Some common formulae/proofs used in data assimilation

Ross Bannister

November 18, 2022

Contents

1 The Sherman-Morrison-Woodbury formula

The S-M-W formula (1) is commonly used in data assimilation. Here we quote it and then demonstrate that it is correct.

Aim

Prove that the following identity holds:

$$\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R}+\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\right)^{-1}=\left(\mathbf{P}_{\mathrm{f}}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}. \tag{1}$$

Proof

Take the inverse brackets to the other side of the equation:

$$\left(\mathbf{P}_{\mathrm{f}}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\stackrel{?}{=}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{R}+\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\right).$$

Take the $\mathbf{P}^{\mathrm{f}}$ inside the bracket on the left hand side, and take the $\mathbf{R}^{-1}$ inside the bracket on the right hand side:

$$\left(\mathbf{I}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{P}_{\mathrm{f}}\right)\mathbf{H}^{\mathrm{T}}\stackrel{?}{=}\mathbf{H}^{\mathrm{T}}\left(\mathbf{I}+\mathbf{R}^{-1}\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\right).$$

Subtract $\mathbf{H}^{\mathrm{T}}$ from each side:

$$\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\stackrel{?}{=}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}. \qquad \checkmark$$

2 The Kalman gain as a maximum a-posteriori

A gain operator translates an innovation to an analysis increment. Here we will derive the Kalman gain operator from the standard cost function used in data assimilation. The cost function is

$$J[\delta\mathbf{x}] = \frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{P}_{\mathrm{f}}^{-1}\delta\mathbf{x} + \frac{1}{2}\left(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{b}}) - \mathbf{H}\delta\mathbf{x}\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{b}}) - \mathbf{H}\delta\mathbf{x}\right), \tag{2}$$

where $\mathbf{x} = \mathbf{x}_{\mathrm{b}} + \delta\mathbf{x}$ and $\mathbf{h}(x) \approx \mathbf{h}(\mathbf{x}_{\mathrm{b}}) + \mathbf{H}\delta\mathbf{x}$. The first derivative is the column vector $\nabla_{\delta\mathbf{x}}J$ of derivatives with element i $\partial J/\partial\delta x_i$:

$$\nabla_{\delta\mathbf{x}}J = \mathbf{P}_{\mathrm{f}}^{-1}\delta\mathbf{x} - \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{b}}) - \mathbf{H}\delta\mathbf{x}\right). \tag{3}$$

At the analysis (the point $\delta\mathbf{x} = \delta\mathbf{x}_{\mathrm{a}}$ where the cost function is at a minimum), this gradient is zero:

$$\mathbf{P}_{\mathrm{f}}^{-1}\delta\mathbf{x}_{\mathrm{a}} - \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{b}}) - \mathbf{H}\delta\mathbf{x}_{\mathrm{a}}\right) = 0.$$

Factorising the $\delta\mathbf{x}_{\mathrm{a}}$:

$$\left(\mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)\delta\mathbf{x}_{\mathrm{a}} = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{b}})\right).$$

Rearranging gives the analysis increment:

$$\begin{aligned}
\delta\mathbf{x}_{\mathrm{a}} = \mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}} &= \left(\mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{b}})\right) \\
&= \mathbf{K}\left(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{b}})\right) && (4) \\
\text{where } \mathbf{K} &= \left(\mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}. && (5)
\end{aligned}$$

The Kalman gain is usually given in another form, which is found from (5) using the S-M-W formula (1):

$$\mathbf{K} = \mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\right)^{-1}. \tag{6}$$

An alternative derivation of the Kalman gain is given in Sect. 4, which is based on the 'minimum variance' principle.

