A Recursive Filter for Differential Observations

Bill Triggs
Oxford University Robotics Group,
Department of Engineering Science,
Parks Rd, Oxford OX1 3PJ, U.K.
bill@robots.oxford.ac.uk

1 Introduction

The standard Kalman filter applies to linear systems for which observations of some linear function of the current state are made. This model covers many practical cases, but it is sometimes necessary to deal with observations of *differences* between states at different times. For example a navigator might use dead reconning (odometry) or the relative position of some convenient local landmark to estimate the distance travelled between observations. The absolute value of such measures is meaningless because they have an unknown offset (the zero point of the odometer or the absolute location of the landmark), but the differences between observations at different times provide useful information about *changes* in location.

There are several methods of dealing with relative observations within the standard Kalman filter framework:

Firstly, the state space can be augmented to include state information at previous time steps, so that the observations become a linear function of the augmented state vector. In principle this method can be used to look arbitrarily far back into the past, but since a doubling of the state dimension (naïvely) increases the filtering computation by a factor of eight, this method soon becomes computationally prohibitive.

Secondly, differential observations can be interpreted as observations of *average state velocity* and these velocities can be incorporated into an augmented state vector. But it is important to distinguish between *instantaneous* velocities, which may well be components of a physically meaningful system state, and *average* velocities, which are integral quantities that essentially just re-encode changes in state variables between time steps. Formally, including average velocities in the system state is equivalent to augmenting the state with positions at previous time steps. Confounding average and instantaneous velocity and then fabricating some nominal velocity dynamics to fit it into the Kalman framework is asking for trouble, given the notorious sensitivity of the Kalmen filter to such not-quite-correct assumptions.

This report describes a third method of dealing with relative observations called the differential Kalman filter (dKF)[1]. It does not require explicit state augmentation and it is a rigorous maximum likelihood estimator under the usual linear-quadratic-Gaussian

[1]The trendy capitalization distinguishes it from the DKF or Distributed Kalman Filter

hypotheses of Kalman filtering. However it only applies to observations between *adjacent* time steps and the resulting equations are somewhat more complex than the standard Kalman filter (though still tractable).

The observation model of the dKF is designed for 'random walk' situations such as vehicle odometry, where the errors between time steps are independent and cumulative. Strictly speaking it does not apply to the tracking of fixed landmarks over multiple time steps, which produces correlated differential measurements: for this the state should be augmented with the estimated landmark locations. However the dKF may still be a useful approximation in applications such as real time computer-visual motion estimation where there are many unreliable short-lived landmarks and processing speed is more important than strict optimality.

The dKF can be seen as an expansion of the equations of a state-augmented system. It is possible to avoid explicit state augmentation for observations between adjacent time steps because the system output already depends indirectly on the previous state through the current state prediction: the existing dependencies are complicated but no essentially new ones are introduced. This is no longer true if the output depends on states before the previous one.

The next section formulates the system model to which the dKF applies, then the state and information forms of the dKF are derived using state augmentation. The next two sections provide alternative derivations of the information filter: from first principles; and as an optimal linear estimator (Joseph form). These derivations closely mirror the corresponding Kalman ones. .... For convenience, the matrix identities used are gathered into an appendix.

2 System Model and Estimation Framework

Notation: Vectors are lower case, matrices upper. $\boldsymbol{x}$ is for system state vectors, $\boldsymbol{u}$ for system inputs (controls) and $\boldsymbol{z}$ for system outputs (observations). Capitals denote the corresponding covariance matrix (*e.g.* $\boldsymbol{X} = \mathrm{Cov}(\boldsymbol{x})$). Carets ($\hat{\boldsymbol{x}}$) indicate predictions (estimates made prior to an observation using previous information), and bars ($\bar{\boldsymbol{x}}$) denote updated quantities (posterior estimates incorporating the results of the current observation). Unaccented letters represent the true physical value (often a random variable). Quantities subscripted t are always available at or after the t^{th} observation, $(\ldots t-2, t-1, t\ldots)$ denotes the sequence of times at which observations are made (not necessarily integral or evenly spaced). $\boldsymbol{n} \in \mathcal{N}(\boldsymbol{x},\boldsymbol{X})$ denotes a sample of a Gaussian distribution with mean $\boldsymbol{x}$ and covariance $\boldsymbol{X}$.

We will restrict ourselves to the linear quadratic Gaussian case, as is usual in discussions of Kalman filtering: linear system dynamics and output model, quadratic cost functions and Gaussian state distributions. The dKF can be also be used as a linear approximation to a nonlinear system as in the Extended Kalman Filter, with all of the usual caveats and restrictions.

The system dynamical and output (observation) models are taken to be

$$\boxed{\begin{array}{rclll} \boldsymbol{x}_t &=& \boldsymbol{F}_t\,\boldsymbol{x}_{t-1} + \boldsymbol{G}_t\,\boldsymbol{u}_{t-1} + \boldsymbol{n}_{x_t} & \boldsymbol{n}_{x_t} \in \mathcal{N}(\boldsymbol{0},\boldsymbol{Q}_t) & \text{(dynamics)} \\ \boldsymbol{z}_t &=& \boldsymbol{H}_t\,\boldsymbol{x}_t - \check{\boldsymbol{H}}_t\,\boldsymbol{x}_{t-1} + \boldsymbol{n}_{z_t} & \boldsymbol{n}_{z_t} \in \mathcal{N}(\boldsymbol{0},\boldsymbol{R}_t) & \text{(observation)} \end{array}}$$

Here $\boldsymbol{x}_{t-1}$ is the true system state at the previous time step, $\boldsymbol{u}_{t-1}$ the control applied then, $\boldsymbol{x}_t$ the resulting system state and z_t the observed system output at the current time step. $\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{H}_t$ and $\check{\boldsymbol{H}}_t$ are given gain matrices and $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ are given covariance matrices. The dynamic and output noises $\boldsymbol{n}_{x_t}$ and $\boldsymbol{n}_{z_t}$ are assumed to be zero-mean Gaussian perturbations independent of each other and of all the previous states and perturbations. The state dynamics and the independence of the perturbations imply that the system evolution is a Markov process. The only change from the usual Kalman case is that the output may depend on the previous state as well as the current one, via the matrix $\check{\boldsymbol{H}}_t$. The direct and indirect dependence of z_t on $\boldsymbol{x}_{t-1}$ is summarized by

$$\boldsymbol{H}'_t \equiv \boldsymbol{H}_t \boldsymbol{F}_t - \check{\boldsymbol{H}}_t$$

If the observations are exactly differential we have $\check{\boldsymbol{H}}_t \equiv \boldsymbol{H}_t$, $\boldsymbol{H}'_t \equiv \boldsymbol{H}_t(\boldsymbol{F}_t - \boldsymbol{1})$.

