Chapter 7 One-Dimensional Search Methods

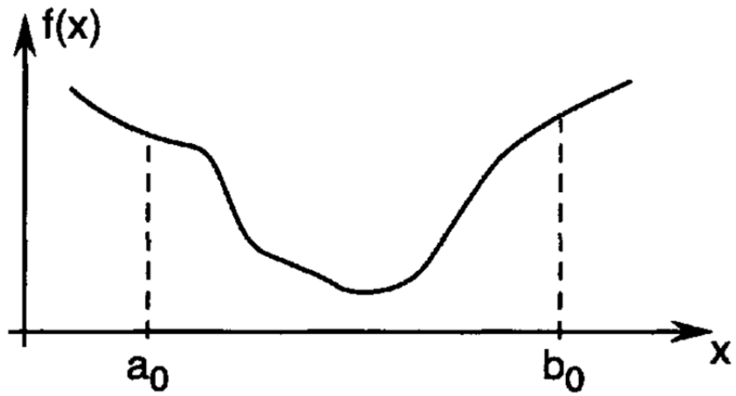
An Introduction to Optimization

Spring, 2014

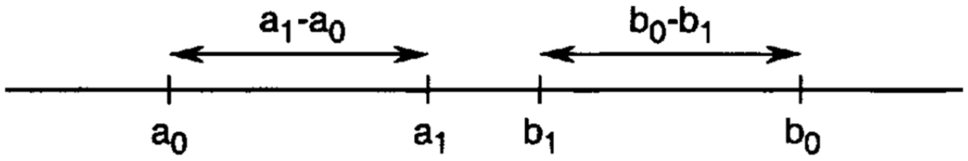
Wei-Ta Chu

Golden Section Search

- Determine the minimizer of a function $f: R \to R$ over a closed interval, say $[a_0, b_0]$. The only assumption is that the objective function is ***unimodal***, which means that it has only one local minimizer.
- The method is based on evaluating the objective function at different points in the interval. We choose these points in such a way that an approximation to the minimizer may be achieved in as few evaluations as possible.
- Narrow the range progressively until the minimizer is "boxed in" with sufficient accuracy.



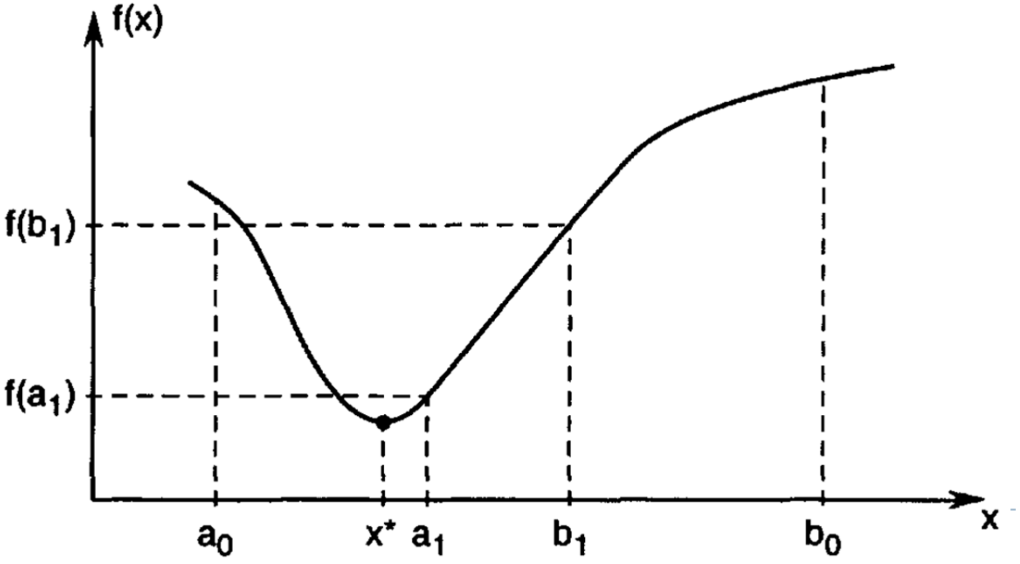
Golden Section Search

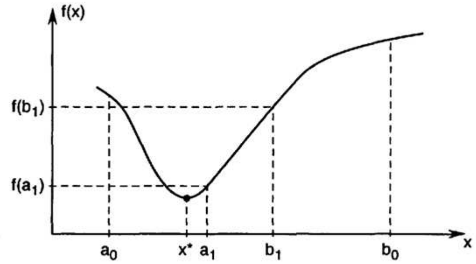


- We have to evaluate f at two intermediate points. We choose the intermediate points in such a way that the reduction in the range is symmetric.