---

3 The analysis error covariance matrix

The background, analysis, and observations are all imperfect quantities. First write each in terms of a truth and an error, and note the definitions of the background, analysis, and observation error covariance matrices:

$$\begin{aligned}
\mathbf{x}_{\mathrm{b}} &= \mathbf{x}_{\mathrm{t}} + \boldsymbol{\varepsilon}_{\mathrm{b}}, & \mathbf{P}_{\mathrm{f}} &= \left\langle \boldsymbol{\varepsilon}_{\mathrm{b}}\boldsymbol{\varepsilon}_{\mathrm{b}}^{\mathrm{T}}\right\rangle, && (7a) \\
\mathbf{x}_{\mathrm{a}} &= \mathbf{x}_{\mathrm{t}} + \boldsymbol{\varepsilon}_{\mathrm{a}}, & \mathbf{P}_{\mathrm{a}} &= \left\langle \boldsymbol{\varepsilon}_{\mathrm{a}}\boldsymbol{\varepsilon}_{\mathrm{a}}^{\mathrm{T}}\right\rangle, && (7b) \\
\mathbf{y} &= \mathbf{h}(\mathbf{x}_{\mathrm{t}}) + \boldsymbol{\varepsilon}_{\mathrm{o}}, & \mathbf{R} &= \left\langle \boldsymbol{\varepsilon}_{\mathrm{o}}\boldsymbol{\varepsilon}_{\mathrm{o}}^{\mathrm{T}}\right\rangle. && (7c)
\end{aligned}$$

Substitute these into the gain result (4):

$$\begin{aligned}
\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}} &= \mathbf{K}\left(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{b}})\right) \\
\mathbf{x}_{\mathrm{t}} + \boldsymbol{\varepsilon}_{\mathrm{a}} - (\mathbf{x}_{\mathrm{t}} + \boldsymbol{\varepsilon}_{\mathrm{b}}) &= \mathbf{K}\left(\mathbf{h}(\mathbf{x}_{\mathrm{t}}) + \boldsymbol{\varepsilon}_{\mathrm{o}} - \mathbf{h}(\mathbf{x}_{\mathrm{t}} + \boldsymbol{\varepsilon}_{\mathrm{b}})\right) \\
&\approx \mathbf{K}\left(\mathbf{h}(\mathbf{x}_{\mathrm{t}}) + \boldsymbol{\varepsilon}_{\mathrm{o}} - \mathbf{h}(\mathbf{x}_{\mathrm{t}}) - \mathbf{H}\boldsymbol{\varepsilon}_{\mathrm{b}}\right) \\
\boldsymbol{\varepsilon}_{\mathrm{a}} - \boldsymbol{\varepsilon}_{\mathrm{b}} &\approx \mathbf{K}\left(\boldsymbol{\varepsilon}_{\mathrm{o}} - \mathbf{H}\boldsymbol{\varepsilon}_{\mathrm{b}}\right) && (8) \\
\text{or } \boldsymbol{\varepsilon}_{\mathrm{a}} &\approx \boldsymbol{\varepsilon}_{\mathrm{b}} + \mathbf{K}\left(\boldsymbol{\varepsilon}_{\mathrm{o}} - \mathbf{H}\boldsymbol{\varepsilon}_{\mathrm{b}}\right).
\end{aligned}$$

Analysis error covariance is $\left\langle \varepsilon_{\mathrm{a}} \varepsilon_{\mathrm{a}}^{\mathrm{T}} \right\rangle$. We develop this and assume that the background and observation errors are uncorrelated:

