Vector Calculus

Last modified: 2026-09-28 23:45:45 (PDT)

(adapted from Fieller (2016), Section 7.2)

This section covers derivatives of functions of vectors and matrices. Linear algebra prerequisites — including vectors, matrices, transpose, dot product, and quadratic forms — are covered in Linear Algebra.

Let $\tilde{x}$ and $\tilde{\beta}$ be column vectors of length p (see column vector and dot product).

Definition 0.1 (Vector derivative)**.** If $f(\tilde{\beta})$ is a scalar-valued function of a $p \times 1$ vector $\tilde{\beta}$, such as $f(\tilde{\beta}) = \tilde{x}^\top \tilde{\beta}$, then its **vector derivative** is:

$$\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta}) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} f(\tilde{\beta}) \\ \frac{\partial}{\partial \beta_2} f(\tilde{\beta}) \\ \vdots \\ \frac{\partial}{\partial \beta_p} f(\tilde{\beta}) \end{bmatrix}$$

Video lecture

Hutchinson's Gradients Refresher (17 min) covers gradients and partial derivatives, the ideas behind this section (Hutchinson, n.d.). The login for the video site is posted on Canvas.

Definition 0.2 (Row-vector derivative)**.** If $f(\tilde{\beta})$ is a scalar-valued function of a $p \times 1$ vector $\tilde{\beta}$, such as $f(\tilde{\beta}) = \tilde{x}^\top \tilde{\beta}$, then its **row-vector derivative** is:

$$\frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta}) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} f(\tilde{\beta}) & \frac{\partial}{\partial \beta_2} f(\tilde{\beta}) & \cdots & \frac{\partial}{\partial \beta_p} f(\tilde{\beta}) \end{bmatrix}$$

Theorem 0.1 (Row and column derivatives are transposes)**.**

$$\frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta}) = \left(\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta})\right)^\top$$

$$\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta}) = \left(\frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta})\right)^\top$$

Proof

Proof. By Definition 0.1 and Definition 0.2, entry j of both $\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta})$ and $\frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta})$ is $\frac{\partial}{\partial \beta_j} f(\tilde{\beta})$; the first is a $p \times 1$ column and the second a $1 \times p$ row with the same entries in the same order, so each is the transpose of the other. □

Example 0.1 (Row and column derivatives of a linear function)**.** For $f(\tilde{\beta}) = 3\beta_1 + 5\beta_2$:

$$\frac{\partial}{\partial \tilde{\beta}} f(\tilde{\beta}) = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \qquad \frac{\partial}{\partial \tilde{\beta}^\top} f(\tilde{\beta}) = \begin{bmatrix} 3 & 5 \end{bmatrix},$$

and each is the transpose of the other.

Definition 0.3 (Derivative of a vector-valued function)**.** If $\tilde{y} = \tilde{y}(\tilde{\beta}) = (y_1, \dots, y_q)^\top$ is a $q \times 1$ vector-valued function of the $p \times 1$ vector $\tilde{\beta}$, its **derivative with respect to $\tilde{\beta}$** is the $p \times q$ matrix whose (i, j) entry is

$$\left[\frac{\partial}{\partial \tilde{\beta}} \tilde{y}^\top\right]_{ij} \overset{\text{def}}{=} \frac{\partial}{\partial \beta_i} y_j, \qquad i = 1, \dots, p, \quad j = 1, \dots, q.$$

These notes use this *denominator layout* throughout: rows index the entries of $\tilde{\beta}$ (the denominator) and columns index the entries of $\tilde{y}$ (the numerator), so column j is the vector derivative $\frac{\partial}{\partial \tilde{\beta}} y_j$ (Definition 0.1). Both $\frac{\partial}{\partial \tilde{\beta}} \tilde{y}$ and $\frac{\partial}{\partial \tilde{\beta}} \tilde{y}^\top$ denote this $p \times q$ matrix.

Some sources use the transposed, *numerator layout*, in which the derivative is the $q \times p$ Jacobian matrix with (j, i) entry $\frac{\partial}{\partial \beta_i} y_j$. Check a source's layout before combining its formulas with these.