$$a_1 - a_0 = b_0 - b_1 = \rho(b_0 - a_0) \qquad \rho < \frac{1}{2}$$

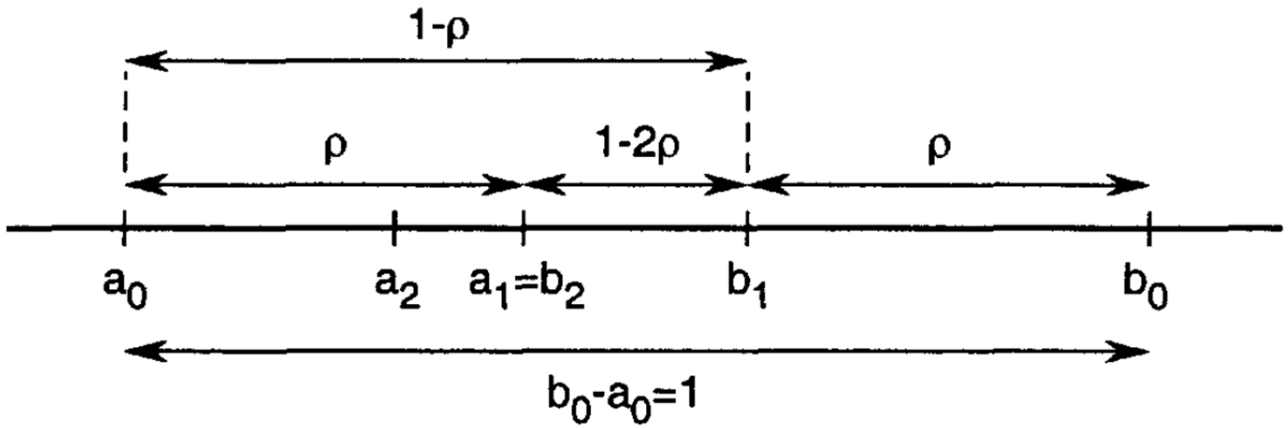
- If $f(a_1) < f(b_1)$, then the minimizer must lie in the range $[a_0, b_1]$
- If $f(a_1) \geq f(b_1)$, then the minimizer is located in the range $[a_1, b_0]$

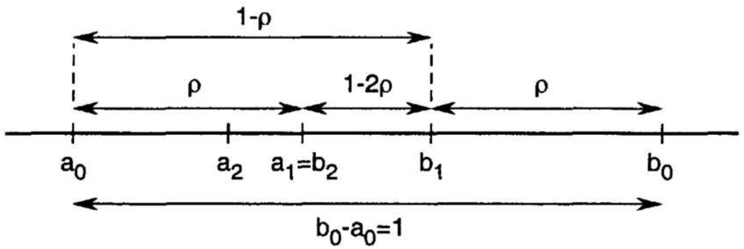




Golden Section Search

- We would like to minimize the number of objective function evaluations.
- Suppose $f(a_1) < f(b_1)$. Then, we know that $x^* \in [a_0, b_1]$. Because a_1 is already in the uncertainty interval and $f(a_1)$ is already known, we can make a_1 coincide with b_2. Thus, only one new evaluation of f at a_2 would be necessary.





Golden Section Search

- Without loss of generality, imagine that the original range $[a_0, b_0]$ is of unit length. Then,

$$\rho(b_1 - a_0) = b_1 - b_2$$

Because $b_1 - a_0 = 1 - \rho$ and $b_1 - b_2 = 1 - 2\rho$

$$\rho(1 - \rho) = 1 - 2\rho$$

$$\rho^2 - 3\rho + 1 = 0 \Rightarrow \rho_1 = \frac{3+\sqrt{5}}{2} \qquad \rho_2 = \frac{3-\sqrt{5}}{2}$$

Because we require $\rho < \frac{1}{2}$, we take $\rho = \frac{3-\sqrt{5}}{2} \approx 0.382$

Observe that

$$1 - \rho = \frac{\sqrt{5}-1}{2} \Rightarrow \frac{\rho}{1-\rho} = \frac{3-\sqrt{5}}{\sqrt{5}-1} = \frac{\sqrt{5}-1}{2} = \frac{1-\rho}{1}$$

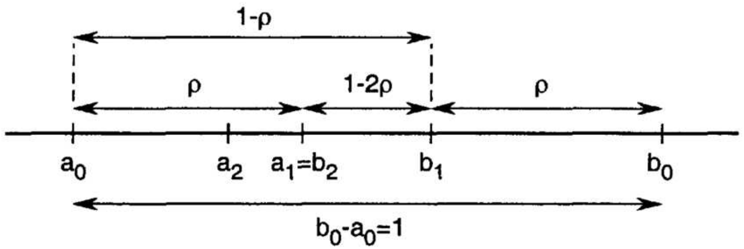
Dividing a range in the ratio of ρ to $1 - \rho$ has the effect that the ratio of the shorter segment to the longer equals to the ratio of the longer to the sum of the two. This rule is called ***golden section***.

Golden Section Search

- The uncertainty range is reduced by the ratio $1 - \rho \approx 0.61803$ at every stage. Hence, N steps of reduction using the golden section method reduces the range by the factor

$$(1 - \rho)^N \approx (0.61803)^N$$

Example



- Use the golden section search to find the value of x that minimizes $f(x) = x^4 - 14x^3 + 60x^2 - 70x$ in the range [0,2]. Locate this value of x to within a range of 0.3.
- After *N* stage the range [0,2] is reduced by $(0.61803)^N$. So we choose *N* so that $(0.61803)^N \leq 0.3/2$. *N*=4 will do.
- Iteration 1. We evaluate f at two intermediate points a_1 and b_1. We have

$$a_1 = a_0 + \rho(b_0 - a_0) = 0.7639$$

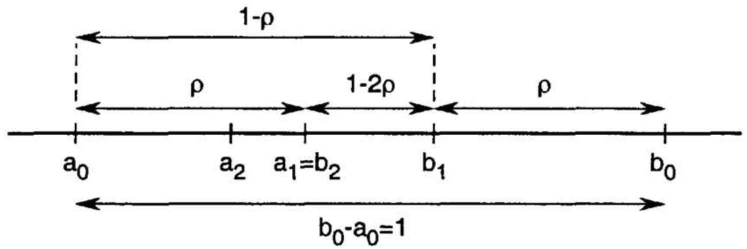
$$\rho = \frac{3-\sqrt{5}}{2}$$

$$b_1 = a_0 + (1-\rho)(b_0 - a_0) = 1.236$$

$$f(a_1) = -24.36$$

$$f(b_1) = -18.96$$

$f(a_1) < f(b_1)$, so the uncertainty interval is reduced to $[a_0, b_1] = [0, 1.236]$