Our aim is to estimate the current system state $\boldsymbol{x}_t$ from our complete set of observations up to and including the current time.

3 The dKF as a Kalman Filter

The optimal filter for our system can be found by augmenting the state to $\binom{x_{t-1}}{x_t}$ so that the traditional Kalman prediction and update equations can be used. These are summarized in our notation in (**??**). Denoting augmented quantities by large sans serif type and noting that the future state dynamics depends only on the $\boldsymbol{x}_t$ components of the augmented state, we have:

$$\begin{aligned}
\hat{\mathsf{x}}_t &\equiv \begin{pmatrix} \bar{\boldsymbol{x}}_{t-1} \\ \hat{\boldsymbol{x}}_t \end{pmatrix} = \mathsf{F}_t\,\bar{\mathsf{x}}_{t-1} + \mathsf{G}_t\,\mathsf{u}_t \\
&= \begin{pmatrix} \boldsymbol{1} \\ \boldsymbol{F}_t \end{pmatrix} \bar{\boldsymbol{x}}_{t-1} + \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{G}_t \end{pmatrix} \boldsymbol{u}_{t-1} \\
\hat{\mathsf{X}}_t &= \mathsf{F}_t\,\bar{\mathsf{X}}_{t-1}\,\mathsf{F}_t^\top + \mathsf{Q}_t \\
&= \begin{pmatrix} \boldsymbol{1} \\ \boldsymbol{F}_t \end{pmatrix} \bar{\boldsymbol{X}}_{t-1} \begin{pmatrix} \boldsymbol{1} & \boldsymbol{F}_t^\top \end{pmatrix} + \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_t \end{pmatrix} \\
&= \begin{pmatrix} \bar{\boldsymbol{X}}_{t-1} & \bar{\boldsymbol{X}}_{t-1}\boldsymbol{F}_t^\top \\ \boldsymbol{F}_t\bar{\boldsymbol{X}}_{t-1} & \boldsymbol{Q}_t + \boldsymbol{F}_t\bar{\boldsymbol{X}}_{t-1}\boldsymbol{F}_t^\top \end{pmatrix} \\
&= \begin{pmatrix} \bar{\boldsymbol{X}}_{t-1}^{-1} + \boldsymbol{F}_t^\top\boldsymbol{Q}_t^{-1}\boldsymbol{F}_t & -\boldsymbol{F}_t^\top\boldsymbol{Q}_t^{-1} \\ -\boldsymbol{Q}_t^{-1}\boldsymbol{F}_t & \boldsymbol{Q}_t^{-1} \end{pmatrix}^{-1} \qquad \text{(by inversion identity (3))} \\
\hat{z}_t &= \mathsf{H}_t\,\hat{\mathsf{x}}_t \\
&= \begin{pmatrix} -\check{\boldsymbol{H}}_t & \boldsymbol{H}_t \end{pmatrix} \begin{pmatrix} \bar{\boldsymbol{x}}_{t-1} \\ \hat{\boldsymbol{x}}_t \end{pmatrix} \\
\hat{\boldsymbol{Z}}_t &= \mathsf{H}_t\,\hat{\mathsf{X}}_t\,\mathsf{H}_t^\top + \boldsymbol{R}_t \\
&= \begin{pmatrix} -\check{\boldsymbol{H}}_t & \boldsymbol{H}_t \end{pmatrix} \begin{pmatrix} \bar{\boldsymbol{X}}_{t-1} & \bar{\boldsymbol{X}}_{t-1}\boldsymbol{F}_t^\top \\ \boldsymbol{F}_t\bar{\boldsymbol{X}}_{t-1} & \boldsymbol{Q}_t + \boldsymbol{F}_t\bar{\boldsymbol{X}}_{t-1}\boldsymbol{F}_t^\top \end{pmatrix} \begin{pmatrix} -\check{\boldsymbol{H}}_t^\top \\ \boldsymbol{H}_t^\top \end{pmatrix} + \boldsymbol{R}_t
\end{aligned}$$

$$
\begin{aligned}
&= \boldsymbol{H'}_t\bar{\boldsymbol{X}}_{t-1}\boldsymbol{H'}_t + \boldsymbol{H}_t\boldsymbol{Q}_t\boldsymbol{H}_t^\top + \boldsymbol{R}_t \\
\mathsf{K}_t &= \hat{\mathsf{X}}_t\,\mathsf{H}_t^\top\,\hat{\boldsymbol{Z}}_t^{-1} \\
&= \begin{pmatrix} \bar{\boldsymbol{X}}_{t-1}\boldsymbol{H'}_t^\top \\ \boldsymbol{F}_t\bar{\boldsymbol{X}}_{t-1}\boldsymbol{H'}_t^\top + \boldsymbol{Q}_t\boldsymbol{H}_t^\top \end{pmatrix} \left(\boldsymbol{H'}_t\bar{\boldsymbol{X}}_{t-1}\boldsymbol{H'}_t + \boldsymbol{H}_t\boldsymbol{Q}_t\boldsymbol{H}_t^\top + \boldsymbol{R}_t\right)^{-1} \\
\bar{\mathsf{x}}_t &= \hat{\mathsf{x}}_t + \mathsf{K}_t\,(z_t - \hat{z}_t) \\
\bar{\mathsf{X}}_t &= \left(1 - \mathsf{K}_t\,\mathsf{H}_t\right)\hat{\mathsf{X}}_t
\end{aligned}
$$

Assuming we are not interested in updating our estimate of the previous state (which is irrelevant to future predictions), we can discard the $\boldsymbol{x}_{t-1}$ components of these equations. The resulting state and output prediction equations are:

State & Output Prediction

$$
\begin{aligned}
\hat{\boldsymbol{x}}_t &= \boldsymbol{F}_t\,\bar{\boldsymbol{x}}_{t-1} + \boldsymbol{G}_t\,\boldsymbol{u}_{t-1} \\
\hat{\boldsymbol{z}}_t &= \boldsymbol{H}_t\,\hat{\boldsymbol{x}}_t - \check{\boldsymbol{H}}_t\,\bar{\boldsymbol{x}}_{t-1} \\
&= \boldsymbol{H'}_t\,\bar{\boldsymbol{x}}_{t-1} + \boldsymbol{H}_t\,\boldsymbol{G}_t\,\boldsymbol{u}_t \\
\hat{\boldsymbol{X}}_t &= \boldsymbol{F}_t\,\bar{\boldsymbol{X}}_{t-1}\,\boldsymbol{F}_t^\top + \boldsymbol{Q}_t \\
\hat{\boldsymbol{Z}}_t &= \boldsymbol{H'}_t\,\bar{\boldsymbol{X}}_{t-1}\,\boldsymbol{H'}_t + \boldsymbol{H}_t\,\boldsymbol{Q}_t\,\boldsymbol{H}_t^\top + \boldsymbol{R}_t
\end{aligned}
$$

