Problem Sheet 2 (3d-Var.)

Ross Bannister, Room 1U11, Dept. of Meteorology, Univ. of Reading
r.n.bannister@reading.ac.uk.

Version 2010.

1. The gradient and Hessian of the 3d-Var. cost function derived from first principles

(a) The background term

The background term of the 3d-Var. cost function in matrix notation is the scalar

$$J_B = \frac{1}{2}(\vec{x} - \vec{x}_B)^{\mathrm{T}} \mathbf{B}^{-1} (\vec{x} - \vec{x}_B) .$$

This is a matrix expression (matrix algebra has implied summations).

(i) Derive the first derivative J_B

Expand-out this matrix expression as a double summation (indices i and j), each running over the n components of $\vec{x}$.

Find the partial derivative of J_B with respect to one component of $\vec{x}$ (say x_k).

Arrange the derivatives with respect to each x_k ($1 \leqslant k \leqslant n$) into a column vector (ie let $\partial J_B / \partial x_k$ be the kth element in the vector).

Given that the matrix $\mathbf{B}^{-1}$ is symmetric, show that the matrix expression

$$\mathbf{B}^{-1} (\vec{x} - \vec{x}_B) ,$$

evaluates to the same column vector. This expression is often denoted as $\nabla_x J_B$.

(ii) Derive the second derivative of J_B

Differentiate again your expanded expression for $\partial J_B / \partial x_k$, with respect to a different component of $\vec{x}$ (say x_l). Show that by arranging the components in a matrix (ie let $\partial^2 J_B / \partial x_k \partial x_l$ form the (k, l)th matrix element), that the result is $\mathbf{B}^{-1}$.

(b) The observation term

The observation term of the 3d-Var. cost function in matrix notation is the scalar

$$J_O = \frac{1}{2}(\vec{y} - \vec{h})^{\mathrm{T}} \mathbf{R}^{-1} (\vec{y} - \vec{h}) ,$$

where $\vec{h}$ is a function of $\vec{x}$ and maps from n-element $\vec{x}$ space to p-element $\vec{y}$ space.

(i) Derive the first derivative of J_O

Expand-out this matrix expression as a double summation (indices q and r), each running over the p components of $\vec{y}$ and $\vec{h}$.

Find the partial derivative of J_O with respect to one component of $\vec{h}$ (say h_i).

Arrange the derivatives with respect to each h_i ($1 \leqslant i \leqslant p$) into a column vector (ie let $\partial J_O / \partial h_i$ be the ith element in the vector).

Given that the matrix $\mathbf{R}^{-1}$ is symmetric, show that the matrix expression

$$-\mathbf{R}^{-1} (\vec{y} - \vec{h}) ,$$

evaluates to the same column vector.

The vector function $\vec{h}$ is a function of $\vec{x}$. Let it be linearized about x_0

$$\vec{h}[\vec{x}] \approx \vec{h}[\vec{x}_0] + \mathbf{H}(\vec{x} - \vec{x}_0),$$
$$= \mathbf{H}\vec{x} + \overrightarrow{\text{constant}},$$

where $\mathbf{H}$ is the Jacobian matrix

$$\mathbf{H} = \left.\frac{\partial \vec{h}[\vec{x}]}{\partial \vec{x}}\right|_{x_0}, \qquad \text{comprising elements} \qquad \mathbf{H}_{ik} = \left.\frac{\partial h_i[\vec{x}]}{\partial x_k}\right|_{x_0}.$$

The generalised chain rule relates partial derivatives of a scalar with respect to x_k to derivatives of the scalar with respect to h_i as follows

$$\frac{\partial J_O}{\partial x_k} = \sum_{i=1}^{p} \frac{\partial h_i}{\partial x_k}\frac{\partial J_O}{\partial h_i}.$$

Apply this result to the partial derivatives with respect to h_i found above.
Arrange the derivatives with respect to each x_k ($1 \leqslant k \leqslant n$) into a column vector (ie let $\partial J_O / \partial x_k$ be the kth element in the vector). Show that the matrix expression

$$-\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\vec{y} - \vec{h}),$$

evaluates to the same column vector. This expression is often denoted as $\nabla_x J_O$.

(ii) Derive the second derivative of J_O
Differentiate again your expanded expression for $\partial J_O / \partial x_k$, with respect to a different component of $\vec{x}$ (say x_l). Show that by arranging the components in a matrix (ie let $\partial^2 J_O / \partial x_k \partial x_l$ form the (k, l)th matrix element), that the result is $\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$.

