M.Sc. Course on Operational Data Assimilation Techniques (MTMD02): Solutions for Part I

1. **Useful formula related to the Sherman-Morrison-Woodbury formula**
   Multiply the identity from the left with $\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$ and from the right with $\mathbf{R} + \mathbf{HBH}^{\mathrm{T}}$

$$(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})\mathbf{BH}^{\mathrm{T}} = \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{R} + \mathbf{HBH}^{\mathrm{T}}).$$

   Take the terms outside of the brackets inside and cancel where appropriate

$$\begin{aligned} \mathbf{B}^{-1}\mathbf{BH}^{\mathrm{T}} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{HBH}^{\mathrm{T}} &= \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{R} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{HBH}^{\mathrm{T}}, \\ \mathbf{H}^{\mathrm{T}} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{HBH}^{\mathrm{T}} &= \mathbf{H}^{\mathrm{T}} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{HBH}^{\mathrm{T}}. \end{aligned}$$

   Each side is identical and so the identity is correct.

---

2. **The Euler-Lagrange equations and the method of representers**
   Substitute the forms (28) and (29) into (23a) and (23b). First (23a):

$$W_{\mathrm{ic}}\left\{\phi_{\mathrm{B}}(x,0) + \sum_{i=1}^{M}\beta_i r_i(x,0) - I(x)\right\} - \sum_{i=1}^{M}\beta_i\alpha_i(x,0) = 0.$$

$$\text{Use (3) } W_{\mathrm{ic}}\sum_{i=1}^{M}\beta_i r_i(x,0) - \sum_{i=1}^{M}\beta_i\alpha_i(x,0) = 0,$$

$$\text{and now use (27a) to rewrite the l.h.s. } W_{\mathrm{ic}}\sum_{i=1}^{M}\beta_i W_{\mathrm{ic}}^{-1}\alpha_i(x,0) - \sum_{i=1}^{M}\beta_i\alpha_i(x,0) = 0,$$

   where l.h.s. equals r.h.s. A similar procedure for (23b):

$$W_{\mathrm{bc}}\left\{\phi_{\mathrm{B}}(0,t) + \sum_{i=1}^{M}\beta_i r_i(0,t) - B(t)\right\} - u\sum_{i=1}^{M}\beta_i\alpha_i(0,t) = 0.$$

$$\text{Use (4) } W_{\mathrm{bc}}\sum_{i=1}^{M}\beta_i r_i(0,t) - u\sum_{i=1}^{M}\beta_i\alpha_i(0,t) = 0,$$

$$\text{and now use (27b) to rewrite the l.h.s. } W_{\mathrm{bc}}\sum_{i=1}^{M}\beta_i W_{\mathrm{bc}}^{-1}u\alpha_i(0,t) - u\sum_{i=1}^{M}\beta_i\alpha_i(0,t) = 0,$$

   where l.h.s. equals r.h.s.

---

3. **Inner product forms**
   For shorthand let $\mathbf{d} = \mathbf{x} - \mathbf{x}_{\mathrm{B}}$, and let $\mathbf{f} = \mathbf{P}^{-1}\mathbf{d}$. Expand the matrix algebra of $\mathbf{d}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{d}$:

$$\mathbf{d}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{d} = \mathbf{d}^{\mathrm{T}}\mathbf{f} = \sum_{i=1}^{n} d_i f_i,$$

   where d_i is the ith element of $\mathbf{d}$ and f_i is the ith element of $\mathbf{f}$. Now,

$$f_i = \sum_{j=1}^{n}(\mathbf{P}^{-1})_{ij}d_j,$$

but since $\mathbf{P}$ is diagonal, $\mathbf{P}^{-1}$ is also diagonal and we can write $(\mathbf{P}^{-1})_{ij} = (\mathbf{P}^{-1})_{ii}\delta_{ij} = \delta_{ij}/(\mathbf{P})_{ii}$ (δ_{ij} is the Kroneker delta-function). First combining the above two results and then substituting the expression involving the delta-function leads to:

$$\mathbf{d}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{d} = \sum_{i=1}^{n} d_i \sum_{j=1}^{n} (\mathbf{P}^{-1})_{ij} d_j = \sum_{i=1}^{n} d_i \sum_{j=1}^{n} \delta_{ij}/(\mathbf{P})_{ii} d_j = \sum_{i=1}^{n} d_i^2/(\mathbf{P})_{ii} = \sum_{i=1}^{n} \left\{(\mathbf{x})_i - (\mathbf{x}_{\mathrm{B}})_i\right\}^2/(\mathbf{P})_{ii}.$$

---

4. **Forward model example**

(a) The state vector $\mathbf{x}$ and the observation vector $\mathbf{y}$ are

$$\mathbf{x} = \begin{pmatrix} T_1^{\mathrm{m}} \\ T_2^{\mathrm{m}} \end{pmatrix}, \qquad \mathbf{y} = (F).$$

(b) The forward operator is

$$\mathbf{h}(\mathbf{x}) = \left(\kappa (T_2^{\mathrm{m}})^4\right),$$

$$\text{the Jacobian is } \mathbf{H} = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 0 & 4\kappa (T_2^{\mathrm{m}})^3 \end{pmatrix},$$

$$\text{and its adjoint is } \mathbf{H}^{\mathrm{T}} = \begin{pmatrix} 0 \\ 4\kappa (T_2^{m})^3 \end{pmatrix}.$$

---

5. **Maximum likelihood solution (MAP)**

(a) Take the logarithm and expand

$$\begin{aligned} J &= c + \frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}), \qquad \left(c = \ln\left[(2\pi)^{p/2}|\mathbf{R}|^{1/2}\right]\right) \\ & \frac{1}{2}\sum_{i'=1}^{p}\left(\mathbf{y}_{i'} - \sum_{j'=1}^{n}\mathbf{H}_{i'j'}\mathbf{x}_{j'}\right)\sum_{i=1}^{p}(\mathbf{R}^{-1})_{i'i}\left(\mathbf{y}_i - \sum_{j=1}^{n}\mathbf{H}_{ij}\mathbf{x}_j\right). \end{aligned}$$

Differentiate w.r.t. $\mathbf{x}_k$ (use the product rule - looks complicated, but is straightforward)

