Finding Gradient by Inspecting First Order Expansion

Mong H. Ng

March 11, 2018

1 Overview

In this note, we will first understand how we can find derivative by inspecting first order in 1-dimensional function. Then, extending this, we apply the same technique in higher dimension. At times, this can be easier than computing individual partial derivatives and stacking them in gradient vector.

2 Background and 1-Dimensional Derivative

From calculus, we know the definition of a 1-dimensional function $f : \mathbb{R} \to \mathbb{R}$ is as followed:

$$f'(x) = \lim_{y \to x, y \neq x} \frac{f(y) - f(x)}{y - x}$$

Using the change of variable $h = y - x$, we obtain an alternate definition of derivative:

$$f'(x) = \lim_{h \to 0, h \neq 0} \frac{f(x+h) - f(x)}{h}$$

We can multiply both side by h, move terms to the left hand side, and obtain

$$\lim_{h \to 0, h \neq 0} [f(x+h) - f(x)] - f'(x)h = 0$$

This equation has a simple interpretation. We are looking for a number $f'(x)$ such that when we are near x, multiplying the changes h by this number will get us very close to $f(x+h) - f(x)$.

As such, we can define the function

$$f(y) \approx f(x) + f'(x)(y - x)$$

and call it the *first-order approximation* of f near x.

3 Finding Derivative by First Order Expansion in 1-Dimension

In general case, the technique of finding derivative by inspecting first order expansion works as followed:

1. We are given a function f(x).
2. For fixed x (after all, derivative works only when we are near x), we evaluate and expand $f(x + \Delta)$ for some small Δ.
3. Lastly, we extract the first order term with respect to Δ and claim it our derivative.

As an (trivial) example, say we want to find the derivative of $f(x) = x^2$. We evaluate

$$f(x + \Delta) = (x + \Delta)^2 = x^2 + 2x\Delta + \Delta^2$$

The term Δ^2 is a second order term, so we ignore it. And now, compare

$$f(y) \approx f(x) + f'(x)(y - x)$$

which we write as

$$f(x + \Delta) \approx f(x) + f'(x)\Delta$$

with

$$f(x + \Delta) = (x + \Delta)^2 \approx f(x) + 2x\Delta$$

By inspection, the derivative $f'(x)$ is simply just $2x$, as expected.

4 Higher Dimension

Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, we can adopt a similar definition

$$\lim_{y \in \mathbf{dom} f, y \neq x, y \rightarrow x} \frac{||f(y) - f(x) - Df(x)(y - x)||_2}{||y - x||_2} = 0$$

where Df(x) is called the Jacobian (derivative in higher dimension). We define gradient $\nabla f(x)$ as the transpose of the Jacobian.

Similarly, we call the following the *first-order approximation*

$$f(y) \approx f(x) + Df(x)(y - x) \tag{1}$$

4.1 Example 1

Let $f(x) = x^T Ax$, we apply the same technique to find the gradient.

$$\begin{aligned}
f(x + \Delta) &= (x + \Delta)^T A(x + \Delta) \\
&= x^T Ax + x^T A\Delta + \Delta^T Ax + \Delta^T \Delta \\
&= f(x) + x^T A\Delta + x^T A^T \Delta + H.O.T && \text{since } \Delta^T Ax = x^T A^T \Delta \\
&\approx f(x) + x^T (A + A^T)\Delta && \text{where we drop the H.O.T.}
\end{aligned}$$

Compare this with equation 1, we again deduce by inspection that the Jacobian $Df(x) = x^T(A + A^T)$. Therefore, the gradient $\nabla f(x) = Df(x)^T = (A + A^T)x$.

5 $\mathbb{R}^{m\times n} \to \mathbb{R}$

For function $f : \mathbb{R}^{m\times n} \to \mathbb{R}$, we can define the Jacobian as followed:

$$\lim_{\Delta\to 0, \Delta\neq 0} \frac{f(X+\Delta) - f(X) + Tr(Df(X)\Delta)}{||\Delta||_F} = 0$$

This definition is consistent with our definition for vector. Indeed, if we think of a matrix $A = \begin{bmatrix} a_1 & a_2 & ... & a_m \end{bmatrix}$ and $B = \begin{bmatrix} b_1 & b_2 & ... & b_n \end{bmatrix}$ as one long column vector

$$\bar{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \bar{B} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Then we have $Tr(AB) = \bar{A}^T \bar{B}$.

As before, we have the following as the first-order approximation.

$$f(X+\Delta) \approx f(X) + Tr(Df(X)\Delta)$$

5.1 Example 2

Let $f(X) = Tr(XAX^T)$, where $A = A^T$ is symmetric, find the Jacobian.

$$\begin{aligned} f(X+\Delta) &= Tr[(X+\Delta)A(X+\Delta)^T] \\ &= Tr(XAX^T) + Tr(XA\Delta^T) + Tr(\Delta AX^T) + h.o.t. \\ &\approx f(X) + Tr(\Delta A^T X^T) + Tr(\Delta AX^T) && \text{since } Tr(XA\Delta^T) = Tr(\Delta A^T X^T) \\ &\approx f(X) + 2Tr(\Delta AX^T) && \text{since } A = A^T \\ &\approx f(X) + 2Tr(AX^T\Delta) && \text{since } Tr(AB) = Tr(BA) \end{aligned}$$

By inspection, we can see that $2AX^T$ is our Jacobian.