The $\boldsymbol{x}_t$ components of K_t and $\bar{\mathsf{X}}_t$ provide state space update equations to complete the recursive filter:

State Space Update

$$
\begin{aligned}
\boldsymbol{L} &\equiv \boldsymbol{F}_t\,\bar{\boldsymbol{X}}_{t-1}\,\boldsymbol{H'}_t^\top + \boldsymbol{Q}_t\,\boldsymbol{H}_t^\top \\
&= \hat{\boldsymbol{X}}_t\,\boldsymbol{H}_t^\top - \boldsymbol{F}_t\,\bar{\boldsymbol{X}}_{t-1}\,\check{\boldsymbol{H}}_t^\top \\
\boldsymbol{K}_t &\equiv \boldsymbol{L}\,\hat{\boldsymbol{Z}}_t^{-1} \\
\bar{\boldsymbol{x}}_t &= \hat{\boldsymbol{x}}_t + \boldsymbol{K}_t\,(z_t - \hat{z}_t) \\
\bar{\boldsymbol{X}}_t &= \hat{\boldsymbol{X}}_t - \boldsymbol{L}\,\hat{\boldsymbol{Z}}_t^{-1}\,\boldsymbol{L}^\top
\end{aligned}
$$

We see that the update has a similar form to the Kalman one, with gain matrix

$$
\begin{aligned}
\boldsymbol{K}_t &\equiv \left(\boldsymbol{F}_t\,\bar{\boldsymbol{X}}_{t-1}\,\boldsymbol{H'}_t^\top + \boldsymbol{Q}_t\,\boldsymbol{H}_t^\top\right)\left(\boldsymbol{H'}_t\,\bar{\boldsymbol{X}}_{t-1}\,\boldsymbol{H'}_t^\top + \boldsymbol{H}_t\,\boldsymbol{Q}_t\,\boldsymbol{H}_t^\top + \boldsymbol{R}_t\right)^{-1} \\
&= \left(\hat{\boldsymbol{X}}_t\,\boldsymbol{H}_t^\top - \boldsymbol{F}_t\,\bar{\boldsymbol{X}}_{t-1}\,\check{\boldsymbol{H}}_t^\top\right)\left(\boldsymbol{H}_t\,\hat{\boldsymbol{X}}_t\,\boldsymbol{H}_t^\top - \boldsymbol{H}_t\,\boldsymbol{F}_t\,\bar{\boldsymbol{X}}_{t-1}\,\check{\boldsymbol{H}}_t^\top - \check{\boldsymbol{H}}_t\,\bar{\boldsymbol{X}}_{t-1}\,\boldsymbol{F}_t^\top\boldsymbol{H}_t^\top + \check{\boldsymbol{H}}_t\,\bar{\boldsymbol{X}}_{t-1}\,\check{\boldsymbol{H}}_t^\top + \boldsymbol{R}_t\right)^{-1}
\end{aligned}
$$

There is also an information form of the dKF. Continuing with the augmented state calculation:

$$
\begin{aligned}
\bar{\mathsf{X}}_t &= \left(\hat{\mathsf{X}}_t^{-1} + \mathsf{H}_t^\top\,\boldsymbol{R}_t^{-1}\,\mathsf{H}_t\right)^{-1} \\
&= \begin{pmatrix} \bar{\boldsymbol{X}}_{t-1}^{-1} + \boldsymbol{F}_t^\top\,\boldsymbol{Q}_t^{-1}\,\boldsymbol{F}_t + \check{\boldsymbol{H}}_t^\top\,\boldsymbol{R}_t^{-1}\,\check{\boldsymbol{H}}_t & -(\boldsymbol{F}_t^\top\,\boldsymbol{Q}_t^{-1} + \check{\boldsymbol{H}}_t^\top\,\boldsymbol{R}_t^{-1}\,\boldsymbol{H}_t) \\ -(\boldsymbol{Q}_t^{-1}\,\boldsymbol{F}_t + \boldsymbol{H}_t^\top\,\boldsymbol{R}_t^{-1}\,\check{\boldsymbol{H}}_t) & \boldsymbol{Q}_t^{-1} + \boldsymbol{H}_t^\top\,\boldsymbol{R}_t^{-1}\,\boldsymbol{H}_t \end{pmatrix}^{-1}
\end{aligned}
$$

$$\mathsf{K}_t = \bar{\mathsf{X}}_t\,\mathsf{H}_t^\top \boldsymbol{R}_t^{-1}$$

Applying inversion identity (2) to $\bar{\mathsf{X}}_t^{-1}$ and once again discarding the $\boldsymbol{x}_{t-1}$ components produces the following information space update equations:

Information Space Update
$$\begin{aligned} \boldsymbol{A} &\equiv \bar{\boldsymbol{X}}_{t-1}^{-1} + \boldsymbol{F}_t^\top \boldsymbol{Q}_t^{-1} \boldsymbol{F}_t + \check{\boldsymbol{H}}_t^\top \boldsymbol{R}_t^{-1} \check{\boldsymbol{H}}_t \\ \boldsymbol{B} &\equiv \boldsymbol{Q}_t^{-1} \boldsymbol{F}_t + \boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} \check{\boldsymbol{H}}_t \\ \bar{\boldsymbol{X}}_t &= \left(\boldsymbol{Q}_t^{-1} + \boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} \boldsymbol{H}_t - \boldsymbol{B}\,\boldsymbol{A}^{-1}\boldsymbol{B}^\top\right)^{-1} \\ \boldsymbol{K}_t &\equiv \bar{\boldsymbol{X}}_t \left(\boldsymbol{H}_t^\top - \boldsymbol{B}\,\boldsymbol{A}^{-1}\check{\boldsymbol{H}}_t^\top\right) \boldsymbol{R}_t^{-1} \\ \bar{\boldsymbol{x}}_t &= \hat{\boldsymbol{x}}_t + \boldsymbol{K}_t\,(z_t - \hat{z}_t) \end{aligned}$$

Once again the equations are similar to the corresponding Kalman information updates, although some additional covariance calculations are required that take place in the information space of the previous state. If the information $\bar{\boldsymbol{X}}_{t-1}^{-1}$ is remembered from the previous time step only two state space matrix inversions are required, for $\boldsymbol{A}^{-1}$ and $\bar{\boldsymbol{X}}_t$. (The standard Kalman information update also requires two inversions for $\hat{\boldsymbol{X}}_t^{-1}$ and $\bar{\boldsymbol{X}}_t$. $\hat{\boldsymbol{X}}_t$ does not appear at all in this version of the dKF).