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{x}_k} &= -\frac{1}{2}\sum_{i'=1}^{p}\sum_{j'=1}^{n}\mathbf{H}_{i'j'}\frac{\partial \mathbf{x}_{j'}}{\partial \mathbf{x}_k}\sum_{i=1}^{p}(\mathbf{R}^{-1})_{i'i}\left(\mathbf{y}_i - \sum_{j=1}^{n}\mathbf{H}_{ij}\mathbf{x}_j\right) \\ & -\frac{1}{2}\sum_{i'=1}^{p}\left(\mathbf{y}_{i'} - \sum_{j'=1}^{n}\mathbf{H}_{i'j'}\mathbf{x}_{j'}\right)\sum_{i=1}^{p}(\mathbf{R}^{-1})_{i'i}\sum_{j=1}^{n}\mathbf{H}_{ij}\frac{\partial \mathbf{x}_j}{\partial \mathbf{x}_k}, \\ &= -\frac{1}{2}\sum_{i'=1}^{p}\sum_{j'=1}^{n}\mathbf{H}_{i'j'}\delta_{j'k}\sum_{i=1}^{p}(\mathbf{R}^{-1})_{i'i}\left(\mathbf{y}_i - \sum_{j=1}^{n}\mathbf{H}_{ij}\mathbf{x}_j\right) \\ & -\frac{1}{2}\sum_{i'=1}^{p}\left(\mathbf{y}_{i'} - \sum_{j'=1}^{n}\mathbf{H}_{i'j'}\mathbf{x}_{j'}\right)\sum_{i=1}^{p}(\mathbf{R}^{-1})_{i'i}\sum_{j=1}^{n}\mathbf{H}_{ij}\delta_{jk}, \\ &= -\frac{1}{2}\sum_{i'=1}^{p}\mathbf{H}_{i'k}\sum_{i=1}^{p}(\mathbf{R}^{-1})_{i'i}\left(\mathbf{y}_i - \sum_{j=1}^{n}\mathbf{H}_{ij}\mathbf{x}_j\right) - \frac{1}{2}\sum_{i'=1}^{p}\left(\mathbf{y}_{i'} - \sum_{j'=1}^{n}\mathbf{H}_{i'j'}\mathbf{x}_{j'}\right)\sum_{i=1}^{p}(\mathbf{R}^{-1})_{i'i}\mathbf{H}_{ik}. \end{aligned}$$

Re-index the second summation $i' \to i$, $i \to i'$, $j' \to j$. This becomes

$$\frac{1}{2}\sum_{i=1}^{p}\left(\mathbf{y}_i - \sum_{j=1}^{n}\mathbf{H}_{ij}\mathbf{x}_j\right)\sum_{i'=1}^{p}(\mathbf{R}^{-1})_{ii'}\mathbf{H}_{i'k}.$$

Furthermore, note that $\mathbf{R}$ (and hence $\mathbf{R}^{-1}$) is symmetric and change the order in which the expressions occur

$$\frac{1}{2}\sum_{i'=1}^{p}\mathbf{H}_{i'k}\sum_{i=1}^{p}(\mathbf{R}^{-1})_{i'i}\left(\mathbf{y}_i-\sum_{j=1}^{n}\mathbf{H}_{ij}\mathbf{x}_j\right).$$

This is identical to the first term and so

$$\frac{\partial J}{\partial \mathbf{x}_k} = -\sum_{i'=1}^{p}\mathbf{H}_{i'k}\sum_{i=1}^{p}(\mathbf{R}^{-1})_{i'i}\left(\mathbf{y}_i-\sum_{j=1}^{n}\mathbf{H}_{ij}\mathbf{x}_j\right) = -\sum_{i'=1}^{p}\mathbf{H}^{\mathrm{T}}_{ki'}\sum_{i=1}^{p}(\mathbf{R}^{-1})_{i'i}\left(\mathbf{y}_i-\sum_{j=1}^{n}\mathbf{H}_{ij}\mathbf{x}_j\right).$$

This is the derivative with respect to just one component, $\mathbf{x}_k$. This is component k of the vector/matrix expression

$$\nabla_{\mathbf{x}}J = -\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Hx}).$$

(b) Setting this to zero (at $\mathbf{x}=\mathbf{x}_{\mathrm{A}}$) gives

$$\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}_{\mathrm{A}})=0, \qquad \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{x}_{\mathrm{A}}=\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}, \qquad \mathbf{x}_{\mathrm{A}}=(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}.$$

Assuming that $\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$ is positive definite, then this $\mathbf{x}_{\mathrm{A}}$ minimizes J (maximizes L) and so is the maximum likelihood solution.

---

6. **Minimum (co)variance solution**

What $\hat{\mathbf{x}}$ gives minimum error that is unbiased? The error in $\hat{\mathbf{x}}$ is $\boldsymbol{\varepsilon}_{\mathbf{x}}=\mathbf{x}-\hat{\mathbf{x}}$, where $\mathbf{x}$ is considered the truth. Let $\mathcal{E}$ denote the expectation value (average of doing an experiment many times).

(a) Define $\mathbf{r}$ as the expectation of $\hat{\mathbf{x}}$

$$\mathbf{r} = E[\hat{\mathbf{x}}]=\mathcal{E}[\mathbf{b}+\mathbf{Ay}]=\mathbf{b}+\mathbf{A}\mathcal{E}[\mathbf{y}]=\mathbf{b}+\mathbf{AH}\mathcal{E}[\mathbf{x}]+\mathbf{A}\mathcal{E}[\boldsymbol{\varepsilon}],$$

where we have used the proposed form of the solution, $\hat{\mathbf{x}}=\mathbf{b}+\mathbf{Ay}$ and the relationship between $\mathbf{y}$ and $\mathbf{x}$, $\mathbf{y}=\mathbf{Hx}+\boldsymbol{\varepsilon}$. For unbiased observations $\mathcal{E}[\boldsymbol{\varepsilon}]=0$ and for an unbiased solution, $\mathcal{E}[\mathbf{x}]=\mathcal{E}[\hat{\mathbf{x}}](=\mathbf{r})$. Thus

$$\mathbf{r}=\mathbf{b}+\mathbf{AHr}, \qquad \mathbf{b}=(\mathbf{I}-\mathbf{AH})\mathbf{r}.$$

