Numerical Optimization

Instructor : Sung Chan Jun

Week #2 : September 06, 2023 (Wednesday Class)

Course Syllabus (Tentative)

Calendar	Description	Remarks
1st week	*Introduction of optimization*	
2nd week	Univariate Optimization	
3rd week	Univariate Optimization	
4th week	Unconstrained Optimization	
5th week	Unconstrained Optimization	
6th week	Constrained Optimization	
7th week	Constrained Optimization, No Class	Oct. 9 (National Holiday)
8th week	Constrained Optimization, Midterm	Oct. 18 (Midterm)

Recall – This Monday

- Optimality Condition : Unconstrained Univariate
  - (Generalization of optimal conditions) Assume objective univariate function $f(x)$ is at least <u>n times continuously differentiable</u>.

    Let $f'(x^*) = f''(x^*) = \ldots = f^{(n-1)}(x^*) = 0$ & $f^{(n)}(x^*) \neq 0$. Then
    - If $f^{(n)}(x^*) > 0$ and n even, x^* is a local minimum.
- Multivariate Calculus
  - Differentiation of function $z = f(x, y) : R^2 \rightarrow R^1$
    - Partial differentiation (in general)

$$\partial f(x, y)/\partial x = \lim_{h_x \rightarrow 0} [f(x + h_x, y) - f(x, y)]/h_x$$

$$\partial f(x, y)/\partial y = \lim_{h_y \rightarrow 0} [f(x, y + h_y) - f(x, y)]/h_y$$

Recall – This Monday

- Multivariate Calculus
  - Differentiation of vector valued function

$$F : R^2 \to R^2 \text{ defined by } F(x, y) = (f_1(x, y), f_2(x, y))$$

- Partial differentiation : $D_x F(x, y) = (\partial f_1(x, y)/\partial x \quad \partial f_2(x, y)/\partial x)^T$
- Derivative matrix DF

$$\lim_{\mathbf{h}=(h_x \ h_y)\to\mathbf{0}} \left\| F(x+h_x, y+h_y) - F(x,y) - DF(x,y)(h_x \ \ h_y)^T \right\| / \left\| (h_x \ \ h_y) \right\| = 0$$

- $DF(x, y) = \begin{pmatrix} \partial f_1/\partial x & \partial f_1/\partial y \\ \partial f_2/\partial x & \partial f_2/\partial y \end{pmatrix}$

Recall – This Monday

- Multivariate Calculus
  - Gradient (grad f, ∇f) : Let $f(\mathbf{x})$ be a scalar valued function $R^n \to R$.
    - $\nabla f(\mathbf{x}) = \nabla f(x_1, x_2, \ldots, x_n) = (\partial f/\partial x_1 \;\; \partial f/\partial x_2 \;\; \ldots \;\; \partial f/\partial x_n)^T$
    - Physical meaning : steepest increasing direction
  - Divergence (div F, $\nabla \cdot F$): Let $F = (f_1, f_2, \ldots, f_n) : R^n \to R^n$
    - $\nabla \cdot F = \partial f_1/\partial x_1 + \partial f_2/\partial x_2 + \ldots + \partial f_n/\partial x_n$
    - Physical meaning : rate of volume change per unit volume
  - 2nd derivative for multivariate $f(\mathbf{x}) = f(x_1, x_2)$: $R^2 \to R$
    - notation : $H(\mathbf{x}) = H(x_1, x_2)$ called 'Hessian'
    - definition : matrix $H(x_1, x_2)$ defined by $\begin{pmatrix} \partial^2 f / \partial x_1^2 & \partial^2 f / \partial x_1 \partial x_2 \\ \partial^2 f / \partial x_2 \partial x_1 & \partial^2 f / \partial x_2^2 \end{pmatrix}$

Recall – This Monday

- Optimality Conditions : Unconstrained Multivariate
  - Assume objective function f(**x**) is at least twice-continuously differentiable.

> • (**NC**) Necessary condition for a local minimum
>
> 1. grad(f(**x**)) = 0
> 2. H(**x**) ("hessian") ≥ 0 i.e H(**x**) is positive semi-definite.
>
> • (**SC**) Sufficient condition for a local minimum
>
> 1. grad(f(**x**)) = 0
> 2. H(**x**) > 0 i.e H(**x**) is positive definite.

- A symmetric real matrix **A** ($n \times n$)
  - is said to be positive definite if $\mathbf{z}^T \mathbf{A} \mathbf{z} > 0$ (strictly positive) for every non-zero vector **z** of real numbers.

Recall – This Monday

- Univariate Optimization

  Minimize $f(x)$ on $x \in R$

  - Conventional strategy

    Due to optimality conditions,

    - first seek points x with $f'(x) = 0$ (stationary points).
    - then check the sign of $f''(x)$ at those points.

  - How to find zero of $f'(x)$? $\Rightarrow$ root finding

    - Conventional techniques for root finding
      - Method of bisection, Newton's method
      - Secant method, Regula falsi method

Optimality conditions for univariate problem

- Necessary condition for a local minimum

  $f'(x^*) = 0$ & $f''(x^*) \geq 0$

- Sufficient condition for a local minimum

  $f'(x^*) = 0$ & $f''(x^*) > 0$

Univariate Optimization

Minimize $f(x)$ on $x \in R$

- Necessary condition for a local minimum

  $f'(x^*) = 0$ & $f''(x^*) \geq 0$
- Sufficient condition for a local minimum

  $f'(x^*) = 0$ & $f''(x^*) > 0$

- Conventional strategy

Due to optimality conditions,

- first seek points x with $f'(x) = 0$ (stationary points).
- then check the sign of $f''(x)$ at those points.

