标量对矩阵求导

$f: \mathbb{R}^{m\times n} \mapsto \mathbb{R}$ $f(X)$

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

Jacobia matrix: $$\frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}$$

矩阵

Identities: vector-by-vector $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

Condition	Expression	Numerator layout, i.e. by y and x^T	Denominator layout, i.e. by y^T and x
a is not a function of **x**	$\frac{\partial \mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{0}$	
	$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{I}$	
A is not a function of **x**	$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{A}$	$\mathbf{A}^\top$
A is not a function of **x**	$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} =$	$\mathbf{A}^\top$	$\mathbf{A}$
a is not a function of **x**, **u** = **u**(**x**)	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	
a = *a*(**x**), **u** = **u**(**x**)	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}\frac{\partial a}{\partial \mathbf{x}}$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial a}{\partial \mathbf{x}}\mathbf{u}^\top$
A is not a function of **x**, **u** = **u**(**x**)	$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{A}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}^\top$
u = **u**(**x**), **v** = **v**(**x**)	$\frac{\partial (\mathbf{u}+\mathbf{v})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	
u = **u**(**x**)	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
u = **u**(**x**)	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$

Identities: scalar-by-vector $\frac{\partial y}{\partial \mathbf{x}} = \nabla_{\mathbf{x}} y$

Condition	Expression	Numerator layout, i.e. by x^T; result is row vector	Denominator layout, i.e. by x; result is column vector
a is not a function of **x**	$\frac{\partial a}{\partial \mathbf{x}} =$	$\mathbf{0}^\top$ [4]	$\mathbf{0}$ [4]
a is not a function of **x**, *u* = *u*(**x**)	$\frac{\partial au}{\partial \mathbf{x}} =$	$a\frac{\partial u}{\partial \mathbf{x}}$	
u = *u*(**x**), *v* = *v*(**x**)	$\frac{\partial (u+v)}{\partial \mathbf{x}} =$	$\frac{\partial u}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}$	
u = *u*(**x**), *v* = *v*(**x**)	$\frac{\partial uv}{\partial \mathbf{x}} =$	$u\frac{\partial v}{\partial \mathbf{x}} + v\frac{\partial u}{\partial \mathbf{x}}$	
u = *u*(**x**)	$\frac{\partial g(u)}{\partial \mathbf{x}} =$	$\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{x}}$	
u = *u*(**x**)	$\frac{\partial f(g(u))}{\partial \mathbf{x}} =$	$\frac{\partial f(g)}{\partial g}\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{x}}$	
u = **u**(**x**), **v** = **v**(**x**)	$\frac{\partial (\mathbf{u}\cdot\mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}^\top\mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^\top\frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^\top\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ • assumes numerator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{u}$ • assumes denominator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$
u = **u**(**x**), **v** = **v**(**x**), **A** is not a function of **x**	$\frac{\partial (\mathbf{u}\cdot\mathbf{A}\mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}^\top\mathbf{A}\mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^\top\mathbf{A}\frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^\top\mathbf{A}^\top\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ • assumes numerator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{A}^\top\mathbf{u}$ • assumes denominator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$
	$\frac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}^T} =$		$\mathbf{H}$, the Hessian matrix[5]
a is not a function of **x**	$\frac{\partial (\mathbf{a}\cdot\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial (\mathbf{x}\cdot\mathbf{a})}{\partial \mathbf{x}} =$ $\frac{\partial \mathbf{a}^\top\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^\top\mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{a}^\top$	$\mathbf{a}$
A is not a function of **x** **b** is not a function of **x**	$\frac{\partial \mathbf{b}^\top\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{b}^\top\mathbf{A}$	$\mathbf{A}^\top\mathbf{b}$
A is not a function of **x**	$\frac{\partial \mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{x}^\top(\mathbf{A}+\mathbf{A}^\top)$	$(\mathbf{A}+\mathbf{A}^\top)\mathbf{x}$
A is not a function of **x** **A** is symmetric	$\frac{\partial \mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$2\mathbf{x}^\top\mathbf{A}$	$2\mathbf{A}\mathbf{x}$
A is not a function of **x**	$\frac{\partial^2 \mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial \mathbf{x}^2} =$	$\mathbf{A}+\mathbf{A}^\top$	
A is not a function of **x** **A** is symmetric	$\frac{\partial^2 \mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial \mathbf{x}^2} =$	$2\mathbf{A}$	
	$\frac{\partial (\mathbf{x}\cdot\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^\top\mathbf{x}}{\partial \mathbf{x}} =$	$2\mathbf{x}^\top$	$2\mathbf{x}$
a is not a function of **x**, **u** = **u**(**x**)	$\frac{\partial (\mathbf{a}\cdot\mathbf{u})}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^\top\mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{a}^\top\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ • assumes numerator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{a}$ • assumes denominator layout of $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$
a, **b** are not functions of **x**	$\frac{\partial\, \mathbf{a}^\top\mathbf{x}\mathbf{x}^\top\mathbf{b}}{\partial\, \mathbf{x}} =$	$\mathbf{x}^\top(\mathbf{a}\mathbf{b}^\top+\mathbf{b}\mathbf{a}^\top)$	$(\mathbf{a}\mathbf{b}^\top+\mathbf{b}\mathbf{a}^\top)\mathbf{x}$
A, **b**, **C**, **D**, **e** are not functions of **x**	$\frac{\partial\, (\mathbf{A}\mathbf{x}+\mathbf{b})^\top\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{e})}{\partial\, \mathbf{x}} =$	$(\mathbf{D}\mathbf{x}+\mathbf{e})^\top\mathbf{C}^\top\mathbf{A}+(\mathbf{A}\mathbf{x}+\mathbf{b})^\top\mathbf{C}\mathbf{D}$	$\mathbf{D}^\top\mathbf{C}^\top(\mathbf{A}\mathbf{x}+\mathbf{b})+\mathbf{A}^\top\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{e})$
a is not a function of **x**	$\frac{\partial\, \|\mathbf{x}-\mathbf{a}\|}{\partial\, \mathbf{x}} =$	$\frac{(\mathbf{x}-\mathbf{a})^\top}{\|\mathbf{x}-\mathbf{a}\|}$	$\frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|}$

