CMPUT 340: Numerical Methods
Fun with Vector Derivatives

Gautham Vasan
vasan@ualberta.ca

January 22rd, 2024

1 Notation

X	Random variables (scalar)
x	Values of random variables or scalar functions
$\mathbf{x}$	Quantities required to be real-valued coloumn vectors (even random variables)
$\mathbf{X}$	Matrices
$\mathbf{X}^T, \mathbf{x}^T$	The transpose of a matrix or vector
$[\mathbf{x}]_i$	The i-th element of the vector.
$[\mathbf{X}]_{i:}$	The i-th row of the matrix. A row vector
$[\mathbf{X}]_{:j}$	The j-th column of the matrix. A column vector
$[\mathbf{X}]_{ij}$	The i,j-th element of the matrix.
$\mathbf{x}_t$	The vector at time t
X_t	The random variable at time t

2 Vector Multiplication

Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix}$ be a column vector, and $\mathbf{v}^\top = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ be a row vector.

The outer product of $\mathbf{u}$ and $\mathbf{v}$ is given by:

$$\mathbf{u}\mathbf{v}^\top = \begin{bmatrix} u_1v_1 & u_1v_2 & \cdots & u_1v_n \\ u_2v_1 & u_2v_2 & \cdots & u_2v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_mv_1 & u_mv_2 & \cdots & u_mv_n \end{bmatrix} \tag{1}$$

3 Matrix Multiplication

Let's consider two matrices A and B with dimensions $m \times n$ and $n \times p$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}$$

The matrix multiplication $C = A \cdot B$ is given by:

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{bmatrix}$$

where each element c_{ij} is computed as:

$$c_{ij} = a_{i1} \cdot b_{1j} + a_{i2} \cdot b_{2j} + \cdots + a_{in} \cdot b_{nj} \tag{2}$$

So, for example, the element at position (i, j) in the result matrix C is c_{ij}.

4 Gradient of a Scalar With Respect to a Vector

Let $f(\mathbf{x})$ be a scalar function where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

The gradient of $f(\mathbf{x})$ with respect to $\mathbf{x}$, denoted as $\nabla f(\mathbf{x})$, is given by:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{3}$$

Each element of the gradient vector is the partial derivative of f with respect to the corresponding variable.

$$[\nabla f(\mathbf{x})]_i = \frac{\partial f}{\partial x_i}$$

5 The Jacobian Matrix

Let $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$ be a vector-valued function where $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$.

The *Jacobian* matrix $J(\mathbf{f}, \mathbf{x})$ is given by:

$$J(\mathbf{f}, \mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{4}$$

The element in the i-th row and j-th column of the Jacobian matrix is:

$$[J(\mathbf{f}, \mathbf{x})]_{i,j} = \frac{\partial f_i}{\partial x_j}$$

Note that the Jacobian is a $m \times n$ matrix.

6 Jacobian Example I

The given vector derivative rule is a product rule for the gradient of a scalar function $f(\mathbf{x})$ times a vector function $\mathbf{g}(\mathbf{x})$. The rule is stated as:

$$\nabla_{\mathbf{x}}(f(\mathbf{x})\mathbf{g}(\mathbf{x})) = f(\mathbf{x})\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\left(\nabla_{\mathbf{x}} f(\mathbf{x})\right)^\top$$

6.1 Proof

Let $f(\mathbf{x})$ be a scalar function and $\mathbf{g}(\mathbf{x})$ be a vector function, where $\mathbf{x}$ is a vector of variables. We assume that $\mathbf{x}$ is of dimension $n \times 1$ and $\mathbf{g}(\mathbf{x})$ is of dimension $m \times 1$. Now, let's identify the dimensions of each components in the R.H.S:

$$\begin{aligned} \mathbf{x} &: n \times 1 \\ f(\mathbf{x}) &: 1 \times 1 \\ \nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}) &: m \times n \\ \mathbf{g}(\mathbf{x}) &: m \times 1 \\ \left(\nabla_{\mathbf{x}} f(\mathbf{x})\right)^\top &: 1 \times n \end{aligned}$$

We can arrive at the proof step by step using the definition of the gradient and the product rule.

The gradient of $f(\mathbf{x})$ is given by Eq. 3:

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Now, let's consider the product $f(\mathbf{x})\mathbf{g}(\mathbf{x})$:

$$f(\mathbf{x})\mathbf{g}(\mathbf{x}) = \begin{bmatrix} f(\mathbf{x})g_1(\mathbf{x}) \\ f(\mathbf{x})g_2(\mathbf{x}) \\ \vdots \\ f(\mathbf{x})g_m(\mathbf{x}) \end{bmatrix}$$

where $g_i(\mathbf{x})$ represents the i-th component of the vector function $\mathbf{g}(\mathbf{x})$.
Now, let's compute the gradient of this product:

$$\nabla_{\mathbf{x}}(f(\mathbf{x})\mathbf{g}(\mathbf{x})) = \begin{bmatrix} \frac{\partial f(\mathbf{x})g_1(\mathbf{x})}{\partial x_1} & \frac{\partial f(\mathbf{x})g_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f(\mathbf{x})g_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial f(\mathbf{x})g_2(\mathbf{x})}{\partial x_1} & \frac{\partial f(\mathbf{x})g_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f(\mathbf{x})g_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{x})g_m(\mathbf{x})}{\partial x_1} & \frac{\partial f(\mathbf{x})g_m(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f(\mathbf{x})g_m(\mathbf{x})}{\partial x_n} \end{bmatrix} \quad \text{Using Jacobian definition in Eq. 4}$$

