How to calculate gradient of $x^T Ax$

I am watching the following video lecture:

https://www.youtube.com/watch?v=G_p4QJrjdOw

In there, he talks about calculating gradient of $x^T Ax$ and he does that using the concept of exterior derivative. The proof goes as follows:

1. $y = x^T Ax$
2. $dy = dx^T Ax + x^T Adx = x^T(A + A^T)dx$ (using trace property of matrices)
3. $dy = (\nabla y)^T dx$ and because the rule is true for all dx
4. $\nabla y = x^T(A + A^T)$

It seems that in step 2, some form of product rule for differentials is applied. I am familiar with product rule for single variable calculus, but I am not understanding how product rule was applied to a multi-variate function expressed in matrix form.

It would be great if somebody could point me to a mathematical theorem that allows Step 2 in the above proof.

Thanks! Ajay

(matrices) (multivariable-calculus) (differential-forms)

edited Jul 16 '14 at 14:46
Mike Bell
191 1 11

asked Sep 3 '13 at 5:13
user855
158 1 7

Step 2 is using product rule rather than chain rule. – Shuhao Cao Sep 3 '13 at 5:16

Thanks. It's product rule indeed. But, where can I find a proof for the product rule for multivariable functions using differentials? I am not finding en.wikipedia.org/wiki/Product_rule convincing because uses the term differential very loosely. Is there a product rule proof using differential form properties? – user855 Sep 3 '13 at 5:40

1 $d(fg) = \sum_{i=1}^n \frac{\partial(fg)}{\partial x_i}dx_i = \sum_{i=1}^n (\frac{\partial f}{\partial x_i}g + f\frac{\partial g}{\partial x_i})dx_i = (\sum_{i=1}^n \frac{\partial f}{\partial x_i}dx_i)g + f(\sum_{i=1}^n \frac{\partial g}{\partial x_i}dx_i = gdf + fdg)$ – user71352 Sep 3 '13 at 5:56

5 Answers

$dy = d(x^T Ax) = d(Ax \cdot x) = d(\sum_{i=1}^n (Ax)_i x_i) = d(\sum_{i=1}^n \sum_{j=1}^n a_{i,j}x_j x_i) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j}x_i dx_j + \sum_{i=1}^n \sum_{j=1}^n a_{i,j}x_j dx_i$

$= \sum_{i=1}^n (Ax)dx_i + \sum_{i=1}^n (Adx)x_i = (dx)^T Ax + x^T Adx = (dx)^T Ax + (dx)^T A^T x$
$= (dx)^T(A + A^T)x$

answered Sep 3 '13 at 5:39
user71352
10.4k 2 7 24

The lecture says that using the properties of external definition of derivative, we can avoid representing it as "sums" and this simplifies the whole calculation – user855 Sep 3 '13 at 5:46

The above comments by others indicate that he applied the product rule for differentials. Unfortunately, I am unable to find a proof for the result of product rule for differentials. Any help there would be great. – user855 Sep 3 '13 at 5:47

I have placed a proof of product rule in the comment section. I hope that helps. – user71352 Sep 3 '13 at 6:16

Yes! Thanks a lot! – user855 Sep 3 '13 at 15:39

Step 2 might be the result of a simple computation. Consider $u(x) = x^T Ax$, then

$$u(x + h) = (x + h)^T A(x + h) = x^T Ax + h^T Ax + x^T Ah + h^T Ah,$$

that is, $u(x+h) = u(x) + x^T(A+A^T)h + r_x(h)$ where $r_x(h) = h^TAh$ (this uses the fact that $h^TAx = x^TA^Th$, which holds because $m = h^TAx$ is a 1×1 matrix hence $m^T = m$).

One sees that $r_x(h) = o(\|h\|)$ when $h \to 0$. This proves that the differential of u at x is the linear function $\nabla u(x) : \mathbb{R}^n \to \mathbb{R}$, $h \mapsto x^T(A+A^T)h$, which can be identified with the unique vector z such that $\nabla u(x)(h) = z^Th$ for every h in $\mathbb{R}^n$, that is, $z = (A+A^T)x$.

edited Sep 3 '13 at 6:09

answered Sep 3 '13 at 5:51
Did
209k 18 154 342

Here's a method which calculates the gradient of x^TAx without using the exterior derivative. I know that this is not what you are after, but it is worth noting how to prove it without the exterior derivative. This also allows for comparison with the exterior derivative method to see how much easier it is.

Let A be $n \times n$, $A = [a_{ij}]$. If $x \in \mathbb{R}^n$, $x = (x_1, \dots, x_n)^T$, then $y = \displaystyle\sum_{i=1}^n \sum_{j=1}^n a_{ij}x_ix_j$.