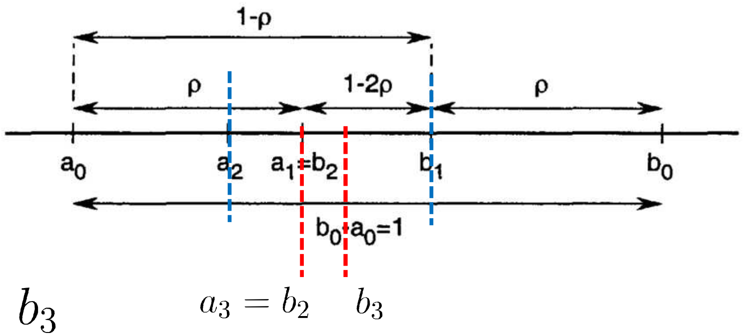
Example

- Iteration 2. We choose b_2 to coincide with a_1 , and f need only be evaluated at one new point,

$$a_2 = a_0 + \rho(b_1 - a_0) = 0.4721$$
$$f(a_2) = -21.10$$
$$f(b_2) = f(a_1) = -24.36$$

Now, $f(b_2) < f(a_2)$, so the uncertainty interval is reduced to $[a_2, b_1] = [0.4721, 1.236]$

Example



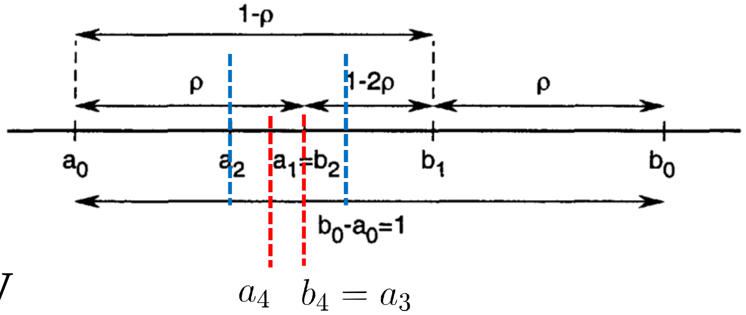
- Iteration 3. We set $a_3 = b_2$ and compute b_3

$$b_3 = a_2 + (1-\rho)(b_1 - a_2) = 0.9443$$

We have

$$f(a_3) = f(b_2) = -24.36$$

$$f(b_3) = -23.59$$



So $f(b_3) > f(a_3)$. Hence, the uncertainty interval is further reduced to $[a_2, b_3] = [0.4721, 0.9443]$

- Iteration 4. We set $b_4 = a_3$ and $a_4 = a_2 + \rho(b_3 - a_2) = 0.6525$

We have

$$f(a_4) = -23.84$$

$$f(b_4) = f(a_3) = -24.36$$

$f(a_4) > f(b_4)$. Thus, the value of x that minimizes f is located in the interval $[a_4, b_3] = [0.6525, 0.9443]$. Note that $b_3 - a_4 = 0.292 < 0.3$

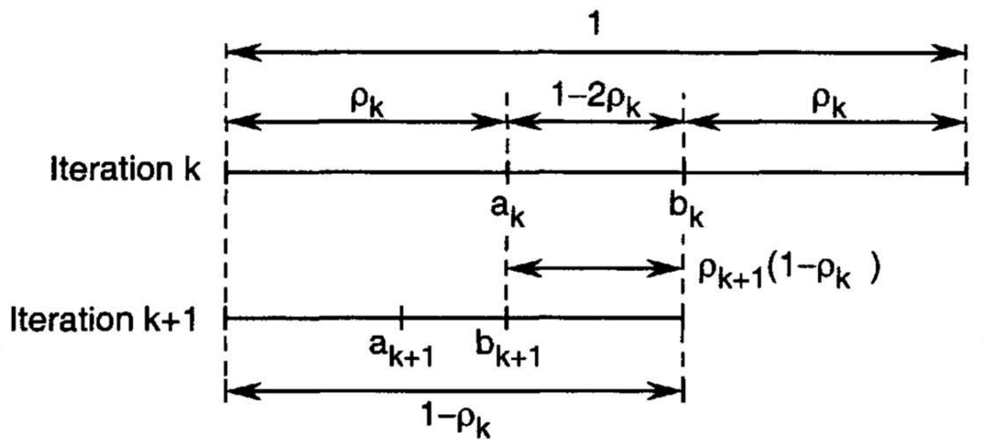
Fibonacci Search

- Suppose now that we are allowed to vary the value ρ from stage to stage.
- As in the golden section search, our goal is to select successive values of ρ_k , $0 \le \rho_k \le 1/2$, such that only one new function evaluation is required at each stage.

$$\rho_{k+1}(1-\rho_k) = 1 - 2\rho_k$$

After some manipulations, we obtain

$$\rho_{k+1} = 1 - \frac{\rho_k}{1-\rho_k}$$



Fibonacci Search

$$\rho_{k+1} = 1 - \frac{\rho_k}{1-\rho_k}$$

- Suppose that we are given a sequence $\rho_1, \rho_2, \ldots$ satisfying the conditions above and we use this sequence in our search algorithm. Then, after *N* iterations, the uncertainty range is reduced by a factor of

$$(1-\rho_1)(1-\rho_2)\cdots(1-\rho_N)$$

- What sequence $\rho_1, \rho_2, \ldots$ minimizes the reduction factor above?
- This is a constrained optimization problem

$$\begin{aligned} &\text{minimize } (1-\rho_1)(1-\rho_2)\cdots(1-\rho_N) \\ &\text{subject to } \rho_{k+1} = 1 - \tfrac{\rho_k}{1-\rho_k}, k = 1, \ldots, N-1 \\ &\qquad\qquad 0 \le \rho_k \le 1/2, k = 1, \ldots, N \end{aligned}$$

