Second order Taylor approximation formula for multivariate functions

— Joaquín Gómez.

— February 13, 2025

Taylor approximation polynomials in one variable functions are derived in the following way.

First, let $f(x)$ be an n^{th} degree polynomial like the following

$$f(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 + ... + \lambda_n x^n$$

Then, to get the first coefficient, modify $f(x)$ to be a linear function by setting $\lambda_k = 0$ for $k = 2, 3, ..., n$, such that $f(x) = \lambda_0 + \lambda_1 x$. Then, differentiate f and get

$$f'(x) = \lambda_1$$

So the approximation becomes

$$f(x) = \lambda_0 + f'(x)x$$

It's necessary to clarify that λ_0 is arbitrary, however if we want to find an approximation formula for $f(x + \Delta x)$ we would require to recall the definition for the derivative, but letting $h = \Delta x$ and not infinitesimal, "small enough"

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} \approx f'(x)$$

Solving for $f(x + \Delta x)$ this function looks like our Taylor approximation $f(x) = \lambda_0 + f'(x)x$

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

Then, we can rewrite our starting polynomial as

$$f(x+\Delta x) = f(x) + f'(x)\Delta x + \lambda_2(\Delta x)^2 + ... + \lambda_n(\Delta x)^n$$

Now, let's repeat the process of setting $\lambda_k = 0$ now for $k = 3, 4, ..., n$. Now, take the second order derivative of the polynomial to get

$$f''(x) \approx 2\lambda_2 \longrightarrow \lambda_2 = \frac{1}{2}f''(x+\Delta x)$$

(Note that, we aren't taking the derivative of the approximation $f(x+\Delta x)$). If you continue this process, you will find that the n^{th} coefficient is given by

$$\lambda_n = \frac{1}{n!}f^{(n)}(x)$$

So, our Taylor expansion is

$$f(x+\Delta x) = \lim_{n\to\infty}\sum_{k=1}^{n}\frac{f^{(n)}(x)}{n!}\Delta x$$

Multivariate case

The multivariate case requires a little bit of setup. Let f be a transformation $f : \mathbb{R}^n \to \mathbb{R}$, a **scalar field** that is differentiable in $x \in \mathbb{R}^n$. The derivative of this function is in terms of a **unit vector** called the *direction* of the derivative, thus is named the "directional derivative" of the function. We define this derivative as

$$f'(\mathbf{x};\mathbf{u}) = \lim_{h\to 0}\frac{f(\mathbf{x}+h\mathbf{u}) - f(\mathbf{x})}{h} \tag{1}$$

This is important to define the gradient. First, the Taylor approximation formula will have the form

$$f(\mathbf{x}+\mathbf{u}) \approx f(\mathbf{x}) + f'(\mathbf{x};\mathbf{u}) \tag{2}$$

for $\mathbf{u}$ sufficiently small. We now will prove that

$$f'(\mathbf{x};\mathbf{u}) = \nabla f(x)\cdot\mathbf{u}$$

First, we can write $\mathbf{u}$ as

$$\mathbf{u} = \sum_{i=1}^{n} u_i e_i \qquad \text{where } \mathbf{u} = (u_1, ..., u_n)$$

Then, we apply a transformation $T(\mathbf{u}) = f'(\mathbf{x}; \mathbf{u})$, this transformation is linear because the directional derivative is linear, this means $T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$.

$$T(\mathbf{u}) = T\left(\sum_{i=1}^{n} u_i e_i\right) = \sum_{i=1}^{n} u_i T(e_i) = \sum_{i=1}^{n} u_i f'(\mathbf{x}; e_i) = \sum_{i=1}^{n} u_i D_k f(\mathbf{x}) \tag{3}$$

The derivatives in the direction of the set of basis vectors $e_1, e_2, ..., e_n$ are the partial derivatives given by $\frac{\partial f}{\partial e_k} = D_k f(\mathbf{x})$ for $k = 1, 2, ..., n$. (these vectors are, for example $e_1 = (1, 0, 0, ..., 0)$. These are the basis vectors of $\mathbb{R}^n$). Then we rewrite (3) as

$$\sum_{i=1}^{n} u_i D_k f(\mathbf{x}) = \nabla f(x) \cdot \mathbf{u}$$

And, we know $\nabla f = (D_1 f(\mathbf{x}), ..., D_n f(\mathbf{x}))$.

So (2) becomes

$$f(\mathbf{x} + \mathbf{u}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot u \tag{4}$$

Why is this true? well, we know by linear algebra that $\nabla f(\mathbf{x}) \cdot u = \|\nabla f(\mathbf{x})\| \|\mathbf{u}\| \cos \theta$, but $\mathbf{u}$ is a unit vector so $\|\mathbf{u}\| = 1$. This product is the norm of the gradient in the direction of $\mathbf{u}$, if both vectors are orthogonal then it's zero, this means that there is no change in that direction.

Finding the second order Taylor approximation is easy, we will generalize the one variable case, thus we need to find the second order derivative

Finding the second order Taylor approximation is easy, we will generalize the one variable case, thus we need to find the second order derivative, for this purpose we will use the right hand side expression for the directional derivative in (3)

$$f'(\mathbf{x}; \mathbf{u}) = \sum_{i=1}^{n} u_i D_k f(\mathbf{x})$$

Then, by the chain rule we know that

$$f''(\mathbf{x};\mathbf{u}) = \sum_{j=1}^{n} u_j \frac{\partial}{\partial e_j}(f'(\mathbf{x};\mathbf{u})) = \sum_{j=1}^{n} u_j D_j(f'(\mathbf{x};\mathbf{u})) \tag{5}$$

This is the second derivative in the direction of the unit vector $\mathbf{u}$. Expanding the expression we find

$$f''(\mathbf{x};\mathbf{u}) = \sum_{j=1}^{n} u_j D_{ij} f(\mathbf{x}) u_i \tag{6}$$

Which can be rewritten in matrix notation as

$$\sum_{j=1}^{n} u_j D_{ij} f(\mathbf{x}) u_i = \mathbf{u}^T H(\mathbf{x})\mathbf{u} \tag{7}$$

where $H(\mathbf{x})$ is the hessian matrix of f, defined as

$$H(\mathbf{x}) = [D_{ij} f(\mathbf{x})]_{i,j=1}^{n} = \begin{bmatrix} \frac{\partial^2 f}{\partial e_1^2} & \frac{\partial^2 f}{\partial e_1 e_2} & \dots & \frac{\partial^2 f}{\partial e_1 e_n} \\ \frac{\partial^2 f}{\partial e_2 e_1} & \frac{\partial^2 f}{\partial e_2^2} & \dots & \frac{\partial^2 f}{\partial e_2 e_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial e_n e_1} & \frac{\partial^2 f}{\partial e_n e_2} & \dots & \frac{\partial^2 f}{\partial e_n^2} \end{bmatrix}$$

We have shown that the second order Taylor approximation for scalar fields is given by

$$f(\mathbf{x}+\mathbf{u}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot u + \frac{1}{2!}\mathbf{u}^T H(\mathbf{x})\mathbf{u} \tag{8}$$

where $H(x)$ is the Hessian matrix of f.

— END — (February 13, 2025)