This is an expression for $\mathbf{b}$ that gives an unbiased solution, i.e.

$$\hat{\mathbf{x}}=(\mathbf{I}-\mathbf{AH})\mathbf{r}+\mathbf{Ay}=\mathbf{r}+\mathbf{A}(\mathbf{y}-\mathbf{Hr}).$$

(b) The a-posteriori error covariance is

$$\begin{aligned}
\mathbf{P}_{\mathrm{A}} &= \mathcal{E}[\boldsymbol{\varepsilon}_{\mathbf{x}}\boldsymbol{\varepsilon}_{\mathbf{x}}^{\mathrm{T}}]=\mathcal{E}[(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^{\mathrm{T}}],\\
&= \mathcal{E}[\{\mathbf{x}-\mathbf{r}-\mathbf{A}(\mathbf{y}-\mathbf{Hr})\}\{\mathbf{x}-\mathbf{r}-\mathbf{A}(\mathbf{y}-\mathbf{Hr})\}^{\mathrm{T}}],\\
&= \mathcal{E}[\{\mathbf{x}-\mathbf{r}\}\{\mathbf{x}-\mathbf{r}\}^{\mathrm{T}}]-\mathcal{E}[\{\mathbf{x}-\mathbf{r}\}(\mathbf{y}-\mathbf{Hr})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}]\\
&\quad -\mathcal{E}[\mathbf{A}(\mathbf{y}-\mathbf{Hr})\{\mathbf{x}-\mathbf{r}\}^{\mathrm{T}}]+\mathcal{E}[\mathbf{A}(\mathbf{y}-\mathbf{Hr})(\mathbf{y}-\mathbf{Hr})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}]
\end{aligned}$$

Note that $\mathbf{y}-\mathbf{Hr}=\mathbf{Hx}+\boldsymbol{\varepsilon}-\mathbf{Hr}=\mathbf{H}(\mathbf{x}-\mathbf{r})+\boldsymbol{\varepsilon}$ and let $\mathbf{P}_{\mathbf{x}}=\mathcal{E}[\{\mathbf{x}-\mathbf{r}\}\{\mathbf{x}-\mathbf{r}\}]$

$$\begin{aligned}
\mathbf{P}_{\mathrm{A}} &= \mathbf{P}_{\mathbf{x}}-\mathcal{E}[\{\mathbf{x}-\mathbf{r}\}(\mathbf{H}(\mathbf{x}-\mathbf{r})+\boldsymbol{\varepsilon})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}]-\mathcal{E}[\mathbf{A}(\mathbf{H}(\mathbf{x}-\mathbf{r})+\boldsymbol{\varepsilon})\{\mathbf{x}-\mathbf{r}\}^{\mathrm{T}}]\\
&\quad +\mathcal{E}[\mathbf{A}(\mathbf{H}(\mathbf{x}-\mathbf{r})+\boldsymbol{\varepsilon})(\mathbf{H}(\mathbf{x}-\mathbf{r})+\boldsymbol{\varepsilon})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}],\\
&= \mathbf{P}_{\mathbf{x}}-\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}-\mathcal{E}[\{\mathbf{x}-\mathbf{r}\}\boldsymbol{\varepsilon}^{\mathrm{T}}]\mathbf{A}^{\mathrm{T}}-\mathbf{AHP}_{\mathbf{x}}-\mathbf{A}\mathcal{E}[\boldsymbol{\varepsilon}\{\mathbf{x}-\mathbf{r}\}^{\mathrm{T}}]\\
&\quad +\mathbf{AHP}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}+\mathbf{AH}\mathcal{E}[(\mathbf{x}-\mathbf{r})\boldsymbol{\varepsilon}^{\mathrm{T}}]\mathbf{A}^{\mathrm{T}}+\mathbf{A}\mathcal{E}[\boldsymbol{\varepsilon}(\mathbf{x}-\mathbf{r})^{\mathrm{T}}]\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}+\mathbf{A}\mathcal{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}]\mathbf{A}^{\mathrm{T}}.
\end{aligned}$$

Assume that $\mathcal{E}[\{\mathbf{x}-\mathbf{r}\}\boldsymbol{\varepsilon}^{\mathrm{T}}]=0$, $\mathcal{E}[\boldsymbol{\varepsilon}\{\mathbf{x}-\mathbf{r}\}^{\mathrm{T}}]=0$ and define $\mathcal{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}]=\mathbf{R}$. The above then simplifies

$$\mathbf{P}_{\mathrm{A}}=\mathbf{P}_{\mathbf{x}}-\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}-\mathbf{AHP}_{\mathbf{x}}+\mathbf{A}(\mathbf{HP}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}}+\mathbf{R})\mathbf{A}^{\mathrm{T}}.$$

(c) What is the trace of $\mathbf{P}_{\mathrm{A}}$?

$$\begin{aligned}
\mathrm{tr}(\mathbf{P}_{\mathrm{A}}) &= \sum_{i=1}^{n} (\mathbf{P}_{\mathrm{A}})_{ii}, \\
&= \sum_{i=1}^{n} (\mathbf{P}_{\mathbf{x}})_{ii} - \sum_{i=1}^{n}\sum_{j=1}^{p} (\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}})_{ij}(\mathbf{A}^{\mathrm{T}})_{ji} - \sum_{i=1}^{n}\sum_{j=1}^{p} \mathbf{A}_{ij}(\mathbf{H}\mathbf{P}_{\mathbf{x}})_{ji} + \sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{p} \mathbf{A}_{ij}(\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})_{jk}(\mathbf{A}^{\mathrm{T}})_{ki}, \\
&= \sum_{i=1}^{n} (\mathbf{P}_{\mathbf{x}})_{ii} - \sum_{i=1}^{n}\sum_{j=1}^{p} (\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}})_{ij}\mathbf{A}_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{p} \mathbf{A}_{ij}(\mathbf{H}\mathbf{P}_{\mathbf{x}})_{ji} + \sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{p} \mathbf{A}_{ij}(\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})_{jk}\mathbf{A}_{ik}.
\end{aligned}$$

