

Line Search Methods

Shefali Kulkarni-Thaker

Consider the following unconstrained optimization problem

$$\min_{x \in R} f(x)$$

Any optimization algorithm starts by an initial point x_0 and performs series of iterations to reach the optimal point x^* . At any k^{th} iteration the next point is given by the sum of old point and the direction in which to search of a next point multiplied by how far to go in that direction. Thus

$$x_{k+1} = x_k + \alpha_k d_k$$

here d_k is a search direction and α_k is a positive scalar determining how far to go in that direction. It is called as the *step length*. Additionally, the new point must be such that the function value (the function which we are optimizing) at that point should be less than or equal to the previous point. This is quite obvious because if we are moving in a different direction where the function value is increasing we are not really moving towards the minimum. Thus,

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) \\ f(x_k + \alpha_k d_k) &\leq f(x_k) \end{aligned}$$

The general optimization algorithm begins with an initial point, finds a descent search direction, determines the step length and checks the termination criteria. So, when we are trying to find the step length, we already know that the direction in which we are going is descent. So, ideally we want to go far enough so that the function reaches its minimum. Thus, given the previous point and a descent search direction, we are trying to find a scalar step length such that the value of the function is minimum in that direction. In its mathematical form we may write,

$$\min_{\alpha \geq 0} f(x + \alpha d)$$

Since, x and d are known, this problem reduces to a univariate, $1-D$, minimization problem. Assuming that f is smooth and continuous, we find its optimum where its first-derivative is zero, i.e. $\nabla f(x + \alpha d) = 0$. All we are doing is trying to find zero of a function (i.e. the point where the curves intersects the x -axis). We will visit some known and some newer zero-finding (or root-finding) techniques.

When, the dimensionality of the problem or the degree of equation increases finding an exact zero is difficult. Usually, we are looking for an interval $0 \in [a, b]$ so that $|a - b| < \epsilon$, where ϵ is an acceptable tolerance.

1 Bisection Method

In bisection method we reduce begin with an interval so that $0 \in [a, b]$ and divide the interval in two halves, i.e. $[a, \frac{a+b}{2}]$ and $[\frac{a+b}{2}, b]$. A next search interval is chosen by comparing and finding which one has zero. This is done by evaluating the sign. The algorithm for this is given as follows: Choose a, b so that $f(a)f(b) < 0$

1. $m = \frac{a+b}{2}$

2. $f(m) = 0$ Stop and return.
3. $f(m)f(a) < 0; b = m$
4. $f(m)f(b) < 0; a = m$
5. $|a - b| < \epsilon$ stop and return.

1.1 Features of bisection

- Guaranteed Convergence.
- Slow as it has linear convergence with $\beta = 0.5$.
- At the k^{th} iteration $\frac{|b-a|}{2^k} = \epsilon$ giving total evaluations as $k = \log_2\left(\frac{|b-a|}{\epsilon}\right)$

1.2 Example

Converging $f(x) = \frac{x}{1+x^2}$ with $[a, b] = [-0.6, 0.75]$ and $\epsilon = 10^{-4}$.

Number	a	$f(a)$	b	$f(b)$	m	$f(m)$
1	-0.6	-0.441176	0.75	0.48	0.075	0.07458
2	-0.6	-0.441176	0.075	0.07458	-0.2625	-0.245578
3	-0.2625	-0.245578	0.075	0.07458	-0.09375	-0.092933
4	-0.09375	-0.092933	0.075	0.07458	-0.009375	-0.009374
5	-0.009375	-0.009374	0.075	0.07458	0.032813	0.032777
6	-0.009375	-0.009374	0.032813	0.032777	0.011719	0.011717
7	-0.009375	-0.009374	0.011719	0.011717	0.001172	0.001172
8	-0.009375	-0.009374	0.001172	0.001172	-0.004102	-0.004101
9	-0.004102	-0.004101	0.001172	0.001172	-0.001465	-0.001465
10	-0.001465	-0.001465	0.001172	0.001172	-0.000146	-0.000146
11	-0.000146	-0.000146	0.001172	0.001172	0.000513	0.000513
12	-0.000146	-0.000146	0.000513	0.000513	0.000183	0.000183
13	-0.000146	-0.000146	0.000183	0.000183	0.000018	0.000018
14	-0.000146	-0.000146	0.000018	0.000018	-0.000064	-0.000064