Example 0.2 (Differentiating a 3×1 function of a 2×1 vector)**.** Let $\tilde{\beta} = (\beta_1, \beta_2)^\top$ $(p = 2)$ and $\tilde{y}(\tilde{\beta}) = (\beta_1^2,\ \beta_1\beta_2,\ 3\beta_2)^\top$ $(q = 3)$. Then

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}} \tilde{y}^\top}_{2\times 3} = \begin{bmatrix} \frac{\partial}{\partial \beta_1}\beta_1^2 & \frac{\partial}{\partial \beta_1}\beta_1\beta_2 & \frac{\partial}{\partial \beta_1}3\beta_2 \\ \frac{\partial}{\partial \beta_2}\beta_1^2 & \frac{\partial}{\partial \beta_2}\beta_1\beta_2 & \frac{\partial}{\partial \beta_2}3\beta_2 \end{bmatrix} = \begin{bmatrix} 2\beta_1 & \beta_2 & 0 \\ 0 & \beta_1 & 3 \end{bmatrix}$$

Definition 0.4 (Constant)**.** A $q \times 1$ vector $\tilde{x}$ is **constant with respect to** the $p \times 1$ vector $\tilde{\beta}$ if its derivative (Definition 0.3) is zero:

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}} \tilde{x}^\top}_{p\times q} = \underset{p\times q}{\underset{\smile}{\mathbf{0}}}$$

Example 0.3 (A constant vector)**.** Let $\tilde{\beta} = (\beta_1, \beta_2)^\top$ and $\tilde{x} = (3, 5)^\top$, so $x_1 = 3$ and $x_2 = 5$ do not depend on $\tilde{\beta}$. Expanding $\frac{\partial}{\partial \tilde{\beta}} \tilde{x}^\top$ into its matrix of scalar partial derivatives (Definition 0.3) and evaluating each entry:

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}} \tilde{x}^\top}_{2\times 2} = \frac{\partial}{\partial \tilde{\beta}} \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \beta_1}x_1 & \frac{\partial}{\partial \beta_1}x_2 \\ \frac{\partial}{\partial \beta_2}x_1 & \frac{\partial}{\partial \beta_2}x_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \beta_1}3 & \frac{\partial}{\partial \beta_1}5 \\ \frac{\partial}{\partial \beta_2}3 & \frac{\partial}{\partial \beta_2}5 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \underset{2\times 2}{\underset{\smile}{\mathbf{0}}}$$

Every entry is the derivative of a constant, so $\frac{\partial}{\partial \tilde{\beta}} \tilde{x}^\top = \underset{2\times 2}{\underset{\smile}{\mathbf{0}}}$ and $\tilde{x}$ is constant with respect to $\tilde{\beta}$ (Definition 0.4).

Theorem 0.2 (Derivative of a dot product)**.** *If $\tilde{x}$ is constant with respect to $\tilde{\beta}$, then:*

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}} (\tilde{x} \cdot \tilde{\beta})}_{p\times 1} = \underbrace{\frac{\partial}{\partial \tilde{\beta}} (\tilde{\beta} \cdot \tilde{x})}_{p\times 1} = \underset{p\times 1}{\underset{\smile}{\tilde{x}}}$$

Proof

Proof.

$$
\begin{aligned}
\frac{\partial}{\partial \tilde{\beta}}(\tilde{x} \cdot \tilde{\beta}) &= \begin{bmatrix} \frac{\partial}{\partial \beta_1}(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p) \\ \frac{\partial}{\partial \beta_2}(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p) \\ \vdots \\ \frac{\partial}{\partial \beta_p}(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p) \end{bmatrix} \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \\
&= \tilde{x}
\end{aligned}
$$

□

Example 0.4 (Derivative of a dot product)**.** Let $\tilde{x} = (3, 5)^\top$ (constant with respect to $\tilde{\beta}$; see Example 0.3) and $\tilde{\beta} = (\beta_1, \beta_2)^\top$. Then $\tilde{x} \cdot \tilde{\beta} = 3\beta_1 + 5\beta_2$, and by Theorem 0.2:

$$
\underbrace{\frac{\partial}{\partial \tilde{\beta}}(\tilde{x} \cdot \tilde{\beta})}_{2\times 1} = \underbrace{\tilde{x}}_{2\times 1} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}
$$

Verifying entry-wise:

$$
\frac{\partial}{\partial \tilde{\beta}}(3\beta_1 + 5\beta_2) = \begin{pmatrix} \frac{\partial}{\partial \beta_1}(3\beta_1 + 5\beta_2) \\ \frac{\partial}{\partial \beta_2}(3\beta_1 + 5\beta_2) \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}
$$

Both methods agree.

