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

1. **Useful formula related to the Sherman-Morrison-Woodbury formula**
   Show that the following identity holds:

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

---

2. **The Euler-Lagrange equations and the method of representers**
   (Here equation numbers refer to those on the Euler-Lagrange handout.) The weak constraint Euler-Lagrange equations, which satisfy the best fit to (18) are:

$$\begin{aligned}
\frac{\partial\hat{\phi}}{\partial t} + u\frac{\partial\hat{\phi}}{\partial x} - F &= W_{\mathrm{e}}^{-1}\hat{\mu}, \\
W_{\mathrm{ic}}\{\hat{\phi}(x,0) - I(x)\} - \hat{\mu}(x,0) &= 0, \\
W_{\mathrm{bc}}\{\hat{\phi}(0,t) - B(t)\} - u\hat{\mu}(0,t) &= 0, \\
W_{\mathrm{ob}}\sum_{i=1}^{p}\{\hat{\phi}(x_i,t_i) - y_i\}\delta(x-x_i)\delta(t-t_i) - \left(\frac{\partial\hat{\mu}}{\partial t} + u\frac{\partial\hat{\mu}}{\partial x}\right) &= 0, \\
\hat{\mu}(x,T) &= 0, \\
\mu(\hat{L},t) &= 0,
\end{aligned}$$

   (see the handout for a definition of the symbols). In the method of representers, (23, 24) are not enforced explicitly. Use the equations on the handout to show that (23, 24) are satisfied by the representer method.

---

3. **Inner product forms**
   The following term is a measure of the total square deviation between the continuous fields $\phi(x)$ and $I(x)$:

$$\int_{x=0}^{L} dx\,\frac{\{\phi(x) - I(x)\}^2}{\sigma_I^2(x)},$$

   where x is position in the domain between $x = 0$ and $x = L$, and $\sigma_I^2(x)$ is the (position dependent) standard deviation. An approximate version of the above is found by discretizing the domain into n points:

$$\sum_{i=1}^{n} \frac{\{(\mathbf{x})_i - (\mathbf{x}_\mathrm{B})_i\}^2}{(\mathbf{P})_{ii}},$$

where $\mathbf{x}$ is the discrete vector representation of $\phi(x)$ (where $(\mathbf{x})_i = \phi(iL/n)$ is the ith element of $\mathbf{x}$) and $\mathbf{x}_\mathrm{B}$ is the discrete representation of $I(x)$ (where $(\mathbf{x}_\mathrm{B})_i = I(iL/n)$ is the ith element of $\mathbf{x}_\mathrm{B}$), and $(\mathbf{P})_{ii} = \sigma_I^2(iL/n)$ represents the ith diagonal element of the diagonal matrix $\mathbf{P}$. Show that this is equal to the following compact form:

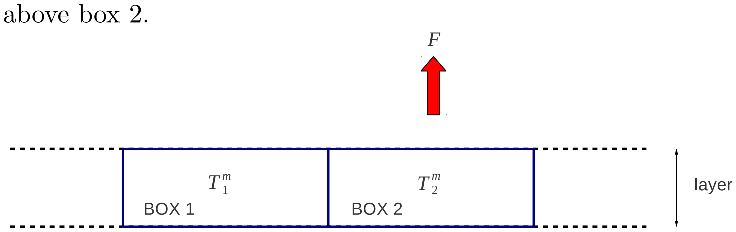
$$(\mathbf{x} - \mathbf{x}_\mathrm{B})^\mathrm{T} \mathbf{P}^{-1} (\mathbf{x} - \mathbf{x}_\mathrm{B}).$$

---

4. **Forward model example**
All bodies at a temperature above absolute zero emit thermal radiation. In this example, the radiation from an atmosphere layer is monitored by satellite. A model represents this layer of the atmosphere with two grid boxes and carries temperature in each, T_1^m and T_2^m (see Fig. below). Radiation flux, F, is related to layer temperature, T, by the Stefan-Boltzmann Law:

$$F = \kappa T^4,$$

where κ is the Stefan-Boltzmann constant. A flux measurement is made above box 2.



