Solutions to problem sheet 2 (3d-Var.)

Ross Bannister, Room 1U11, Dept. of Meteorology, Univ. of Reading

r.n.bannister@reading.ac.uk.

Version 2010.

1. (a) (i) Expanding, $J_B = \frac{1}{2}\sum_{ij}(x_i - x_i^B)\mathbf{B}_{ij}^{-1}(x_j - x_j^B)$ where $\mathbf{B}_{ij}^{-1} = (\mathbf{B}^{-1})_{ij}$ and not $(\mathbf{B}_{ij})^{-1}$.
Differentiate with respect to x_k using the product rule for differentiation

$$\frac{\partial J_B}{\partial x_k} = \frac{1}{2}\left(\sum_{ij}(x_i - x_i^B)\mathbf{B}_{ij}^{-1}\delta_{jk} + \sum_{ij}\delta_{ik}\mathbf{B}_{ij}^{-1}(x_j - x_j^B)\right),$$

$$= \frac{1}{2}\left(\sum_{i}(x_i - x_i^B)\mathbf{B}_{ik}^{-1} + \sum_{j}\mathbf{B}_{kj}^{-1}(x_j - x_j^B)\right).$$

These partial derivatives can be assembled into a vector as shown in the handout, §D.1. In the first term of the above, use the fact that $\mathbf{B}^{-1}$ is symmetric, $\mathbf{B}_{ik}^{-1} = \mathbf{B}_{ki}^{-1}$, and in the second term relabel $j \rightarrow i$

$$\frac{\partial J_B}{\partial x_k} = \frac{1}{2}\left(\sum_{i}(x_i - x_i^B)\mathbf{B}_{ki}^{-1} + \sum_{i}\mathbf{B}_{ki}^{-1}(x_i - x_i^B)\right),$$

$$= \sum_{i}\mathbf{B}_{ki}^{-1}(x_i - x_i^B). \qquad \text{(A)}$$

This is the kth element of the vector $\nabla_x J_B = \mathbf{B}^{-1}(\vec{x} - \vec{x}_B)$.
(ii) Differentiate (A) again, with respect to x_l

$$\frac{\partial^2 J_B}{\partial x_k \partial x_l} = \sum_i \mathbf{B}_{ki}^{-1}\delta_{il} = \mathbf{B}_{kl}^{-1}.$$

This is simply the $(k,\ l)$ matrix element of $\mathbf{B}^{-1}$.

(b) (i) Expanding, $J_O = \sum_{qr}(y_q - h_q)\mathbf{R}_{qr}^{-1}(y_r - h_r)$, where $\mathbf{R}_{qr}^{-1} = (\mathbf{R}^{-1})_{qr}$ and not $(\mathbf{R}_{qr})^{-1}$.
Differentiate with respect to h_i using the product rule for differentiation

$$\frac{\partial J_O}{\partial h_i} = -\frac{1}{2}\left(\sum_{qr}(y_q - h_q)\mathbf{R}_{qr}^{-1}\delta_{ri} + \sum_{qr}\delta_{qi}\mathbf{R}_{qr}^{-1}(y_r - h_r)\right),$$

$$= -\frac{1}{2}\left(\sum_{q}(y_q - h_q)\mathbf{R}_{qi}^{-1} + \sum_{r}\mathbf{R}_{ir}^{-1}(y_r - h_r)\right).$$

These partial derivatives can be assembled into a vector as shown in the handout, §D.1. In the first term of the above, use the fact that $\mathbf{R}^{-1}$ is symmetric, $\mathbf{R}_{qi}^{-1} = \mathbf{R}_{iq}^{-1}$, and in the second term relabel $r \rightarrow q$

$$\frac{\partial J_O}{\partial h_i} = -\frac{1}{2}\left(\sum_{q}(y_q - h_q)\mathbf{R}_{iq}^{-1} + \sum_{q}\mathbf{R}_{iq}^{-1}(y_q - h_q)\right),$$

$$= -\sum_q \mathbf{R}_{iq}^{-1}(y_q - h_q). \qquad \text{(B)}$$

This is the ith element of the vector $\nabla_h J_O = -\mathbf{R}^{-1}(\vec{y} - \vec{h})$. Now use the generalised chain rule with (B) to find $\partial J_O / \partial x_k$

$$\frac{\partial J_O}{\partial x_k} = -\sum_{i=1}^{p}\mathbf{H}_{ik}\sum_q \mathbf{R}_{iq}^{-1}(y_q - h_q),$$

$$= -\sum_{i=1}^{p}\mathbf{H}_{ki}^{\mathrm{T}}\sum_q \mathbf{R}_{iq}^{-1}(y_q - h_q). \qquad \text{(C)}$$