(d) What matrix $\mathbf{A}$ minimizes the trace of $\mathbf{P}_{\mathrm{A}}$? Differentiate this trace w.r.t. an arbitrary element of matrix $\mathbf{A}$, $\mathbf{A}_{\alpha\beta}$ and then set to zero for stationary value

$$\begin{aligned}
\frac{\partial \mathrm{tr}(\mathbf{P}_{\mathrm{A}})}{\partial \mathbf{A}_{\alpha\beta}} &= -\sum_{i=1}^{n}\sum_{j=1}^{p} (\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}})_{ij}\frac{\partial \mathbf{A}_{ij}}{\partial \mathbf{A}_{\alpha\beta}} - \sum_{i=1}^{n}\sum_{j=1}^{p} \frac{\partial \mathbf{A}_{ij}}{\partial \mathbf{A}_{\alpha\beta}}(\mathbf{H}\mathbf{P}_{\mathbf{x}})_{ji} \\
&+ \sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{p} \frac{\partial \mathbf{A}_{ij}}{\partial \mathbf{A}_{\alpha\beta}}(\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})_{jk}\mathbf{A}_{ik} + \sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{p} \mathbf{A}_{ij}(\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})_{jk}\frac{\partial \mathbf{A}_{ik}}{\partial \mathbf{A}_{\alpha\beta}}.
\end{aligned}$$

Note that $\partial \mathbf{A}_{ij}/\partial \mathbf{A}_{\alpha\beta} = \delta_{i\alpha}\delta_{j\beta}$ which makes the above

$$\frac{\partial \mathrm{tr}(\mathbf{P}_{\mathrm{A}})}{\partial \mathbf{A}_{\alpha\beta}} = -(\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}})_{\alpha\beta} - (\mathbf{H}\mathbf{P}_{\mathbf{x}})_{\beta\alpha} + \sum_{k=1}^{p} (\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})_{\beta k}\mathbf{A}_{\alpha k} + \sum_{j=1}^{p} \mathbf{A}_{\alpha j}(\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})_{j\beta}.$$

The first two and the last two terms evaluate to the same values

$$\frac{\partial \mathrm{tr}(\mathbf{P}_{\mathrm{A}})}{\partial \mathbf{A}_{\alpha\beta}} = -2(\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}})_{\alpha\beta} + 2\sum_{j=1}^{p} \mathbf{A}_{\alpha j}(\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})_{j\beta}.$$

This is just element (α, β) of the matrix

$$\frac{\partial \mathrm{tr}(\mathbf{P}_{\mathrm{A}})}{\partial \mathbf{A}} = -2\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + 2\mathbf{A}(\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}).$$

(e) The stationary value is when this is zero which gives

$$\mathbf{A} = \mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}.$$

(f) Putting this together: using the $\mathbf{b}$ found earlier

$$\hat{\mathbf{x}} = \mathbf{b} + \mathbf{A}\mathbf{y} \;=\; (\mathbf{I} - \mathbf{A}\mathbf{H})\mathbf{r} + \mathbf{A}\mathbf{y} = \mathbf{r} + \mathbf{A}(\mathbf{y} - \mathbf{H}\mathbf{r}),$$

and using the $\mathbf{A}$ found above $\hat{\mathbf{x}} = \mathbf{r} + \mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{r})$.

This is equivalent to $\hat{\mathbf{x}} = \mathbf{r} + (\mathbf{P}_{\mathbf{x}}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{r})$.

(g) $\mathbf{r}$ resembles the background and $\mathbf{P}_{\mathbf{x}}$ resembles its error covariance. For the situation with no background, $\mathbf{P}_{\mathbf{x}} \to \infty$ and so $\mathbf{P}_{\mathbf{x}}^{-1} \to 0$

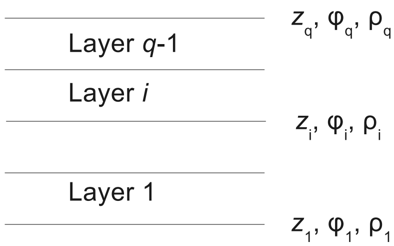
$$\hat{\mathbf{x}} \;=\; \mathbf{r} + (\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{r}) = (\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}.,$$

This result for $\hat{\mathbf{x}}$ is the same as $\mathbf{x}_{\mathrm{A}}$ found in Q. 5.

---

7. **Forward model and null-space example**

(a)



(b) The total column ozone per unit area is the sum of the layerwise ozone amount per unit area summed over the $q-1$ layers

$$\begin{aligned}
\frac{1}{4}\sum_{i=1}^{q-1}(\rho_i+\rho_{i+1})(\phi_i+\phi_{i+1})(z_{i+1}-z_i) &= \frac{1}{4}\sum_{i=1}^{q-1}\phi_i(\rho_i+\rho_{i+1})(z_{i+1}-z_i)+\frac{1}{4}\sum_{i=1}^{q-1}\phi_{i+1}(\rho_i+\rho_{i+1})(z_{i+1}-z_i),\\
&= \frac{1}{4}\sum_{i=1}^{q-1}\phi_i(\rho_i+\rho_{i+1})(z_{i+1}-z_i)+\frac{1}{4}\sum_{i=2}^{q}\phi_i(\rho_{i-1}+\rho_i)(z_i-z_{i-1}).
\end{aligned}$$

Defining $\tilde{\rho}_i=\rho_i+\rho_{i+1}$ and $\Delta z_i=z_{i+1}-z_i$ gives the above as

$$\frac{1}{4}\sum_{i=1}^{q-1}\phi_i\tilde{\rho}_i\Delta z_i+\frac{1}{4}\sum_{i=2}^{q}\phi_i\tilde{\rho}_{i-1}\Delta z_{i-1}=\frac{1}{4}\left\{\phi_1\tilde{\rho}_1\Delta z_1+\sum_{i=2}^{q-1}\phi_i\left[\tilde{\rho}_i\Delta z_i+\tilde{\rho}_{i-1}\Delta z_{i-1}\right]+\phi_q\tilde{\rho}_{q-1}\Delta z_{q-1}\right\}.$$

(c) Linear interpolation for the second measurement gives $(\phi_k+\phi_{k+1})/2$.

