Lecture 2 Part 2: Vectorization of Matrix Functions + Lecture 3 Part 1: Kronecker Products and Jacobians

MIT 18.S096 Matrix Calculus For Machine Learning and Beyond

March 17, 2024

1 Differential

$$f(A+dA) = f(A) + f'(A)[dA] + \text{higher-order stuff}$$

$$df(A;dA) = f'(A)[dA]$$

2 Example

Let f be the following function that maps from $\mathbb{R}^{n\times n}$ to $\mathbb{R}^{n\times n}$:

$$f(A) \;=\; A^2$$

Deriving the difference:

$$\begin{aligned} f(A+dA) - f(A) &= (A+dA)^2 - A^2 \\ &= A^2 + A\,(dA) + (dA)\,A + (dA)^2 - A^2 \\ &= A\,(dA) + (dA)\,A + (dA)^2 \end{aligned}$$

The differential is the part of the difference that's *linear* in dA:

$$df(A;dA) = A\,(dA) + (dA)\,A$$

This can be abbreviated as

$$df = A\,(dA) + (dA)\,A.$$

We can represent this result as matrix multiplication:

$$\begin{aligned} df &= A\,(dA) + (dA)\,A \\ &= A\,(dA)I + I(dA)\,A \\ \text{vec}(df) &= \text{vec}(A\,(dA)I + I(dA)\,A) \\ &= \text{vec}(A\,(dA)I) + \text{vec}(I(dA)\,A) \\ &= (I\otimes A)\,\text{vec}(dA) + (A^T\otimes I)\text{vec}(dA) \\ &= (I\otimes A + A^T\otimes I)\,\text{vec}(dA) \end{aligned}$$

3 Example

Let f be the following function that maps from $\mathbb{R}^{n\times n}$ to $\mathbb{R}^{n\times n}$:

$$f(A) \;=\; A^3$$

Deriving the difference:

$$\begin{aligned}
& f(A+dA)-f(A) \\
= \;& (A+dA)^3 - A^3 \\
= \;& (A+dA)^2(A+dA) - A^3 \\
= \;& (A+dA)^2 A + (A+dA)^2\, dA - A^3 \\
= \;& [A^2 + A\,(dA) + (dA)A + (dA)^2]A + [A^2 + A\,(dA) + (dA)\,A + (dA)^2]dA - A^3 \\
= \;& A^3 + A(dA)A + (dA)A^2 + (dA)^2A + A^2(dA) + A(dA)^2 + (dA)A(dA) + (dA)^3 - A^3 \\
= \;& A(dA)A + (dA)A^2 + (dA)^2A + A^2(dA) + A(dA)^2 + (dA)A(dA) + (dA)^3
\end{aligned}$$

The differential is the part of the difference that's *linear* in dA:

$$df(A;dA) \;=\; A(dA)A + (dA)A^2 + A^2(dA)$$

This can be abbreviated as

$$df = A(dA)A + (dA)A^2 + A^2(dA)$$

We can represent this result as matrix multiplication:

$$\begin{aligned}
df &= A(dA)A + (dA)A^2 + A^2(dA) \\
&= A(dA)A + I(dA)A^2 + A^2(dA)I \\
\operatorname{vec}(df) &= (A^T \otimes A + (A^2)^T \otimes I + I \otimes A^2)\operatorname{vec}(dA)
\end{aligned}$$