Vector Derivatives
(and Application to Differentiating the Cost Function)

Ross Bannister, December 2000/August 2001

1. Vector derivative

Let $\vec{y} = \mathbf{A}\vec{x}$, where $\mathbf{A}$ is a matrix. Prove that the vector derivative $\mathrm{d}\vec{y} / d\vec{x} = \mathbf{A}$.

Expanding out the linear operator expression,

$$y_i = \sum_k A_{ik} x_k. \tag{1.1}$$

With the vector derivative, $\mathrm{d} / \mathrm{d}\vec{x}$ defined as the row vector,

$$\frac{\mathrm{d}}{\mathrm{d}\vec{x}} \equiv \begin{pmatrix} \partial / \partial x_1 & \partial / \partial x_2 & \partial / \partial x_3 & \dots \end{pmatrix}, \tag{1.2}$$

the definition of $\mathrm{d}\vec{y} / d\vec{x}$ is,

$$\frac{\mathrm{d}\vec{y}}{\mathrm{d}\vec{x}} \equiv \left(\left(\frac{\mathrm{d}}{\mathrm{d}\vec{x}} \right)^T \vec{y}^T \right)^T, \tag{1.3}$$

$$= \left(\begin{pmatrix} \partial / \partial x_1 \\ \partial / \partial x_2 \\ \partial / \partial x_3 \\ \dots \end{pmatrix} \begin{pmatrix} y_1 & y_2 & y_3 & \dots \end{pmatrix} \right)^T,$$

$$= \begin{pmatrix} \partial y_1 / \partial x_1 & \partial y_1 / \partial x_2 & \partial y_1 / \partial x_3 & \dots \\ \partial y_2 / \partial x_1 & \partial y_2 / \partial x_2 & \partial y_2 / \partial x_3 & \dots \\ \partial y_3 / \partial x_1 & \partial y_3 / \partial x_2 & \partial y_3 / \partial x_3 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}. \tag{1.4}$$

Differentiating Eq. (1.1) with respect to an arbitrary component gives us information regarding the matrix elements of Eq. (1.4),

$$\frac{\partial y_i}{\partial x_j} = A_{ij}. \tag{1.4}$$

Thus, by the definition of Eq. (1.3), we have proved that $\partial\vec{y} / \partial\vec{x} = \mathbf{A}$.

2. First and second derivative of a quadratic scalar (cost function)

Consider the cost function as used in variational data assimilation. This is comprised of a background term and an observation term. Respectively these are:

$$J[\vec{x}] = \frac{1}{2}(\vec{x}^B - \vec{x})^T \mathbf{B}^{-1}(\vec{x}^B - \vec{x}) + \frac{1}{2}(\vec{y} - \vec{H}[\vec{x}])^T \mathbf{E}^{-1}(\vec{y} - \vec{H}[\vec{x}]), \qquad (2.1)$$

where $\vec{x}^B$ is the background term with error covariance matrix $\mathbf{B}$, $\vec{x}$ is the variational state, $\vec{y}$ is the vector of observations with error covariance matrix $\mathbf{Q}$ and $\vec{H}[\vec{x}]$ is the forward model. This is a vector functional of the variational state and is the model's prediction of the observations. When the cost function is minimized, the variational state is known as the analysis. The cost function is usually written in incremental form, that is with the variational state written as an increment to a linearization state:

$$J[\vec{x}_0 + \delta\vec{x}] = \frac{1}{2}(\delta\vec{x}^B - \delta\vec{x})^T \mathbf{B}^{-1}(\delta\vec{x}^B - \delta\vec{x}) + \frac{1}{2}(\vec{y} - \vec{H}[\vec{x}_0 + \delta\vec{x}])^T \mathbf{E}^{-1}(\vec{y} - \vec{H}[\vec{x}_0 + \delta\vec{x}]), \qquad (2.2)$$

where

$$\vec{x} = \vec{x}_0 + \delta\vec{x} \qquad \text{and} \qquad (2.3)$$

$$\vec{x}^B = \vec{x}_0 + \delta\vec{x}^B \qquad (2.4)$$

have been substituted into Eq. (2.1) to obtain Eq. (2.2).

In order to differentiate Eq. (2.1), we find it easiest to first expand the matrix notation into explicit summation form:

$$J[\vec{x}_0 + \delta\vec{x}] = \frac{1}{2}\sum_{ij}(\delta x_i^B - \delta x_i)\mathbf{B}_{ij}^{-1}(\delta x_j^B - \delta x_j) + \frac{1}{2}\sum_{ij}(y_i - H_i[\vec{x}_0 + \delta\vec{x}])\mathbf{E}_{ij}^{-1}(y_j - H_j[\vec{x}_0 + \delta\vec{x}]). \qquad (2.5)$$