(d) $\mathbf{H}=\begin{pmatrix}\frac{1}{4}\tilde{\rho}\Delta z_1 & \frac{1}{4}[\tilde{\rho}_2\Delta z_2+\tilde{\rho}_1\Delta z_1] & \frac{1}{4}[\tilde{\rho}_3\Delta z_3+\tilde{\rho}_2\Delta z_2] & \frac{1}{4}\tilde{\rho}_3\Delta z_3\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\end{pmatrix}$ (for $\mathbf{x}=(\phi_1\,\phi_2\,\phi_3\,\phi_4)^{\mathrm{T}}$).

(e)

$$\mathbf{R}^{-1}=\begin{pmatrix}0.04 & 0\\ 0 & 0.25\end{pmatrix},\qquad \mathbf{H}=\begin{pmatrix}2.5 & 4.25 & 3 & 1\\ 0.5 & 0.5 & 0 & 0\end{pmatrix},$$

$$\mathbf{R}^{-1}\mathbf{H}=\begin{pmatrix}0.1 & 0.17 & 0.12 & 0.04\\ 0.1 & 0.1 & 0 & 0\end{pmatrix},\qquad \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}=\begin{pmatrix}0.3 & 0.475 & 0.3 & 0.1\\ 0.475 & 0.7725 & 0.51 & 0.17\\ 0.3 & 0.51 & 0.36 & 0.12\\ 0.1 & 0.17 & 0.12 & 0.04\end{pmatrix}.$$

There are two eigenvectors that have zero eigenvalues:

$$\mathbf{v}_1=\begin{pmatrix}0.654\\ -0.654\\ 0.381\\ 0\end{pmatrix},\qquad \mathbf{v}_2=\begin{pmatrix}0.079\\ -0.079\\ -0.272\\ 0.956\end{pmatrix},$$

(each component given to 3 d.p.). These define the null space.

(f) The matrix to calculate now is $\mathbf{B}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$

$$\mathbf{B}^{-1}=\frac{1}{781}\begin{pmatrix}189 & -186 & 98 & -24\\ -186 & 369 & -270 & 98\\ 98 & -270 & 369 & -186\\ -24 & 98 & -186 & 189\end{pmatrix},\quad \mathbf{B}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}=\begin{pmatrix}0.542 & 0.237 & 0.425 & 0.069\\ 0.237 & 1.245 & 0.164 & 0.295\\ 0.425 & 0.164 & 0.832 & -0.118\\ 0.069 & 0.295 & -0.118 & 0.282\end{pmatrix},$$

which has eigenvalues (no zero eigenvalues): $\lambda_1=1.509,\lambda_2=0.992,\lambda_3=0.291,\lambda_4=0.110$.

---

8. **Relationship between covariance and correlation**
   Expand the matrix notation

$$\mathbf{COV}_{ij} = \sum_{k=1}^{n}\sum_{l=1}^{n} \Sigma_{ik}\mathbf{COV}_{kl}\Sigma_{lj},$$

$$\begin{aligned} \text{but } \Sigma_{ik} = \sigma_i\delta_{ik} \text{ leading to } \mathbf{COV}_{ij} &= \sum_{k=1}^{n}\sum_{l=1}^{n} \sigma_i\delta_{ik}\mathbf{COR}_{kl}\sigma_l\delta_{lj} = \sigma_i\mathbf{COR}_{ij}\sigma_j, \\ \text{hence } \mathbf{COR}_{ij} &= \frac{\mathbf{COV}_{ij}}{\sigma_i\sigma_j.}. \end{aligned}$$

---

9. **Structure functions**
   One way of doing this is to check that $\mathbf{Pv} = \sum_{i=1}^{n} \mathbf{p}_i v_i$ by expanding each side. The jth component of the left hand side is $\sum_{i=1}^{n} p_{ji}v_i$ (where p_{ji} is the jith matrix element of $\mathbf{P}$) and the jth component of the right hand side is $\sum_{i=1}^{n} (\mathbf{p}_i)_j v_i$ (where $(\mathbf{p}_i)_j$ is the jth component of $\mathbf{p}_i$. Now, $(\mathbf{p}_i)_j = p_{ji}$, making the left and right hand sides equal.

---

10. **Assimilation of a single observation in VAR to probe the background error covariance structure**
    The OI formula for the analysis increment is

$$\mathbf{x}^{\mathrm{A}} - \mathbf{x}^{\mathrm{B}} = \mathbf{BH}^{\mathrm{T}}(\mathbf{R} + \mathbf{HBH}^{\mathrm{T}})^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}^{\mathrm{B}})).$$

$$\text{A single direct observation at grid-point k means that } \mathbf{y} = y, \qquad \mathbf{h}(\mathbf{x}^{\mathrm{B}}) = \mathbf{x}_k^{\mathrm{B}},$$

(both scalars). $\mathbf{R}$ and $\mathbf{HBH}^{\mathrm{T}}$ are also both scalars. For the single observation, $\mathbf{H}$ will be the $1 \times n$ matrix with all elements zero except the one for element k: $\mathbf{H} = (0\,0\,0\,1\,0\,0\,0)$.

$$\text{First find } \mathbf{BH}^{\mathrm{T}}\text{: } \mathbf{BH}^{\mathrm{T}} = \begin{pmatrix} \mathbf{B}_{1k} \\ \mathbf{B}_{2k} \\ \vdots \\ \mathbf{B}_{kk} \\ \vdots \\ \mathbf{B}_{n-1,k} \\ \mathbf{B}_{nk} \end{pmatrix}. \text{ Now find } \mathbf{HBH}^{\mathrm{T}}\text{: } \mathbf{HBH}^{\mathrm{T}} = \mathbf{B}_{kk}.$$

Let $\mathbf{R} = \sigma_y^2$ (the variance of the single observation). Then, putting this all together

$$\mathbf{x}^{\mathrm{A}} - \mathbf{x}^{\mathrm{B}} = \begin{pmatrix} \mathbf{B}_{1k} \\ \mathbf{B}_{2k} \\ \vdots \\ \mathbf{B}_{kk} \\ \vdots \\ \mathbf{B}_{n-1,k} \\ \mathbf{B}_{nk} \end{pmatrix} \frac{y - \mathbf{x}_k^{\mathrm{B}}}{\sigma_y^2 + \mathbf{B}_{kk}}.$$