If the observations fall into several independent groups so that $\boldsymbol{R}_t$ is (block) diagonal, they can be combined by accumulating the independent contributions to $\boldsymbol{H}_t^\top\boldsymbol{R}_t^{-1}\boldsymbol{H}_t$, $\check{\boldsymbol{H}}_t^\top\boldsymbol{R}_t^{-1}\boldsymbol{H}_t$, $\check{\boldsymbol{H}}_t^\top\boldsymbol{R}_t^{-1}\check{\boldsymbol{H}}_t$, $\boldsymbol{H}_t^\top\boldsymbol{R}_t^{-1}z_t$, and $\check{\boldsymbol{H}}_t^\top\boldsymbol{R}_t^{-1}z_t$, and then evaluating $\boldsymbol{A}$, $\boldsymbol{B}$, $\bar{\boldsymbol{X}}_t$, and $\bar{\boldsymbol{x}}_t$ as usual. This is similar to the distributed information filter, although (if $\check{\boldsymbol{H}}_t \neq \boldsymbol{H}_t$) more terms must be accumulated.

If $\check{\boldsymbol{H}}_t = \boldsymbol{0}$ it is easy to check that the filter reduces to the usual Kalman one, by expanding the $\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^\top$ in the covariance update and using inversion identity (6). If $\check{\boldsymbol{H}}_t \neq \boldsymbol{0}$ it is rather messy to show that the state and information forms of the filter are equivalent. By far the simplest method is to go back to the augmented state space and use the standard Kalman proofs: transfer identity (9) for $\mathsf{K}_t = \bar{\mathsf{X}}_t\,\mathsf{H}_t^\top\,\mathsf{R}_t^{-1} = \hat{\mathsf{X}}_t\,\mathsf{H}_t^\top\,\hat{\mathsf{Z}}_t^{-1}$ and inversion identity (5) for $\bar{\mathsf{X}}_t = \left(\hat{\mathsf{X}}_t^{-1} + \mathsf{H}_t^\top\,\mathsf{R}_t^{-1}\mathsf{H}_t\right)^{-1} = \hat{\mathsf{X}}_t - \hat{\mathsf{X}}_t\,\mathsf{H}_t^\top\,\hat{\mathsf{Z}}_t^{-1}\,\mathsf{H}_t\,\hat{\mathsf{X}}_t$.

4 Derivation of dKF from First Principles

It is straightforward and rather illuminating to derive the dKF from first principles. We begin with a short discussion of Markov systems, formulate a noise model for the system, and then examine the evolution of the state distribution assuming an initial Gaussian distribution.

Suppose our system has the Markov property that the current state value summarizes everything about the past that is relevant to the future: the future behaviour of the system is independent of how it got to the current state, given the state it got to. This is what 'state' means: the minimum set of variables now that is sufficient to predict as much of the future as can be predicted from the past.

If the current state is unknown but distributed according to some particular probability distribution, the Markov property implies that this distribution (*i.e.* the value of the probability density at each point in state space) summarizes everything we currently know about the system that is relevant to the future. Any given state evolves into a state distribution according to the specified system dynamics. A distribution over the current states simply evolves into a sum of these distributions weighted by the current state probabilities. (This is the Fundamental Assumption of Probability Theory: the linear combination of independent probabilities or 'Independent Universes' principle). The Markov sufficiency of the current state implies that current state distributions are sufficient statistics for all future predictions. This is true for all Markov systems no matter how nonlinear their dynamics or complicated their state distributions.

Hence, for Markov systems we can do no better than to estimate the current state distribution from the past observations, then ignore the observations and work with the estimated state distribution from then on. If we have already managed to do this for the previous time step, the previous distribution and the system dynamics give the correct prior distribution at the current time step, and conditioning on the current observation produces the correct posterior distribution for the current state. So for Markov systems we can proceed recursively, calculating the state distribution time step by time step.

Suppose that at some time the state distribution turns out to be Gaussian. For an affine system each stage of the updating process preserves normality so the following distribution is also Gaussian and the Gaussian form is preserved from then on. (By the Central Limit argument the Gaussian form is also stable: for affine systems other reasonable initial distributions will evolve towards it on average, so long as things remain well conditioned). Because a Gaussian is determined by its mean and variance, the calculations with state distributions can then be reduced to calculations with mean vectors and covariance matrices: for affine Markov systems there is always something analogous to the Kalman filter.

All of the above applies to our system, so let us assume that we have already determined a distribution $\mathcal{N}(\bar{\boldsymbol{x}}_{t-1}, \bar{\boldsymbol{X}}_{t-1})$ for the previous state $\boldsymbol{x}_{t-1}$. This uncertainty in our knowledge is combined with the two independent sources of physical uncertainty (the system dynamical and output noises) as follows:

$$
\begin{aligned}
\boldsymbol{x}_{t-1} &= \bar{\boldsymbol{x}}_{t-1} + \boldsymbol{n}_{\bar{x}_{t-1}} & \boldsymbol{n}_{\bar{x}_{t-1}} \in \mathcal{N}(\boldsymbol{0}, \bar{\boldsymbol{X}}_{t-1}) \\
\boldsymbol{x}_t &= \boldsymbol{F}_t \boldsymbol{x}_{t-1} + \boldsymbol{G}_t \boldsymbol{u}_{t-1} + \boldsymbol{n}_{x_t} & \boldsymbol{n}_{x_t} \in \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_t) \\
&= (\boldsymbol{F}_t \bar{\boldsymbol{x}}_{t-1} + \boldsymbol{G}_t \boldsymbol{u}_{t-1}) + \boldsymbol{F}_t \boldsymbol{n}_{\bar{x}_{t-1}} + \boldsymbol{n}_{x_t} & \\
&= \hat{\boldsymbol{x}}_t + \boldsymbol{F}_t \boldsymbol{n}_{\bar{x}_{t-1}} + \boldsymbol{n}_{x_t} & \\
\boldsymbol{z}_t &= \boldsymbol{H}_t \boldsymbol{x}_t - \check{\boldsymbol{H}}_t \boldsymbol{x}_{t-1} + \boldsymbol{n}_{z_t} & \boldsymbol{n}_{z_t} \in \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_t) \\
&= \left(\boldsymbol{H}_t \hat{\boldsymbol{x}}_t - \check{\boldsymbol{H}}_t \bar{\boldsymbol{x}}_{t-1}\right) + \left(\boldsymbol{H}_t \boldsymbol{F}_t - \check{\boldsymbol{H}}_t\right) \boldsymbol{n}_{\bar{x}_{t-1}} + \boldsymbol{H}_t \boldsymbol{n}_{x_t} + \boldsymbol{n}_{z_t} & \\
&= \hat{\boldsymbol{z}}_t + \boldsymbol{H}'_t \boldsymbol{n}_{\bar{x}_{t-1}} + \boldsymbol{H}_t \boldsymbol{n}_{x_t} + \boldsymbol{n}_{z_t} &
\end{aligned}
$$