Fibonacci Search

- The ***Fibonacci sequence*** $F_1, F_2, F_3, \ldots$ is defined as follows. Let $F_{-1} = 0$ and $F_0 = 1$. Then, for $k \geq 0$

$$F_{k+1} = F_k + F_{k-1}$$

- Some values of elements in the Fibonacci sequence

F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
1	2	3	5	8	13	21	34

- It turns out the solution to the optimization problem above is

$$\begin{aligned}
\rho_1 &= 1 - \frac{F_N}{F_{N+1}} \\
\rho_2 &= 1 - \frac{F_{N-1}}{F_N} \\
&\vdots \\
\rho_k &= 1 - \frac{F_{N-k+1}}{F_{N-k+2}} \\
&\vdots \\
\rho_N &= 1 - \frac{F_1}{F_2}
\end{aligned}$$

Fibonacci Search

- The resulting algorithm is called the ***Fibonacci search method***.
- In this method, the uncertainty range is reduced by the factor

$$(1-\rho_1)(1-\rho_2)\cdots(1-\rho_N) = \frac{F_N}{F_{N+1}}\frac{F_{N-1}}{F_N}\cdots\frac{F_1}{F_2} = \frac{F_1}{F_{N+1}} = \frac{1}{F_{N+1}}$$

- The reduction factor is less than that of the golden section method.
- There is an anomaly in the final iteration, because

$$\rho_N = 1 - \frac{F_1}{F_2} = \frac{1}{2}$$

- Recall that we need two intermediate points at each stage, one comes from a previous iteration and another is a new evaluation point. However, with $\rho_N = \frac{1}{2}$, the two intermediate points coincide in the middle of the uncertainty interval, and thus we cannot further reduce the uncertainty range.

Fibonacci Search

- To get around this problem, we perform the new evaluation for the last iteration using $\rho_N = \frac{1}{2} - \epsilon$, where ϵ is a small number.
- The new evaluation point is just to the left or right of the midpoint of the uncertainty interval.
- As a result of the modification, the reduction in the uncertainty range at the last iteration may be either

$$1 - \rho_N = \frac{1}{2}$$

or

$$1 - (\rho_N - \epsilon) = \frac{1}{2} + \epsilon = \frac{1+2\epsilon}{2}$$

depending on which of the two points has the smaller objective function value. Therefore, in the worst case, the reduction factor in the uncertainty range for the Fibonacci method is $\dfrac{1+2\epsilon}{F_{N+1}}$

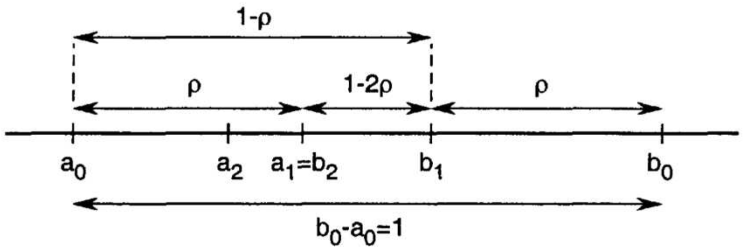
Example

- Consider the function $f(x) = x^4 - 14x^3 + 60x^2 - 70x$. Use the Fibonacci search method to find the value of x that minimizes f over the range [0,2]. Locate this value of x to within the range 0.3.
- After *N* steps the range is reduced by $(1 + 2\epsilon)/F_{N+1}$ in the worst case. We need to choose *N* such that

$$\frac{1 + 2\epsilon}{F_{N+1}} \leq \frac{\text{final range}}{\text{initial range}} = 0.3/2 = 0.15$$

- Thus, we need $F_{N+1} \geq \frac{1+2\epsilon}{0.15}$
- If we choose $\epsilon \leq 0.1$, then *N*=4 will do.

Example



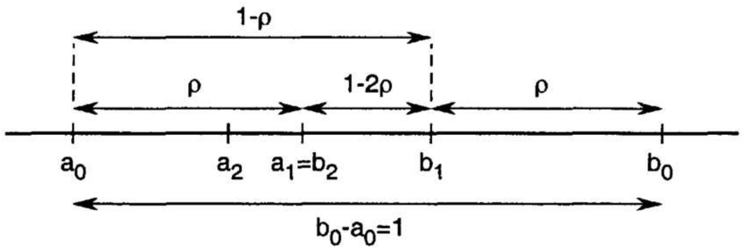
- Iteration 1. We start with

$$1-\rho_1 = \frac{F_4}{F_5} = \frac{5}{8}$$

  We then compute

$$a_1 = a_0 + \rho_1(b_0 - a_0) = \frac{3}{4}$$
$$b_1 = a_0 + (1-\rho_1)(b_0 - a_0) = \frac{5}{4}$$
$$f(a_1) = -24.34$$
$$f(b_1) = -18.65$$
$$f(a_1) < f(b_1)$$

