## Insight into the Gradient Formula

The second formula in Taylor's Theorem (Theorem 1) can seem puzzling. Let's see where it comes from.

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)p\,dt$$

This is just the Fundamental Theorem of Calculus, $g(1) - g(0) = \int_0^1 g'(t)dt$, applied to a vector function.

1. Define $g(t) = \nabla f(x+tp)$ for $t \in [0,1]$.
2. Apply the theorem: $g(1) = \nabla f(x+p)$ and $g(0) = \nabla f(x)$

   So, $\nabla f(x+p) - \nabla f(x) = \displaystyle\int_0^1 g'(t)dt$.

Let's differentiate $g(t) = \nabla f(x+tp)$ with respect to t: $g'(t) = \dfrac{d}{dt}\nabla f(x+tp)$ Using the multivariate chain rule:

$$g'(t) = \underbrace{\nabla^2 f(x+tp)}_{\text{Derivative of } \nabla f} \cdot \overbrace{p}^{\text{Derivative of } x+tp}$$

Substituting this back into the integral gives the final result.



### Comments

Let's break down the second formula from the Theorem 1, as its origin is not immediately obvious. The formula is a beautiful application of single-variable calculus to a multivariate setting.

The key is to reduce the problem to one dimension. We are interested in the change of the gradient, $\nabla f$, as we move from point $x$ to point $x+p$ along a straight line. We can parameterize this path using a variable $t \in [0,1]$, where our position is $x+tp$.

We define a new vector function of a single variable, $g(t) = \nabla f(x+tp)$. This function simply tells us the gradient of f at a point t-fraction of the way along the vector p.

Now, we can apply the Fundamental Theorem of Calculus to $g(t)$, which states that the total change in g from $t=0$ to $t=1$ is the integral of its rate of change, $g'(t)$. In other words, $g(1) - g(0) = \displaystyle\int_0^1 g'(t)dt$.

Substituting our definitions back in, $g(1) = \nabla f(x+p)$ and $g(0) = \nabla f(x)$. The final piece is to find $g'(t)$. Using the chain rule, the derivative of $\nabla f(x+tp)$ with respect to t is the derivative of the outer function ($\nabla f$) evaluated at the inner function ($x+tp$), multiplied by the derivative of the inner function. The derivative of the gradient vector $\nabla f$ is the Hessian matrix $\nabla^2 f$, and the derivative of $x+tp$ with respect to t is simply the vector p.

Thus, the rate of change of the gradient along the path is $\nabla^2 f(x+tp)p$. Integrating this rate of change from the start to the end gives the total change in the gradient. The formula simply states:

Final Gradient = Initial Gradient + Total Accumulated Change.