Using the fact that the uncertainties $\boldsymbol{n}_{\bar{x}_{t-1}}$, $\boldsymbol{n}_{x_t}$ and $\boldsymbol{n}_{z_t}$ are independent zero-mean Gaussians, the system dynamical and output mappings produce Gaussian distributions

$\boldsymbol{x}_t \in \mathcal{N}(\hat{\boldsymbol{x}}_t, \hat{\boldsymbol{X}}_t)$ and $z_t \in \mathcal{N}(\hat{z}_t, \hat{\boldsymbol{Z}}_t)$ for the state $\boldsymbol{x}_t$ and the output z_t where (as before):

$$\boxed{\begin{array}{rcl} \multicolumn{3}{c}{\textbf{\textit{State \& Output Prediction}}} \\ \hat{\boldsymbol{x}}_t &=& \boldsymbol{F}_t \bar{\boldsymbol{x}}_{t-1} + \boldsymbol{G}_t \boldsymbol{u}_{t-1} \\ \hat{z}_t &=& \boldsymbol{H}_t \hat{\boldsymbol{x}}_t - \check{\boldsymbol{H}}_t \bar{\boldsymbol{x}}_{t-1} \\ &=& \boldsymbol{H}'_t \bar{\boldsymbol{x}}_{t-1} + \boldsymbol{H}_t \boldsymbol{G}_t \boldsymbol{u}_t \\ \hat{\boldsymbol{X}}_t &=& \boldsymbol{F}_t \bar{\boldsymbol{X}}_{t-1} \boldsymbol{F}_t^\top + \boldsymbol{Q}_t \\ \hat{\boldsymbol{Z}}_t &=& \boldsymbol{H}'_t \bar{\boldsymbol{X}}_{t-1} \boldsymbol{H}'_t + \boldsymbol{H}_t \boldsymbol{Q}_t \boldsymbol{H}_t^\top + \boldsymbol{R}_t \end{array}}$$

Of course $\boldsymbol{x}_{t-1}$, $\boldsymbol{x}_t$ and z_t are correlated and we must be careful to take account of this. Defining the joint deviation vector

$$\begin{aligned} \mathsf{y} &\equiv \begin{pmatrix} \delta\bar{\boldsymbol{x}}_{t-1} \\ \delta\hat{\boldsymbol{x}}_t \\ \delta\hat{z}_t \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{x}_{t-1} - \bar{\boldsymbol{x}}_{t-1} \\ \boldsymbol{x}_t - \hat{\boldsymbol{x}}_t \\ z_t - \hat{z}_t \end{pmatrix} \\ &= \begin{pmatrix} 1 & & \\ \boldsymbol{F}_t & 1 & \\ \boldsymbol{H}'_t & \boldsymbol{H}_t & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{n}_{\bar{x}_{t-1}} \\ \boldsymbol{n}_{x_t} \\ \boldsymbol{n}_{z_t} \end{pmatrix} \end{aligned}$$

we have the following joint covariance and information matrices:

$$\begin{aligned} \mathsf{Y} &\equiv \langle \mathsf{y}\mathsf{y}^\top \rangle \\ &= \begin{pmatrix} 1 & & \\ \boldsymbol{F}_t & 1 & \\ \boldsymbol{H}'_t & \boldsymbol{H}_t & 1 \end{pmatrix} \begin{pmatrix} \bar{\boldsymbol{X}}_{t-1} & & \\ & \boldsymbol{Q}_t & \\ & & \boldsymbol{R}_t \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{F}_t^\top & \boldsymbol{H}'^\top_t \\ & 1 & \boldsymbol{H}_t^\top \\ & & 1 \end{pmatrix} \\ &= \begin{pmatrix} \bar{\boldsymbol{X}}_{t-1} & \bar{\boldsymbol{X}}_{t-1}\boldsymbol{F}_t^\top & \bar{\boldsymbol{X}}_{t-1}\boldsymbol{H}'^\top_t \\ \boldsymbol{F}_t\bar{\boldsymbol{X}}_{t-1} & \boldsymbol{F}_t\bar{\boldsymbol{X}}_{t-1}\boldsymbol{F}_t^\top + \boldsymbol{Q}_t & \boldsymbol{F}_t\bar{\boldsymbol{X}}_{t-1}\boldsymbol{H}'^\top_t + \boldsymbol{Q}_t\boldsymbol{H}_t^\top \\ \boldsymbol{H}'_t\bar{\boldsymbol{X}}_{t-1} & \boldsymbol{H}'_t\bar{\boldsymbol{X}}_{t-1}\boldsymbol{F}_t^\top + \boldsymbol{H}_t\boldsymbol{Q}_t & \boldsymbol{H}'_t\bar{\boldsymbol{X}}_{t-1}\boldsymbol{H}'^\top_t + \boldsymbol{H}_t\boldsymbol{Q}_t\boldsymbol{H}_t^\top + \boldsymbol{R}_t \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \mathsf{Y}^{-1} &= \begin{pmatrix} 1 & -\boldsymbol{F}_t^\top & \check{\boldsymbol{H}}_t^\top \\ & 1 & -\boldsymbol{H}_t^\top \\ & & 1 \end{pmatrix} \begin{pmatrix} \bar{\boldsymbol{X}}_{t-1}^{-1} & & \\ & \boldsymbol{Q}_t^{-1} & \\ & & \boldsymbol{R}_t^{-1} \end{pmatrix} \begin{pmatrix} 1 & & \\ -\boldsymbol{F}_t & 1 & \\ \check{\boldsymbol{H}}_t & -\boldsymbol{H}_t & 1 \end{pmatrix} \\ &= \begin{pmatrix} \bar{\boldsymbol{X}}_{t-1}^{-1} + \boldsymbol{F}_t^\top\boldsymbol{Q}_t^{-1}\boldsymbol{F}_t + \check{\boldsymbol{H}}_t^\top\boldsymbol{R}_t^{-1}\check{\boldsymbol{H}}_t & -\left(\boldsymbol{F}_t^\top\boldsymbol{Q}_t^{-1} + \check{\boldsymbol{H}}_t^\top\boldsymbol{R}_t^{-1}\boldsymbol{H}_t\right) & \check{\boldsymbol{H}}_t^\top\boldsymbol{R}_t^{-1} \\ -\left(\boldsymbol{Q}_t^{-1}\boldsymbol{F}_t + \boldsymbol{H}_t^\top\boldsymbol{R}_t^{-1}\check{\boldsymbol{H}}_t\right) & \boldsymbol{Q}_t^{-1} + \boldsymbol{H}_t^\top\boldsymbol{R}_t^{-1}\boldsymbol{H}_t & -\boldsymbol{H}_t^\top\boldsymbol{R}_t^{-1} \\ \boldsymbol{R}_t^{-1}\check{\boldsymbol{H}}_t & -\boldsymbol{R}_t^{-1}\boldsymbol{H}_t & \boldsymbol{R}_t^{-1} \end{pmatrix} \end{aligned}$$

