Optimization from Calculus

Chris Cornwell

September 11 and 16, 2025

Outline

Approximations from derivatives

Stationary points

Outline

Approximations from derivatives

Stationary points

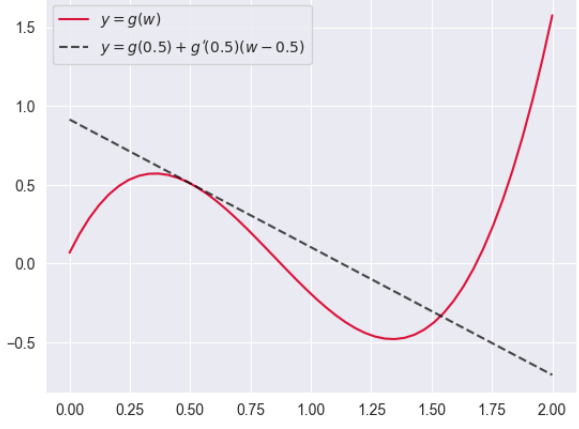
Linear approximations

$g : \mathbb{R} \to \mathbb{R}$, a twice-differentiable function (at least). Write w for input to g, so $g(w)$.

Linear approximation to $g(w)$. At a point $(v, g(v))$ on its graph, function whose graph is the tangent line:

$$h(w) = g(v) + g'(v)(w - v).$$

- Also called first order Taylor series approximation (or Taylor polynomial) of g.
- Approximates values of g, for inputs near v.



Second order (quadratic) approximations

To approximate g better near v (and in a larger interval around v), we can use the second order Taylor polynomial – which incorporates both first and second derivative information. This is the function:

$$h(w) = g(v) + g'(v)(w - v) + \frac{1}{2}g''(v)(w - v)^2.$$

Note: if $g'(v) = 0$ and $g''(v) > 0$, then $h(w) \geq g(v)$. So (where the approximation is quite good, sufficiently near v), values of the original function $g(w)$ should be larger or equal to $g(v)$ (A local minimum).

The opposite is true if $g'(v) = 0$ and $g''(v) < 0$.

Extending approximations to multiple variables

We may extend the first and second order approximations to the setting of multiple variables. This requires the use of the gradient where, if $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \dots & w_N \end{bmatrix}^T$, then

$$\nabla g = [\frac{\partial g}{\partial w_1} \ \frac{\partial g}{\partial w_2} \ \dots \ \frac{\partial g}{\partial w_N}]^T.$$

The linear approximation is $h(\mathbf{w}) = g(\mathbf{v}) + \nabla g(\mathbf{v})^T(\mathbf{w} - \mathbf{v})$. (Note that this *is* a linear function; its graph is a translate of the subspace of $\mathbb{R}^{N+1}$ that is normal to $\begin{bmatrix} -\nabla g(\mathbf{v})^T, & 1 \end{bmatrix}^T$, the translation is by a vector whose dot product with $\begin{bmatrix} -\nabla g(\mathbf{v})^T, & 1 \end{bmatrix}^T$ is equal to $g(\mathbf{v}) - \nabla g(\mathbf{v})^T\mathbf{v}$.)

Outline

Approximations from derivatives

Stationary points

Stationary points versus (local) minimal

First order condition