MATHEMATICS

Differentiate $f(x) = x^T Ax$

Asked 8 years, 7 months ago Active 2 months ago Viewed 37k times

16

13

Calculate the differential of the function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = x^T Ax$$

with A symmetric. Also, differentiate this function with respect to x^T.

How exactly does this work in the case of vectors and matrices? Could anyone please help me out?

matrices multivariable-calculus derivatives quadratic-forms

Share Cite Follow

edited Sep 15 '19 at 16:40 Rodrigo de Azevedo 1

asked Feb 23 '13 at 15:59 dreamer 3,029 6 36 73

And yes, I will soon try to learn to use Latex :). – dreamer Feb 23 '13 at 16:03

2 I've edited your math formatting, could you look through it and see that it is still correct? – Arthur Feb 23 '13 at 16:06

1 Write math between $...$, you can find symbols etc. here: codecogs.com/latex/eqneditor.php – Kasper Feb 23 '13 at 16:06

1 try a 2×2 case explicitly and see if you can guess the general form of answer. – Maesumi Feb 23 '13 at 16:09

1 This is the composition of the linear map $x \longmapsto (x,x)$ and the bilinear map $(x,y) \longmapsto x^t Ay$. You can use the chain rule. – Julien Feb 23 '13 at 16:10

4 Answers

Active Oldest Votes

20

There is another way to solve the problem:

Let $\mathbf{x}^{n\times 1} = (x_1, \dots, x_n)'$ be a vector, the derivative of $\mathbf{y} = f(\mathbf{x})$ with respect to the vector $\mathbf{x}$ is defined by

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Let

$$\begin{aligned} \mathbf{y} &= f(\mathbf{x}) \\ &= \mathbf{x}' A\mathbf{x} \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \\ &= \sum_{i=1}^n a_{i1} x_i x_1 + \sum_{j=1}^n a_{1j} x_1 x_j + \sum_{i=2}^n \sum_{j=2}^n a_{ij} x_i x_j \\ \frac{\partial f}{\partial x_1} &= \sum_{i=1}^n a_{i1} x_i + \sum_{j=1}^n a_{1j} x_j \end{aligned}$$

$$= \sum_{i=1}^{n} a_{1i}x_i + \sum_{i=1}^{n} a_{1i}x_i \; [\text{since } a_{1i} = a_{1i}]$$

$$= 2\sum_{i=1}^{n} a_{1i}x_i$$

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} 2\sum_{i=1}^{n} a_{1i}x_i \\ \vdots \\ 2\sum_{i=1}^{n} a_{ni}x_i \end{pmatrix}$$

$$= 2\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= 2A\mathbf{x}$$

Share Cite Follow

edited Jun 12 '20 at 10:38
Community Bot
1

answered Feb 23 '13 at 20:05
Argha
4,353 1 27 47

Thanks for showing me this way as well :). Gives me more options :) – dreamer Feb 24 '13 at 10:01

1 The only thing that is slightly unclear to me is how x'Ax becomes the double summation (aijxixj). Why is the order reversed here? I mean, why arent the a's in the middle anymore? @Argha – dreamer Feb 24 '13 at 10:04

Note that $\mathbf{x}'A\mathbf{x} = (x_1, \dots, x_n) \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{11} & a_{12} & \dots & a_{1n} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and simply multipling we get required result. Also note order of $\mathbf{x}'$ is $1 \times n$ and order of A is $n \times n$. So order of $\mathbf{x}'A\mathbf{x}$ is 1×1. – Argha Feb 24 '13 at 11:24

Note that $a_{ij}\, x_i\, x_j \equiv x_i\, a_{ij}\, x_j$. So there is no problem at all. – Argha Feb 24 '13 at 11:36

1 On the first summation of the line that says [since a_1i = a_1i, how did you swap the indices from the previous step? – Michael Chav Sep 28 '17 at 23:40

As a start, things work "as usual": You calculate the difference between $f(x+h)$ and $f(x)$ and check how it depends on h, looking for a dominant linear part as $h \to 0$. Here, $f(x+h) = (x+h)^T A(x+h) = x^T Ax + h^T Ax + x^T Ah + h^T Ah = f(x) + 2x^T Ah$, so

14

$$+ h^T Ah$$

$f(x+h)-f(x)=2x^TA\cdot h+h^TAh$. The first summand is linear in h with a factor $2x^TA$, the second summand is quadratic in h, i.e. goes to 0 faster than the first / is negligible against the first for small h. So the row vector $2x^TA$ is our derivative (or transposed: $2Ax$ is the derivative with respect to x^T).