The distribution function for the joint variable y prior to any measurement is therefore proportional to

$$\mathsf{e}^{-\frac{1}{2}\mathsf{y}^\top\mathsf{Y}^{-1}\mathsf{y}}$$

Now suppose we make an observation z_t. This fixes the value of $\delta\hat{z}_t = z_t - \hat{z}_t$, and in effect freezes the $\delta\hat{z}_t$ component of y at the observed value. The unnormalized conditional distribution of the remaining components of y given the observation z_t is given by the remaining free terms of $\mathsf{y}^\top \mathsf{Y}^{-1} \mathsf{y}$:

$$\begin{aligned}\mathsf{y}^\top \mathsf{Y}^{-1} \mathsf{y} &= \begin{pmatrix} \delta\bar{x}_{t-1}^\top & \delta x_t^\top \end{pmatrix} \begin{pmatrix} A & -B^\top \\ -B & C \end{pmatrix} \begin{pmatrix} \delta\bar{x}_{t-1} \\ \delta x_t \end{pmatrix} \\ &\quad -2 \begin{pmatrix} \delta\bar{x}_{t-1}^\top & \delta x_t^\top \end{pmatrix} \begin{pmatrix} -\check{H}_t^\top \\ H_t^\top \end{pmatrix} R_t^{-1} \delta\hat{z}_t + \delta\hat{z}_t^\top R_t^{-1} \delta\hat{z}_t\end{aligned}$$

where

$$\begin{pmatrix} A & -B^\top \\ -B & C \end{pmatrix} \equiv \begin{pmatrix} \bar{X}_{t-1}^{-1} + F_t^\top Q_t^{-1} F_t + \check{H}_t^\top R_t^{-1} \check{H}_t & -\left(F_t^\top Q_t^{-1} + \check{H}_t^\top R_t^{-1} H_t\right) \\ -\left(Q_t^{-1} F_t + H_t^\top R_t^{-1} \check{H}_t\right) & Q_t^{-1} + H_t^\top R_t^{-1} H_t \end{pmatrix}$$

Hence, the normalized posterior distribution of $\begin{pmatrix} \delta\bar{x}_{t-1} \\ \delta\hat{x}_t \end{pmatrix}$ is a Gaussian with covariance:

$$\begin{aligned}\begin{pmatrix} A & -B^\top \\ -B & C \end{pmatrix}^{-1} &= \begin{pmatrix} A^{-1} + A^{-1} B^\top \bar{X}_t B A^{-1} & A^{-1} B^\top \bar{X}_t \\ \bar{X}_t B A^{-1} & \bar{X}_t \end{pmatrix} \\ \bar{X}_t &\equiv \left(C - B A^{-1} B^\top\right)^{-1}\end{aligned}$$

(by inversion identity (2)) and mean and maximum likelyhood value:

$$\left\langle \begin{pmatrix} \delta\bar{x}_{t-1} \\ \delta x_t \end{pmatrix} \right\rangle = \begin{pmatrix} A & -B^\top \\ -B & C \end{pmatrix}^{-1} \begin{pmatrix} -\check{H}_t^\top \\ H_t^\top \end{pmatrix} R_t^{-1} \delta\hat{z}_t$$

These conditional means correct the means $\bar{x}_{t-1}$ and $\hat{x}_t$ of the joint prior state distribution for the effect of the observation. As before assume that we are not interested in updating our estimate of the previous state (which is irrelevant to future system behaviour given our new estimate of the current state by the Markov property), and discard the $\delta\bar{x}_{t-1}$ components of these equations. The result is the information form of the state update equations (as before):

Information Space Update

$$\begin{aligned} A &\equiv \bar{X}_{t-1}^{-1} + F_t^\top Q_t^{-1} F_t + \check{H}_t^\top R_t^{-1} \check{H}_t \\ B &\equiv Q_t^{-1} F_t + H_t^\top R_t^{-1} \check{H}_t \\ \bar{X}_t &= \left(Q_t^{-1} + H_t^\top R_t^{-1} H_t - B A^{-1} B^\top\right)^{-1} \\ K_t &\equiv \bar{X}_t \left(H_t^\top - B A^{-1} \check{H}_t^\top\right) R_t^{-1} \\ \bar{x}_t &= \hat{x}_t + K_t (z_t - \hat{z}_t) \end{aligned}$$

5 Optimal Linear Estimation and Joseph Form of the dKF

It is also possible to obtain the dKF very directly as an optimal least-squares linear estimator (*cf.* [?]). It turns out that optimal least squares estimators for linear Gaussian systems are always *linear* estimators ([?]), so we hypothesize the general linear form and optimize it.

Consider all linear estimators of the form:

$$\bar{x}(K_x, K_u, K_z) \equiv K_x \bar{x}_{t-1} + K_u u_{t-1} + K_z z_t$$

for gain matrices K_x, K_u and K_z. The discussion of uncertainties in section (4) showed that

$$\begin{aligned} x_{t-1} &= \bar{x}_{t-1} + n_{\bar{x}_{t-1}} && n_{\bar{x}_{t-1}} \in \mathcal{N}(0, \bar{X}_{t-1}) \\ x_t &= \hat{x}_t + F_t n_{\bar{x}_{t-1}} + n_{x_t} && n_{x_t} \in \mathcal{N}(0, Q_t) \\ z_t &= \hat{z}_t + H'_t n_{\bar{x}_{t-1}} + H_t n_{x_t} + n_{z_t} && n_{z_t} \in \mathcal{N}(0, R_t) \end{aligned}$$

where $n_{\bar{x}_{t-1}}$, n_{x_t} and n_{z_t} are independent zero-mean Gaussian perturbations. This implies:

$$\begin{aligned} & x_t - \bar{x}(K_x, K_u, K_z) \\ &\quad = (\hat{x}_t + F_t n_{\bar{x}_{t-1}} + n_{x_t}) - (K_x \bar{x}_{t-1} + K_u u_{t-1} + K_z (\hat{z}_t + H'_t n_{\bar{x}_{t-1}} + H_t n_{x_t} + n_{z_t})) \\ &\quad = (F_t - K_x - K_z H'_t)\, \bar{x}_{t-1} + (G_t - K_u - K_z H_t G_t)\, u_{t-1} \\ &\quad\quad + (F_t - K_z H'_t)\, n_{\bar{x}_{t-1}} + (1 - K_z H_t)\, n_{x_t} + K_z n_{z_t} \end{aligned}$$