(a) Write down the model state vector $\mathbf{x}$ and the observation vector $\mathbf{y}$.

(b) What is the forward operator $\mathbf{h}$, the Jacobian $\mathbf{H}$, and its adjoint $\mathbf{H}^\mathrm{T}$?

---

5. **Maximum likelihood solution (MAP)**
The following function gives the likelihood that the n-element vector $\mathbf{x}$ is the state of the system given p observations in $\mathbf{y}$:

$$L(\mathbf{x}|\mathbf{y}) = \frac{1}{(2\pi)^{p/2}|\mathbf{R}|^{1/2}} \exp -\frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}),$$

where $\mathbf{R}$ is the observation error covariance matrix and $\mathbf{H}$ is the forward operator matrix.

(a) Define J as minus the natural logarithm of L and expand the vector/matrix notation of the expression for J.

(b) Differentiate J in this form with respect to one of the components of $\mathbf{x}$ and then find a vector/matrix form for $\nabla_{\mathbf{x}} J = (\partial J/\partial x_1, \dots \partial J/\partial x_k, \dots, \partial J/\partial x_n)^{\mathrm{T}}$.

(c) Find the $\mathbf{x}$ that makes J stationary, i.e. $\nabla_{\mathbf{x}} J = 0$ (call this $\mathbf{x}_{\mathrm{A}}$).

---

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

(a) Assume that the best estimate, $\hat{\mathbf{x}}$ has the form $\hat{\mathbf{x}} = \mathbf{b} + \mathbf{A}\mathbf{y}$ where $\mathbf{b}$ is an n-element vector, $\mathbf{A}$ is an $n \times p$ matrix, and $\mathbf{y}$ is the p-element vector of observations. Also note that $\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}$. By defining $\mathbf{r} = \mathcal{E}[\hat{\mathbf{x}}]$, show that the following expression for $\mathbf{b}$ follows for an unbiased solution with an unbiased set of observations

$$\mathbf{b} = (\mathbf{I} - \mathbf{A}\mathbf{H})\mathbf{r}.$$

(Note that an unbiased solution means that $\mathcal{E}[\mathbf{x}] = \mathcal{E}[\hat{\mathbf{x}}] = \mathbf{r}$, and unbiased observations means that $\mathcal{E}[\boldsymbol{\varepsilon}] = 0$.)

(b) Define the a-posteriori error as $\boldsymbol{\varepsilon}_{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$. Use the result from (a) to show that the a-posteriori error covariance, $\mathbf{P}_{\mathrm{A}} = \mathcal{E}[\boldsymbol{\varepsilon}_{\mathbf{x}} \boldsymbol{\varepsilon}_{\mathbf{x}}^{\mathrm{T}}]$ is

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

where $\mathbf{P}_{\mathbf{x}} = \mathcal{E}[\{\mathbf{x} - \mathbf{r}\}\{\mathbf{x} - \mathbf{r}\}]$, $\mathbf{R} = \mathcal{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}]$, and where it is assumed that $\mathcal{E}[\{\mathbf{x} - \mathbf{r}\}\boldsymbol{\varepsilon}^{\mathrm{T}}] = 0$ and $\mathcal{E}[\boldsymbol{\varepsilon}\{\mathbf{x} - \mathbf{r}\}^{\mathrm{T}}] = 0$.

(c) The trace of a matrix is defined as the sum of its diagonal elements. By expanding the matrix notation of the result for $\mathbf{P}_{\mathrm{A}}$ in (b), give an expression for the trace of $\mathbf{P}_{\mathrm{A}}$ (to cut down on the algebra, the terms $\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}}$, $\mathbf{H}\mathbf{P}_{\mathbf{x}}$ and $(\mathbf{H}\mathbf{P}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \mathbf{R})$ do not need to be expanded, i.e. each may be considered a single matrix).

(d) Differentiate this trace w.r.t. an arbitrary element of matrix $\mathbf{A}$, e.g. $\mathbf{A}_{\alpha\beta}$ and then form a matrix expression for the matrix that makes up all such derivatives.