- The range is reduced to $[a_0, b_1] = [0, \frac{5}{4}]$



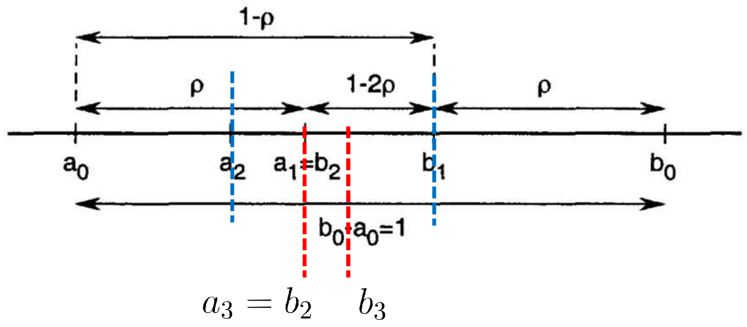
Example

- Iteration 2. We have

$$1 - \rho_2 = \frac{F_3}{F_4} = \frac{3}{5}$$

$$a_2 = a_0 + \rho_2(b_1 - a_0) = \frac{1}{2}$$
$$b_2 = a_1 = \frac{3}{4}$$
$$f(a_2) = -21.69$$
$$f(b_2) = f(a_1) = -24.34$$
$$f(a_2) > f(b_2)$$

so the range is reduced to $[a_2, b_1] = [\frac{1}{2}, \frac{5}{4}]$



Example

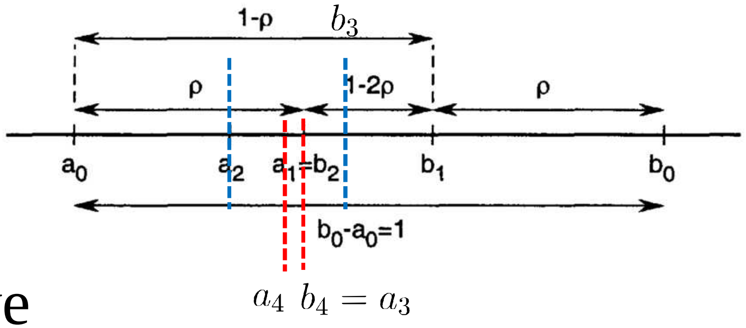
- Iteration 3. We compute

$$1 - \rho_3 = \frac{F_2}{F_3} = \frac{2}{3}$$

$$\begin{aligned} a_3 &= b_2 = \tfrac{3}{4} \\ b_3 &= a_2 + (1-\rho_3)(b_1 - a_2) = 1 \\ f(a_3) &= f(b_2) = -24.34 \\ f(b_3) &= -23 \\ f(a_3) &< f(b_3) \end{aligned}$$

The range is reduced to $[a_2, b_3] = [\frac{1}{2}, 1]$

Example



- Iteration 4. We choose $\epsilon = 0.05$. We have

$$1 - \rho_4 = \frac{F_1}{F_2} = \frac{1}{2}$$

$$a_4 = a_2 + (\rho_4 - \epsilon)(b_3 - a_2) = 0.725$$

$$b_4 = a_3 = \frac{3}{4}$$

$$f(a_4) = -24.27$$

$$f(b_4) = f(a_3) = -24.34$$

$$f(a_4) > f(b_4)$$

The range is reduced to $[a_4, b_3] = [0.725, 1]$

- Note that $b_3 - a_4 = 0.275 < 0.3$

Newton's Method

- In the problem of minimizing a function f of a single variable x
- Assume that at each measurement point $x^{(k)}$ we can calculate $f(x^{(k)})$, $f'(x^{(k)})$, and $f''(x^{(k)})$.
- We can fit a quadratic function through $x^{(k)}$ that matches its first and second derivatives with that of the function f.

$$q(x) = f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) + \frac{1}{2} f''(x^{(k)})(x - x^{(k)})^2$$

- Note that $q(x^{(k)}) = f(x^{(k)})$, $q'(x^{(k)}) = f'(x^{(k)})$, and $q''(x^{(k)}) = f''(x^{(k)})$
- Instead of minimizing f, we minimize its approximation q. The first order necessary condition for a minimizer of q yields

$$0 = q'(x) = f'(x^{(k)}) + f''(x^{(k)})(x - x^{(k)})$$

setting $x = x^{(k+1)}$, we obtain

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$$

Example

- Using Newton's method, find the minimizer of $f(x) = \frac{1}{2}x^2 - \sin x$
  The initial value is $x^{(0)} = 0.5$. The required accuracy is $\epsilon = 10^{-5}$
  in the sense that we stop when $|x^{(k+1)} - x^{(k)}| < \epsilon$
- We compute $f'(x) = x - \cos x \qquad f''(x) = 1 + \sin x$
- Hence,

$$x^{(1)} = 0.5 - \frac{0.5 - \cos 0.5}{1 + \sin 0.5} = 0.5 - \frac{-0.3775}{1.479} = 0.7552$$

- Proceeding in a similar manner, we obtain

$$x^{(2)} = x^{(1)} - \frac{f'(x^{(1)})}{f''(x^{(1)})} = x^{(1)} - \frac{0.02710}{1.685} = 0.7391$$

$$x^{(3)} = x^{(2)} - \frac{f'(x^{(2)})}{f''(x^{(2)})} = x^{(2)} - \frac{9.461 \times 10^{-5}}{1.673} = 0.7390$$

$$x^{(4)} = x^{(3)} - \frac{f'(x^{(3)})}{f''(x^{(3)})} = x^{(3)} - \frac{1.17 \times 10^{-9}}{1.673} = 0.7390 \qquad |x^{(4)} - x^{(3)}| < \epsilon = 10^{-5}$$