Theorem 0.3 (Product rule for dot-products)**.** *If $\tilde{a} = \tilde{a}(\tilde{x})$ and $\tilde{b} = \tilde{b}(\tilde{x})$ are differentiable $p \times 1$ vector functions of $\tilde{x}$, then:*

$$
\frac{\partial}{\partial \underbrace{\tilde{x}}_{p\times 1}} \underbrace{\tilde{a}}_{p\times 1} \cdot \underbrace{\tilde{b}}_{p\times 1} = \left(\frac{\partial}{\partial \underbrace{\tilde{x}}_{p\times 1}} \underbrace{\tilde{a}^\top}_{1\times p} \right) \underbrace{\tilde{b}}_{p\times 1} + \left(\frac{\partial}{\partial \underbrace{\tilde{x}}_{p\times 1}} \underbrace{\tilde{b}^\top}_{1\times p} \right) \underbrace{\tilde{a}}_{p\times 1}
$$

Proof

Proof. Entry-wise, for $i = 1, \dots, p$:

$$
\begin{aligned}
\left[\frac{\partial}{\partial \tilde{x}}(\tilde{a}\cdot\tilde{b})\right]_i &= \frac{\partial}{\partial x_i}\sum_{k=1}^{p} a_k b_k \\
&= \sum_{k=1}^{p}\left(b_k \frac{\partial}{\partial x_i}a_k + a_k \frac{\partial}{\partial x_i}b_k\right) \\
&= \left[\left(\frac{\partial}{\partial \tilde{x}}\tilde{a}^\top\right)\tilde{b}\right]_i + \left[\left(\frac{\partial}{\partial \tilde{x}}\tilde{b}^\top\right)\tilde{a}\right]_i
\end{aligned}
$$

□

Example 0.5 (Example of the dot-product rule)**.** Apply Theorem 0.3 with the vector $\tilde{\beta} = (\beta_1, \beta_2)^\top$ in the role of $\tilde{x}$. Let $\tilde{a}(\tilde{\beta}) = (\beta_1, \beta_1\beta_2)^\top$ and $\tilde{b}(\tilde{\beta}) = (\beta_2, \beta_1)^\top$. Then:

$$\tilde{a}\cdot\tilde{b} = \beta_1\cdot\beta_2 + \beta_1\beta_2\cdot\beta_1 = \beta_1\beta_2 + \beta_1^2\beta_2$$

By direct calculation:

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}}(\tilde{a}\cdot\tilde{b})}_{2\times 1} = \frac{\partial}{\partial \tilde{\beta}}(\beta_1\beta_2 + \beta_1^2\beta_2) = \begin{pmatrix}\beta_2 + 2\beta_1\beta_2 \\ \beta_1 + \beta_1^2\end{pmatrix}$$

By the product rule (Theorem 0.3), using $\underbrace{\frac{\partial}{\partial \tilde{\beta}}\tilde{a}^\top}_{2\times 2} = \begin{pmatrix}1 & \beta_2 \\ 0 & \beta_1\end{pmatrix}$ and $\underbrace{\frac{\partial}{\partial \tilde{\beta}}\tilde{b}^\top}_{2\times 2} = \begin{pmatrix}0 & 1 \\ 1 & 0\end{pmatrix}$:

$$
\begin{aligned}
\underbrace{\frac{\partial}{\partial \tilde{\beta}}(\tilde{a}\cdot\tilde{b})}_{2\times 1} &= \underbrace{\begin{pmatrix}1 & \beta_2 \\ 0 & \beta_1\end{pmatrix}}_{2\times 2}\underbrace{\begin{pmatrix}\beta_2 \\ \beta_1\end{pmatrix}}_{2\times 1} + \underbrace{\begin{pmatrix}0 & 1 \\ 1 & 0\end{pmatrix}}_{2\times 2}\underbrace{\begin{pmatrix}\beta_1 \\ \beta_1\beta_2\end{pmatrix}}_{2\times 1} \\
&= \begin{pmatrix}\beta_2 + \beta_1\beta_2 \\ \beta_1^2\end{pmatrix} + \begin{pmatrix}\beta_1\beta_2 \\ \beta_1\end{pmatrix} \\
&= \begin{pmatrix}\beta_2 + 2\beta_1\beta_2 \\ \beta_1^2 + \beta_1\end{pmatrix}
\end{aligned}
$$

Both methods agree.

Theorem 0.4 (Derivative of a linear map)**.** *If* $\mathbf{A}$ *is an* $m \times p$ *matrix that is constant with respect to* $\tilde{\beta}$*, then:*

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}}(\mathbf{A}\tilde{\beta})}_{p\times m} = \underbrace{\mathbf{A}^\top}_{p\times m}$$

Proof

Proof. For entry (i, j), where row i indexes the denominator $\tilde{\beta}$ (see Definition 0.3) and column j indexes the numerator $\mathbf{A}\tilde{\beta}$:

$$\begin{aligned}\left[\frac{\partial}{\partial \tilde{\beta}}(\mathbf{A}\tilde{\beta})\right]_{ij} &= \frac{\partial}{\partial \beta_i}(\mathbf{A}\tilde{\beta})_j \\ &= \frac{\partial}{\partial \beta_i}\sum_{k=1}^{p} a_{jk}\beta_k \\ &= a_{ji} \\ &= [\mathbf{A}^\top]_{ij}\end{aligned}$$

