^{\mathrm{T}})^{-1} W \mathcal{Y}^{-1}. \tag{9}$$

Look at $W \mathcal{Y}^{-1} W^{\mathrm{T}}$. The row $W = [\,0 \quad \ldots \quad 0 \quad A_k\,]$ is zero until the last block. The last corner of $\mathcal{Y}^{-1}$ is $P_{k|k-1}$, found above.

Therefore $W\mathcal{Y}^{-1}W^{\mathrm{T}}$ reduces immediately to $A_k P_{k|k-1} A_k^{\mathrm{T}}$. Similarly the last block of $\mathcal{Y}^{-1}W^{\mathrm{T}}$ is $P_{k|k-1}A_k^{\mathrm{T}}$.

Now concentrate on the *last block row* in equation (9). Factoring out $\mathcal{Y}^{-1}$ on the right, this last row of $(\mathcal{A}_k^{\mathrm{T}}\mathcal{A}_k)^{-1}$ is

$$\begin{aligned}\big(I - P_{k|k-1}A_k^{\mathrm{T}}(I + A_k P_{k|k-1} A_k^{\mathrm{T}})^{-1} A_k\big)&(\text{last row of } \mathcal{Y}^{-1}) \\ &= (I - K_k A_k)\big(\text{last row of } (\mathcal{S}_k^{\mathrm{T}}\mathcal{S}_k)^{-1}\big). \qquad (10)\end{aligned}$$

In particular the final entry of this last row is the corner entry $P_{k|k}$ in $(\mathcal{A}_k^{\mathrm{T}}\mathcal{A}_k)^{-1}$:

$$P_{k|k} = (I - K_k A_k) P_{k|k-1}. \qquad (11)$$

This is the second half of the Kalman filter. It gives the error covariance $P_{k|k}$ from $P_{k|k-1}$. The new P uses the observation equation $\boldsymbol{b}_k = A_k \boldsymbol{x}_k + \boldsymbol{e}_k$, whereas the old P doesn't use it. The matrix K that was introduced to simplify the algebra in (10) is Kalman's *gain matrix*:

$$K_k = P_{k|k-1} A_k^{\mathrm{T}} \big(I + A_k P_{k|k-1} A_k^{\mathrm{T}} \big)^{-1}. \tag{12}$$

This identity matrix I will change to $\Sigma_{e,k}$ when the covariance of the observation error is accounted for. You will see in a later equation for $\hat{\boldsymbol{x}}_{k|k}$ why we concentrated on the whole last row of $(\mathcal{A}_k^{\mathrm{T}} \mathcal{A}_k)^{-1}$, as well as its final entry $P_{k|k}$ in (11).