Single-Variable and Multivariate Optimization

Zvi Rosen
Department of Mathematics

September 30, 2016

Task

Given a function f from $\mathbb{R} \to \mathbb{R}$, find the local and global optima of the function.

Definition

Local Minimum (Maximum) is a point x_0 for which there exists $\epsilon > 0$ so that if $|x - x_0| < \epsilon$, then $f(x) \geq f(x_0)$ (resp. $f(x) \leq f(x_0)$).
Global Minimum (Maximum) is a point x_0 so that for any $x \in \mathbb{R}$, $f(x) \geq f(x_0)$ (resp. $f(x) \leq f(x_0)$.

Two techniques

Two types of optimization techniques:

1. Gradient methods: Uses the derivative of the function.
2. Direct methods: Uses only function evaluation.

Single-variable gradient method

Wherever the function f has a local optimum x_0, $f'(x_0) = 0$.

- If $f''(x_0) > 0$ then the function is convex or concave up. The local optimum is a minimum.
- If $f''(x_0) < 0$, then the function is concave, or concave down.
  The local optimum is a maximum.
- If $f''(x_0) = 0$, then it may be a maximum, minimum or an inflection point.

Single-variable example

Minimize $f(x) = e^x - x^3 - x$.
We compute $f'(x) = e^x - 3x^2 - 1$, and $f''(x) = e^x - 6x$.
One obvious root is 0, which has $f''(0) > 0$ so it is a maximum.
Then we look for a root, using any root-finding method. E.g. the`fzero` method of MATLAB.
The root 3.7823 has positive $f''(x)$ so it is a minimum.
Further argument shows that it is the global minimum.

Gradient Descent

Let $f(x, y)$ be a function of two variables. The gradient is:

$$\nabla f(x, y) = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$$

The gradient is a vector pointing up, the negative gradient points down.
Iterate:

$$x_{n+1} = x_n - \lambda_n \nabla f(x, y)$$

Gradient Descent Example

$$f(x,y) = 2x^2 + 2y^2 - 6x - 6y + 9$$

$$\nabla f(x,y) = \begin{bmatrix} 4x - 6 \\ 4y - 6 \end{bmatrix}$$

Set $\lambda = .1$. Start at $(0,0)$.

Direct Methods: Golden-Section Search

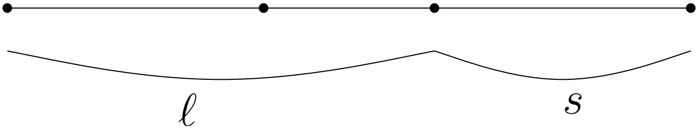
Golden-section has a lot in common with the bisection method. For optimization, computing the sign compared to the brackets is not enough. We need to find the subinterval where the function is largest (resp. smallest).
So, at each step, we take two points inside the interval. Then, we take the point with the larger value, and the two points on either side, and iterate.

Direct Methods: Golden-Section Search

Why is it called golden-section?

We want to save time by only computing one new function value at each iteration.

In particular, we want $\frac{\ell}{s} = \frac{s}{\ell - s}$, with $\ell + s = 1$.



What if our bracket contains multiple local optima?

Golden section will not work.

Parabolic Interpolation

Begin with three points $x_1 x_2, x_3$, two of which bracket the true maximum (resp. minimum). Suppose x_1 is the left endpoint and x_2 is the right endpoint.

There is a unique parabola that contains the three points. Find its maximum. Then add that point as x_4. For the next iteration, use the three consecutive points for which f(midpoint) is largest.

Parabolic Interpolation

Begin with three points $x_1 x_2, x_3$, two of which bracket the true maximum (resp. minimum). Suppose x_1 is the left endpoint and x_2 is the right endpoint.

There is a unique parabola that contains the three points. Find its maximum. Then add that point as x_4. For the next iteration, use the three consecutive points for which f(midpoint) is largest.

Parabolic Interpolation - Example

Maximize $f(x) = \sin(x)$. Let $x_1 = 0, x_2 = 1, x_3 = 2$. The corresponding sine values are $0, 0.84, 0.91$.
The parabola has formula

$$y = (x-0)(x-1)\frac{0.91}{2} + (x-0)(x-2)\frac{0.84}{-1} + (x-1)(x-2)\frac{0}{2}$$

$$= 1.225x - 0.385x^2.$$

The maximum of this parabola is reached at $x_4 = 1.225/(2 \cdot 0.385) = 1.59$. The value of sine at this point is 0.9998. In the next iteration, we would use x_2, x_4, and x_3.

Nelder-Mead Algorithm

Implemented in MATLAB using `fminsearch`.
See gif.