□

Example 0.6 (Derivative of a linear map)**.** Let $\mathbf{A} = \begin{pmatrix} 2 & 3 \end{pmatrix}$ (1×2) and $\tilde{\beta} = (\beta_1, \beta_2)^\top$. Then $\mathbf{A}\tilde{\beta} = 2\beta_1 + 3\beta_2$, and by Theorem 0.4:

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}}(\mathbf{A}\tilde{\beta})}_{2\times 1} = \underbrace{\mathbf{A}^\top}_{2\times 1} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Theorem 0.5 (Vector-derivative of a matrix-vector product)**.** *If* $\mathbf{A}$ *is an* $m \times q$ *matrix that is constant with respect to* $\tilde{\beta}$*, and* $\tilde{v} = \tilde{v}(\tilde{\beta})$ *is a* $q \times 1$ *vector that depends on the* $p \times 1$ *vector* $\tilde{\beta}$*, then:*

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}}(\mathbf{A}\tilde{v})}_{p\times m} = \underbrace{\left(\frac{\partial}{\partial \tilde{\beta}}\tilde{v}\right)}_{p\times q}\underbrace{\mathbf{A}^\top}_{q\times m}$$

This result generalizes Theorem 0.4, which is the special case $\tilde{v} = \tilde{\beta}$ *(so that* $\frac{\partial}{\partial \tilde{\beta}}\tilde{\beta} = \mathbf{I}$ *and* $\frac{\partial}{\partial \tilde{\beta}}(\mathbf{A}\tilde{\beta}) = \mathbf{A}^\top$*).*

Proof

Proof. For entry (i, j), where row i indexes the denominator $\tilde{\beta}$ and column j indexes the numerator $\mathbf{A}\tilde{v}$ (see Definition 0.3):

$$
\begin{aligned}
\left[\frac{\partial}{\partial\tilde{\beta}}(\mathbf{A}\tilde{v})\right]_{ij} &= \frac{\partial}{\partial\beta_i}(\mathbf{A}\tilde{v})_j \\
&= \frac{\partial}{\partial\beta_i}\sum_{k=1}^{q} a_{jk}v_k \\
&= \sum_{k=1}^{q} a_{jk}\frac{\partial}{\partial\beta_i}v_k \\
&= \sum_{k=1}^{q}\left[\frac{\partial}{\partial\tilde{\beta}}\tilde{v}\right]_{ik}\left[\mathbf{A}^\top\right]_{kj} \\
&= \left[\left(\frac{\partial}{\partial\tilde{\beta}}\tilde{v}\right)\mathbf{A}^\top\right]_{ij}
\end{aligned}
$$

□

Example 0.7 (Vector-derivative of a matrix-vector product)**.** Let $\mathbf{A} = \begin{pmatrix}2 & 3\end{pmatrix}$ (1×2, constant) and $\tilde{v}(\tilde{\beta}) = (\beta_1^2, \beta_2^2)^\top$. Then $\mathbf{A}\tilde{v} = 2\beta_1^2 + 3\beta_2^2$. By Theorem 0.5:

$$
\begin{aligned}
\underbrace{\frac{\partial}{\partial\tilde{\beta}}(\mathbf{A}\tilde{v})}_{2\times 1} &= \begin{pmatrix}2\beta_1 & 0 \\ 0 & 2\beta_2\end{pmatrix}\begin{pmatrix}2 \\ 3\end{pmatrix} \\
&= \begin{pmatrix}4\beta_1 \\ 6\beta_2\end{pmatrix}
\end{aligned}
$$

Theorem 0.6 (Vector-derivative of a product of matrices)**.** *If* $\mathbf{A}$ *($\ell \times m$) and* $\mathbf{B}$ *($m \times q$) are constant with respect to* $\tilde{\beta}$*, and* $\tilde{v} = \tilde{v}(\tilde{\beta})$ *is a* $q \times 1$ *vector that depends on the* $p \times 1$ *vector* $\tilde{\beta}$*, then:*

$$
\underbrace{\frac{\partial}{\partial\tilde{\beta}}(\mathbf{A}\mathbf{B}\tilde{v})}_{p\times\ell} = \underbrace{\left(\frac{\partial}{\partial\tilde{\beta}}\tilde{v}\right)}_{p\times q}\underbrace{\mathbf{B}^\top}_{q\times m}\underbrace{\mathbf{A}^\top}_{m\times\ell}
$$

