Incomplete mini-overview of optimisers

Michael Clerx

June 2, 2020

Abstract

Brief overview of derivative-base methods, because some concepts come up again later. Then longer overview of derivative-free methods. Then some meta-methods, that use other methods. Throughout we assume we're minimising a scalar function on a real-valued parameter space.

1 Root-finding methods

These methods are concerned with the specific problem of finding the *roots* of a function.

1.1 Derivative-based root-finding methods

1.1.1 Newton's method

Requires $\boldsymbol{x} \in \mathbb{R}^n$, $f(\boldsymbol{x}) \to \mathbb{R}$, and a twice-differentiable function so that $f'(\boldsymbol{x})$ and $f''(\boldsymbol{x})$ both exist and can be computed.

One-dimension Starting from a point x_k, we approximate f near x_k with a Taylor series, using a free variable t:

$$f(x_k + t) \approx f(x_k) + f'(x_k)t + \frac{1}{2}f''(x_k)t^2 \tag{1}$$

and we approximate the root of f as the root of this parabola, which we find by looking where the derivative is zero:

$$0 = \frac{d}{dt}\left(f(x_k) + f'(x_k)t + \frac{1}{2}f''(x_k)t^2\right) = f'(x_k) + f''(x_k)t \tag{2}$$

$$t = -\frac{f'(x_k)}{f''(x_k)} \tag{3}$$

Finally, turn this into an iterative procedure:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \tag{4}$$

Modified Newton's method Introduce some damping, by setting a step-size $0 < \gamma \leq 1$:

$$x_{k+1} = x_k - \gamma\frac{f'(x_k)}{f''(x_k)} \tag{5}$$

Multivariate Let $f'(\boldsymbol{x})$ be the gradient of $f(\boldsymbol{x})$, and H be the Hessian $H_{i,j} = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]$. Then

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \gamma H^{-1}(\boldsymbol{x}_k) f'(\boldsymbol{x}_k) \tag{6}$$

.

1.1.2 Halley's method

https://en.wikipedia.org/wiki/Halley%27s_method

Housholder's methods Generalised this idea, Newton = 1, Halley = 2

1.1.3 More methods

1956 Muller's method https://www.jstor.org/stable/2001916 https://en.wikipedia.org/wiki/Muller%27s_method 1979 Ridders' method https://doi.org/10.1109/TCS.1979.1084580 https://en.wikipedia.org/wiki/Ridders%27_method

1.2 Derivative-free root-finding methods

1.2.1 Bisection method

https://en.wikipedia.org/wiki/Bisection_method

1.2.2 Regular falsi

https://en.wikipedia.org/wiki/Regula_falsi

1.2.3 Secant method

Predates Newton's method, but can be seen as finite-difference approximation to it.

https://en.wikipedia.org/wiki/Secant_method

2 Derivative-based optimisation methods

These methods assume we know the derivative, and sometimes the second derivative, of the function we're trying to minimise.

2.1 Line-search based methods

At each iteration, these methods choose a direction in parameter space, and then search for a minimum along that line.

They require $\boldsymbol{x} \in \mathbb{R}^n$, $f(\boldsymbol{x}) \to \mathbb{R}$ and the ability to calculate or approximate $f'(\boldsymbol{x})$ and $f''(\boldsymbol{x})$.

1. Estimate a direction $\boldsymbol{p}_k$
2. Estimate a step size α_k that minimises $f(\boldsymbol{x_k} + \alpha_k \boldsymbol{p}_k)$

3. Set $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$

Note that step two requires solving a 1-dimensional optimisation sub-problem. This can be solved in an *exact* or an *inexact* way. Some guarantees on convergence can be made if α_k is chosen so that the *Wolfe conditions* are satisfied (Wolfe, 1969, 1971).

2.1.1 BFGS: Broyden–Fletcher–Goldfarb–Shanno algorithm

L-BFGS: Limited memory BFGS Slight variation, removes need to store entire Hessian matrix.

2.2 Trust-region based methods

3 Derivative-free optimisation methods

Example reference Nelder and Mead (1965)

4 Meta-metods

References

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. The Computer Journal 7, 308–313.

Wolfe, P., 1969. Convergence conditions for ascent methods. SIAM review 11, 226–235.

Wolfe, P., 1971. Convergence conditions for ascent methods. ii: Some corrections. SIAM review 13, 185–188.