This is the kth element of the vector $\nabla_x J_O = -\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\vec{y} - \vec{h})$.
(ii) Differentiate (C) again, with respect to x_l

$$\frac{\partial^2 J_O}{\partial x_k \partial x_l} = \sum_{i=1}^{p}\mathbf{H}_{ki}^{\mathrm{T}}\sum_q \mathbf{R}_{iq}^{-1}\frac{\partial h_q}{\partial x_l},$$

$$= \sum_{i=1}^{p}\mathbf{H}_{ki}^{\mathrm{T}}\sum_q \mathbf{R}_{iq}^{-1}\mathbf{H}_{ql}.$$

This is the $(k,\ l)$ matrix element of $\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$.

2. (a) The linearized OI formula is

$$\vec{x}_A = \vec{x}_B + \mathbf{K}(\vec{y} - \vec{h}[\vec{x}_t] - \mathbf{H}\vec{\varepsilon}_B).$$

Take away $\vec{x}_t$ from each side

$$\vec{\varepsilon}_A = \vec{\varepsilon}_B + \mathbf{K}(\vec{\varepsilon}_y - \mathbf{H}\vec{\varepsilon}_B).$$

(b) Use this to evaluate $\mathbf{P}_A = \langle\vec{\varepsilon}_A\vec{\varepsilon}_A^{\mathrm{T}}\rangle$

$$\mathbf{P}_A = \langle\{\mathbf{K}(\vec{\varepsilon}_y - \mathbf{H}\vec{\varepsilon}_B) + \vec{\varepsilon}_B\}\{\mathbf{K}(\vec{\varepsilon}_y - \mathbf{H}\vec{\varepsilon}_B) + \vec{\varepsilon}_B\}^{\mathrm{T}}\rangle,$$

$$= \langle\mathbf{K}(\vec{\varepsilon}_y - \mathbf{H}\vec{\varepsilon}_B)(\vec{\varepsilon}_y - \mathbf{H}\vec{\varepsilon}_B)^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} + \vec{\varepsilon}_B\vec{\varepsilon}_B^{\mathrm{T}} +$$

$$\mathbf{K}(\vec{\varepsilon}_y - \mathbf{H}\vec{\varepsilon}_B)\vec{\varepsilon}_B^{\mathrm{T}} + \vec{\varepsilon}_B(\vec{\varepsilon}_y - \mathbf{H}\vec{\varepsilon}_B)^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\rangle,$$

$$= \mathbf{K}\langle\vec{\varepsilon}_y\vec{\varepsilon}_y^{\mathrm{T}}\rangle\mathbf{K}^{\mathrm{T}} + \mathbf{K}\mathbf{H}\langle\vec{\varepsilon}_B\vec{\varepsilon}_B^{\mathrm{T}}\rangle\mathbf{H}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} -$$

$$\mathbf{K}\langle\vec{\varepsilon}_y\vec{\varepsilon}_B^{\mathrm{T}}\rangle\mathbf{H}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} - \mathbf{K}\mathbf{H}\langle\vec{\varepsilon}_B\vec{\varepsilon}_y^{\mathrm{T}}\rangle\mathbf{K}^{\mathrm{T}} +$$

$$\langle\vec{\varepsilon}_B\vec{\varepsilon}_B^{\mathrm{T}}\rangle + \mathbf{K}\langle\vec{\varepsilon}_y\vec{\varepsilon}_B^{\mathrm{T}}\rangle - \mathbf{K}\mathbf{H}\langle\vec{\varepsilon}_B\vec{\varepsilon}_B^{\mathrm{T}}\rangle +$$

$$\langle\vec{\varepsilon}_B\vec{\varepsilon}_y^{\mathrm{T}}\rangle\mathbf{K}^{\mathrm{T}} - \langle\vec{\varepsilon}_B\vec{\varepsilon}_B^{\mathrm{T}}\rangle\mathbf{H}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}.$$

Substitute the outer products ($\langle\vec{\varepsilon}_A\vec{\varepsilon}_A^{\mathrm{T}}\rangle$, $\langle\vec{\varepsilon}_B\vec{\varepsilon}_B^{\mathrm{T}}\rangle$ and $\langle\vec{\varepsilon}_y\vec{\varepsilon}_y^{\mathrm{T}}\rangle$) with the respective error covariances. It is usual to assume that the observation errors are uncorrelated with the background errors, ie, $\langle\vec{\varepsilon}_y\vec{\varepsilon}_B^{\mathrm{T}}\rangle = 0$ and $\langle\vec{\varepsilon}_B\vec{\varepsilon}_y^{\mathrm{T}}\rangle = 0$

$$\mathbf{P}_A = \mathbf{KRK}^{\mathrm{T}} + \mathbf{KHBH}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} + \mathbf{B} - \mathbf{KHB} - \mathbf{BH}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}.$$