$$
\begin{aligned}
\left\langle \varepsilon_{\mathrm{a}} \varepsilon_{\mathrm{a}}^{\mathrm{T}} \right\rangle &= \left\langle \left[\varepsilon_{\mathrm{b}} + \mathbf{K}\left(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{b}}\right)\right)\right] \left[\varepsilon_{\mathrm{b}} + \mathbf{K}\left(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{b}}\right)\right)\right]^{\mathrm{T}} \right\rangle \\
&= \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle + \left\langle \varepsilon_{\mathrm{b}} \left(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{b}}\right)\right)^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \right\rangle + \left\langle \mathbf{K}\left(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{b}}\right)\right) \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle + \left\langle \mathbf{K}\left(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{b}}\right)\right) \left(\varepsilon_{\mathrm{o}} - \mathbf{H}\varepsilon_{\mathrm{b}}\right)\right)^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \right\rangle \\
&= \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle + \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{\mathrm{o}}^{\mathrm{T}} \right\rangle \mathbf{K}^{\mathrm{T}} - \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle \mathbf{H}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} + \mathbf{K} \left\langle \varepsilon_{\mathrm{o}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle - \mathbf{K}\mathbf{H} \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle + \\
&\quad \mathbf{K} \left\langle \varepsilon_{\mathrm{o}} \varepsilon_{\mathrm{o}}^{\mathrm{T}} \right\rangle \mathbf{K}^{\mathrm{T}} - \mathbf{K} \left\langle \varepsilon_{\mathrm{o}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle \mathbf{H}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} - \mathbf{K}\mathbf{H} \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{0}^{\mathrm{T}} \right\rangle \mathbf{K}^{\mathrm{T}} + \mathbf{K}\mathbf{H} \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle \mathbf{H}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \\
&= \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle - \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle \mathbf{H}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} - \mathbf{K}\mathbf{H} \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle + \\
&\quad \mathbf{K} \left\langle \varepsilon_{\mathrm{o}} \varepsilon_{\mathrm{o}}^{\mathrm{T}} \right\rangle \mathbf{K}^{\mathrm{T}} + \mathbf{K}\mathbf{H} \left\langle \varepsilon_{\mathrm{b}} \varepsilon_{\mathrm{b}}^{\mathrm{T}} \right\rangle \mathbf{H}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}}.
\end{aligned}
$$

The expectations evaluate to the error covariance matrices defined above, giving an expression that can be factorised:

$$
\begin{aligned}
\mathbf{P}_{\mathrm{a}} &= \mathbf{P}_{\mathrm{f}} - \mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} - \mathbf{K}\mathbf{H}\mathbf{P}_{\mathrm{f}} + \mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}} + \mathbf{K}\mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}}. \\
&= \mathbf{P}_{\mathrm{f}} - \mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} - \mathbf{K}\mathbf{H}\mathbf{P}_{\mathrm{f}} + \mathbf{K}\left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}}\right) \mathbf{K}^{\mathrm{T}}.
\end{aligned} \tag{9}
$$

We know that the Kalman gain has the form (6), which can be substituted into the above (remembering that the covariance matrices are symmetric, i.e. $\mathbf{P}_{\mathrm{f}} = \mathbf{P}_{\mathrm{f}}^{\mathrm{T}}$ and $\mathbf{R} = \mathbf{R}^{\mathrm{T}}$):

$$
\begin{aligned}
\mathbf{P}_{\mathrm{a}} &= \mathbf{P}_{\mathrm{f}} - \mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}}\right)^{-1} \mathbf{H}\mathbf{P}^{\mathrm{f}} - \\
&\quad \mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}}\right)^{-1} \mathbf{H}\mathbf{P}_{\mathrm{f}} + \\
&\quad \mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}}\right)^{-1} \left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}}\right) \left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}}\right)^{-1} \mathbf{H}\mathbf{P}^{\mathrm{f}} \\
&= \mathbf{P}_{\mathrm{f}} - 2\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}}\right)^{-1} \mathbf{H}\mathbf{P}^{\mathrm{f}} + \mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}}\right)^{-1} \mathbf{H}\mathbf{P}^{\mathrm{f}} \\
&= \mathbf{P}_{\mathrm{f}} - \mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}}\right)^{-1} \mathbf{H}\mathbf{P}^{\mathrm{f}} \\
&= \left[\mathbf{I} - \mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}} \mathbf{H}^{\mathrm{T}}\right)^{-1} \mathbf{H}\right] \mathbf{P}_{\mathrm{f}} \\
&= \left[\mathbf{I} - \mathbf{K}\mathbf{H}\right] \mathbf{P}_{\mathrm{f}}.
\end{aligned} \tag{10}
$$

