Some vector algebra and the generalized chain rule

Ross Bannister
Data Assimilation Research Centre, University of Reading, UK
Last updated 10/06/10

1. Introduction and notation

As we shall see in these notes, the chain rule can be applied to vector as well as scalar derivatives. We will derive the relevant expressions useful in the theory of variational data assimilation and inverse modelling. The results lead us to the concept of adjoint variables and adjoint operators. In section 1 we review the standard notation used in linear algebra.

Vectors and vector derivatives

As is usual notation, scalars and vectors are distinguished from each other by writing vectors in bold. A vector $\boldsymbol{x}$ is, by convention, a column vector (here with n elements)

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}. \tag{1.1}$$

The vector derivative operator operates on some function of $\boldsymbol{x}$ where each element of the vector derivative is the derivative with respect to each element of $\boldsymbol{x}$ as follows

$$\nabla_x = \left(\frac{\partial}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} = \begin{pmatrix} \partial / \partial x_1 \\ \partial / \partial x_2 \\ \dots \\ \partial / \partial x_n \end{pmatrix}. \tag{1.2}$$

$\partial / \partial \boldsymbol{x}$ is a row vector by convention, but the transpose symbol in (1.2), written as a superscript "T" makes rows into columns and vice-versa. The 'nabla' version of the derivative, ∇_x, is usually a column vector by convention.

Matrices and the transpose instruction

A matrix, $\mathbf{A}$, contains element A_{ij} in row i and column j. The transpose is thus

$$A_{ij} = A^{\mathrm{T}}_{ji}. \tag{1.3}$$

Since a vector is a special case of a matrix with either just one row or one column (depending on whether the vector is row or column vector), the transpose instruction here makes row vectors into column vectors, and vice-versa. In particular, for vector derivatives

$$\nabla_x = \left(\frac{\partial}{\partial \boldsymbol{x}}\right)^{\mathrm{T}}. \tag{1.4}$$

The inner product

The combination $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ (an inner product) is a scalar. It is found by the summation ($\boldsymbol{x}$ and $\boldsymbol{y}$ must be vectors of the same number of elements, n)

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \sum_i x_i y_i. \tag{1.5}$$

The outer product

The outer product is written $\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$ and yields a matrix

$$\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} = \begin{pmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_n \\ x_2y_1 & x_2y_2 & \dots & x_2y_n \\ \dots & \dots & \dots & \dots \\ x_ny_1 & x_ny_2 & x_n & x_ny_n \end{pmatrix}. \tag{1.6}$$

The number of elements of $\boldsymbol{x}$ and $\boldsymbol{y}$ need not be the same for the outer product. For $\boldsymbol{x}$ with n elements and $\boldsymbol{y}$ with m elements, the outer product as defined above will be a $n \times m$ matrix.

Matrix operators

A matrix acts on one vector to give another vector. The following action

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x}, \tag{1.7}$$

is valid if the number of rows of $\mathbf{A}$ (n) is the same as the number of elements in $\boldsymbol{y}$ and the number of columns of $\mathbf{A}$ (m) is the same as the number of elements in $\boldsymbol{x}$. Equation (1.7) is shorthand for

$$y_i = \sum_{j=1}^{m} A_{ij}x_j \qquad (i = 1, n). \tag{1.8}$$

This action is like performing many inner products, one for each row of $\mathbf{A}$. In this respect, the matrix operator is sometimes used as a transformation (or change of basis) where each row of $\mathbf{A}$ represents the row vector for a member of the new basis.

Generally, the matrix elements can be thought of as the partial derivatives

$$A_{ij} = \frac{\partial y_i}{\partial x_j}, \tag{1.9}$$

and the whole matrix can be written as

$$\mathbf{A} = \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}. \tag{1.10}$$

The inner and outer products and matrix operators applied with vectors and vector derivatives can be used in innovative ways to write compact multi-variable expressions. Such expressions are used, e.g., in data assimilation.