---

11. **Ensemble covariance in matrix form**
    Let us define $\boldsymbol{\eta}_{\mathrm{B}}^{(i)} = \mathbf{x}_{\mathrm{B}}^{(i)} - \langle\mathbf{x}\rangle$for shorthand. The standard expression for the error covariance matrix of these N perturbations is:

$$\mathbf{P}_{(N)}^{\mathrm{f}} = \frac{1}{N-1}\sum_{i=1}^{N} \boldsymbol{\eta}_{\mathrm{B}}^{(i)}\boldsymbol{\eta}_{\mathrm{B}}^{(i)\mathrm{T}}, \text{ which has matrix elements } \left(\mathbf{P}_{(N)}^{\mathrm{f}}\right)_{kl} = \frac{1}{N-1}\sum_{i=1}^{N} \left(\boldsymbol{\eta}_{\mathrm{B}}^{(i)}\right)_k \left(\boldsymbol{\eta}_{\mathrm{B}}^{(i)}\right)_l.$$

If we can show that $\mathbf{X}\mathbf{X}^{\mathrm{T}}/(N-1)$ has these matrix elements then we have answered the question. The k,lth matrix element of $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ is:

$$\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)_{kl} = \sum_{i=1}^{N} \left(\mathbf{X}\right)_{ki} \left(\mathbf{X}^{\mathrm{T}}\right)_{il} = \sum_{i=1}^{N} \left(\mathbf{X}\right)_{ki} \left(\mathbf{X}\right)_{li}.$$

From the definition of $\mathbf{X}$, the matrix element $(\mathbf{X})_{ki}$ is the kth element of the ith ensemble member: $(\mathbf{X})_{ki} = (\boldsymbol{\eta}_{\mathrm{B}}^{(i)})_k$. Putting this into the outer product expression above gives:

$$\left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)_{kl} = \sum_{i=1}^{N} \left(\boldsymbol{\eta}_{\mathrm{B}}^{(i)}\right)_k \left(\boldsymbol{\eta}_{\mathrm{B}}^{(i)}\right)_l = \sum_{i=1}^{N} \left(\boldsymbol{\eta}_{\mathrm{B}}^{(i)}\right)_k \left(\boldsymbol{\eta}_{\mathrm{B}}^{(i)}\right)_l.$$

Dividing by $N-1$ then concludes that the matrix elements of $\mathbf{X}\mathbf{X}^{\mathrm{T}}/(N-1)$ are the same as those of the standard expression. If all matrix elements are the same, then the matrices are the same.

---

12. **Implied covariances**

(a) Using the information given, the background error covariance matrices in each space are defined as the following outer product expectations:

$$\mathbf{B}_{\delta\mathbf{x}} = \left\langle \delta\mathbf{x}\delta\mathbf{x}^{\mathrm{T}} \right\rangle, \qquad \mathbf{B}_{\delta\boldsymbol{\chi}} = \left\langle \delta\boldsymbol{\chi}\delta\boldsymbol{\chi}^{\mathrm{T}} \right\rangle.$$

Substituting in the CVT into the first definition, taking the CVT outside of the averaging brackets, and then using the second definition gives:

$$\mathbf{B}_{\delta\mathbf{x}} = \left\langle \mathbf{U}\delta\boldsymbol{\chi}(\mathbf{U}\delta\boldsymbol{\chi})^{\mathrm{T}} \right\rangle = \left\langle \mathbf{U}\delta\boldsymbol{\chi}\delta\boldsymbol{\chi}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}} \right\rangle = \mathbf{U}\left\langle \delta\boldsymbol{\chi}\delta\boldsymbol{\chi}^{\mathrm{T}} \right\rangle \mathbf{U}^{\mathrm{T}} = \mathbf{U}\mathbf{B}_{\delta\boldsymbol{\chi}}\mathbf{U}^{\mathrm{T}}.$$

(b) From the lectures, the CVT is chosen to make $\mathbf{U}^{\mathrm{T}}\mathbf{B}_{\delta\mathbf{x}}^{-1}\mathbf{U} = \mathbf{I}$ (the matrix $\mathbf{B}_{\delta\mathbf{x}}$ is denoted $\mathbf{B}$ in the lectures). Re-arranging this gives:

$$\begin{aligned} \mathbf{U}^{\mathrm{T}}\mathbf{B}_{\delta\mathbf{x}}^{-1}\mathbf{U} &= \mathbf{I}, \\ \mathbf{U}^{-\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{B}_{\delta\mathbf{x}}^{-1}\mathbf{U}\mathbf{U}^{-1} &= \mathbf{U}^{-\mathrm{T}}\mathbf{U}^{-1}, \\ \mathbf{B}_{\delta\mathbf{x}}^{-1} &= \mathbf{U}^{-\mathrm{T}}\mathbf{U}^{-1}, \\ \therefore \mathbf{B}_{\delta\mathbf{x}} &= \mathbf{U}\mathbf{U}^{\mathrm{T}}. \end{aligned}$$

This means that solving the (simpler) problem of minimizing $J[\delta\boldsymbol{\chi}]$ where $\delta\boldsymbol{\chi}$ has unit-matrix background error covariances is equivalent to minimizing $J[\delta\mathbf{x}]$ with background error covariance $\mathbf{B}_{\delta\mathbf{x}}$. $\mathbf{B}_{\delta\mathbf{x}}$ is the implied background error covariance matrix. This is consistent with putting $\mathbf{B}_{\delta\boldsymbol{\chi}} = \mathbf{I}$ in part (a).

---

13. **The generalized chain rule**

Write down the information given in expanded form $(\nabla_{\mathbf{v}_{\mathrm{B}}} f)_i = \dfrac{\partial f}{\partial(\mathbf{v}_{\mathrm{B}})_i}$, $\qquad (\mathbf{v}_{\mathrm{B}})_i = \sum_j \mathbf{N}_{ij}(\mathbf{v}_{\mathrm{A}})_j$.

Furthermore, the gradient with respect to $\mathbf{v}_{\mathrm{A}}$is $(\nabla_{\mathbf{v}_{\mathrm{A}}} f)_j = \dfrac{\partial f}{\partial(\mathbf{v}_{\mathrm{A}})_j}$,

and the generalized chain rule relates the derivatives with respect to each variable

$$\frac{\partial f}{\partial(\mathbf{v}_{\mathrm{A}})_j} = \sum_i \frac{\partial(\mathbf{v}_{\mathrm{B}})_i}{\partial(\mathbf{v}_{\mathrm{A}})_j}\frac{\partial f}{\partial(\mathbf{v}_{\mathrm{B}})_i}.$$

From the information given, the following is found $\dfrac{\partial(\mathbf{v}_{\mathrm{B}})_i}{\partial(\mathbf{v}_{\mathrm{A}})_j} = \mathbf{N}_{ij} = (\mathbf{N}^{\mathrm{T}})_{ji}$.