Share Cite Follow

answered Feb 23 '13 at 16:13

Hagen von Eitzen

1

Thank you. This is also what I tried. However, what confused me is that the question mentions that you should differentiate with respect to xT. From your answer, I see that you took the transpose of the 'ordinary' derivative. Does that imply that the ordinary derive is always taken with respect to x so that you can just take the transpose when you differentiate with respect to xT? – dreamer Feb 23 '13 at 16:21

2 Are you sure that $x^TAh=h^TAx$? Why? – mavavilj Nov 17 '16 at 13:24

@mavavij it's not. The dimensions don't necessarily check out. – Michael Chav Sep 28 '17 at 23:28

3 Actually, it's because A is symmetric. – Michael Chav Sep 28 '17 at 23:43

@Hagen von Eitzen's answer is certainly the fastest route here, but since you asked, here is a chain rule.

3

Here are two useful facts about linear and bilinear bounded maps from normed vectors spaces to normed vector spaces.

If f is linear and bounded, then trivially:

$$df_x(h)=f(h).$$

And if g is bilinear and bounded ($\|g(h,k)\|\le C\|h\|\|k\|$), we have

$$dg_{(x,y)}(h,k)=g(x,k)+g(h,y).$$

Now take $f(x)=(x,x)$ and $g(x,y)=x^tAy$. The former is linear and bounded, the latter is bilinear and bounded.

So, by the chain rule, $g\circ f(x)=x^tAx$ is differentiable and

$$d(g\circ f)_x(h)=dg_{f(x)}\circ df_x(h)=dg_{(x,x)}(h,h)=x^tAh+h^tAx.$$

This is true for any matrix A. Now if A is symmetric, this can be simplified since

$$x^tAh + h^tAx = x^tAh + h^tA^tx = x^tAh + (Ah)^tx = 2x^tAh.$$

Removing h, this gives

$$d(g \circ f)_x = 2x^tA.$$

Share Cite Follow

edited Feb 23 '13 at 16:41

answered Feb 23 '13 at 16:34

Julien
42.1k 3 68 149

Thank you. The other answer is indeed quicker but I am glad that I know now how to do it in this way as well. Much appreciated :). – dreamer Feb 23 '13 at 17:03

2 @user48288 You're welcome. And I am sure these general facts about bounded linear and bilinear maps will prove useful sooner or later. – Julien Feb 23 '13 at 17:04

Can I know in detail? I want to know this, but it can be hard to understand. – jakeoung Apr 15 '15 at 23:37

0

Here is relationship between directional derivative whenever f is differentiable.

$f'(p:v)$ denotes the derivative of f at p in the direction of v.

Let $f: U \subset \mathbb{R}^n \to \mathbb{R}$ and $p \in U, v \in \mathbb{R}^n$. Suppose that f is differentiable at p. Then we have

$$df_p(v) = f(p;v) = \lim_{t \to 0} \frac{f(\sigma(t)) - f(p)}{t}$$

for any differentiable curve $\sigma : (-\epsilon, \epsilon) \to U$ such that $\sigma(0) = p$ and $\sigma'(0) = v$.

In our case $f(x) = x^TAx$ and $\sigma(t) = x + th$,

$$f'(x;h) = \lim_{t \to 0} \frac{(x+th)^TA(x+th) - x^TAx}{t}$$

$$f'(x;h) = x^T Ah + h^T Ax$$

Since A is symmetric and we have the follwing:

$$f'(x;h) = x^T Ah + h^T Ax = x^T Ah + x^T A^T h$$

$$f'(x;h) = x^T (A + A^T)h$$

So the differential/gradient is simply $2x^T A$.

$$f'(x;h) = 2x^T Ah$$

Share Cite Follow

answered Jul 19 at 8:38

Tutankhamun

356 1 3 12