The last two terms in the above are equal (this is proved by substituting with $\mathbf{K}$)

$$\mathbf{P}_A = \mathbf{K}(\mathbf{R} + \mathbf{HBH}^{\mathrm{T}})\mathbf{K}^{\mathrm{T}} + (\mathbf{I} - 2\mathbf{KH})\mathbf{B}.$$

Substitute with $\mathbf{K}$ into the second occurrence of $\mathbf{K}$ only

$$\mathbf{P}_A = \mathbf{K}(\mathbf{R} + \mathbf{HBH}^{\mathrm{T}})(\mathbf{R} + \mathbf{HBH}^{\mathrm{T}})^{-1}\mathbf{HB} +$$

$$(\mathbf{I}-2\mathbf{KH})\mathbf{B},$$

$$= \mathbf{KHB} + (\mathbf{I}-2\mathbf{KH})\mathbf{B} = (\mathbf{I}-\mathbf{KH})\mathbf{B}.$$

The analysis error covariance is found to be the background error covariance reduced by **KHB** due to the introduction of observational information (combining obs. with the background state reduces uncertainty).

(c) Substituting with **K** and using the S-M-W identity in Q5 gives

$$\mathbf{P}_A = \{\mathbf{I}-\mathbf{BH}^\mathrm{T}(\mathbf{R}+\mathbf{HBH}^\mathrm{T})^{-1}\mathbf{H}\}\mathbf{B},$$

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

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

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

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

$$\{\mathbf{I}+\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{HB}-\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{HB}\},$$

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

3. (a) Acting with the inverse matrix on the forward matrix should give the identity

$$\frac{1}{ac-b^2}\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}\begin{pmatrix} a & b \\ b & c \end{pmatrix} =$$

$$\frac{1}{ac-b^2}\begin{pmatrix} ca-b^2 & cb-bc \\ -ba+ab & -b^2+ac \end{pmatrix} = \mathbf{I}.$$

(b) The inverse cannot be found if the determinant ($ac-b^2$ for this 2 × 2 matrix) is zero.

(c) The matrix is singular.

(d) A singular Hessian matrix means that the cost function is flat (zero curvature) in at least one direction in phase space. It then has no unique minimum and has an infinitely large error (the error covariance is the inverse of the Hessian - see Q2) in that (or those) direction(s).

4. (a) The total column ozone in a column is the sum over the $n-1$ layers

$$\frac{1}{4}\sum_{i=1}^{n-1}(\rho_i+\rho_{i+1})(\phi_i+\phi_{i+1})(z_{i+1}-z_i).$$

(b) The observation covariance matrix and its inverse

$$\mathbf{R} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}, \qquad \mathbf{R}^{-1} = \begin{pmatrix} \sigma_1^{-2} & 0 \\ 0 & \sigma_2^{-2} \end{pmatrix}.$$

(ie since **R** is diagonal we can simply invert the diagonal elements). Note that the diagonal elements of **R** are *variances*, which are the squares of the *standard deviations* specified.

(c) The innovation vector is

obs − background model obs

$$= \begin{pmatrix} y_1 - \frac{1}{4}\sum_{i=1}^{n-1}(\rho_i+\rho_{i+1})(\phi_i^B+\phi_{i+1}^B)(z_{i+1}-z_i) \\ y_2 - \frac{1}{4}\sum_{i=1}^{n-1}(\rho_i+\rho_{i+1})(\phi_i^B+\phi_{i+1}^B)(z_{i+1}-z_i) \end{pmatrix}.$$

(d) There are $2n$ elements to the Jacobian. We are asked to work out six of them. First work out the sensitivity of the kth observation to the j model ozone concentration

$$\frac{\partial h_k}{\partial \phi_j} = \frac{1}{4}\sum_{i=1}^{n-1}(\rho_i+\rho_{i+1})\left(\frac{\partial \phi_i}{\partial \phi_j}+\frac{\partial \phi_{i+1}}{\partial \phi_j}\right)(z_{i+1}-z_i)$$

$$= \frac{1}{4}\sum_{i=1}^{n-1}(\rho_i+\rho_{i+1})\left(\delta_{i,j}+\delta_{i+1,j}\right)(z_{i+1}-z_i)$$

$$= \frac{1}{4}(\rho_j+\rho_{j+1})(z_{j+1}-z_j)(1-\delta_{j,n}) +$$

$$\frac{1}{4}(\rho_{j-1}+\rho_j)(z_j-z_{j-1})(1-\delta_{j,1})$$

$(1-\delta_{j,n})$ excludes $j=n$,

$(1-\delta_{j,1})$ excludes $j=1$.