$$f'(x^{(4)}) = -8.6 \times 10^{-6} \approx 0 \qquad f''(x^{(4)}) = 1.763 > 0$$

We can assume that $x^* \approx x^{(4)}$ is a strict minimizer

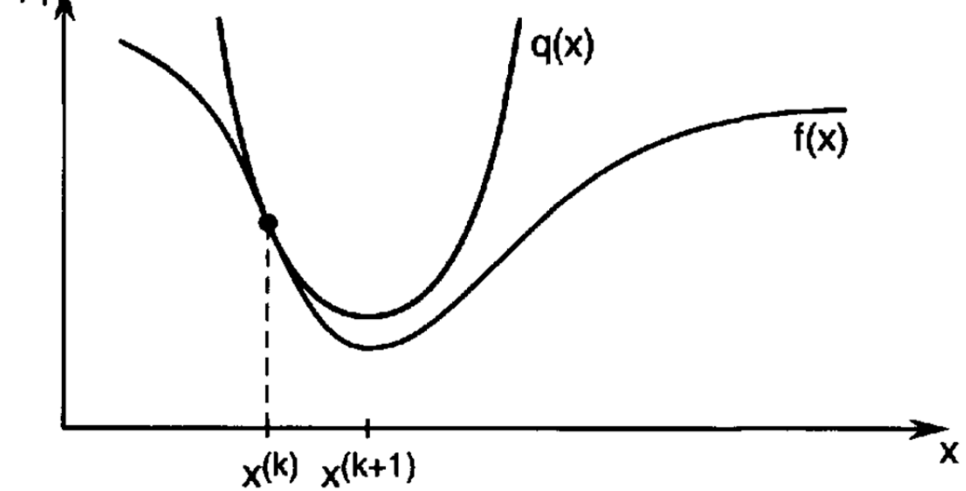
Corollary 6.1

Newton's Method

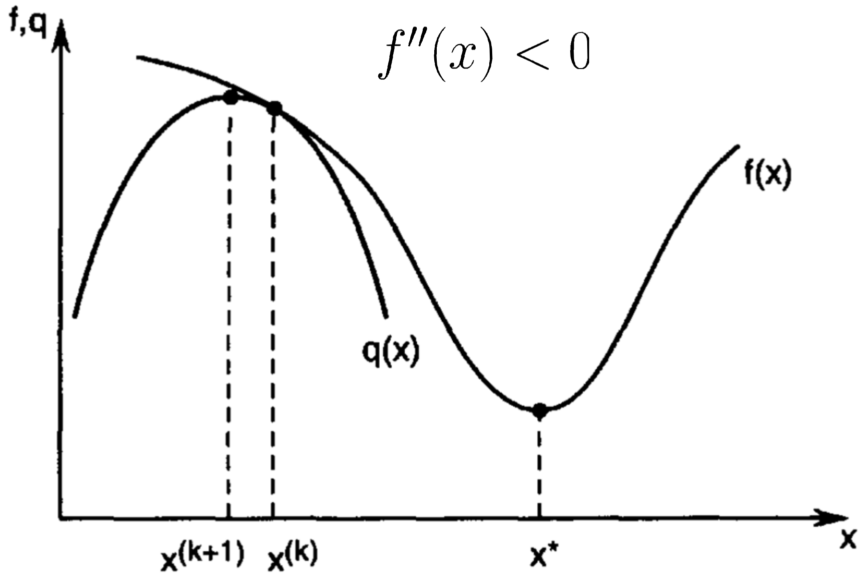
- Newton's method works well if $f''(x) > 0$ everywhere. However, if $f''(x) < 0$ for some x , Newton's method may fail to converge to the minimizer.
- Newton's method can also be viewed as a way to drive the first derivative of f to zero. If we set $g(x) = f'(x)$, then we obtain

$$x^{(k+1)} = x^{(k)} - \frac{g(x^{(k)})}{g'(x^{(k)})}$$

$f''(x) > 0$



$f''(x) < 0$



Example

- We apply Newton's method to improve a first approximation, $x^{(0)} = 12$, to the root of the equation $g(x) = x^3 - 12.2x^2 + 7.45x + 42 = 0$
- We have $g'(x) = 3x^2 - 24.4x + 7.45$
- Performing two iterations yields

$$x^{(1)} = 12 - \frac{102.6}{146.65} = 11.33$$

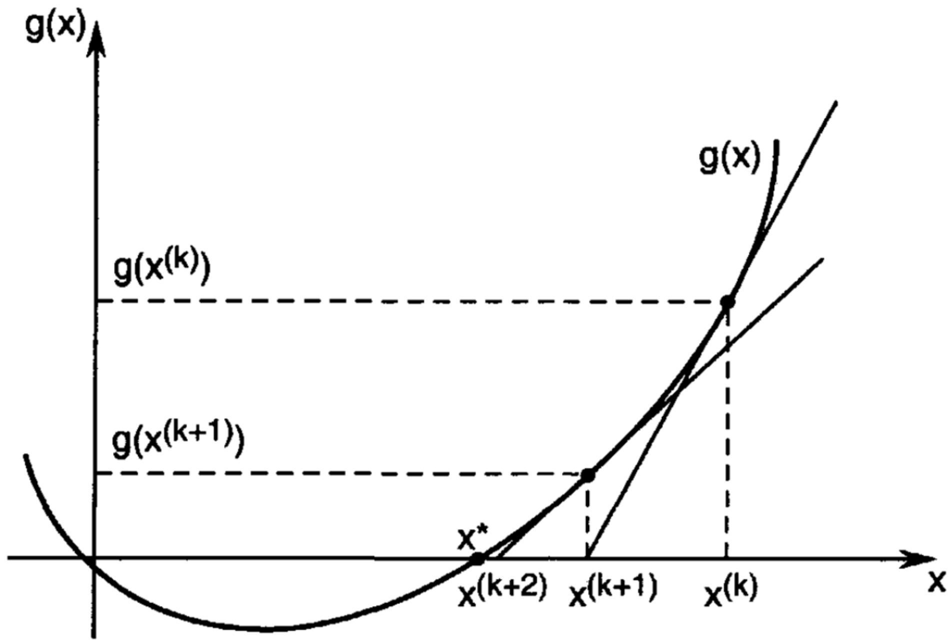
$$x^{(2)} = 11.33 - \frac{14.73}{116.11} = 11.21$$