In Eq. (2.5), we note that a term like $\mathbf{B}_{ij}^{-1}$ is the (i, j) th component of $\mathbf{B}^{-1}$ (and not the reciprocal of the (i, j) th component of $\mathbf{B}$). Differentiate with respect to the k th component of $\delta\vec{x}$:

$$\frac{\partial J}{\partial \delta x_k} = \frac{1}{2}\sum_{ij}\left\{(\delta x_i^B - \delta x_i)\mathbf{B}_{ij}^{-1}(-\delta_{jk}) + (-\delta_{ik})\mathbf{B}_{ij}^{-1}(\delta x_j^B - \delta x_j)\right\} +$$
$$\frac{1}{2}\sum_{ij}\left\{(y_i - H_i[\vec{x}_0 + \delta\vec{x}])\mathbf{E}_{ij}^{-1}\left(-\frac{\partial H_j[\vec{x}_0 + \delta\vec{x}]}{\partial \delta x_k}\right) + \left(-\frac{\partial H_i[\vec{x}_0 + \delta\vec{x}]}{\partial \delta x_k}\right)\mathbf{E}_{ij}^{-1}(y_j - H_j[\vec{x}_0 + \delta\vec{x}])\right\}, \qquad (2.6)$$

where δ_{ij} is the kronecker delta function. Exploiting the fact that the error covariance matrices are symmetric, and using the following substitution,

$$\frac{\partial H_i[\vec{x}_0 + \delta\vec{x}]}{\partial \delta x_k} = \mathbf{H}_{ik}, \qquad (2.7)$$

(which is effectively a linearization of the forward model), we arrive at,

$$\frac{\partial J}{\partial \delta x_k} = -\sum_j \mathbf{B}_{kj}^{-1}(\delta x_j^B - \delta x_j) - \sum_{ij}\mathbf{H}_{ik}\mathbf{E}_{ij}^{-1}(y_j - H_j[\vec{x}_0 + \delta\vec{x}]). \qquad (2.8)$$

This can now be compactly written back in matrix notation by assembling a column vector of partial derivatives from all values of the index k,

$$\nabla_{\delta x} J = \left(\frac{\mathrm{d}J}{\mathrm{d}\delta\vec{x}}\right)^T = -\mathbf{B}^{-1}(\delta\vec{x}^B - \delta\vec{x}) - \mathbf{H}^T\mathbf{E}^{-1}(\vec{y} - \vec{H}[\vec{x}_0 + \delta\vec{x}]) . \tag{2.9}$$

The first derivative is thus a vector. The second derivative is a matrix and is called the Hessian. In order to find the Hessian matrix, **H***e*, we return to Eq. (2.8) and differentiate with respect to another component of the variational state, δx_l,

$$\frac{\partial^2 J}{\partial\delta x_k \partial\delta x_l} = \sum_j \mathbf{B}_{kj}^{-1}(-\delta_{jl}) + \sum_{ij}\mathbf{H}_{ik}\mathbf{E}_{ij}^{-1}(y_j - H_{jl}) , \tag{2.10}$$

$$= \mathbf{B}_{kl}^{-1} + \sum_{ij}\mathbf{H}_{ik}\mathbf{E}_{ij}^{-1}H_{jl}, \tag{2.11}$$

which may, once more, be written in the compact matrix notation,

$$\mathbf{H}e = \left(\frac{\mathrm{d}}{\mathrm{d}\delta\vec{x}}\right)^T \frac{\mathrm{d}J}{\mathrm{d}\delta\vec{x}} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{E}^{-1}\mathbf{H} \tag{2.12}$$

Sometimes, for reasons of preconditioning, the variational procedure is performed in a space which is a transformation of the $\delta\vec{x}$ space used above. Let a new $\delta\vec{v}$ space be defined by the transformations,

$$\delta\vec{x} = \mathbf{U}\delta\vec{v} \qquad \text{and} \tag{2.13}$$

$$\delta\vec{v} = \mathbf{T}\delta\vec{x}, \tag{2.14}$$

where $\mathbf{T} = \mathbf{U}^{-1}$. These operators may represent either rotations or scalings (or both). The first and second derivatives with respect to $\delta\vec{v}$ are shown below. Also are the $\delta\vec{x}$ derivatives summarized from above.

$\delta\vec{x}$ derivatives

$$\nabla_{\delta x} J = \left(\frac{\mathrm{d}J}{\mathrm{d}\delta\vec{x}}\right)^T = -\mathbf{B}^{-1}(\delta\vec{x}^B - \delta\vec{x}) - \mathbf{H}^T\mathbf{E}^{-1}(\vec{y} - \vec{H}[\vec{x}_0 + \delta\vec{x}])$$

$$\mathbf{H}e = \left(\frac{\mathrm{d}}{\mathrm{d}\delta\vec{x}}\right)^T \frac{\mathrm{d}J}{\mathrm{d}\delta\vec{x}} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{E}^{-1}\mathbf{H}$$

$\delta\vec{v}$ derivatives

$$\nabla_{\delta v} J = \left(\frac{\mathrm{d}J}{\mathrm{d}\delta\vec{v}}\right)^T = -\mathbf{U}^T\mathbf{B}^{-1}\mathbf{U}(\delta\vec{v}^B - \delta\vec{v}) - \mathbf{U}^T\mathbf{H}^T\mathbf{E}^{-1}(\vec{y} - \vec{H}[\vec{x}_0 + \mathbf{U}\delta\vec{v}])$$

$$\mathbf{H}e = \left(\frac{\mathrm{d}}{\mathrm{d}\delta\vec{v}}\right)^T \frac{\mathrm{d}J}{\mathrm{d}\delta\vec{v}} = \mathbf{U}^T\mathbf{B}^{-1}\mathbf{U} + \mathbf{U}^T\mathbf{H}^T\mathbf{E}^{-1}\mathbf{H}\mathbf{U}$$