Proof

Proof. Apply Theorem 0.5 with the constant $\ell \times q$ matrix $\mathbf{AB}$, then use the transpose of a product:

$$
\begin{aligned}
\frac{\partial}{\partial \tilde{\beta}}(\mathbf{A}\mathbf{B}\tilde{v}) &= \left(\frac{\partial}{\partial \tilde{\beta}}\tilde{v}\right)(\mathbf{A}\mathbf{B})^\top \quad \text{(derivative of a matrix-vector product)} \\
&= \left(\frac{\partial}{\partial \tilde{\beta}}\tilde{v}\right)\mathbf{B}^\top\mathbf{A}^\top \quad \text{(transpose of a product)}
\end{aligned}
$$

□

Example 0.8. Let $\mathbf{A} = \begin{pmatrix} 1 & 0 \end{pmatrix}$ (1×2), $\mathbf{B} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ (2×2), and $\tilde{v}(\tilde{\beta}) = \tilde{\beta}$ where $\tilde{\beta} = (\beta_1, \beta_2)^\top$. Then $\mathbf{A}\mathbf{B}\tilde{v} = 2\beta_1$, and:

$$
\underbrace{\frac{\partial}{\partial \tilde{\beta}}(\mathbf{A}\mathbf{B}\tilde{v})}_{2\times 1} = \underbrace{\left(\frac{\partial}{\partial \tilde{\beta}}\tilde{\beta}\right)}_{2\times 2}\underbrace{\mathbf{B}^\top}_{2\times 2}\underbrace{\mathbf{A}^\top}_{2\times 1} = \mathbf{I}_2 \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}
$$

Corollary 0.1 (Derivative of a dot product, transpose-product form)**.** *If $\tilde{x}$ is constant with respect to $\tilde{\beta}$, then:*

$$
\underbrace{\frac{\partial}{\partial \tilde{\beta}}(\underbrace{\tilde{x}^\top}_{1\times p}\underbrace{\tilde{\beta}}_{p\times 1})}_{p\times 1} = \underbrace{\tilde{x}}_{p\times 1}
$$

This vector derivative formula looks a lot like non-vector calculus, except that you have to transpose the coefficient: in scalar calculus $\frac{\partial}{\partial x}(cx) = c$, but here the coefficient $\tilde{x}^\top$ (a row vector) becomes $\tilde{x}$ (a column vector) in the result.

ℹ Proof

Proof. **Using Theorem 0.2:**
Since $\tilde{x}^\top\tilde{\beta} = \tilde{x} \cdot \tilde{\beta}$ (see dot product), and $\tilde{x}$ is constant with respect to $\tilde{\beta}$:

$$
\frac{\partial}{\partial \tilde{\beta}}(\tilde{x}^\top\tilde{\beta}) = \frac{\partial}{\partial \tilde{\beta}}(\tilde{x} \cdot \tilde{\beta}) = \tilde{x}
$$

by Theorem 0.2. □

ℹ Proof

Proof. **Using Theorem 0.5:**
Since $\tilde{x}$ is constant with respect to $\tilde{\beta}$, $\mathbf{A} = \tilde{x}^\top$ is a constant $1 \times p$ matrix. Applying

Theorem 0.5 with $\tilde{v} = \tilde{\beta}$ (so $\frac{\partial}{\partial \tilde{\beta}} \tilde{\beta} = \mathbf{I}$):

$$\begin{aligned}
\frac{\partial}{\partial \tilde{\beta}} (\tilde{x}^\top \tilde{\beta}) &= \left(\frac{\partial}{\partial \tilde{\beta}} \tilde{\beta} \right) (\tilde{x}^\top)^\top \\
&= \mathbf{I} \cdot \tilde{x} \\
&= \tilde{x}
\end{aligned}$$

□

Example 0.9 (Derivative of a transpose product)**.** Let $\tilde{x} = (3, 5)^\top$ and $\tilde{\beta} = (\beta_1, \beta_2)^\top$. Then $\tilde{x}^\top \tilde{\beta} = 3\beta_1 + 5\beta_2$, and by Corollary 0.1:

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}} \left(\underbrace{\tilde{x}^\top}_{1\times 2} \underbrace{\tilde{\beta}}_{2\times 1} \right)}_{2\times 1} = \underbrace{\tilde{x}}_{2\times 1} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$$

Theorem 0.7 (Derivative of a quadratic form)**.** *For a quadratic form (see quadratic form), if* $\mathbf{S}$ *is a symmetric* $p \times p$ *matrix that is constant with respect to* $\tilde{\beta}$*, then:*

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}} (\tilde{\beta}^\top \mathbf{S} \tilde{\beta})}_{p\times 1} = 2\underbrace{\mathbf{S}\tilde{\beta}}_{p\times 1}$$