Now, apply the product rule to each component in this Jacobian:

$$[\nabla_{\mathbf{x}}(f(\mathbf{x})\mathbf{g}(\mathbf{x}))]_{i,j} = \frac{\partial}{\partial x_j}(f(\mathbf{x})g_i(\mathbf{x})) = f(\mathbf{x})\frac{\partial g_i}{\partial x_j} + \frac{\partial f}{\partial x_j} g_i(\mathbf{x}) \quad \text{(Product Rule)}$$

Now that we have a general rule for each element in the resultant matrix, let's derive a unified form for a single row:

$$[\nabla_{\mathbf{x}}(f(\mathbf{x})\mathbf{g}(\mathbf{x}))]_{i,:} = f(\mathbf{x})(\nabla_{\mathbf{x}} g_i(\mathbf{x}))^\top + g_i(\mathbf{x})\left(\nabla_{\mathbf{x}} f(\mathbf{x})\right)^\top$$

Please note that the transpose operation is applied in the above context because, in our notations, the derivative of a scalar with respect to a vector results in a column vector. However, the expression we are dealing with is seeking a row vector representation.

Finally, using Eq. 1 and 4, we can arrive at the general rule:

$$\nabla_{\mathbf{x}}(f(\mathbf{x})\mathbf{g}(\mathbf{x})) = f(\mathbf{x})\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\left(\nabla_{\mathbf{x}} f(\mathbf{x})\right)^\top$$

7 Jacobian Example II

Show that

$$\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{h}(\mathbf{x})) = \nabla_{\mathbf{r}}\mathbf{g}(\mathbf{r})\nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x})\Big|_{\mathbf{r}=\mathbf{h}(\mathbf{x})}.$$

Note that in the above $\mathbf{x}$ is $n \times 1$, $\mathbf{h}$ is $p \times 1$ and $\mathbf{g}$ is $m \times 1$. Then $\nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x})$ is $p \times n$ and $\nabla_{\mathbf{r}}\mathbf{g}(\mathbf{r})$ is $m \times p$.

7.1 Proof

To prove the given expression for the element-wise gradients, we can use the chain rule of calculus. Let's denote the elements of vectors $\mathbf{x}$, $\mathbf{h}(\mathbf{x})$, and $\mathbf{j}$ as x_i, $h_i(\mathbf{x})$, and j_i respectively.

$$\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{h}(\mathbf{x})) = \begin{bmatrix} \frac{\partial g_1(\mathbf{h}(\mathbf{x}))}{\partial x_1} & \frac{\partial g_1(\mathbf{h}(\mathbf{x}))}{\partial x_2} & \cdots & \frac{\partial g_1(\mathbf{h}(\mathbf{x}))}{\partial x_n} \\ \frac{\partial g_2(\mathbf{h}(\mathbf{x}))}{\partial x_1} & \frac{\partial g_2(\mathbf{h}(\mathbf{x}))}{\partial x_2} & \cdots & \frac{\partial g_2(\mathbf{h}(\mathbf{x}))}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m(\mathbf{h}(\mathbf{x}))}{\partial x_1} & \frac{\partial g_m(\mathbf{h}(\mathbf{x}))}{\partial x_2} & \cdots & \frac{\partial g_m(\mathbf{h}(\mathbf{x}))}{\partial x_n} \end{bmatrix} \quad \text{Using Jacobian definition in Eq. 4}$$

Now, consider the i, j-th element of the expression:

$$[\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{h}(\mathbf{x}))]_{i,j} = \frac{\partial g_i(\mathbf{h}(\mathbf{x}))}{\partial x_j} = \sum_{k=1}^{p} \frac{\partial g_i(\mathbf{r})}{\partial r_k}\frac{\partial h(\mathbf{x})_k}{\partial x_j}\Bigg|_{\mathbf{r}=\mathbf{h}(\mathbf{x})} \quad \text{Chain rule}$$

Remember that the k^{th} component of $\mathbf{h}(\mathbf{x})$ depends on x_j. Hence, we need to account for the partial derivative from every element of $\mathbf{h}(\mathbf{x})$

$$\nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x}) = J(\mathbf{h}, \mathbf{x}) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \cdots & \frac{\partial h_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_p}{\partial x_1} & \frac{\partial h_p}{\partial x_2} & \cdots & \frac{\partial h_p}{\partial x_n} \end{bmatrix} \tag{5}$$

$$\nabla_{\mathbf{r}}\mathbf{g}(\mathbf{r})\Bigg|_{\mathbf{r}=\mathbf{h}(\mathbf{x})} = J(\mathbf{g}, \mathbf{r}) = \begin{bmatrix} \frac{\partial g_1}{\partial r_1} & \frac{\partial g_1}{\partial r_2} & \cdots & \frac{\partial g_1}{\partial r_p} \\ \frac{\partial g_2}{\partial r_1} & \frac{\partial g_2}{\partial r_2} & \cdots & \frac{\partial g_2}{\partial r_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial r_1} & \frac{\partial g_m}{\partial r_2} & \cdots & \frac{\partial g_m}{\partial r_p} \end{bmatrix} \tag{6}$$

Using Eq. 5 and 6, we get:

$$\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{h}(\mathbf{x})) = \nabla_{\mathbf{r}}\mathbf{g}(\mathbf{r})\nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x})\Big|_{\mathbf{r}=\mathbf{h}(\mathbf{x})}.$$

Verify that the matrix multiplication of 5 and 6 can give the expression of each individual element we showed earlier using the chain rule.

8 References

- Matrices eClass Wiki
- Derviatives eClass Wiki
- Matrix cookbook
- Jacobian Wiki
- Chain rule from Khan Academy