Substituting this into the chain rule gives $(\nabla_{\mathbf{v}_\mathrm{A}} f)_j = \sum_i (\mathbf{N}^\mathrm{T})_{ji} (\nabla_{\mathbf{v}_\mathrm{B}} f)_i$,

which is just the expanded form of $\nabla_{\mathbf{v}_\mathrm{A}} f = \mathbf{N}^\mathrm{T} \nabla_{\mathbf{v}_\mathrm{B}} f$.

This is a useful result and will be used in Q. 14.

---

14. **Gradient and Hessian of the cost function w.r.t the control variable**

(a) Expand J_B

$$J_\mathrm{B} = \frac{1}{2} \sum_{i=1}^{n} \delta\chi_i^2.$$

Hence, the ith component of the first derivative (w.r.t. $\delta\chi$) is

$$\frac{\partial J_\mathrm{B}}{\partial \delta\chi_j} = \frac{1}{2} \sum_{i=1}^{n} \frac{\partial \delta\chi_i^2}{\partial \delta\chi_j} = \sum_{i=1}^{n} \delta\chi_i \frac{\partial \delta\chi_i}{\partial \delta\chi_j} = \sum_{i=1}^{n} \delta\chi_i \delta_{ij} = \delta\chi_j,$$

which is the jth component of the column vector $\nabla_{\delta\boldsymbol{\chi}} J_\mathrm{B}$ by definition. The second derivative

$$\frac{\partial^2 J_\mathrm{B}}{\partial \delta\chi_j \partial \delta\chi_k} = \frac{\partial \delta\chi_j}{\partial \delta\chi_k} = \delta_{ij}, \text{ which is the } (i,j)\text{th component of the matrix } \mathbf{I}.$$

(b) Expand J_O (do this for general - non-diagonal - $\mathbf{R}_t$)

$$J_\mathrm{O}(t) = \frac{1}{2} \sum_{i,j=1}^{p} (\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t))_i (\mathbf{R}_t^{-1})_{ij} (\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t))_j.$$

Hence, the kth component of the first derivative (w.r.t. $\delta\mathbf{y}^\mathrm{m}(t)$) is (use the differentiation product rule)

$$\begin{aligned}
\frac{\partial J_\mathrm{O}(t)}{\partial \delta\mathbf{y}^\mathrm{m}(t)_k} &= \frac{1}{2} \sum_{i,j=1}^{p} \frac{\partial}{\partial \delta\mathbf{y}^\mathrm{m}(t)_k} \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right)_i (\mathbf{R}_t^{-1})_{ij} \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right)_j , \\
&= -\frac{1}{2} \sum_{i,j=1}^{p} \left\{ \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right)_i (\mathbf{R}_t^{-1})_{ij} \frac{\partial \delta\mathbf{y}^\mathrm{m}(t)_j}{\partial \delta\mathbf{y}^\mathrm{m}(t)_k} + \frac{\partial \delta\mathbf{y}^\mathrm{m}(t)_i}{\partial \delta\mathbf{y}^\mathrm{m}(t)_k} (\mathbf{R}_t^{-1})_{ij} \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right)_j \right\}, \\
&= -\frac{1}{2} \sum_{i,j=1}^{p} \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right)_i (\mathbf{R}_t^{-1})_{ij} \delta_{jk} - \frac{1}{2} \sum_{i,j=1}^{p} \delta_{ik} (\mathbf{R}_t^{-1})_{ij} \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right)_j , \\
&= -\frac{1}{2} \sum_{i=1}^{p} \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right)_i (\mathbf{R}_t^{-1})_{ik} - \frac{1}{2} \sum_{j=1}^{p} (\mathbf{R}_t^{-1})_{kj} \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right)_j ,
\end{aligned}$$

The first summation can be re-indexed $i \to j$ and the symmetric property of $\mathbf{R}_t^{-1}$ used

$$\frac{\partial J_\mathrm{O}(t)}{\partial \delta\mathbf{y}^\mathrm{m}(t)_k} = -\frac{1}{2} \sum_{j=1}^{p} \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right)_j (\mathbf{R}_t^{-1})_{kj} - \frac{1}{2} \sum_{j=1}^{p} (\mathbf{R}_t^{-1})_{kj} \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right)_j .$$

Both summations are shown to be equal, hence

$$\frac{\partial J_\mathrm{O}(t)}{\partial \delta\mathbf{y}^\mathrm{m}(t)_k} = -\sum_{j=1}^{p} (\mathbf{R}_t^{-1})_{kj} (\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t))_j.$$

This is the kth component of the column vector

$$\nabla_{\delta\mathbf{y}^\mathrm{m}(t)} J_\mathrm{O}(t) = -\mathbf{R}_t^{-1} \left(\delta\mathbf{y}(t) - \delta\mathbf{y}^\mathrm{m}(t)\right).$$

Differentiating again w.r.t. $\delta\mathbf{y}^{\mathrm{m}}(t)_l$ gives

$$\begin{aligned}\frac{\partial^2 J_{\mathrm{O}}(t)}{\partial\delta\mathbf{y}^{\mathrm{m}}(t)_k\partial\delta\mathbf{y}^{\mathrm{m}}(t)_l} &= -\sum_{j=1}^{p}\frac{\partial}{\partial\delta\mathbf{y}^{\mathrm{m}}(t)_l}(\mathbf{R}_t^{-1})_{kj}\left(\delta\mathbf{y}(t)-\delta\mathbf{y}^{\mathrm{m}}(t)\right)_j, \\ &= \sum_{j=1}^{p}(\mathbf{R}_t^{-1})_{kj}\frac{\partial\delta\mathbf{y}^{\mathrm{m}}(t)_j}{\partial\delta\mathbf{y}^{\mathrm{m}}(t)_l} = \sum_{j=1}^{p}(\mathbf{R}_t^{-1})_{kj}\delta_{jl} = (\mathbf{R}_t^{-1})_{kl},\end{aligned}$$

which is the (k,l)th component of the matrix $\dfrac{\partial^2 J_{\mathrm{O}}(t)}{\partial\delta\mathbf{y}^{\mathrm{m}}(t)^2} = \mathbf{R}_t^{-1}$.

