Kalman Smoothing via Auxiliary Outputs

John Kormylo

When transition matrix Φ is invertable, from equation (34) in the tutorial we see that

$$\hat{\mathbf{x}}(k|N) = \Phi^{-1}\left[\hat{\mathbf{x}}(k+1|N) - \Gamma\hat{\mathbf{u}}(k|N)\right] \tag{1}$$

which can be computed recursively for $k = N, N-1, \ldots, 0$ using

$$\hat{\mathbf{u}}(k|N) = \mathbf{b}(k) + Q\Gamma'\mathbf{r}(k+1) \tag{2}$$

and where $\mathbf{r}(k)$ is also computed recursively during the second pass using equation (33) from the tutorial.

This has the advantage that it does not require storing $\hat{\mathbf{x}}(k|k-1)$ or $P(k|k-1)$ for all $k = 1, 2, \ldots, N$ between the first and second passes. But besides requiring that Φ be invertable, if any of the eigenvalues of Φ are inside the unit circle, then the corresponding eigenvalue in Φ^{-1} will be outside the unit circle and the recursion in (1) will be numerically unstable.

If Φ is of rank $n-1$ (where n is the dimension of the state vector), then there will exists vectors $\mathbf{f}$ and $\mathbf{g}$ such that $[\Phi + \mathbf{g}\mathbf{f}']$ is invertable. The proof becomes obvious if you perform a singular value decomposition of matrix Φ.

Create an auxiliary output of our state vector model of the form

$$y(k) = \mathbf{f}'\mathbf{x}(k) \tag{3}$$

so that the state vector model can be reformulated as

$$\mathbf{x}(k+1) = [\Phi + \mathbf{g}\mathbf{f}']\mathbf{x}(k) + \Gamma\mathbf{u}(k) - \mathbf{g}y(k) \tag{4}$$

and therefore (1) can be replaced by

$$\hat{\mathbf{x}}(k|N) = [\Phi + \mathbf{g}\mathbf{f}']^{-1}\left[\hat{\mathbf{x}}(k+1|N) - \Gamma\hat{\mathbf{u}}(k|N) + \mathbf{g}\hat{y}(k|N)\right] \tag{5}$$

where

$$\hat{y}(k|N) = \hat{y}(k|k-1) + \mathbf{f}'P(k|k-1)\mathbf{r}(k) \quad . \tag{6}$$

So instead of saving vector $\hat{\mathbf{x}}(k|k-1)$ and matrix $P(k|k-1)$ we merely have to save scalar $\hat{y}(k|k-1)$ and vector $P(k|k-1)\mathbf{f}$ between the two passes.

Of particular interest is when $\mathbf{f}$ and $\mathbf{g}$ are chosen from the singular value decomposition of Φ so that

$$\Phi\mathbf{f} = \mathbf{0} = \Phi'\mathbf{g}$$

and

$$\mathbf{f}'\mathbf{f} = 1 = \mathbf{g}'\mathbf{g} \quad .$$

In this case, $y(k)$ contains the portion of $\mathbf{x}(k)$ which is about to be lost forever by the recursion

$$\mathbf{x}(k+1) = \Phi\mathbf{x}(k) + \Gamma\mathbf{u}(k)$$

and therefore cannot be reconstructed from future observations. Also, one can show that

$$[\Phi + \mathbf{g}\mathbf{f}']^{-1} = A + \mathbf{f}\mathbf{g}' \tag{7}$$

where A is the pseudo-inverse of Φ based on its sigular value decomposition, in which case

$$A\mathbf{g} = \mathbf{0} = A'\mathbf{f} \quad \text{and} \quad \Phi A = I - \mathbf{g}\mathbf{g}' \quad .$$

Substituting (7) into (5) we have

$$\hat{\mathbf{x}}(k|N) = A\left[\hat{\mathbf{x}}(k+1|N) - \Gamma\hat{\mathbf{u}}(k|N)\right] + \mathbf{f}\hat{y}(k|N) \tag{8}$$

since

$$\mathbf{g}'\left[\hat{\mathbf{x}}(k+1|N) - \Gamma\hat{\mathbf{u}}(k|N)\right] = \mathbf{g}'\Phi\hat{\mathbf{x}}(k|N) = \mathbf{0} \quad .$$

Multiplying both sides of (8) by $\mathbf{f}'$ gives us

$$\mathbf{f}'\hat{\mathbf{x}}(k|N) = \hat{y}(k|N) \tag{9}$$

while multiplying by Φ yields

$$\Phi\hat{\mathbf{x}}(k|N) = \hat{\mathbf{x}}(k+1|N) - \Gamma\hat{\mathbf{u}}(k|N) \tag{10}$$

as expected.

