A Brief Note on Notation

Conrad Schiff

Sep. 21, 2011

The aim of this note is to address the vector notation used in the *Orbit Determination* class delivered by Russell Carpenter on Sep. 15, 2011 and in the process compare it to the index notation used by the author.

1 Problem Setup

One of the main points of the class on that day was the manipulation of vector- and scalar-valued functions of vector inputs. Specifically, the class was asked to assume a vector x of dimension N and vector-valued function $y = Hx$ where y is of dimension M and H is an $M \times N$ matrix relating the two. Next the class considered the scalar-valued function J defined by

$$J(x) = \frac{1}{2}(y - Hx)^T (y - Hx).$$

Derivatives of both expressions with respect to x were then considered in the process of obtaining the minimum of J.

2 Vector Notation

At the heart of the vector notation is the need to distinguish between row- and column-arrays of the vector. The starting point seems to be that each vector will be regarded as a column array, in keeping with the usual algebra of matrices and in complete agreement with the expression above for J. While producing compact expressions, the user of this notation often produces expressions of the form

$$\frac{\partial q}{\partial x}$$

where some additional consideration must be used. After all, what does it mean to divide a column vector by another column vector. The understanding is that the 'division by a vector' is interpretted by allowing q-components to pick out the rows and the x-components to pick out the columns (row-and-column as read from top to bottom). Assuming that x is dimension 2 and q is dimension

3, an explicit representation would be

$$\frac{\partial q}{\partial x} = \begin{bmatrix} \frac{\partial q_1}{\partial x_1} & \frac{\partial q_1}{\partial x_2} \\ \frac{\partial q_2}{\partial x_1} & \frac{\partial q_2}{\partial x_2} \\ \frac{\partial q_3}{\partial x_1} & \frac{\partial q_3}{\partial x_2} \end{bmatrix}.$$

With this convention, the one would expect that the derivative of J would have a row-form expression as

$$\frac{\partial J}{\partial x} = \begin{bmatrix} \frac{\partial J}{\partial x_1} & \frac{\partial J}{\partial x_2} \end{bmatrix}$$

and it seems that the answer found in the notes

$$\frac{\partial J}{\partial x} = (y - Hx)^T(-H)$$

supports this as $(y - Hx)^T$ is an M-dimensional row vector which yields a N-dimensional row vector when right-multiplied by H.

All seems well, but the mechanics of getting from start to finish seem awkward. To demonstrate my concern, I will mimic, as best as I can, the steps the student should follow.

$$\begin{aligned} \frac{\partial J}{\partial x} &= \frac{\partial}{\partial x}(y - Hx)^T(y - Hx) \\ &= \left[\frac{\partial}{\partial x}(y - Hx)^T\right](y - Hx) + (y - Hx)^T\left[\frac{\partial}{\partial x}(y - Hx)\right] \\ &= \frac{\partial x^T}{\partial x}(-H)(y - Hx) + (y - Hx)^T(-H)\frac{\partial x}{\partial x}. \end{aligned}$$

The second terms in the expression presents no difficulty as long as we apply the rule above for $\partial x/\partial x$ to determine that it is the $N \times N$ identity matrix. But the interpretation of the first term has at least two new notational hurdles. The first is how shall we interpret

$$\frac{\partial x^T}{\partial x}?$$

The second is that the term $H(y - Hx)$ consists of an $M \times N$ matrix left-multiplying a $M \times 1$ column vector, which violates the usual rule of matrix multiplication. At this point, an experienced user of this notation can see his way out of this dilemma but I don't think the new user will feel so comfortable. Of course, a set of new rules can be made to rigorously beat this problem down but what results is so many rules that I think it begins to look like a bunch of arbitrary notational conventions with little or no justification.

3 Index Notation

While harder to learn the index notation route requires fewer rules and always gets the right answer. I won't go into a great deal of exposition but will simply show my computation done during the class. Start by defining $q_i = y_i - H_{ik}x_k$

$$
\begin{aligned}
\frac{\partial J}{\partial x} &\equiv \frac{\partial J}{\partial x_j} \\
&= \frac{1}{2}\frac{\partial}{\partial x_j} q_i q_i \\
&= \frac{1}{2}\frac{\partial q_i}{\partial x_j} q_i + \frac{1}{2} q_i \frac{\partial q_i}{\partial x_j} \\
&= q_i \frac{\partial q_i}{\partial x_j} \\
&= q_i \frac{\partial}{\partial x_j}(y_i - H_{ik}x_k) \\
&= q_i(-H_{ik})\frac{\partial x_k}{\partial x_j} \\
&= q_i(-H_{ik})\delta_{kj} \\
&= q_i(-H_{ij}) \\
&\equiv (y - Hx)^T(-H).
\end{aligned}
$$

The above computation was actually done with more steps than are usual but the only rules required for unambiguous interpretation are the summation convention $q_i q_i = q_1^2 + q_2^2 + ...$ (something like the .* operator in Matlab), and the definition of the Kronecker delta $H_{ij}\delta_{jk} = H_{ik}$ (which is essentially HI).