2. Chain rule for scalar functions (first derivative)

Consider a scalar that is a function of the n elements of $\boldsymbol{x}$, $J(\boldsymbol{x})$. Its derivative with respect to the vector $\boldsymbol{x}$ is the vector

$$\nabla_x J \ = \ \left(\frac{\partial J}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} \ = \ \begin{pmatrix} \partial J / \partial x_1 \\ \partial J / \partial x_2 \\ \cdots \\ \partial J / \partial x_n \end{pmatrix}. \tag{2.1}$$

An important question is: what is $\partial J / \partial \boldsymbol{x}'$ in the case that the two sets of variables $\boldsymbol{x}$ and $\boldsymbol{x}'$ are related via the transformation

$$\boldsymbol{x} \ = \ \mathbf{A}\boldsymbol{x}' \ ? \tag{2.2}$$

$\mathbf{A}$ is sometimes referred to as a *Jacobian*, and has matrix elements $A_{ij} \ = \ \partial x_i / \partial x'_j$ (as Eq. (1.9)). It shall be assumed for now that all elements A_{ij} are real (see below for the modification required when A_{ij} are complex). Let us write an equation for the derivative of J with respect to $\boldsymbol{x}'_i$, expressed explicitly via the chain rule

$$\frac{\partial J}{\partial x'_i} \ = \ \sum_j \frac{\partial x_j}{\partial x'_i}\frac{\partial J}{\partial x_j}, \tag{2.3}$$

$$= \ \sum_j \mathbf{A}_{ji}\frac{\partial J}{\partial x_j}. \tag{2.4}$$

Expressions for derivatives with respect to each component of $\boldsymbol{x}'$ can be assembled into a vector. It can be checked that the following, when expanded using Eqs. (1.3), (1.7) and (1.8), is equivalent to the above

$$\left(\frac{\partial J}{\partial \boldsymbol{x}'}\right)^{\mathrm{T}} \ = \ \mathbf{A}^{\mathrm{T}}\left(\frac{\partial J}{\partial \boldsymbol{x}}\right)^{\mathrm{T}}. \tag{2.5}$$

This is the generalised chain rule for vector derivatives in the case when the operator $\mathbf{A}$ is real. A column derivative with respect to a vector, such as $(\partial J / \partial \boldsymbol{x})^{\mathrm{T}}$, is often called an *adjoint variable*. The operator $\mathbf{A}^{\mathrm{T}}$ (as distinct from the forward operator $\mathbf{A}$, as defined in Eq. (2.2)) is similarly called the *adjoint operator*. It is important to note that the adjoint of an operator is not generally its inverse: While $\mathbf{A}$ transmits information from $\boldsymbol{x}'$ to $\boldsymbol{x}$ (Eq. (2.2)), $\mathbf{A}^{\mathrm{T}}$ transmits information in the reverse direction, <u>but for adjoint variables</u>.

Using Eq. (1.10), Eq. (2.5) can be written as

$$\left(\frac{\partial J}{\partial \boldsymbol{x}'}\right)^{\mathrm{T}} \ = \ \left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}'}\right)^{\mathrm{T}}\left(\frac{\partial J}{\partial \boldsymbol{x}}\right)^{\mathrm{T}}. \tag{2.6}$$

This has the same appearance as the chain rule for single variable functions (now with vectors in the place of scalars) and is a convenient way of remembering the multi-variable result.

Equation 2.5 must be modified when $\mathbf{A}$ is a complex operator. Understanding the modification requires analysis of the chain rule for scalars only. Consider the complex scalar expression

$$x = Zx', \tag{2.7}$$

where $x = x_R + ix_I$, $x' = x'_R + ix'_I$, $Z = Z_R + iZ_I$, and $i = \sqrt{-1}$. The chain rule for this case does not translate to $d/dx' = Z\,d/dx$ in this case. To see what the result should be, expand Eq. (2.7) into its components

$$\begin{aligned} x_R &= Z_R x'_R - Z_I x'_I, \\ x_I &= Z_I x'_R + Z_R x'_I. \end{aligned} \tag{2.8}$$

The minus sign in the first line of Eqs. (2.8) is important. The chain rule has real and imaginary parts as follows

$$\begin{aligned} \frac{\partial}{\partial x'_R} &= \frac{\partial x_R}{\partial x'_R}\frac{\partial}{\partial x_R} + \frac{\partial x_I}{\partial x'_R}\frac{\partial}{\partial x_I}, \\ \frac{\partial}{\partial x'_I} &= \frac{\partial x_R}{\partial x'_I}\frac{\partial}{\partial x_R} + \frac{\partial x_I}{\partial x'_I}\frac{\partial}{\partial x_I}. \end{aligned} \tag{2.9}$$

Four partial derivatives can be found from Eqs. (2.8) to give

$$\begin{aligned} \frac{\partial}{\partial x'_R} &= Z_R\frac{\partial}{\partial x_R} + Z_I\frac{\partial}{\partial x_I}, \\ \frac{\partial}{\partial x'_I} &= -Z_I\frac{\partial}{\partial x_R} + Z_R\frac{\partial}{\partial x_I}. \end{aligned} \tag{2.10}$$