If the rank m of Φ is less than $n-1$, one must include additional auxilliary outputs, one for each zero valued singular value of Φ. This corresponds to replacing the scalar $y(k)$ with a vector $\mathbf{y}(k)$ of dimension $n-m$.

Stability

Let us assume that Φ is invertable, but the algorithm in (1) is numerically unstable. One can show that an algorithm of the form

$$\hat{\mathbf{x}}(k|N) = [I - \mathbf{g}\mathbf{f}']\Phi^{-1}\left[\hat{\mathbf{x}}(k+1|N) - \Gamma\hat{\mathbf{u}}(k|N)\right] + \mathbf{g}\hat{y}(k|N) \tag{12}$$

can be made stable if $\mathbf{f}$ and $\mathbf{g}$ satisfy observability and controlability requirements. In fact, any numerical error can be completely removed in n steps if

$$\mathbf{f}'\Phi^k\mathbf{g} = \begin{cases} 1 & :k = 0 \\ 0 & :k = 1, 2, \ldots, n-1 \end{cases} \tag{13}$$

as shown in the Appendix. Note that multiplying both sides of (12) by $\mathbf{f}'$ gives us (9) as before.

Given $\mathbf{f}$ one can compute $\mathbf{g}$ using

$$\mathbf{g} = \begin{pmatrix} \mathbf{f}' \\ \mathbf{f}'\Phi \\ \mathbf{f}'\Phi^2 \\ \vdots \\ \mathbf{f}'\Phi^{n-1} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{14}$$

provided the matrix is invertable (which incidentally is the definition of observablilty).

Any $\mathbf{f}$ which satisfies the observability requirement will produce good results. A logical choice would be the vector associated with the minimum singular value of Φ. Unfortunately, there is no simple solution for the optimal $\mathbf{f}$ (unless $\mathbf{g}$ is given). The best one can do is to assume a quadratic objective function and perform nonlinear optimization for a specific Φ matrix.

Let assume that the error covariance due to a numerical error $\mathbf{e}$ in $\hat{\mathbf{x}}(k|N)$ can be represented as

$$\mathcal{E}\{\mathbf{e}\mathbf{e}'\} = P(0) = I \tag{15}$$

since numerical errors tend to be random, uncorrelated and equally likely to occur anywhere. The absolute magnitude in this case doesn't matter. As this error propagates through (12) the resulting error covariances are given by

$$P(k) = (I - \mathbf{g}\mathbf{f}')\Phi^{-1}P(k-1){\Phi'}^{-1}(I - \mathbf{f}\mathbf{g}') \tag{16}$$

for $k = 1, 2, \ldots, n$. After n steps we assume the error has been totally eliminated.

Let us now assume an objective function of the form

$$\begin{aligned} J = & \sum_{k=1}^{n} \mathrm{tr}\,\{P(k)\} \\ & + \sum_{k=1}^{n} \mathrm{tr}\left\{S'(k)\left[(I - \mathbf{g}\mathbf{f}')\Phi^{-1}P(k-1){\Phi'}^{-1}(I - \mathbf{f}\mathbf{g}') - P(k)\right]\right\} \\ & + \mathbf{v}'\begin{pmatrix} \mathbf{f}'\mathbf{g} - 1 \\ \mathbf{f}'\Phi\mathbf{g} \\ \vdots \\ \mathbf{f}'\Phi^{n-1}\mathbf{g} \end{pmatrix} \end{aligned} \tag{17}$$

where $\mathbf{v}$ and the $S(k)$ are Lagrange multipliers. Minimizing J with respect to $\mathbf{v}$ gives us (13), while miminizing with respect to $S(k)$ gives us (16). Minimizing J with respect to $P(k)$ gives us

$$S(k) = I + {\Phi'}^{-1}(I - \mathbf{f}\mathbf{g}')S(k+1)(I - \mathbf{g}\mathbf{f}')\Phi^{-1} \tag{18}$$

starting from $S(n) = I$. Minimizing J with respect to $\mathbf{g}$ gives us

$$\mathbf{v} = \begin{pmatrix} \mathbf{f} & \Phi'\mathbf{f} & \ldots & {\Phi'}^{n-1}\mathbf{f} \end{pmatrix}^{-1} \sum_{k=1}^{n} 2S(k)(I - \mathbf{g}\mathbf{f}')\Phi^{-1}P(k-1){\Phi'}^{-1}\mathbf{f} \tag{19}$$

since $S(k) = S'(k)$ and $P(k-1) = P'(k-1)$. All that remains is to compute the gradient

$$\frac{dJ}{d\mathbf{f}} = \begin{pmatrix} \mathbf{g} & \Phi\mathbf{g} & \ldots & \Phi^{n-1}\mathbf{g} \end{pmatrix}\mathbf{v} - \sum_{k=1}^{n} 2\Phi^{-1}P(k-1){\Phi'}^{-1}(I - \mathbf{f}\mathbf{g}')S(k)\mathbf{g} \tag{20}$$

to use with a nonlinear optimization algorithm.