(c) This is done by simple substitution.

(d) This is done by simple substitution.

(e) The total gradient of the cost function is

$$\nabla_{\delta\boldsymbol{\chi}}J = \nabla_{\delta\boldsymbol{\chi}}J_{\mathrm{B}} + \sum_{t=0}^{T}\nabla_{\delta\boldsymbol{\chi}}J_{\mathrm{O}}(t) = \delta\boldsymbol{\chi} - \mathbf{U}^{\mathrm{T}}\sum_{t=0}^{T}\mathbf{M}_{t\leftarrow 0}^{\mathrm{T}}\mathbf{H}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\left(\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}\delta\boldsymbol{\chi}\right).$$

The total Hessian of the cost function is

$$\frac{\partial^2 J}{\partial\delta\boldsymbol{\chi}^2} = \frac{\partial^2 J_{\mathrm{B}}}{\partial\delta\boldsymbol{\chi}^2} + \sum_{t=0}^{T}\frac{\partial^2 J_{\mathrm{O}}(t)}{\partial\delta\boldsymbol{\chi}^2}\mathbf{I} + \sum_{t=0}^{T}\mathbf{U}^{\mathrm{T}}\mathbf{M}_{t\leftarrow 0}^{\mathrm{T}}\mathbf{H}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}.$$

---

15. **Efficient form of the 4D-VAR gradient**

(a) Writing out contributions to the summation on separate lines gives

$$\begin{aligned}\nabla_{\delta\boldsymbol{\chi}}J_{\mathrm{O}} = -\mathbf{U}^{\mathrm{T}}\{&\mathbf{H}_0^{\mathrm{T}}\mathbf{R}_0^{-1}\mathbf{r}(0) + \\ &\mathbf{M}_{1\leftarrow 0}^{\mathrm{T}}\mathbf{H}_1^{\mathrm{T}}\mathbf{R}_1^{-1}\mathbf{r}(1) + \\ &\mathbf{M}_{1\leftarrow 0}^{\mathrm{T}}\mathbf{M}_{2\leftarrow 1}^{\mathrm{T}}\dots\mathbf{M}_{T-2\leftarrow T-3}^{\mathrm{T}}\times \\ &\quad\mathbf{M}_{T-1\leftarrow T-2}^{\mathrm{T}}\mathbf{H}_{T-1}^{\mathrm{T}}\mathbf{R}_{T-1}^{-1}\mathbf{r}(T-1) + \\ &\mathbf{M}_{1\leftarrow 0}^{\mathrm{T}}\mathbf{M}_{2\leftarrow 1}^{\mathrm{T}}\dots\mathbf{M}_{T-1\leftarrow T-2}^{\mathrm{T}}\mathbf{M}_{T\leftarrow T-1}^{\mathrm{T}}\mathbf{H}_T^{\mathrm{T}}\mathbf{R}_T^{-1}\mathbf{r}(T)\}.\end{aligned}$$

(b) $\mathbf{M}_{1\leftarrow 0}^{\mathrm{T}}$ is used in all but the first line, $\mathbf{M}_{2\leftarrow 1}^{\mathrm{T}}$ is used in all but the first and second lines, etc. Identifying the adjoint matrices that are common to many lines provides the 'trick' that is used to write the efficient form of the gradient.

---

16. **The NMC method**

(a) The information given in the question in mathematical form is

$$\mathbf{B} = \langle\boldsymbol{\eta}^{48}\boldsymbol{\eta}^{48\mathrm{T}}\rangle, \qquad \mathbf{B} = \langle\boldsymbol{\eta}^{24}\boldsymbol{\eta}^{24\mathrm{T}}\rangle, \qquad \langle\boldsymbol{\eta}^{48}\boldsymbol{\eta}^{24\mathrm{T}}\rangle = 0.$$

The error covariance of the forecast difference is, by substitution

$$\langle(\mathbf{x}^{\mathrm{f}48}-\mathbf{x}^{\mathrm{f}24})(\mathbf{x}^{\mathrm{f}48}-\mathbf{x}^{\mathrm{f}24})^{\mathrm{T}}\rangle = \langle(\boldsymbol{\eta}^{48}-\boldsymbol{\eta}^{24})(\boldsymbol{\eta}^{48}-\boldsymbol{\eta}^{24})^{\mathrm{T}}\rangle$$

$$= \langle\boldsymbol{\eta}^{48}\boldsymbol{\eta}^{48\mathrm{T}}\rangle - \langle\boldsymbol{\eta}^{48}\boldsymbol{\eta}^{24\mathrm{T}}\rangle - \langle\boldsymbol{\eta}^{24}\boldsymbol{\eta}^{48\mathrm{T}}\rangle + \langle\boldsymbol{\eta}^{24}\boldsymbol{\eta}^{24\mathrm{T}}\rangle = \langle\boldsymbol{\eta}^{48}\boldsymbol{\eta}^{48\mathrm{T}}\rangle + \langle\boldsymbol{\eta}^{24}\boldsymbol{\eta}^{24\mathrm{T}}\rangle = 2\mathbf{B}.$$

(b) None of the assumptions is likely to be valid. Forecast errors are expected to grow with the length of the forecast and so the assumption that each of $\langle\boldsymbol{\eta}^{48}\boldsymbol{\eta}^{48\mathrm{T}}\rangle$ and $\langle\boldsymbol{\eta}^{24}\boldsymbol{\eta}^{48\mathrm{T}}\rangle$ are the same will not be true. Forecast errors are also likely to be correlated in time (e.g. an error in one part of the atmosphere at 24 hours is likely to be correlated with an error at 48 hours, especially downstream of the flow). Additionally, the **B**-matrix is meant to represent the error covariance of forecasts of a particular range, e.g. 6 or 12 hours. The NMC method uses two forecasts of different lengths and neither of 6 or 12 hours in range. [N.B. the reason why the NMC method usually uses a difference between the forecasts of 24 hours is to cancel out diurnal biases.]