The $\bar{x}_{t-1}$ and u_{t-1} terms must vanish if the estimator is to be unbiassed:

$$\begin{aligned} K_x &= F_t - K_z H'_t \\ K_u &= G_t - K_z H_t G_t \end{aligned}$$

Hence

$$\begin{aligned} x_t - \bar{x}(K_z) &= (F_t - K_z H'_t)\, n_{\bar{x}_{t-1}} + (1 - K_z H_t)\, n_{x_t} + K_z n_{z_t} \\ \bar{X}(K_z) &\equiv \langle (x_t - \bar{x}(K_z))(x_t - \bar{x}(K_z))^\top \rangle \\ &= (F_t - K_z H'_t)\, \bar{X}_{t-1}\, (F_t - K_z H'_t)^\top + (1 - K_z H_t)\, Q_t\, (1 - K_z H_t)^\top + K_z R_t K_z^\top \end{aligned}$$

We minimize this prediction covariance with respect to K_z to find the optimal predictor. Noting that the minimum of a quadratic cost function $(A_0 - A_1 K - K A_1^\top + K A_2 K)$ is at $K = A_1 A_2^{-1}$ and has value $A_0 - A_1 A_2^{-1} A_1^\top$, we immediately find that the optimal gain and covariance are the dKF ones in information form:

$$\begin{aligned} K_z^{\text{opt}} &= K_t \equiv \left(F_t \bar{X}_{t-1} H'^\top_t + Q_t H_t^\top \right) \left(H'_t \bar{X}_{t-1} H'_t + H_t Q_t H_t^\top + R_t \right)^{-1} \\ \bar{X}(K_t) &= (F_t - K_t H'_t)\, \bar{X}_{t-1}\, (F_t - K_t H'_t)^\top + (1 - K_t H_t)\, Q_t\, (1 - K_t H_t)^\top + K_t R_t K_t^\top \\ &= Q_t + F_t \bar{X}_{t-1} F_t^\top \\ &\quad - \left(F_t \bar{X}_{t-1} H'^\top_t + Q_t H_t^\top \right) \left(H'_t \bar{X}_{t-1} H'_t + H_t Q_t H_t^\top + R_t \right)^{-1} \left(F_t \bar{X}_{t-1} H'^\top_t + Q_t H_t^\top \right)^\top \\ &= \bar{X}_t \end{aligned}$$

The first form of $\bar{X}_t$ is known as the Joseph form: it is an explicit sum of positive definite matrices and is very stable under round-off errors in the gain because it is stationary with respect to K_t. The more usual second form is shorter, but it is no longer stationary with respect to errors in K_t. It may also involve the difference of two nearly equal matrices, so it is intrinsically less stable and in numerical calculations may become non-positive-definite.

6 Other Forms of the dKF

The state and information space forms given above are probably the most useful forms of the dKF for practical applications, but many equivalent forms can be derived. We will only give a few of them here, with brief notes on their derivation.

The augmented state Kalman update formula $\bar{\mathsf{X}}_t^{-1}\,\bar{\mathsf{x}}_t = \hat{\mathsf{X}}_t^{-1}\,\hat{\mathsf{x}}_t + \mathsf{H}_t^\top \boldsymbol{R}_t^{-1}\,\mathsf{z}_t$ produces the following information space state update rule:

$$\begin{aligned}
\boldsymbol{A} &\equiv \left(\boldsymbol{Q}_t^{-1}\boldsymbol{F}_t + \boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1}\check{\boldsymbol{H}}_t\right)\left(\bar{\boldsymbol{X}}_{t-1}^{-1} + \boldsymbol{F}_t^\top \boldsymbol{Q}_t^{-1}\boldsymbol{F}_t + \check{\boldsymbol{H}}_t^\top \boldsymbol{R}_t^{-1}\check{\boldsymbol{H}}_t\right)^{-1} \\
\bar{\boldsymbol{X}}_t^{-1}\bar{\boldsymbol{x}}_t &= \boldsymbol{A}\,\bar{\boldsymbol{X}}_{t-1}^{-1}\bar{\boldsymbol{x}}_{t-1} + \left(\boldsymbol{I} - \boldsymbol{A}\,\boldsymbol{F}_t^\top\right)\boldsymbol{Q}_t^{-1}\boldsymbol{G}_t\,\boldsymbol{u}_{t-1} + \left(\boldsymbol{H}_t^\top - \boldsymbol{A}\,\check{\boldsymbol{H}}_t\right)\boldsymbol{R}_t^{-1}\boldsymbol{z}_t
\end{aligned}$$

Starting from the information forms of the gain $\boldsymbol{K}_t$ and covariance $\bar{\boldsymbol{X}}_t$, defining the 'retrodicted' covariance

$$\check{\boldsymbol{X}}_{t-1}^{-1} \equiv \bar{\boldsymbol{X}}_{t-1}^{-1} + \boldsymbol{F}_t^\top \boldsymbol{Q}_t^{-1}\boldsymbol{F}_t$$

and using the transfer identity (9) in the form

$$\left(\check{\boldsymbol{X}}_{t-1}^{-1} + \check{\boldsymbol{H}}_t^\top \boldsymbol{R}_t^{-1}\check{\boldsymbol{H}}_t\right)^{-1}\check{\boldsymbol{H}}_t^\top = \check{\boldsymbol{X}}_{t-1}\check{\boldsymbol{H}}_t^\top\left(\boldsymbol{R}_t + \check{\boldsymbol{H}}_t\check{\boldsymbol{X}}_{t-1}\check{\boldsymbol{H}}_t^\top\right)^{-1}\boldsymbol{R}_t$$