Notice that, because of the minus sign, $\mathbf{P}_{\mathrm{a}}$ is a smaller valued matrix than $\mathbf{P}_{\mathrm{f}}$. $\mathbf{P}_{\mathrm{a}}$ has been found on the basis of the gain having form (6). There is an alternative form of this analysis error covariance matrix, which uses form (5) of the Kalman gain. Substitute (5) into (10) and develop:

$$
\begin{aligned}
\mathbf{P}_{\mathrm{a}} &= \left[\mathbf{I} - \left(\mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right] \mathbf{P}_{\mathrm{f}} \\
&= \left(\mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right)^{-1} \left[\left(\mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right) - \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right] \mathbf{P}_{\mathrm{f}} \\
&= \left(\mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right)^{-1} \left[\mathbf{P}_{\mathrm{f}}^{-1}\right] \mathbf{P}_{\mathrm{f}} \\
&= \left(\mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{H}\right)^{-1}.
\end{aligned} \tag{11}
$$

---

4 The Kalman gain as a 'minimum variance' gain

The update equation with an unspecified gain, $\mathbf{G}$ is

$$
\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}} = \mathbf{G}(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{b}})), \tag{12}
$$

which is rather like (8) in Sect. 3. As was done in Sect. 3 we use this to define an error equation, and then an error covariance equation using definitions (7). Assuming that forecast and observation errors are uncorrelated, the analysis error covariance equation is the same as (9) in Sect. 3 but with $\mathbf{K} \to \mathbf{G}$:

$$\begin{aligned} \mathbf{P}_{\mathrm{a}} &= \mathbf{P}_{\mathrm{f}} - \mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}} - \mathbf{G}\mathbf{H}\mathbf{P}_{\mathrm{f}} + \mathbf{G}\left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\right)\mathbf{G}^{\mathrm{T}} \\ &= \mathbf{P}_{\mathrm{f}} - 2\mathbf{G}\mathbf{H}\mathbf{P}_{\mathrm{f}} + \mathbf{G}\left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\right)\mathbf{G}^{\mathrm{T}}. \end{aligned}$$

Note that each term has to be symmetric, so the second and third terms of the first line must be identical, so they have been combined in the second line. The total variance is the trace of the above:

$$\mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right) = \mathrm{Tr}\left(\mathbf{P}_{\mathrm{f}}\right) - 2\mathrm{Tr}\left(\mathbf{G}\mathbf{H}\mathbf{P}_{\mathrm{f}}\right) + \mathrm{Tr}\left(\mathbf{G}(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}})\mathbf{G}^{\mathrm{T}}\right). \tag{13}$$

We will now find the particular gain matrix that minimises this trace. We do this by finding the matrix $\partial\mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)/\partial\mathbf{G}$, which is a matrix that is the same shape as $\mathbf{G}$. We will decompose (13) into its components, and then find a particular matrix element of the derivative matrix. Decomposing (13) into its components (and letting $\mathbf{C} \equiv \mathbf{H}\mathbf{P}_{\mathrm{f}}$ and $\mathbf{D} \equiv \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}$ as shorthand):

$$\mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right) = \mathrm{Tr}\left(\mathbf{P}_{\mathrm{f}}\right) - 2\sum_{i,j} G_{ij}C_{ji} + \sum_{i,j,k} G_{ij}D_{jk}G_{ik}. \tag{14}$$

We will consider the derivative $\partial\mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)/\partial G_{i'j'}$ as contributions from the three terms in (14). Labelling the three contributions with subscripts in brackets, the first term has zero contribution as it does not depend upon $\mathbf{G}$:

$$\frac{\partial\mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)_{(1)}}{\partial\mathbf{G}} = \mathbf{0}. \tag{15}$$

The second term can be developed as follows:

$$\begin{aligned} \frac{\partial\mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)_{(2)}}{\partial G_{i'j'}} &= -2\sum_{i,j}\delta_{ii'}\delta_{jj'}C_{ji} = -2C_{j'i'} \\ &= -2(\mathbf{H}\mathbf{P}_{\mathrm{f}})_{j'i'} = -2(\mathbf{H}\mathbf{P}_{\mathrm{f}})^{\mathrm{T}}_{i'j'} \\ \text{so } \frac{\partial\mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)_{(2)}}{\partial\mathbf{G}} &= -2\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}. \end{aligned} \tag{16}$$