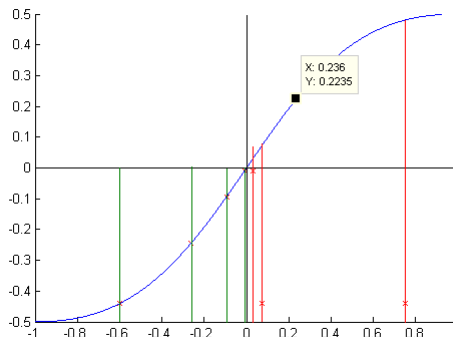


Figure 1: The bisection method

If the function is smooth and continuous, the speed of convergence can be improved by using the derivative information

2 Newton's Method

In Newton's method does a linear approximation of the function and finding the x -intercept of that approximation, thereby improving the performance of the bisection method. Linear approximation can be done by using Taylor's series.

$$\begin{aligned} f'(x_{k+1}) &= f(x_k) + f'(x_k)(x_{k+1} - x_k) \\ f'(x_{k+1}) &= 0 \\ x_{k+1} &= x_k - \frac{f(x_k)}{f'(x_k)} \end{aligned}$$

2.1 Features of Newton's method

- Newton's method has a quadratic convergence when the chosen point is close enough to zero. If the derivative is zero at the root, it has only local quadratic convergence.
- Numerical difficulties occur when the first-derivative is zero.
- If a poor starting point is chosen the method may fail to converge or diverge.
- If it is difficult to find analytical derivation, secant method may be used.

For a smooth, continuous function when proper starting point is chosen, Newton's method can be real fast. The convergence of $f(x) = \frac{x}{1+x^2}$ with $x_0 = 0.5$ and $\epsilon = 10^{-4}$.

X	$f(x)$
1	0.5
2	-0.333333
3	0.083333
4	-0.001166
5	0

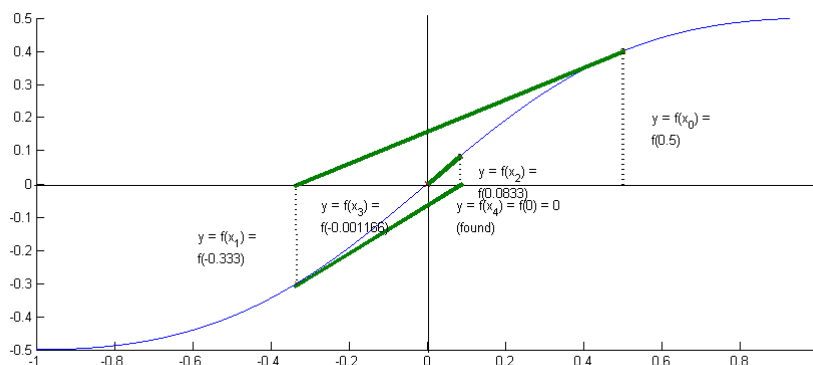
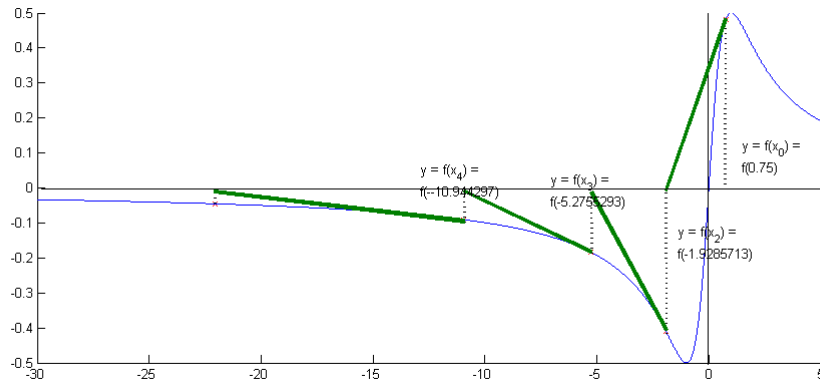


Figure 2: Newton's method converging with $x_0 = 0.5$

The divergence of $f(x) = \frac{x}{1+x^2}$ with $x_0 = 0.75$ and $\epsilon = 10^{-4}$.