(for $\boldsymbol{K}_t$) and inversion identity (4) on $\bar{\boldsymbol{X}}_t$, simplifying the $\boldsymbol{R}_t$ term, applying transfer identity (9) in the form

$$\hat{\boldsymbol{X}}_t^{-1}\boldsymbol{F}_t\bar{\boldsymbol{X}}_{t-1} = \boldsymbol{Q}_t^{-1}\boldsymbol{F}_t\check{\boldsymbol{X}}_{t-1}$$

and using inversion identity (5) on the leading $\hat{\boldsymbol{X}}_t$ term produces:

$$\begin{aligned}
\boldsymbol{B} &\equiv \boldsymbol{H}_t^\top - \hat{\boldsymbol{X}}_t^{-1}\boldsymbol{F}_t\bar{\boldsymbol{X}}_{t-1}\check{\boldsymbol{H}}_t^\top \\
&= \boldsymbol{H}_t^\top - \boldsymbol{Q}_t^{-1}\boldsymbol{F}_t\check{\boldsymbol{X}}_{t-1}\check{\boldsymbol{H}}_t^\top \\
\boldsymbol{K}_t &= \bar{\boldsymbol{X}}_t\boldsymbol{B}\left(\boldsymbol{R}_t + \check{\boldsymbol{H}}_t\check{\boldsymbol{X}}_{t-1}\check{\boldsymbol{H}}_t^\top\right)^{-1} \\
\bar{\boldsymbol{X}}_t^{-1} &= \hat{\boldsymbol{X}}_t^{-1} + \boldsymbol{B}\left(\boldsymbol{R}_t + \check{\boldsymbol{H}}_t\check{\boldsymbol{X}}_{t-1}\check{\boldsymbol{H}}_t^\top\right)^{-1}\boldsymbol{B}^\top \\
&= \boldsymbol{Q}_t^{-1} - \boldsymbol{Q}_t^{-1}\boldsymbol{F}_t^\top\check{\boldsymbol{X}}_{t-1}\boldsymbol{F}_t\boldsymbol{Q}_t^{-1} + \boldsymbol{B}\left(\boldsymbol{R}_t + \check{\boldsymbol{H}}_t\check{\boldsymbol{X}}_{t-1}\check{\boldsymbol{H}}_t^\top\right)^{-1}\boldsymbol{B}^\top
\end{aligned}$$

These equations clearly reduce to the usual information form of the Kalman filter if $\check{\boldsymbol{H}}_t$ vanishes.

A Useful Identities

This section presents some useful matrix identities. It is easy to check that the inverse of a nonsingular block 2×2 matrix with nonsingular diagonal blocks can be expressed as follows:

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} \alpha^{-1} & -\alpha^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}\alpha^{-1} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\alpha^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \end{pmatrix} \tag{1}$$

$$= \begin{pmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\delta^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\delta^{-1} \\ -\delta^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \delta^{-1} \end{pmatrix} \tag{2}$$

where

$$\begin{aligned}
\boldsymbol{\alpha} &\equiv \boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C} \\
\boldsymbol{\delta} &\equiv \boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}
\end{aligned}$$

A more symmetric form of this is:

$$\begin{pmatrix} \boldsymbol{A}+\boldsymbol{BDC} & -\boldsymbol{BD} \\ -\boldsymbol{DC} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A}^{-1} & \boldsymbol{A}^{-1}\boldsymbol{B} \\ \boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{D}^{-1}+\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} \end{pmatrix} \tag{3}$$

Replacing $\boldsymbol{A}$ with $\boldsymbol{A}^{-1}$ and $\boldsymbol{D}$ with $\pm\boldsymbol{D}$ and optionally multiplying on the left and right by $\boldsymbol{A}^{-1}$, the equivalence of the top left hand corners of (1) and (2) yields the following equivalent *inversion identities*[2]:

$$\left(\boldsymbol{A}^{-1}-\boldsymbol{BD}^{-1}\boldsymbol{C}\right)^{-1} = \boldsymbol{A}+\boldsymbol{AB}\left(\boldsymbol{D}-\boldsymbol{CAB}\right)^{-1}\boldsymbol{CA} \tag{4}$$

$$\left(\boldsymbol{A}^{-1}+\boldsymbol{BD}^{-1}\boldsymbol{C}\right)^{-1} = \boldsymbol{A}-\boldsymbol{AB}\left(\boldsymbol{D}+\boldsymbol{CAB}\right)^{-1}\boldsymbol{CA} \tag{5}$$

$$\left(\boldsymbol{A}-\boldsymbol{ABD}^{-1}\boldsymbol{CA}\right)^{-1} = \boldsymbol{A}^{-1}+\boldsymbol{B}\left(\boldsymbol{D}-\boldsymbol{CAB}\right)^{-1}\boldsymbol{C} \tag{6}$$

$$\left(\boldsymbol{A}+\boldsymbol{ABD}^{-1}\boldsymbol{CA}\right)^{-1} = \boldsymbol{A}^{-1}-\boldsymbol{B}\left(\boldsymbol{D}+\boldsymbol{CAB}\right)^{-1}\boldsymbol{C} \tag{7}$$

Using the top right corners of (1) and (2) or factorizing $\boldsymbol{B}+\boldsymbol{BD}^{-1}\boldsymbol{CAB}$ and cross-multiplying produces the following *transfer identities* that move a compound term through a rectangular $\boldsymbol{B}$ matrix:

$$\left(\boldsymbol{A}^{-1}+\boldsymbol{BD}^{-1}\boldsymbol{C}\right)\boldsymbol{AB} = \boldsymbol{BD}^{-1}\left(\boldsymbol{D}+\boldsymbol{CAB}\right) \tag{8}$$

$$\boldsymbol{AB}\left(\boldsymbol{D}+\boldsymbol{CAB}\right)^{-1} = \left(\boldsymbol{A}^{-1}+\boldsymbol{BD}^{-1}\boldsymbol{C}\right)^{-1}\boldsymbol{BD}^{-1} \tag{9}$$

$$\boldsymbol{B}\left(\boldsymbol{D}+\boldsymbol{CAB}\right)^{-1}\boldsymbol{D} = \boldsymbol{A}^{-1}\left(\boldsymbol{A}^{-1}+\boldsymbol{BD}^{-1}\boldsymbol{C}\right)^{-1}\boldsymbol{B} \tag{10}$$

From $\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}=\boldsymbol{A}^{-1}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{B}^{-1}$ or from (8) with $\boldsymbol{A}=\boldsymbol{B}=\boldsymbol{C}$ we get:

$$\left(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}\right)^{-1} = \boldsymbol{A}\left(\boldsymbol{A}+\boldsymbol{B}\right)^{-1}\boldsymbol{B}=\boldsymbol{B}\left(\boldsymbol{A}+\boldsymbol{B}\right)^{-1}\boldsymbol{A} \tag{11}$$

$$\left(\boldsymbol{A}+\boldsymbol{AB}^{-1}\boldsymbol{A}\right)^{-1} = \boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{A}+\boldsymbol{B}\right)^{-1}=\left(\boldsymbol{A}+\boldsymbol{B}\right)^{-1}\boldsymbol{BA}^{-1} \tag{12}$$

[2]Mnemonic: Matrices and inverses in products occur in cyclic order ABDCA..., *i.e.*clockwise around the 2×2 matrix, and clockwise around each of the 2×2 blocks in the tridiagonal case.