The third term requires the product rule for differentiation and can be developed as follows:

$$\begin{aligned} \frac{\partial\mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)_{(3)}}{\partial G_{i'j'}} &= \sum_{i,j,k}\frac{\partial G_{ij}}{\partial G_{i'j'}}D_{jk}G_{ik} + \sum_{i,j,k}G_{ij}D_{jk}\frac{\partial G_{ik}}{\partial G_{i'j'}} \\ &= \sum_{i,j,k}\delta_{ii'}\delta_{jj'}D_{jk}G_{ik} + \sum_{i,j,k}G_{ij}D_{jk}\delta_{ii'}\delta_{kj'} \\ &= \sum_{k}D_{j'k}G_{i'k} + \sum_{j}G_{i'j}D_{jj'}. \end{aligned}$$

Matrix $\mathbf{D}$ is symmetric so we can swap the indices (do this in the first term only and then replace summation variable $k \to j$), which proves that the two terms are identical:

$$\begin{aligned} \frac{\partial\mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)_{(3)}}{\partial G_{i'j'}} &= \sum_{j}D_{jj'}G_{i'j} + \sum_{j}G_{i'j}D_{jj'} = 2\sum_{j}G_{i'j}D_{jj'} = 2(\mathbf{G}\mathbf{D})_{i'j'} \\ &= 2\left(\mathbf{G}(\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})\right)_{i'j'} \\ \text{so } \frac{\partial\mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)_{(3)}}{\partial\mathbf{G}} &= 2\mathbf{G}(\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}). \end{aligned} \tag{17}$$

Adding-up the three terms gives the total derivative:

$$\begin{aligned}\frac{\partial \mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)}{\partial \mathbf{G}} &= \frac{\partial \mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)_{(1)}}{\partial \mathbf{G}} + \frac{\partial \mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)_{(2)}}{\partial \mathbf{G}} + \frac{\partial \mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)_{(3)}}{\partial \mathbf{G}} \\ &= \mathbf{0} - 2\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + 2\mathbf{G}(\mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}). \qquad (18)\end{aligned}$$

Setting this to zero for the turning point gives:

$$\mathbf{G} = \mathbf{K} = \mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\right)^{-1}, \qquad (19)$$

which is the Kalman gain, as derived by the maximum a-posteriori, (6) in Sect. 2.

Does this correspond to a maximum or a minimum total variance? In order to find out, we return to the expanded forms of the derivatives above and differentiate again. The first two derivative terms, (15) and (16), have zero contribution as they do not depend on $\mathbf{G}$. Differentiating the third term:

$$\frac{\partial^2 \mathrm{Tr}\left(\mathbf{P}_{\mathrm{a}}\right)_{(3)}}{\partial G_{i'j'}\partial G_{i''j''}} = 2\sum_j \delta_{i'i''}\delta_{jj''}D_{jj'} = 2\delta_{i'i''}D_{j''j'} = 2\delta_{i'i''}D_{j'j''}. \qquad (20)$$

This is a four-dimensional matrix (a hyper cuboid shape), or alternatively a matrix of matrices. Its elements are zero valued, except where $i' = i''$ (so a block diagonal matrix of matrices), and the matrix stored at that point is $\mathbf{D}$. $\mathbf{D}$ is a positive-definite matrix and so (20) is also positive-definite. This means that the choice $\mathbf{G} = \mathbf{K}$ represents the minimum total variance.