2. Proof that the analysis error covariance matrix is the inverse Hessian

Let x_0 represent the unknown 'true' state of the atmosphere. The analysis, $\vec{x}_A$, background, $\vec{x}_B$ and observations, $\vec{y}$ may then be defined as the true state plus unknown errors $\vec{\varepsilon}_A$, $\vec{\varepsilon}_B$ and $\vec{\varepsilon}_y$ respectively, which have error covariance matrices $\mathbf{P}_A$, $\mathbf{B}$ and $\mathbf{R}$ respectively

$$\vec{x}_A = \vec{x}_t + \vec{\varepsilon}_A, \qquad \mathbf{P}_A = \langle \vec{\varepsilon}_A \vec{\varepsilon}_A^{\mathrm{T}} \rangle,$$
$$\vec{x}_B = \vec{x}_t + \vec{\varepsilon}_B, \qquad \mathbf{B} = \langle \vec{\varepsilon}_B \vec{\varepsilon}_B^{\mathrm{T}} \rangle,$$
$$\vec{y} = \vec{h}[\vec{x}_t] + \vec{\varepsilon}_y, \qquad \mathbf{R} = \langle \vec{\varepsilon}_y \vec{\varepsilon}_y^{\mathrm{T}} \rangle,$$

where angled brackets denote the mean value. The optimal interpolation (OI) formula, which approximates the variational analysis, is

$$\vec{x}_A = \vec{x}_B + \mathbf{K}(\vec{y} - \vec{h}[\vec{x}_B]),$$
$$\text{where } \mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}})^{-1}.$$

The observation operator $\vec{h}[\vec{x}_B]$ may be linearized about x_t

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

(a) Substitute this linearization into the OI formula and derive an expression for $\vec{\varepsilon}_A$.

(b) Show that the matrix $\mathbf{P}_A$ (as defined above) has the form

$$\mathbf{P}_A = (\mathbf{I} - \mathbf{KH})\,\mathbf{B},$$

where **I** is the identity matrix.

(c) Use the definition of **K** and the Sherman-Morrison-Woodbury identity shown in Q5 to show that $\mathbf{P}_A$ can have the form

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

which is the inverse of the Hessian matrix, $\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$.

3. Finding the inverse of a symmetric matrix

The formula for the inverse of a 2×2 symmetric matrix is

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

(a) Confirm that this result is the inverse matrix.

(b) Under what circumstance can the inverse not be evaluated?

(c) Is the matrix said to be *singular* or *non-singular* under this circumstance?

(d) Interpret the meaning of the case when the Hessian matrix is singular.

4. Forward model example and its adjoint - total column amount

In chemical data assimilation, the aim is to assimilate observed chemical concentrations into a chemical transport model (CTM) of the atmosphere. A common observation product produced from nadir viewing (downward looking) satellites is a so-called 'total column' amount. This is a non-local quantity and requires a forward model. Variational assimilation can use the information contained in such a measurement via the forward model and its adjoint.

A transport model may represent ozone concentrations on a set of n vertical height levels. The level height, z_i, the air density ρ_i, and the ozone mass mixing ratio, ϕ_i are stored on each level. Level 1 is the Earth's surface and level n is well above the ozone layer.

(a) Given that the amount of ozone per unit horizontal area in one of the $n - 1$ *layers* of the model 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 mixing ratio $\times$ layer thickness), write down the total column ozone per unit area according to the model state.

(b) A particular nadir viewing satellite contains two separate instruments from which the total column ozone can be deduced. One measurement from each instrument is made at the same time. These measurements have the following characteristics:

Measurement	Value	Standard deviation
1	y_1	σ_1
2	y_2	σ_2

Given that errors of the two measurements are uncorrelated, write down the inverse of the observation error covariance matrix, $\mathbf{R}^{-1}$.

(c) The two measurements are to be used to improve the model's representation of ozone via the '3d-Var.' procedure. Given that the background model ozone values are ϕ_i^B ($i = 1, n$), write down the two-element innovation vector.

(d) The Jacobian, **H** says how sensitive the model observations are to changes in model values. Give expressions for the following six matrix elements of **H**

$$\left.\begin{pmatrix} \partial h_1 / \partial \phi_1 & \partial h_1 / \partial \phi_2 & \dots & \partial h_1 / \partial \phi_n \\ \partial h_2 / \partial \phi_1 & \partial h_2 / \partial \phi_2 & \dots & \partial h_2 / \partial \phi_n \end{pmatrix}\right|_{x_B},$$

where $\vec{x}_B$ is the vector representing model background ozone values, ie

$$\vec{x}_B = (\phi_1^B, \phi_2^B, \dots , \phi_n^B)^{\mathrm{T}}.$$

(Note that the two rows of **H** are identical as both measurements are of the same thing!)

(e) Write down the gradient $\nabla_x J$. Explain why the background term does not contribute to the gradient on the first iteration of 3d-Var.

(f) Write down the Hessian for this problem. Let the inverse of the background error covariance matrix be:

$$\mathbf{B}^{-1} = \begin{pmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2n} \\ \dots & \dots & \dots & \dots \\ \beta_{n1} & \beta_{n2} & \dots & \beta_{nn} \end{pmatrix}.$$

Check that the observation component of the Hessian is an $n \times n$ matrix.

(g) How can we use the Hessian, evaluated at the cost function minimum, to estimate the analysis errors?

(h) What property of the Hessian implies that the cost function is convex (has a minimum with respect to all variables)?

5. Proof of matrix identity

Prove that the following matrix identity (called the Sherman-Morrison-Woodbury formula) holds

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

6. Assimilation of a single observation in Var. to probe the background error covariance structure

By using the equivalence between the optimal interpolation formula and Var., show that the assimilation of a single direct observation results in analysis increments that are proportional to a column of the background error covariance matrix.