Identities: vector-by-scalar $\frac{\partial \mathbf{y}}{\partial x}$

Condition	Expression	Numerator layout, i.e. by y, result is column vector	Denominator layout, i.e. by y^T, result is row vector
a is not a function of *x*	$\frac{\partial \mathbf{a}}{\partial x} =$	$\mathbf{0}$[4]	
a is not a function of *x*, **u** = **u**(*x*)	$\frac{\partial a\mathbf{u}}{\partial x} =$	$a\frac{\partial \mathbf{u}}{\partial x}$	
A is not a function of *x*, **u** = **u**(*x*)	$\frac{\partial \mathbf{A}\mathbf{u}}{\partial x} =$	$\mathbf{A}\frac{\partial \mathbf{u}}{\partial x}$	$\frac{\partial \mathbf{u}}{\partial x}\mathbf{A}^\top$
u = **u**(*x*)	$\frac{\partial \mathbf{u}^\top}{\partial x} =$	$\left(\frac{\partial \mathbf{u}}{\partial x}\right)^\top$	
u = **u**(*x*), **v** = **v**(*x*)	$\frac{\partial (\mathbf{u}+\mathbf{v})}{\partial x} =$	$\frac{\partial \mathbf{u}}{\partial x} + \frac{\partial \mathbf{v}}{\partial x}$	
u = **u**(*x*), **v** = **v**(*x*)	$\frac{\partial (\mathbf{u}^\top\times\mathbf{v})}{\partial x} =$	$\left(\frac{\partial \mathbf{u}}{\partial x}\right)^\top\times\mathbf{v}+\mathbf{u}^\top\times\frac{\partial \mathbf{v}}{\partial x}$	$\frac{\partial \mathbf{u}}{\partial x}\times\mathbf{v}+\mathbf{u}^\top\times\left(\frac{\partial \mathbf{v}}{\partial x}\right)^\top$
u = **u**(*x*)	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial x} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial x}$ Assumes consistent matrix layout; see below.	$\frac{\partial \mathbf{u}}{\partial x}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
u = **u**(*x*)	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial x} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial x}$ Assumes consistent matrix layout; see below.	$\frac{\partial \mathbf{u}}{\partial x}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$
U = **U**(*x*), **v** = **v**(*x*)	$\frac{\partial (\mathbf{U}\times\mathbf{v})}{\partial x} =$	$\frac{\partial \mathbf{U}}{\partial x}\times\mathbf{v}+\mathbf{U}\times\frac{\partial \mathbf{v}}{\partial x}$	$\mathbf{v}^\top\times\left(\frac{\partial \mathbf{U}}{\partial x}\right)^\top+\frac{\partial \mathbf{v}}{\partial x}\times\mathbf{U}^\top$

$$\frac{\partial \beta^T \vec{x}}{\partial \vec{x}} = \beta \qquad \frac{\partial x^T x}{\partial x} = 2x$$

$$\frac{\partial x^T A x}{\partial x} = (A+A^T)x$$

$tr(a) = a$

$tr(AB) = tr(BA)$

$tr(ABC) = tr(CAB) = tr(BCA)$

$$\frac{\partial tr(AB)}{A} = B^T$$

$tr(A) = tr(A^T)$

$$\frac{\partial tr(ABA^TC)}{A} = CAB + C^TAB^T$$