(e) Show that the matrix $\mathbf{A}$ that makes the trace of $\mathbf{P}_{\mathrm{A}}$ stationary is

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

(f) Put together the results for $\mathbf{b}$ in (a) and for $\mathbf{A}$ in (e) to show that the minimum variance solution is

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

and use the result from Q. 1 to give

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

(g) By considering $\mathbf{r}$ as a-priori information and $\mathbf{P_x}$ as its error covariance, show that the resulting minimum variance solution is the same as the MAP found in Q. 5 in the limit $\mathbf{P_x} \to \infty$ (i.e. no a-priori information available).

---

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

In chemical data assimilation, the aim is to use observations to help determine the field of chemical concentrations in the atmosphere. An observation commonly produced from nadir viewing (downward looking) satellites is a so-called 'total column' amount. This is an indirect observation of the concentration field and requires a non-local forward model for it to be assimilated.

Let the model representation be a 1-D column comprising q vertical levels. The level height, z_i, the air density ρ_i, and the ozone mass mixing ratio, ϕ_i are stored on level i. Level 1 is the Earth's surface and level q is well above the ozone layer.

(a) Draw a picture showing the q levels and the quantities stored on each.

(b) There are $q-1$ layers, and the amount of ozone per unit horizontal area in one the ith layer is approximated by $(\rho_i + \rho_{i+1})(\phi_i + \phi_{i+1})(z_{i+1} - z_i)/4$ (ie average density $\times$ average ozone mass mixing ratio $\times$ layer thickness). Show that the total column ozone per unit area according to the model state is

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

where $\tilde{\rho}_i = \rho_i + \rho_{i+1}$ and $\Delta z_i = z_{i+1} - z_i$.

(c) In addition to the total column measurement, a balloon directly measures ozone mass mixing ratio half-way between levels k and $k+1$. Write down the value of the observation expected from the model state (assume linear interpolation).

(d) The total column and in-situ observations are to be assimilated with no a-priori data. Let there be $q = 4$ levels and let $k = 1$. Write down the 2×4 observation operator matrix, $\mathbf{H}$.

(e) To put numbers to the problem and using dimensionless units, let $\tilde{\rho}_1\Delta z_1 = 10$, $\tilde{\rho}_2\Delta z_2 = 7$, $\tilde{\rho}_3\Delta z_3 = 5$ and $\tilde{\rho}_3\Delta z_3 = 4$. Furthermore, let the standard deviation of the total column observation error be 5 and the standard deviation of the in-situ observation error be 2. Write down the 4×4 matrix $\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$, where $\mathbf{R}$ is the diagonal observation error covariance matrix.

(f) Identify the null space of this observation system (you will probably find it easier to do the computations using a computer package - e.g. Mathematica, Maple, IDL, MatLab, NAG, Eispack or the web site Wolfram Alpha).

(g) An a-priori state is introduced that has the following error covariance matrix

$$\mathbf{B} = \begin{pmatrix} 9 & 6 & 2 & 0 \\ 6 & 9 & 6 & 2 \\ 2 & 6 & 9 & 6 \\ 0 & 2 & 6 & 9 \end{pmatrix}.$$

Show that the combined observation/a-priori system has no null space.

---

8. **Relationship between covariance and correlation**
The relationship between covariance **COV** and correlation **COR** is

$$\mathbf{COV} = \mathbf{\Sigma}\mathbf{COR}\mathbf{\Sigma},$$

where $\mathbf{\Sigma}$ is $\mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ and where σ_i are the standard deviations. Show that this general form leads to

$$\mathbf{COR}_{ij} = \frac{\mathbf{COV}_{ij}}{\sigma_i \sigma_j},$$

for individual elements.

---

9. **Structure functions**
Given that $\mathbf{u} = \mathbf{P}\mathbf{v}$, where $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $\mathbf{P} \in \mathbb{R}^{n \times n}$ show that $\mathbf{u}$ may be written as

$$\mathbf{u} = \sum_{i=1}^{n} \mathbf{p}_i v_i,$$

where $\mathbf{p}_i$ is the ith column of $\mathbf{P}$ and v_i is the ith component of $\mathbf{v}$.

