Lecture 1 Part 2: Derivatives as Linear Operator

MIT 18.S096 Matrix Calculus For Machine Learning and Beyond

March 17, 2024

Table of contents

1 Differential

In the scalar-to-scalar case:

$$f(x+dx) \;=\; f(x)+f'(x)\,dx+(\text{higher order terms})$$

When $\boldsymbol{x}\in\mathbb{R}^n$ and $f\colon\mathbb{R}^n\to\mathbb{R}$:

$$f(\boldsymbol{x}+d\boldsymbol{x})=f(\boldsymbol{x})+f'(\boldsymbol{x})d\,\boldsymbol{x}+(\text{higher order terms})$$

We see that the derivative $f'(\boldsymbol{x})$ must be a row vector in order for $f'(\boldsymbol{x})d\,\boldsymbol{x}$ to be a scalar. We define the gradient ∇f as the column vector $f'(\boldsymbol{x})^T$.

2 Example

Let $f\colon\mathbb{R}^n\to\mathbb{R}$ be the function below:

$$f(\boldsymbol{x}) \;=\; \boldsymbol{x}^TA\boldsymbol{x}.$$

Calculating the difference, the differential and the gradient:

$$\begin{aligned}
f(\boldsymbol{x}+d\boldsymbol{x})-f(\boldsymbol{x}) &= (\boldsymbol{x}+d\boldsymbol{x})^TA(\boldsymbol{x}+d\boldsymbol{x})-\boldsymbol{x}^TA\boldsymbol{x}\\
&= \boldsymbol{x}^TA\boldsymbol{x}+\boldsymbol{x}^TA(d\boldsymbol{x})+(d\boldsymbol{x})^TA\boldsymbol{x}+(d\boldsymbol{x})^TA(d\boldsymbol{x})-\boldsymbol{x}^TA\boldsymbol{x}\\
&= \boldsymbol{x}^TA(d\boldsymbol{x})+(d\boldsymbol{x})^TA\boldsymbol{x}+(d\boldsymbol{x})^TA(d\boldsymbol{x})\\
df &= \boldsymbol{x}^TA(d\boldsymbol{x})+(d\boldsymbol{x})^TA\boldsymbol{x}\\
&= \boldsymbol{x}^TA(d\boldsymbol{x})+\boldsymbol{x}^TA^T(d\boldsymbol{x})\\
&= \boldsymbol{x}^T(A+A^T)(d\boldsymbol{x})\\
\nabla f &= (A+A^T)\boldsymbol{x}
\end{aligned}$$