The x and x' adjoint variables for complex derivatives are defined as

$$\begin{aligned} \frac{d}{dx} &= \frac{\partial}{\partial x_R} + i\frac{\partial}{\partial x_I}, \\ \frac{d}{dx'} &= \frac{\partial}{\partial x'_R} + i\frac{\partial}{\partial x'_I}. \end{aligned} \tag{2.11}$$

The following is actually the correct chain rule for the complex case

$$\frac{d}{dx'} = Z^*\frac{d}{dx}, \tag{2.12}$$

where Z^* is the complex conjugate of Z. Substituting Eqs. (2.11) into Eq. (2.12) and then separating into real and imaginary parts gives the same result as Eqs. (2.10). This confirms that Eq. (2.12) is the correct chain rule for the complex case. In the complex vector case, Eq. (2.5) then becomes

$$\left(\frac{\partial J}{\partial \boldsymbol{x}'}\right)^{\mathrm{T}} = \mathbf{A}^{\dagger}\left(\frac{\partial J}{\partial \boldsymbol{x}}\right)^{\mathrm{T}}, \tag{2.13}$$

where the dagger is the conventional shorthand for transpose and complex conjugate.

3. Chain rule for scalar functions (second derivative)

The second derivative with respect to the original variable, $\boldsymbol{x}$, can be written in matrix form as

$$\frac{\partial^2 J}{\partial \boldsymbol{x}^2} = \left(\frac{\partial}{\partial \boldsymbol{x}}\right)^{\mathrm{T}}\left(\frac{\partial}{\partial \boldsymbol{x}}\right)J, \tag{3.1}$$

$$= \begin{pmatrix} \partial^2 J / \partial x_1^2 & \partial^2 J / \partial x_1 \partial x_2 & \dots & \partial^2 J / \partial x_1 \partial x_n \\ \partial^2 J / \partial x_2 \partial x_1 & \partial^2 J / \partial x_2^2 & \dots & \partial^2 J / \partial x_2 \partial x_n \\ \dots & \dots & & \dots \\ \partial^2 J / \partial x_n \partial x_1 & \partial^2 J / \partial x_n \partial x_2 & \dots & \partial^2 J / \partial x_n^2 \end{pmatrix}. \tag{3.2}$$

Although the right hand side of Eq. (3.1) resembles an inner product (scalar), the 'row' property of derivative vectors (mentioned in section 1) means that this is actually an outer product.

Again, imposing the transformation Eq. (2.2), the result, Eq. (2.5), can be used to rewrite the second derivative matrix in terms of the new, primed variables

$$\frac{\partial^2 J}{\partial \boldsymbol{x}'^2} = \left(\frac{\partial}{\partial \boldsymbol{x}'}\right)^{\mathrm{T}} \left(\frac{\partial}{\partial \boldsymbol{x}'}\right) J, \tag{3.3}$$

$$\mathbf{A}^{\mathrm{T}} \left(\frac{\partial}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} \left(\frac{\partial}{\partial \boldsymbol{x}}\right) \mathbf{A} J, \tag{3.4}$$

$$= \mathbf{A}^{\mathrm{T}} \frac{\partial^2 J}{\partial \boldsymbol{x}^2} \mathbf{A}. \tag{3.5}$$

4. Chain Rule for Vector Functions (First Derivative)

If the function itself is a vector, $\boldsymbol{f}(\boldsymbol{x})$, then the derivative is a matrix

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \begin{pmatrix} \partial f_1 / \partial x_1 & \partial f_1 / \partial x_2 & \dots & \partial f_1 / \partial x_n \\ \partial f_2 / \partial x_1 & \partial f_2 / \partial x_2 & \dots & \partial f_2 / \partial x_n \\ \dots & \dots & & \dots \\ \partial f_m / \partial x_1 & \partial f_m / \partial x_2 & \dots & \partial f_m / \partial x_n \end{pmatrix}, \tag{4.1}$$

where the number of components of $\boldsymbol{f}$ (m) is not necessarily the same as the number of components of $\boldsymbol{x}$ (n). Making the same transformation of the independent variable as in section 2, Eq. (2.2), and using the result of Eq. (2.5), allows one to write the derivative in terms of the primed variables as

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}'} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \mathbf{A}. \tag{4.2}$$

All of the results, Eqs. (2.5), (3.5) and (4.2) follow from only one explicit use of the chain rule (in section 2).