Proof

Proof. Expanding entry-wise, $\tilde{\beta}^\top \mathbf{S} \tilde{\beta} = \sum_{j=1}^p \sum_{k=1}^p s_{jk} \beta_j \beta_k$. Differentiating component-wise with respect to β_i for $i = 1, \dots, p$:

$$\begin{aligned}
\left[\frac{\partial}{\partial \tilde{\beta}} (\tilde{\beta}^\top \mathbf{S} \tilde{\beta}) \right]_i &= \frac{\partial}{\partial \beta_i} \sum_{j=1}^p \sum_{k=1}^p s_{jk} \beta_j \beta_k && \text{(expand quadratic form)} \\
&= \sum_{k=1}^p s_{ik} \beta_k + \sum_{j=1}^p s_{ji} \beta_j && \text{(product rule for } \beta_i \beta_k\text{)} \\
&= [\mathbf{S}\tilde{\beta}]_i + [\mathbf{S}^\top \tilde{\beta}]_i && \text{(matrix-vector multiplication definition)} \\
&= [(\mathbf{S} + \mathbf{S}^\top)\tilde{\beta}]_i && \text{(linearity of matrix multiplication)}
\end{aligned}$$

When $\mathbf{S}$ is symmetric ($\mathbf{S} = \mathbf{S}^\top$), $\mathbf{S} + \mathbf{S}^\top = 2\mathbf{S}$, so $\frac{\partial}{\partial \tilde{\beta}} (\tilde{\beta}^\top \mathbf{S} \tilde{\beta}) = 2\mathbf{S}\tilde{\beta}$. □

This operation is like taking the derivative of cx^2 with respect to x in non-vector calculus.

Example 0.10 (Derivative of a quadratic form)**.** Let $\mathbf{S} = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$ (2×2, symmetric and constant) and $\tilde{\beta} = (\beta_1, \beta_2)^\top$. Then $\tilde{\beta}^\top \mathbf{S} \tilde{\beta} = 3\beta_1^2 + 2\beta_1\beta_2 + 2\beta_2^2$. By Theorem 0.7:

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}}(\tilde{\beta}^\top \mathbf{S} \tilde{\beta})}_{2\times 1} = 2\mathbf{S}\tilde{\beta} = 2\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 6\beta_1 + 2\beta_2 \\ 2\beta_1 + 4\beta_2 \end{pmatrix}$$

Differentiating component-wise directly:

$$\begin{pmatrix} \frac{\partial}{\partial \beta_1}(3\beta_1^2 + 2\beta_1\beta_2 + 2\beta_2^2) \\ \frac{\partial}{\partial \beta_2}(3\beta_1^2 + 2\beta_1\beta_2 + 2\beta_2^2) \end{pmatrix} = \begin{pmatrix} 6\beta_1 + 2\beta_2 \\ 2\beta_1 + 4\beta_2 \end{pmatrix}$$

Both methods agree.

Corollary 0.2 (Derivative of a simple quadratic form)**.**

$$\underbrace{\frac{\partial}{\partial \tilde{\beta}}(\tilde{\beta}^\top \tilde{\beta})}_{p\times 1} = \underbrace{2\tilde{\beta}}_{p\times 1}$$

Proof

Proof. Applying Theorem 0.7 with $\mathbf{S} = \mathbf{I}_{p\times p}$ (which is symmetric and constant with respect to $\tilde{\beta}$):

$$\begin{aligned} \frac{\partial}{\partial \tilde{\beta}}(\tilde{\beta}^\top \tilde{\beta}) &= \frac{\partial}{\partial \tilde{\beta}}(\tilde{\beta}^\top \mathbf{I}_{p\times p} \tilde{\beta}) && \text{(rewrite with identity matrix)} \\ &= 2\mathbf{I}_{p\times p}\tilde{\beta} && \text{(derivative of a quadratic form, with } \mathbf{S} = \mathbf{I}_{p\times p}\text{)} \\ &= 2\tilde{\beta} && \text{(identity matrix property)} \end{aligned}$$

□

This vector derivative is like taking the derivative of x^2.

Example 0.11 (Derivative of a sum of squares)**.** Let $\tilde{\beta} = (\beta_1, \beta_2)^\top$, so $\tilde{\beta}^\top \tilde{\beta} = \beta_1^2 + \beta_2^2$. By Corollary 0.2:

$$\underbrace{\frac{\partial}{\partial\tilde{\beta}}(\tilde{\beta}^\top\tilde{\beta})}_{2\times 1} = 2\tilde{\beta} = \begin{pmatrix} 2\beta_1 \\ 2\beta_2 \end{pmatrix}$$

Direct partial differentiation yields the same column vector.