---

5 The Hessian

The Hessian is the matrix of second derivatives of the cost function (2). The incremental cost function, linearised about $\mathbf{x}_{\mathrm{b}}$ is given as (2), where $\mathbf{x} = \mathbf{x}_{\mathrm{b}} + \delta\mathbf{x}$. The first derivative is the column vector $\nabla_{\delta\mathbf{x}}J$ of derivatives with element i $\partial J/\partial\delta x_i$:

$$\nabla_{\delta\mathbf{x}}J = \mathbf{P}_{\mathrm{f}}^{-1}\delta\mathbf{x} - \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{h}(\mathbf{x}_{\mathrm{b}}) - \mathbf{H}\delta\mathbf{x}\right).$$

The second derivative is the matrix $\nabla^2_{\delta\mathbf{x}}J$ of second derivatives with elements i, j $\partial^2 J/(\partial x_i \partial x_j)$:

$$\nabla^2_{\delta\mathbf{x}}J = \mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}. \qquad (21)$$

Note that this is the inverse of the analysis error covariance matrix (11).

---

6 A further identity

Aim

Prove that the following identity holds:

$$\left(\mathbf{R} + \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\right)^{-1} = \mathbf{R}^{-1}\left(\mathbf{R} - \mathbf{H}\mathbf{P}_{\mathrm{a}}\mathbf{H}^{\mathrm{T}}\right)\mathbf{R}^{-1}. \qquad (22)$$

Proof

First note the form of the S-M-W formula (1). Since $\mathbf{P}_\mathrm{f}$ in this identity is really an arbitrary invertible square matrix, we may replace $\mathbf{P}_\mathrm{f} \to -\mathbf{P}_\mathrm{a}$:

$$-\mathbf{P}_\mathrm{a}\mathbf{H}^\mathrm{T}\left(\mathbf{R}-\mathbf{H}\mathbf{P}_\mathrm{a}\mathbf{H}^\mathrm{T}\right)^{-1} = \left(\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}-\mathbf{P}_\mathrm{a}^{-1}\right)^{-1}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}, \tag{23}$$

or rearranged:

$$-\left(\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}-\mathbf{P}_\mathrm{a}^{-1}\right)\mathbf{P}_\mathrm{a}\mathbf{H}^\mathrm{T} = \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\left(\mathbf{R}-\mathbf{H}\mathbf{P}_\mathrm{a}\mathbf{H}^\mathrm{T}\right). \tag{24}$$

We also know the relationship between $\mathbf{P}^\mathrm{a}$ and $\mathbf{P}^\mathrm{f}$ (11). Inverting (11) gives $\mathbf{P}_\mathrm{a}^{-1}$ (which is the Hessian (21)):

$$\mathbf{P}_\mathrm{a}^{-1} = \mathbf{P}_\mathrm{f}^{-1}+\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}. \tag{25}$$

Now develop (22) to show that it is correct (start by left multiplying by $\mathbf{H}^\mathrm{T}$):

$$\begin{aligned}
\mathbf{H}^\mathrm{T}\left(\mathbf{R}+\mathbf{H}\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T}\right)^{-1} &\stackrel{?}{=} \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\left(\mathbf{R}-\mathbf{H}\mathbf{P}_\mathrm{a}\mathbf{H}^\mathrm{T}\right)\mathbf{R}^{-1} \\
\text{use (24):} &\stackrel{?}{=} -\left(\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}-\mathbf{P}_\mathrm{a}^{-1}\right)\mathbf{P}_\mathrm{a}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1} \\
\text{use (25):} &\stackrel{?}{=} -\left(\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}-\mathbf{P}_\mathrm{f}^{-1}-\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}\right)\left(\mathbf{P}_\mathrm{f}^{-1}+\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1} \\
&\stackrel{?}{=} \mathbf{P}_\mathrm{f}^{-1}\left(\mathbf{P}_\mathrm{f}^{-1}+\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1} \\
\text{use (1):} &\stackrel{?}{=} \mathbf{P}_\mathrm{f}^{-1}\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T}\left(\mathbf{R}+\mathbf{H}\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T}\right)^{-1} \\
&\stackrel{?}{=} \mathbf{H}^\mathrm{T}\left(\mathbf{R}+\mathbf{H}\mathbf{P}_\mathrm{f}\mathbf{H}^\mathrm{T}\right)^{-1}. \qquad \checkmark
\end{aligned}$$