---

10. **Assimilation of a single observation in VAR to probe the background error covariance structure**
Use the equivalence between the optimal interpolation formula and VAR to show that the assimilation of a single direct observation of a grid-point value results in an analysis increment that is proportional to a column of the background error covariance matrix.

---

11. **Ensemble covariance in matrix form**
Let the ith member of an ensemble of N states be the n-element vector $\mathbf{x}_{\mathrm{B}}^{(i)}$, which has ensemble mean $\langle \mathbf{x} \rangle = (1/N) \sum_{i=1}^{N} \mathbf{x}_{\mathrm{B}}^{(i)}$. Let the $n \times N$ matrix $\mathbf{X}$ have columns comprising ensemble perturbations from the ensemble mean, and scaled in the following way:

$$\mathbf{X} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow & \uparrow \\ \mathbf{x}_{\mathrm{B}}^{(1)} - \langle \mathbf{x} \rangle & \mathbf{x}_{\mathrm{B}}^{(2)} - \langle \mathbf{x} \rangle & \cdots & \mathbf{x}_{\mathrm{B}}^{(N-1)} - \langle \mathbf{x} \rangle & \mathbf{x}_{\mathrm{B}}^{(N)} - \langle \mathbf{x} \rangle \\ \downarrow & \downarrow & & \downarrow & \downarrow \end{pmatrix}.$$

Show that the simple expression $\mathbf{X}\mathbf{X}^{\mathrm{T}}/(N-1)$ is the error covariance matrix sampled from this ensemble.

---

12. **Implied covariances**

   (a) Consider a change of variable from $\delta\mathbf{x}$ to $\delta\boldsymbol{\chi}$ via the control variable transform $\delta\mathbf{x} = \mathbf{U}\delta\boldsymbol{\chi}$, and given that the background error covariance matrix in $\delta\mathbf{x}$-space is $\mathbf{B}_{\delta\mathbf{x}}$ and the background error covariance matrix in $\delta\boldsymbol{\chi}$-space is $\mathbf{B}_{\delta\boldsymbol{\chi}}$. Show that the relationship between the background error covariances is as follows:

   $$\mathbf{B}_{\delta\mathbf{x}} = \mathbf{U}\mathbf{B}_{\delta\boldsymbol{\chi}}\mathbf{U}^{\mathrm{T}}.$$

   (b) If the cost function is minimized with $\delta\boldsymbol{\chi}$ as the control variable, whose background errors are assumed to be $\mathbf{I}$, show that the implied background error covariance matrix in terms of $\delta\mathbf{x}$ is $\mathbf{B}_{\delta\mathbf{x}} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$. Is this consistent with the result of part (a) of this question?

---

13. **The generalized chain rule**
Let some scalar be a functional of a vector, $\mathbf{v}_{\mathrm{B}}$, i.e. $f(\mathbf{v}_{\mathrm{B}})$ and let the vector $\nabla_{\mathbf{v}_{\mathrm{B}}} f$ be the column vector of partial derivatives of f with respect to each component of $\mathbf{v}_{\mathrm{B}}$. Let $\mathbf{v}_{\mathrm{B}}$ be related to another vector $\mathbf{v}_{\mathrm{A}}$ by the linear relationship $\mathbf{v}_{\mathrm{B}} = \mathbf{N}\mathbf{v}_{\mathrm{A}}$. Use the chain rule to show that the corresponding relationship between the gradient vectors is

$$\nabla_{\mathbf{v}_{\mathrm{A}}} f = \mathbf{N}^{\mathrm{T}} \nabla_{\mathbf{v}_{\mathrm{B}}} f.$$

---

14. **Gradient and Hessian of the cost function w.r.t the control variable**
The incremental cost function written in terms of the control variable is

$$\begin{aligned} J[\delta\boldsymbol{\chi}] &= J_{\mathrm{B}} + \sum_{t=0}^{T} J_{\mathrm{O}}(t), \\ &= \frac{1}{2}\delta\boldsymbol{\chi}^{\mathrm{T}}\delta\boldsymbol{\chi} + \frac{1}{2}\sum_{t=0}^{T}(\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}\delta\boldsymbol{\chi})^{\mathrm{T}}\mathbf{R}_t^{-1}(\delta\mathbf{y}(t) - \mathbf{H}_t\mathbf{M}_{t\leftarrow 0}\mathbf{U}\delta\boldsymbol{\chi}), \end{aligned}$$

using the notation as outlined in the lectures.