Theorem 0.8 (Vector chain rule)**.** *Let $\tilde{x}$ be a $p \times 1$ vector, let $\tilde{y} = \tilde{g}(\tilde{x})$ be a $q \times 1$ vector-valued function of $\tilde{x}$, and let $z = f(\tilde{y})$ be a scalar-valued function of $\tilde{y}$, where $\tilde{g}$ and f have continuous partial derivatives. Then $z = f(\tilde{g}(\tilde{x}))$, as a function of $\tilde{x}$, satisfies*

$$\underbrace{\frac{\partial z}{\partial \tilde{x}}}_{p\times 1} = \underbrace{\frac{\partial \tilde{y}}{\partial \tilde{x}}}_{p\times q} \underbrace{\frac{\partial z}{\partial \tilde{y}}}_{q\times 1}$$

where $\frac{\partial \tilde{y}}{\partial \tilde{x}}$ is the derivative of Definition 0.3 and $\frac{\partial z}{\partial \tilde{x}}$ and $\frac{\partial z}{\partial \tilde{y}}$ are vector derivatives (Definition 0.1).

See https://quickfem.com/finite-element-analysis/, specifically https://quickfem.com/wp-content/uploads/IFEM.AppF_.pdf

See also https://en.wikipedia.org/wiki/Gradient#Relationship_with_Fr%C3%A9chet_derivative

This chain rule is like the univariate chain rule, but the order matters now. The version presented here is for the gradient (column vector); the total derivative (row vector) would be the transpose of the gradient.

Corollary 0.3 (Vector chain rule for quadratic forms)**.** *If $\tilde{\varepsilon} = \tilde{\varepsilon}(\tilde{\beta})$ is an $n \times 1$ vector-valued function of the $p \times 1$ vector $\tilde{\beta}$, with continuous partial derivatives, then*

$$\underbrace{\frac{\partial}{\partial\tilde{\beta}}\left(\tilde{\varepsilon}(\tilde{\beta}) \cdot \tilde{\varepsilon}(\tilde{\beta})\right)}_{p\times 1} = \underbrace{\left(\frac{\partial}{\partial\tilde{\beta}}\tilde{\varepsilon}(\tilde{\beta})\right)}_{p\times n} \underbrace{\left(2\tilde{\varepsilon}(\tilde{\beta})\right)}_{n\times 1}$$

Proof

Proof. Apply Theorem 0.8 with $\tilde{x} = \tilde{\beta}$, $\tilde{y} = \tilde{\varepsilon}$, $q = n$, and $z = f(\tilde{\varepsilon}) = \tilde{\varepsilon} \cdot \tilde{\varepsilon} = \tilde{\varepsilon}^\top\tilde{\varepsilon}$:

$$\frac{\partial}{\partial \tilde{\beta}}(\tilde{\varepsilon} \cdot \tilde{\varepsilon}) = \left(\frac{\partial}{\partial \tilde{\beta}}\tilde{\varepsilon}\right)\frac{\partial}{\partial \tilde{\varepsilon}}(\tilde{\varepsilon}^\top \tilde{\varepsilon}) \quad \text{(vector chain rule)}$$

$$= \left(\frac{\partial}{\partial \tilde{\beta}}\tilde{\varepsilon}\right)(2\tilde{\varepsilon}) \qquad \text{(derivative of a simple quadratic form, in } \tilde{\varepsilon}\text{)}$$

The second step is Corollary 0.2, applied to the $n \times 1$ vector $\tilde{\varepsilon}$ in place of $\tilde{\beta}$. □

Example 0.12 (Derivative of the residual sum of squares)**.** Let $\tilde{y}$ $(n \times 1)$ and $\mathbf{X}$ $(n \times p)$ be constant with respect to $\tilde{\beta}$, and let $\tilde{\varepsilon}(\tilde{\beta}) = \tilde{y} - \mathbf{X}\tilde{\beta}$ be the vector of residuals. By Theorem 0.4, $\frac{\partial}{\partial \tilde{\beta}}(\mathbf{X}\tilde{\beta}) = \mathbf{X}^\top$, and $\frac{\partial}{\partial \tilde{\beta}}\tilde{y} = \mathbf{0}_{p\times n}$ because $\tilde{y}$ is constant, so $\frac{\partial}{\partial \tilde{\beta}}\tilde{\varepsilon} = -\mathbf{X}^\top$ $(p \times n)$. By Corollary 0.3:

$$\frac{\partial}{\partial \tilde{\beta}}(\tilde{\varepsilon} \cdot \tilde{\varepsilon}) = (-\mathbf{X}^\top)(2\tilde{\varepsilon}) \qquad \text{(vector chain rule for quadratic forms)}$$

$$= -2\,\mathbf{X}^\top\left(\tilde{y} - \mathbf{X}\tilde{\beta}\right) \quad \text{(substitute } \tilde{\varepsilon} = \tilde{y} - \mathbf{X}\tilde{\beta}\text{)}$$

Setting this $p \times 1$ vector to $\tilde{0}$ gives the normal equations $\mathbf{X}^\top\mathbf{X}\tilde{\beta} = \mathbf{X}^\top\tilde{y}$ of least squares.