This result is independent of k because the two measurements $k = 1,2$ have the same forward model. Substitute this result to find elements of the Jacobian.

$$k=1,2, j=1: \frac{\partial h_k}{\partial x_1} = \frac{1}{4}(\rho_1+\rho_2)(z_2-z_1),$$

$$k=1,2, j=2: \frac{\partial h_k}{\partial x_2} = \frac{1}{4}(\rho_2+\rho_3)(z_3-z_2) +$$

$$\frac{1}{4}(\rho_1+\rho_2)(z_2-z_1),$$

$$k=1,2, j=n: \frac{\partial h_k}{\partial x_n} = \frac{1}{4}(\rho_{n-1}+\rho_n)(z_n-z_{n-1}).$$

The observation operator is linear and so the Jacobian is independent of ϕ values.

(e) $J = \nabla_x J_B + \nabla_x J_O = \mathbf{B}^{-1}(\vec{x}-\vec{x}_B) - \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}(\vec{y}-\vec{h}[\vec{x}])$.

On the first iteration of Var. , $\vec{x} = \vec{x}_B$, and so the contribution to the first term is zero.

(f) The Hessian is the sum of the observation and background second derivatives

$$\text{Hessian} = \mathbf{B}^{-1}+\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}$$

$$= \begin{pmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2n} \\ \dots & \dots & \dots & \dots \\ \beta_{n1} & \beta_{n2} & \dots & \beta_{nn} \end{pmatrix} +$$

$$\begin{pmatrix} \partial h_1/\partial\phi_1 & \partial h_2/\partial\phi_1 \\ \partial h_1/\partial\phi_2 & \partial h_2/\partial\phi_2 \\ \dots & \dots \\ \partial h_1/\partial\phi_n & \partial h_2/\partial\phi_n \end{pmatrix} \begin{pmatrix} \sigma_1^{-2} & 0 \\ 0 & \sigma_2^{-2} \end{pmatrix}$$

$$\times \begin{pmatrix} \partial h_1/\partial\phi_1 & \partial h_1/\partial\phi_2 & \dots & \partial h_1/\partial\phi_n \\ \partial h_2/\partial\phi_1 & \partial h_2/\partial\phi_2 & \dots & \partial h_2/\partial\phi_n \end{pmatrix}$$

(g) The inverse of the Hessian evaluated at the minimum of the cost function is the covariance matrix for the least squares fitting procedure (ie the error covariance matrix of the analysis). The diagonal elements of Hessian^{-1} are the analysis variances, σ_i^2 (the square of the analysis errors). In operational data assimilation however n is too large to allow explicit calculation of the Hessian.

(h) The Hessian should be positive definite, ie $\vec{z}^{\mathrm{T}}$ Hessian$^{-1}\,\vec{z} > 0 \;\forall\vec{z} \neq 0$. If $\vec{z}$ is a unit eigenvector of the Hessian for example (with eigenvalue λ), then the positive definite test simplifies to

$$\vec{z}^{\mathrm{T}}\ \text{Hessian}^{-1}\,\vec{z} = \vec{z}^{\mathrm{T}}\vec{z}\lambda = \lambda > 0.$$

This will be satisfied providing that all eigenvalues are positive. Compare this in 1d where for a function to have a minimum, $\mathrm{d}^2 J / \mathrm{d}x^2 > 0$.

5. Simply take the matrices that are outside of the brackets inside. Allow matrices to cancel with inverse matrices and preserve the ordering.

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

$$\mathbf{H}^T + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T \stackrel{?}{=} \mathbf{H}^T + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T$$

6. Hints:

- An analysis increment field is $\vec{x}_A - \vec{x}_B$.
- A single observation means that $\vec{y}$ is a scalar, y. Also $\mathbf{R}$ is a scalar, σ_O^2.
- Let the observation be of the quantity stored at the kth position of the state vector, x_k. The forward operator is a row vector full of '0's, apart from '1' at position k, $\mathbf{H} = (0 \dots\ 1 \dots\ 0)$.