   (a) Expand the vector notation to show that the gradient and Hessian of the J_{B} term (w.r.t. $\delta\boldsymbol{\chi}$) are

   $$\nabla_{\delta\boldsymbol{\chi}} J_{\mathrm{B}} = \delta\boldsymbol{\chi}, \qquad \frac{\partial^2 J_{\mathrm{B}}}{\partial \delta\boldsymbol{\chi}^2} = \mathbf{I}.$$

(b) Let $\delta \mathbf{y}^{\mathrm{m}}(t)$ be defined as follows

$$\delta \mathbf{y}^{\mathrm{m}}(t) = \mathbf{H}_t \mathbf{M}_{t\leftarrow 0} \mathbf{U} \delta \boldsymbol{\chi},$$

making the $J_{\mathrm{O}}(t)$ term become

$$J_{\mathrm{O}}(t) = \frac{1}{2} (\delta \mathbf{y}(t) - \delta \mathbf{y}^{\mathrm{m}}(t))^{\mathrm{T}} \mathbf{R}_t^{-1} (\delta \mathbf{y}(t) - \delta \mathbf{y}^{\mathrm{m}}(t)).$$

Expand the vector notation to show that the gradient and Hessian of the J_{B} term (w.r.t. $\delta \mathbf{y}^{\mathrm{m}}(t)$) are

$$\nabla_{\delta \mathbf{y}^{\mathrm{m}}(t)} J_{\mathbf{O}}(t) = -\mathbf{R}_t^{-1} \left(\delta \mathbf{y}(t) - \delta \mathbf{y}^{\mathrm{m}}(t) \right), \qquad \frac{\partial^2 J_{\mathrm{O}}(t)}{\partial \delta \mathbf{y}^{\mathrm{m}}(t)^2} = \mathbf{R}_t^{-1}.$$

(c) Use the generalized chain rule in Q. 13 with $f = J_{\mathrm{O}}(t)$, $\mathbf{v}_{\mathrm{B}} = \mathbf{y}^{\mathrm{m}}(t)$, $\mathbf{v}_{\mathrm{A}} = \delta \boldsymbol{\chi}$ and $\mathbf{N} = \mathbf{H}_t \mathbf{M}_{t\leftarrow 0} \mathbf{U}$ to show that the gradient of the $J_{\mathrm{O}}(t)$ term (w.r.t. $\delta \chi$) is

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

(d) The generalized chain rule in Q. 13 can be extended to second derivatives (when f is a quadratic function) in the following way

$$\text{given} \qquad \nabla_{\mathbf{v}_{\mathrm{A}}} = \mathbf{N}^{\mathrm{T}} \nabla_{\mathbf{v}_{\mathrm{B}}}, \qquad \frac{\partial^2}{\partial \mathbf{v}_{\mathrm{A}}^2} = \nabla_{\mathbf{v}_{\mathrm{A}}} \nabla_{\mathbf{v}_{\mathrm{A}}}^{\mathrm{T}} = \mathbf{N}^{\mathrm{T}} \nabla_{\mathbf{v}_{\mathrm{B}}} \nabla_{\mathbf{v}_{\mathrm{B}}}^{\mathrm{T}} \mathbf{N} = \mathbf{N}^{\mathrm{T}} \frac{\partial^2}{\partial \mathbf{v}_{\mathrm{B}}^2} \mathbf{N}.$$

Use this to show that the Hessian of the $J_{\mathrm{O}}(t)$ term (w.r.t. $\delta \boldsymbol{\chi}$) is

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

(e) Use (a), (c) and (d) to give expressions for the total gradient and total Hessian of the cost function w.r.t. $\delta \boldsymbol{\chi}$.