Newton's Method

- Newton's method for solving equations of the form $g(x) = 0$ is also referred to as ***Newton's method of tangents***.
- If we draw a tangent to $g(x)$ at the given point $x^{(k)}$, then the tangent line intersects the *x*-axis at the point $x^{(k+1)}$, which we expect to be closer to the root x^* of $g(x) = 0$.
- Note that the slope of $g(x)$ at $x^{(k)}$ is

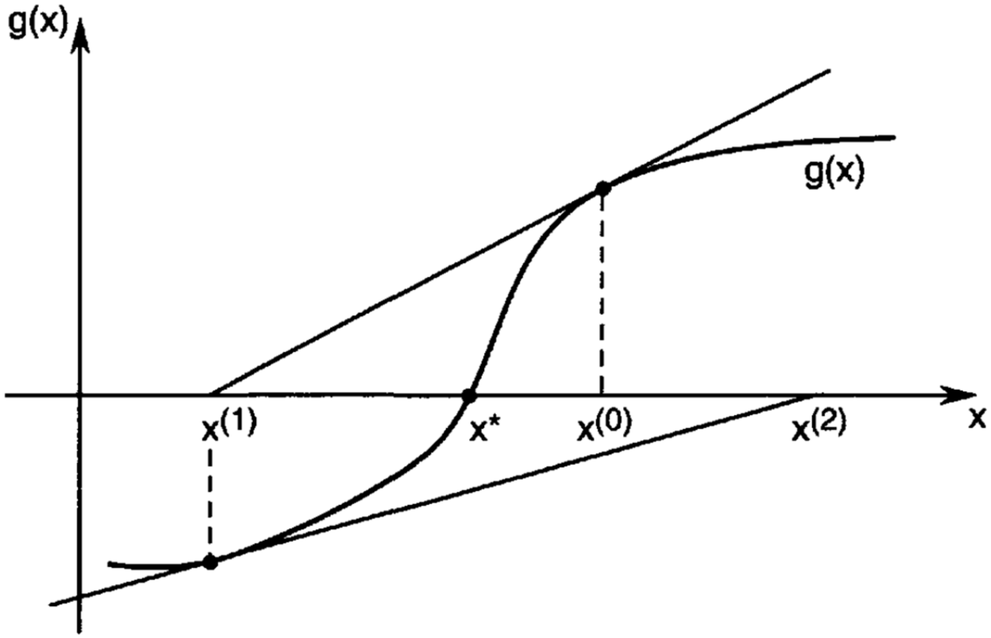
$$g'(x^{(k)}) = \frac{g(x^{(k)})}{x^{(k)} - x^{(k+1)}}$$

$$\Rightarrow \quad x^{(k+1)} = x^{(k)} - \frac{g(x^{(k)})}{g'(x^{(k)})}$$



Newton's Method

- Newton's method of tangents may fail if the first approximation to the root is such that the ratio $g(x^{(0)})/g'(x^{(0)})$ is not small enough.
- Thus, an initial approximation to the root is very important.



Secant Method

- Newton's method for minimizing f uses second derivatives of f

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$$

- If the second derivative is not available, we may attempt to approximate it using first derivative information. We may approximate $f''(x^{(k)})$ with

$$f''(x^{(k)}) = \frac{f'(x^{(k)}) - f'(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}$$

- Using the foregoing approximation of the second derivative , we obtain the algorithm

$$x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f'(x^{(k)}) - f'(x^{(k-1)})} f'(x^{(k)})$$

called the ***secant method***.

Secant Method

$$x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f'(x^{(k)}) - f'(x^{(k-1)})} f'(x^{(k)})$$

- Note that the algorithm requires two initial points to start it, which we denote $x^{(-1)}$ and $x^{(0)}$. The secant algorithm can be represented in the following equivalent form:

$$x^{(k+1)} = \frac{f'(x^{(k)})x^{(k-1)} - f'(x^{(k-1)})x^{(k)}}{f'(x^{(k)}) - f'(x^{(k-1)})}$$

- Like Newton's method, the secant method does not directly involve values of $f(x^{(k)})$. Instead, it tries to drive the derivative f' to zero.
- In fact, as we did for Newton's method, we can interpret the secant method as an algorithm for solving equations of the form $g(x) = 0$.

$$g(x) = f'(x)$$

Secant Method

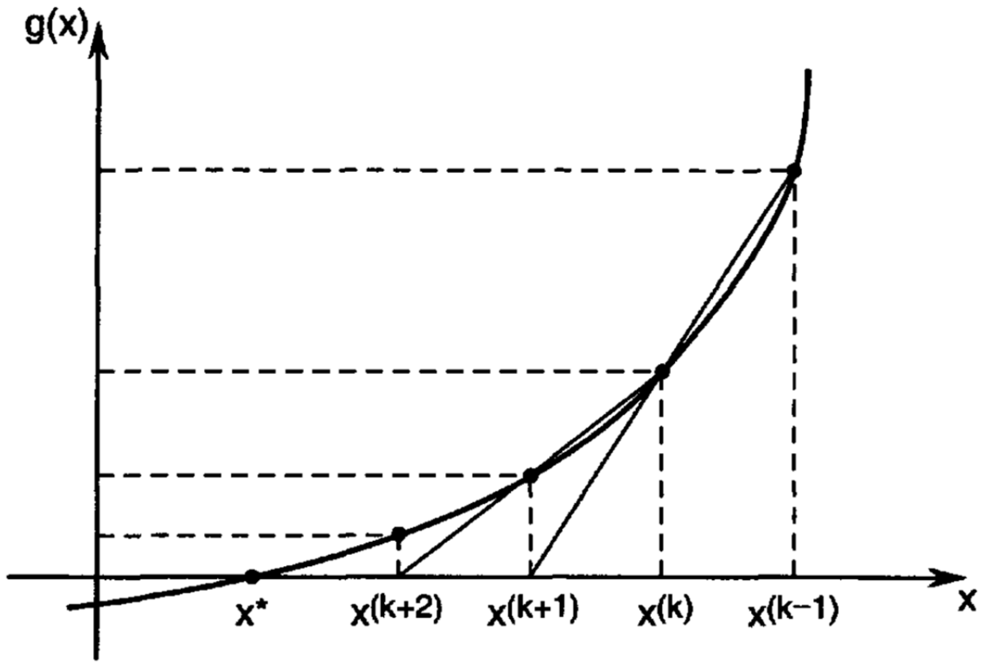
- The secant algorithm for finding a root of the equation $g(x) = 0$ takes the form