1	0.75
2	-1.928571
3	-5.275529
4	-10.944297
5	-22.072876
6	-44.236547
⋮	⋮
20	-725265.55757

Figure 3: Newton's method diverging with $x_0 = 0.75$

Newton's method to find the minimum uses the first derivative of the approximation an equalling it to zero.

$$\begin{aligned}
 f'(x_{k+1}) &= f'(x_k) + f''(x_k)(x_{k+1} - x_k) \\
 f''(x_{k+1}) &= f''(x_k) + f'''(x_k)(x_{k+1} - x_k) \\
 f''(x_{k+1}) &= 0 \\
 x_{k+1} &= x_k - \frac{f'(x_k)}{f''(x_k)}
 \end{aligned}$$

3 Secant Method

When f' is expensive or cumbersome to calculate, one can use secant's method to approximate the derivative. The derivation of this method comes by replacing first derivative in the newton's method by its approximation (finite differentiation), i.e $f'(x_k) = \frac{f_k - f_{k-1}}{x_k - x_{k-1}}$, where $f_k = f(x_k)$

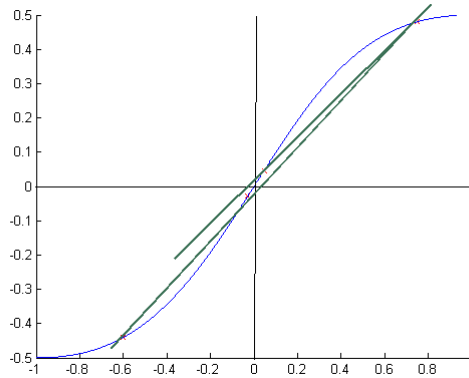
$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f_k - f_{k-1}} f_k$$

Just like Newton's method the secant's method to find the minimum is given by:

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f'_k - f'_{k-1}} f'_k$$

Convergence of secant method is super-linear with $\beta = 1.618$. The table below shows the secant method convergence for $f = \frac{x}{1+x^2}$ with -0.6 and 0.75 as initial points

Number	x_{k-1}	x_k	$f(x_{k-1})$	$f(x_k)$
1	-0.6	0.75	-0.441176	0.48
2	0.75	0.046552	0.48	0.046451
3	0.046552	-0.028817	0.046451	-0.028793
4	-0.028817	0.000024	-0.028793	0.000024

Figure 4: Secant's method converging with $x_0 = -0.6$; $x_1 = 0.75$

4 Golden Section Method

The table below shows the Golden Section method convergence for $f = \frac{x}{1+x^2}$ with -0.6 and 0.75 as initial points

Number	a	b	$f(a)$	$f(b)$
0	-0.6	0.75	-0.08375	0.222146
1	-0.6	0.234346	-0.441176	0.222146
2	-0.6	-0.084346	-0.441176	-0.08375
3	-0.6	-0.281308	-0.441176	-0.26068
\vdots	\vdots	\vdots	\vdots	\vdots
20	-0.6	-0.599911	-0.441176	-0.441146

$$x_{min} = -0.599966 \quad f(x_{min}) = -0.441165$$

4.1 Another example for golden-section

Consider the function:

$$f(x_1, x_2) = (x_2 - x_1^2)^2 + e^{x_1^2}$$

You are at the point (0,1). Find the minimum of the function in the direction (line) $(1, 2)^T$ using the Golden-Section line-search algorithm on the step-length interval $[0, 1]$. Stop when the length of the interval is less than 0.2. Note: step-length interval could be described by the parameter t , and, so, all the points along the direction $(1, 2)^T$ can be expressed as $(0, 1) + t \cdot (1, 2)$.