---

15. **Efficient form of the 4D-VAR gradient**
The total gradient of J_{O} is (see Q. 14)

$$\nabla_{\delta \boldsymbol{\chi}} J_{\mathbf{O}} = -\mathbf{U}^{\mathrm{T}} \sum_{t=0}^{T} \mathbf{M}_{t\leftarrow 0}^{\mathrm{T}} \mathbf{H}_t^{\mathrm{T}} \mathbf{R}_t^{-1} \mathbf{r}(t),$$

where $\mathbf{r}(t) = \delta \mathbf{y}(t) - \mathbf{H}_t \mathbf{M}_{t\leftarrow 0} \mathbf{U} \delta \boldsymbol{\chi}$. This is inefficient to evaluate because it involves acting with the $T+1$ matrices $\mathbf{M}_{t\leftarrow 0}^{\mathrm{T}}$. We will now investigate a more efficient form of this summation.

(a) Consider each adjoint model operator in the form

$$\mathbf{M}_{t\leftarrow 0}^{\mathrm{T}} = \mathbf{M}_{1\leftarrow 0}^{\mathrm{T}} \mathbf{M}_{2\leftarrow 1}^{\mathrm{T}} \ldots \mathbf{M}_{t-1\leftarrow t-2}^{\mathrm{T}} \mathbf{M}_{t\leftarrow t-1}^{\mathrm{T}},$$

and write each contribution to the time summation on a separate line (e.g. write $t=0$, $t=1$, $t=2$, $t=T-1$ and $t=T$ lines explicitly).

(b) Use the fact that many $\mathbf{M}_{j\leftarrow i}^{\mathrm{T}}$ operators are shared to write the gradient algorithm in the following way.

i. Define $\lambda(T) = \mathbf{H}_T^{\mathrm{T}}\mathbf{R}_T^{-1}\mathbf{r}(T)$.

ii. Apply the following for $t = T-1 \to 0$:

$$\lambda(t) = \mathbf{H}_t^{\mathrm{T}}\mathbf{R}_t^{-1}\mathbf{r}(t) + \mathbf{M}_{t+1\leftarrow t}^{\mathrm{T}}\lambda(t+1).$$

iii. The required gradient is then $\nabla_{\delta\mathbf{\chi}} J_{\mathbf{O}} = -\mathbf{U}^{\mathrm{T}}\lambda(0)$.

---

16. **The NMC method**

The NMC method uses the difference between two forecasts of different lengths, but valid at the same time as a proxy for forecast error (e.g. a forecast of 48-hours, $\mathbf{x}^{\mathrm{f48}}$, and a forecast of 24-hours, $\mathbf{x}^{\mathrm{f24}}$). Note the following forms of $\mathbf{x}^{\mathrm{f48}}$ and $\mathbf{x}^{\mathrm{f24}}$

$$\mathbf{x}^{\mathrm{f48}} = \mathbf{x}^{\mathrm{truth}} + \boldsymbol{\eta}^{48}, \qquad \mathbf{x}^{\mathrm{f24}} = \mathbf{x}^{\mathrm{truth}} + \boldsymbol{\eta}^{24},$$

where $\mathbf{x}^{\mathrm{truth}}$ is the true state and $\boldsymbol{\eta}^{t}$ is the forecast error.

(a) By assuming that $\boldsymbol{\eta}^{48}$ and $\boldsymbol{\eta}^{24}$ each have the same covariance, $\mathbf{B}$, and are uncorrelated, show that the following form of $\mathbf{B}$ results

$$\mathbf{B} = \frac{1}{2}\left\langle(\mathbf{x}^{\mathrm{f48}} - \mathbf{x}^{\mathrm{f24}})(\mathbf{x}^{\mathrm{f48}} - \mathbf{x}^{\mathrm{f24}})^{\mathrm{T}}\right\rangle,$$

where $\langle\rangle$ is the average over a population of such forecast differences.

(b) Discuss possible flaws in the NMC method.