Then we have

$$\begin{aligned}
\frac{\partial y}{\partial x_k} &= \sum_{i \neq k} \frac{\partial}{\partial x_k}\left(\sum_{j=1}^n a_{ij}x_ix_j\right) + \frac{\partial}{\partial x_k}\left(\sum_{j=1}^n a_{kj}x_kx_j\right) \\
&= \sum_{i \neq k}\left(\frac{\partial}{\partial x_k}\left(\sum_{j \neq k} a_{ij}x_ix_j\right) + \frac{\partial}{\partial x_k}(a_{ik}x_ix_k)\right) + \sum_{j \neq k}\frac{\partial}{\partial x_k}(a_{kj}x_kx_j) + \frac{\partial}{\partial x_k}(a_{kk}x_k^2) \\
&= \sum_{i \neq k} a_{ik}x_i + \sum_{j \neq k} a_{kj}x_j + 2a_{kk}x_k \\
&= \sum_{i=1}^n a_{ik}x_i + \sum_{j=1}^n a_{kj}x_j \\
&= (x^TA)_k + (Ax)_k
\end{aligned}$$

where $(x^TA)_k$ is the k^{th} component of the row vector x^TA and $(Ax)_k$ is the k^{th} component of the column vector Ax. By taking the transpose of Ax we obtain the row vector x^TA^T which has the same k^{th} component as Ax does. Therefore $\dfrac{\partial y}{\partial x_k} = (x^TA)_k + (x^TA^T)_k$. Therefore

$$\nabla y = x^TA + x^TA^T = x^T(A+A^T).$$

answered Sep 3 '13 at 5:58
Michael Albanese
38k 9 51 159

Another approach is to use a multivariable product rule. Suppose g and h are differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, and $f(x) = \langle g(x), h(x)\rangle$ for all $x \in \mathbb{R}^n$. Then if $\Delta x \in \mathbb{R}^n$ is small we have

$$\begin{aligned}
f(x+\Delta x) &= \langle g(x+\Delta x), h(x+\Delta x)\rangle \\
&\approx \langle g(x) + g'(x)\Delta x, h(x) + h'(x)\Delta x\rangle \\
&= \langle g(x), h(x)\rangle + \langle h(x), g'(x)\Delta x\rangle + \langle g(x), h'(x)\Delta x\rangle + \text{small term} \\
&\approx f(x) + \langle g'(x)^Th(x), \Delta x\rangle + \langle h'(x)^Tg(x), \Delta x\rangle \\
&= f(x) + \langle g'(x)^Th(x) + h'(x)^Tg(x), \Delta x\rangle.
\end{aligned}$$

This suggests that

$$\nabla f(x) = g'(x)^Th(x) + h'(x)^Tg(x).$$

This is our multivariable product rule. (This derivation could be made into a rigorous proof by keeping track of error terms.)

In the case where $g(x) = x$ and $h(x) = Ax$, we see that

$$\begin{aligned}
\nabla f(x) &= Ax + A^Tx \\
&= (A+A^T)x.
\end{aligned}$$

answered Sep 3 '13 at 6:24
littleO
12.2k 2 14 44

I have some difficulties in understanding the derivations below the row ends with "+ small term" in the equation array, could you give me some hints? Thank you! – craftsman.don Oct 27 '13 at 13:34

The exterior derivative has nothing to do here. How could a student understand such a proof ! "Did" gave a good answer.

The gradient $\nabla(f)$ - of a function $f : E \to \mathbb{R}$ - is defined, modulo a dot product . on the vector-space E, by the formula: $\nabla(f)(x).h = Df_x(h)$ where Df_x is the derivative of f in x.

Example 1: $f : x \in \mathbb{R}^n \to x^TAx \in \mathbb{R}$. $Df_x(h) = h^TAx + x^TAh = x^T(A + A^T)h$ (it's the derivative of a non-commutative product !); we consider the dot product $u.v = u^Tv$. Thus $Df_x(h) = ((A + A^T)x).h$ and $\nabla(f)(x) = (A + A^T)x$, that is $\nabla(f) = A + A^T$.

Example 2: $f : X \in \mathcal{M}_n(\mathbb{R}) \to Trace(X^TAX) \in \mathbb{R}$. Since $Trace$ is a linear function, $Df_X(H) = Trace(H^TAX + X^TAH) = Trace(X^T(A + A^T)H)$; we consider the dot product $U.V = Trace(U^TV)$. Thus $Df_X(H) = ((A + A^T)X).H$ and $\nabla(f)(X) = (A + A^T)X$, that is $\nabla(f) = (A + A^T) \otimes I$. (Kronecker product).

Example 3 (more difficult): $f : X \in \mathcal{M}_n(\mathbb{R}) \to \det(X) \in \mathbb{R}$.
$Df_X(H) = Trace(adjoint(X)H) = adjoint(X)^T.H$ and $\nabla(f)(X) = adjoint(X)^T$.

Example 4: $f : X \in \mathcal{M}_n(\mathbb{R}) \to X^TAX \in \mathcal{M}_n(\mathbb{R})$. $Df_X(H) = H^TAX + X^TAH$. Here the gradient of f does not exist. In a pinch, we can define n^2 gradients, the $\nabla(f_{i,j})$ (componentwise) but these functions have no geometric meanings.

answered Sep 4 '13 at 17:43
loup blanc
7,232 2 5 19

What is $M_n(\mathbb{R})$ .. Manifold? – user855 Sep 4 '13 at 18:31

$\mathcal{M}_n(\mathbb{R})$ is the vector-space of $n \times n$ matrices with entries in $\mathbb{R}$. – loup blanc Sep 4 '13 at 18:50