Three Pass Algorithm

Of course, an even simpler solution is to use a third pass of the form

$$\hat{\mathbf{x}}(k+1|N) = \Phi\hat{\mathbf{x}}(k|N) + \Gamma\hat{\mathbf{u}}(k|N)$$

starting from

$$\hat{\mathbf{x}}(0|N) = \hat{\mathbf{x}}(0) + P(0)\mathbf{r}(0)$$

and storing $\hat{\mathbf{u}}(k|N)$ between the second and third passes. Then again, this is numerically stable only if all the eigenvalues of Φ are inside the unit circle.

Appendix

Optimal Control Tutorial

Consider an nth order state vector model of the form

$$\mathbf{x}(k+1) = \Phi \mathbf{x}(k) - \mathbf{g}a(k) \tag{A-1}$$

where matrix Φ is invertable. Our goal is to find a sequence $a(k)$ which will force $\mathbf{x}(n) = \mathbf{0}$ for a given initial state $\mathbf{x}(0)$.

(When matrix Φ is not invertable, there is still an optimal control sequence, but it takes less than n steps to achieve and the solution is not unique.)

One can easily show that

$$\mathbf{x}(n) = \Phi^n \mathbf{x}(0) - \begin{pmatrix} \Phi^{n-1}\mathbf{g} & \Phi^{n-2}\mathbf{g} & \ldots & \Phi\mathbf{g} & \mathbf{g} \end{pmatrix} \begin{pmatrix} a(0) \\ a(1) \\ \vdots \\ a(n-1) \end{pmatrix} \tag{A-2}$$

and setting $\mathbf{x}(n) = \mathbf{0}$ gives us

$$\begin{pmatrix} a(0) \\ a(1) \\ \vdots \\ a(n-1) \end{pmatrix} = \begin{pmatrix} \Phi^{n-1}\mathbf{g} & \Phi^{n-2}\mathbf{g} & \ldots & \Phi\mathbf{g} & \mathbf{g} \end{pmatrix}^{-1} \Phi^n \mathbf{x}(0) \tag{A-3}$$

assuming that the matrix is invertable (which is the definition of controlablity).

Now let us find a vector $\mathbf{f}$ such that

$$\mathbf{f}'\mathbf{x}(0) = a(0) \quad \forall \mathbf{x}(0) \quad . \tag{A-4}$$

From (A-3) we can write

$$\mathbf{f}'\mathbf{x}(0) = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} \Phi^{n-1}\mathbf{g} & \Phi^{n-2}\mathbf{g} & \ldots & \Phi\mathbf{g} & \mathbf{g} \end{pmatrix}^{-1} \Phi^n \mathbf{x}(0) \quad \forall \mathbf{x}(0)$$

and therefore

$$\mathbf{f}' = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} \Phi^{n-1}\mathbf{g} & \Phi^{n-2}\mathbf{g} & \ldots & \Phi\mathbf{g} & \mathbf{g} \end{pmatrix}^{-1} \Phi^n \quad . \tag{A-5}$$

From (A-5) we can write

$$\mathbf{f}'\Phi^{-n} \begin{pmatrix} \Phi^{n-1}\mathbf{g} & \Phi^{n-2}\mathbf{g} & \ldots & \Phi\mathbf{g} & \mathbf{g} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \end{pmatrix}$$

or equivalently

$$\mathbf{f}'\Phi^{-k}\mathbf{g} = \begin{cases} 1 & : k = 1 \\ 0 & : k = 2, 3, \ldots, n \end{cases} \tag{A-6}$$

In fact, every step of the optimal control sequence is given by

$$a(k) = \mathbf{f}'\mathbf{x}(k) \quad \forall k = 0, 1, \ldots, n-1.$$

and therefore we can rewrite (A-1) as

$$\mathbf{x}(k+1) = [\Phi - \mathbf{g}\mathbf{f}']\mathbf{x}(k) \quad . \tag{A-7}$$

From (A-6) and (A-7) we see that

$$\mathbf{f}'\Phi^{k-n}\mathbf{x}(n) = \mathbf{f}'\Phi^{-1}\mathbf{x}(k+1) = 0 \quad \forall k = 0, 1, \ldots, n-1.$$

or equivalently

$$\begin{pmatrix} \mathbf{f}' \\ \mathbf{f}'\Phi \\ \vdots \\ \mathbf{f}'\Phi^{n-1} \end{pmatrix} \Phi^{-n}\mathbf{x}(n) = \mathbf{0} \quad . \tag{A-8}$$

When the matrix in (A-8) is invertable (which is the definition of observability), then (A-8) requires that $\mathbf{x}(n) = \mathbf{0}$ when it is generated using (A-7).