Definition 0.5 (Matrix derivative)**.** For a scalar-valued function $f(\mathbf{X})$ of an $m \times n$ matrix $\mathbf{X}$, the **matrix derivative** is the $m \times n$ matrix whose (i, j) entry is the partial derivative of f with respect to the (i, j) entry of $\mathbf{X}$:

$$\left[\frac{\partial}{\partial \mathbf{X}} f\right]_{ij} = \frac{\partial}{\partial X_{ij}} f$$

Example 0.13 (The matrix derivative of a trace)**.** Let $\mathbf{X}$ be a 2×2 matrix and $f(\mathbf{X}) = \operatorname{tr}(\mathbf{X}) = X_{11} + X_{22}$ (see trace). Then $\frac{\partial}{\partial X_{ij}} f = 1$ if $i = j$ and 0 otherwise, so:

$$\frac{\partial}{\partial \mathbf{X}} f = \mathbf{I}_2$$

Theorem 0.9 (Matrix derivative of the trace of a matrix product)**.** *If* $\mathbf{A}$ *$(r \times m)$ and* $\mathbf{B}$ *$(n \times r)$ are constant with respect to the $m \times n$ matrix* $\mathbf{X}$*, then:*

$$\underbrace{\frac{\partial}{\partial \mathbf{X}} \operatorname{tr}(\mathbf{AXB})}_{m\times n} = \underbrace{\mathbf{A}^\top}_{m\times r}\underbrace{\mathbf{B}^\top}_{r\times n}$$

The trace makes $\operatorname{tr}(\mathbf{AXB})$ *a scalar, so its matrix derivative is again an $m \times n$ matrix. The derivative of the matrix product* $\mathbf{AXB}$ *itself (without the trace) is a fourth-order* *tensor, which is why this result is stated for the scalar* $\mathrm{tr}(\mathbf{AXB})$.

ℹ Proof

Proof. Write A_{kl}, X_{kl}, and B_{kl} for the entries of $\mathbf{A}$, $\mathbf{X}$, and $\mathbf{B}$. For entry (i, j):

$$\begin{aligned}
\left[\frac{\partial}{\partial \mathbf{X}} \mathrm{tr}(\mathbf{AXB})\right]_{ij} &= \frac{\partial}{\partial X_{ij}} \sum_{a=1}^{r} \sum_{b=1}^{m} \sum_{c=1}^{n} A_{ab} X_{bc} B_{ca} && \text{(trace of the } r \times r \text{ product } \mathbf{AXB}) \\
&= \sum_{a=1}^{r} A_{ai} B_{ja} && \text{(only the terms with } b = i,\ c = j \text{ depend on } X_{ij}) \\
&= \sum_{a=1}^{r} [\mathbf{A}^\top]_{ia} [\mathbf{B}^\top]_{aj} && \text{(definition of the transpose)} \\
&= [\mathbf{A}^\top \mathbf{B}^\top]_{ij} && \text{(definition of matrix multiplication)}
\end{aligned}$$

□

Example 0.14 (Differentiating a weighted trace)**.** Let $\mathbf{A} = \mathbf{I}_2$ (2×2) and $\mathbf{B} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ (2×2). Then $\mathrm{tr}(\mathbf{AXB}) = 2X_{11} + 3X_{22}$, and:

$$\underbrace{\frac{\partial}{\partial \mathbf{X}} \mathrm{tr}(\mathbf{AXB})}_{2\times 2} = \underbrace{\mathbf{A}^\top}_{2\times 2} \underbrace{\mathbf{B}^\top}_{2\times 2} = \mathbf{I}_2 \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

Additional resources

See also the Linear Algebra and Vector Calculus references.

- Hua Zhou's lecture notes for "UCLA Biostat 216 - Mathematical Methods for Biostatistics" (2023 Fall)

References

Fieller, Nick. 2016. *Basics of Matrix Algebra for Statistics with R*. Chapman; Hall/CRC. https://doi.org/10.1201/9781315370200.

Hutchinson, Brian. n.d. *DATA 471/571 (Machine Learning) and CSCI 481/581 (Deep Learning) Video Lectures*. Western Washington University. Accessed September 28, 2026. https://facultyweb.cs.wwu.edu/~hutchib2/video_lectures/data371/.