$$x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{g(x^{(k)}) - g(x^{(k-1)})} g(x^{(k)})$$

or equivalently,

$$x^{(k+1)} = \frac{g(x^{(k)})x^{(k-1)} - g(x^{(k-1)})x^{(k)}}{g(x^{(k)}) - g(x^{(k-1)})}$$

- In this figure, unlike Newton's method, the secant method uses the "secant" between the $(k-1)$th and kth points to determine the $(k+1)$th point.



Example

- We apply the secant method to find the root of the equation

$$g(x) = x^3 - 12.2x^2 + 7.45x + 42 = 0$$

- We perform two iterations, with starting points $x^{(-1)} = 13$ and $x^{(0)} = 12$. We obtain

$$x^{(1)} = 11.40$$
$$x^{(2)} = 11.25$$

Example

- Suppose that the voltage across a resistor in a circuit decays according to the model $V(t) = e^{-Rt}$, where $V(t)$ is the voltage at time t and R is the resistance value.
- Given measurements $V_1, ..., V_n$ of the voltage at times $t_1, ..., t_n$, respectively, we wish to find the best estimate of R. By the best estimate we mean the value of R that minimizes the total squared error between the measured voltages and the voltages predicted by the model.
- We derive an algorithm to find the best estimate of R using the secant method. The objective function is

$$f(R) = \sum_{i=1}^{n} (V_i - e^{-Rt_i})^2$$

Example

$$f(R) = \sum_{i=1}^{n} (V_i - e^{-Rt_i})^2$$

- Hence, we have

$$f'(R) = 2\sum_{i=1}^{n} (V_i - e^{-Rt_i}) e^{-Rt_i} t_i$$

- The secant algorithm for the problem is

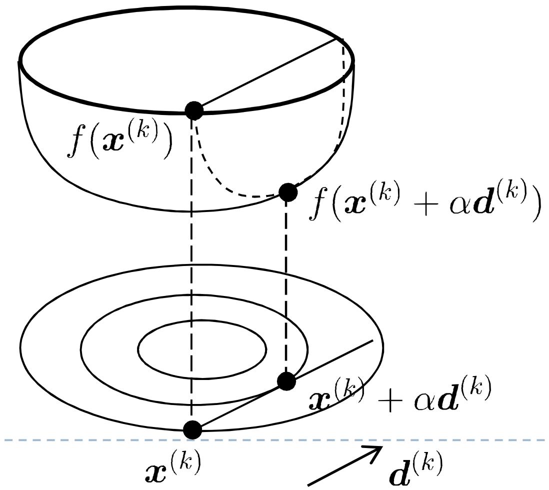
$$R_{k+1} = R_k - \frac{R_k - R_{k-1}}{\sum_{i=1}^{n} (V_i - e^{-R_k t_i}) e^{-R_k t_i} t_i - (V_i - e^{-R_{k-1} t_i}) e^{-R_{k-1} t_i} t_i}$$
$$\times \sum_{i=1}^{n} (V_i - e^{-R_k t_i}) e^{-R_k t_i} t_i$$

Remarks on Line Search Methods

- Iterative algorithms for solving such optimization problems involve a ***line search*** at every iteration.
- Let $f : R^n \to R$ be a function that we wish to minimize. Iterative algorithms for finding a minimizer of f are of the form

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}$$

where $\boldsymbol{x}^{(0)}$ is a given initial point and $\alpha_k \geq 0$ is chosen to minimized $\phi_k(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$. The vector $\boldsymbol{d}^{(k)}$ is called the ***search direction***.



The secant method

$$x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f'(x^{(k)}) - f'(x^{(k-1)})} f'(x^{(k)})$$

Remarks on Line Search Methods

- Note that choice of α_k involves a one-dimensional minimization. This choice ensures that under appropriate conditions, $f(\boldsymbol{x}^{(k+1)}) < f(\boldsymbol{x}^{(k)})$.
- We may, for example, use the secant method to find α_k. In this case, we need the derivative of ϕ_k

$$\phi_k'(\alpha) = \boldsymbol{d}^{(k)T} \bigtriangledown f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$$

- This is obtained by the chain rule. Therefore, applying the secant method for the line search requires the gradient $\bigtriangledown f$, the initial search point $\boldsymbol{x}^{(k)}$, and the search direction $\boldsymbol{d}^{(k)}$

Remarks on Line Search Methods

- Line search algorithms used in practice are much more involved than the one-dimensional search methods.
  - Determining the value of α_k that exactly minimizes ϕ_k may be computationally demanding; even worse, the minimizer of ϕ_k may not even exist.
  - Practical experience suggests that it is better to allocate more computation time on iterating the optimization algorithm rather than performing exact line searches.

Homework

- Exercises 6.8, 6.12, 6.16
- Exercises 7.2(d), 7.10(b)

- Hand over your homework at the class of Mar. 26.