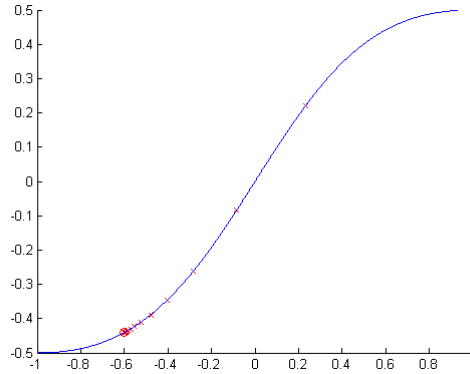


Figure 5: Golden Section method converging for $f = \frac{x}{1+x^2}$ with $a = -0.6$; $b = 0.75$

The equation to find new point is defined as:

$$x_{k+1} = x_k + \alpha_k d_k$$

Here, $d_k = (1, 2)$ and $x_0 = (0, 1)$

In the initial step length interval $[0, 1]$, let $\alpha_1 = 0.618$ and $\alpha_2 = 0.382$. Thus,

$$x_1 = (0, 1)^T + 0.382(1, 2)^T = (0.382, 1.764)^T ; f(x_1) = 3.7753$$

$$x_2 = (0, 1)^T + 0.618(1, 2)^T = (0.618, 2.236)^T ; f(x_2) = 4.9027$$

Since $f(x_1) < f(x_2)$, the new step-length interval is $[0, 0.618]$

$$\alpha_3 = 0.618 * 0.382 = 0.2361$$

$$x_3 = (0, 1)^T + 0.2361(1, 2)^T = (0.2361, 1.4722)^T ; f(x_3) = 3.0637$$

The new step-length interval is $[0, 0.2361]$

Using the Golden Section search method, the new step length is $\alpha_4 = 0.2361 * 0.618 = 0.1459$

$$x_4 = (0, 1)^T + 0.1459(1, 2)^T = (0.1459, 1.2918)^T ; f(x_4) = 2.6357$$

The new step-length interval is $[0, 0.1459]$

$$\alpha_5 = 0.1459 * 0.618 = 0.0902$$

$$x_5 = (0, 1)^T + 0.0902(1, 2)^T = (0.0902, 1.1804)^T ; f(x_5) = 2.3824$$

The new step-length interval is $[0, 0.0902]$

$$\alpha_6 = 0.0902 * 0.618 = 0.0557$$

$$x_6 = (0, 1)^T + 0.0557(1, 2)^T = (0.0557, 1.1114)^T ; f(x_6) = 2.2314$$

Step-length Interval	$\alpha_i = \alpha_{i-1} * 0.618$	x_1	x_2	$f(x_1, x_2) = (x_2 - x_1^2)^2 + e^{x_1^2}$
[0, .0557]	0.0344	0.0344	1.0688	2.1410
[0, .0344]	0.0213	0.0213	1.0426	2.0865
[0, .0213]	0.0132	0.0132	1.0264	2.0533
[0, .0132]	0.0081	0.0081	1.0162	2.0326
[0, .0081]	0.0050	0.0050	1.0100	2.0201
[0, .0050]	0.0031	0.0031	1.0062	2.0124
[0, .00031]	0.0002	0.0002	1.0004	2.0008
[0, .0002]	0.0001	0.0001	1.0002	2.0004
[0, .0001]	0.00006	0.00006	1.00012	2.0002
[0, .00006]	0.000034	0.00004	1.00008	2.0002

5 References

1. Jorge Nocedal, Stephen J. Wright: Numerical Optimization, Springer Series in Operations Research (1999).
2. E. de Klerk, C. Roos, and T. Terlaky, Nonlinear Optimization (pdf) (ps), 1999-2004, Delft.
3. <http://de2de.synechism.org/c3/sec36.pdf>