Univariate Optimization

How to find zero of f'(x)? $\Rightarrow$ root finding

- Conventional techniques for root finding
    - Method of bisection
    - Newton's method
    - Secant method, Regula falsi method

Univariate Optimization:
Root Finding - Method of Bisection

- Method of bisection

  - Interval [a, b] is given such that $f(a)f(b) < 0$.
  - Step 1. compute $f(x)$ at the midpoint $x = (a + b)/2$.
  - Step 2. if $f(x) = 0$ or $(b - a) < TOL$, then terminate.

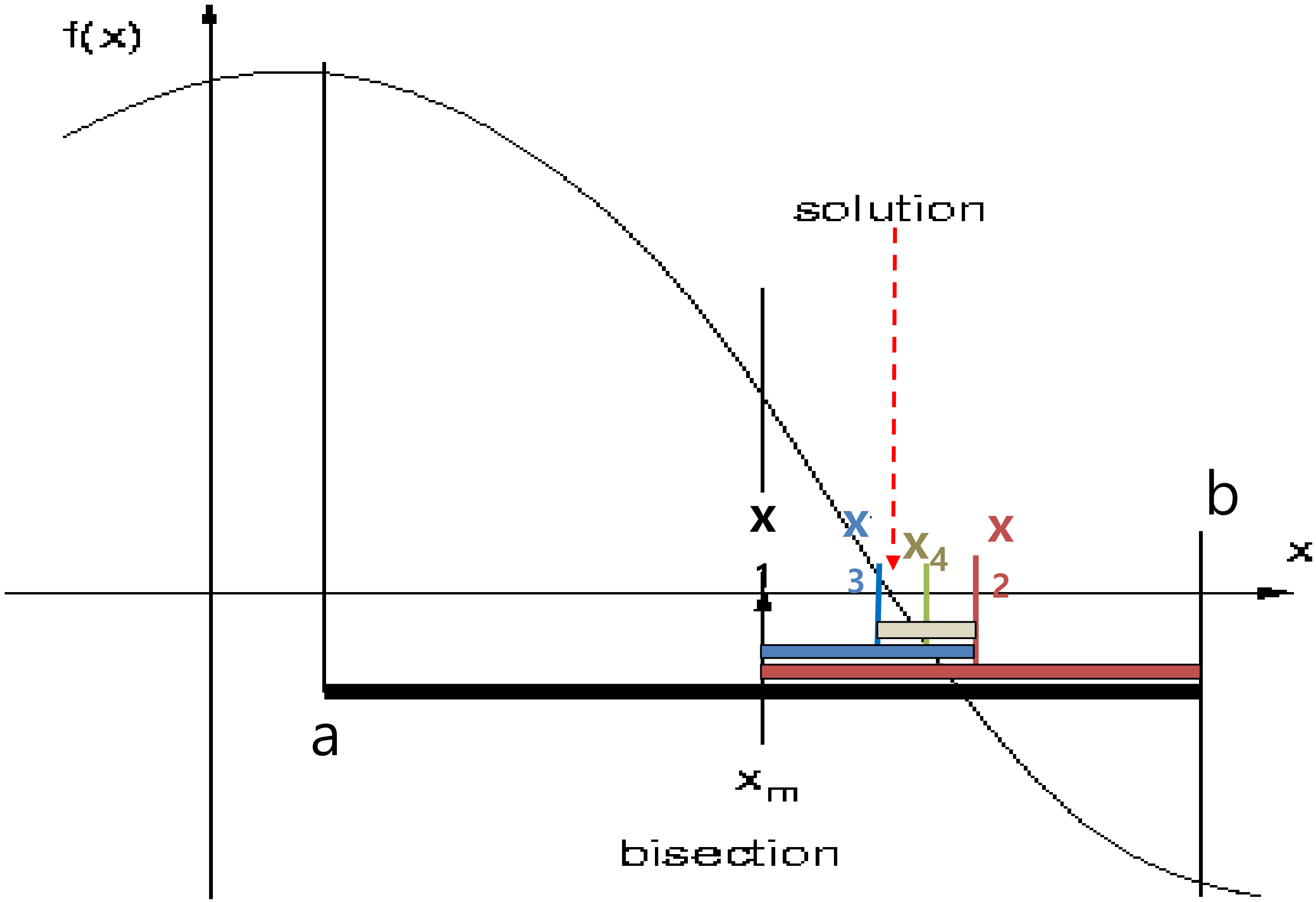
    if $f(x)f(a) < 0$, then $b := x$,

    else $a := x$.
  - Step 3. Go to Step 1.

To make sure the existence of root in an interval [a, b]

- Pros : guaranteed to converge to zero
- Cons
  - too slow (convergence rate ½)
  - relative magnitude of f(x) is not taken account.

Method of Bisection



Convergence Rate

- Assume sequence $\{x_k\}$ converges to x^*
  - The sequence $\{x_k\}$ converges with order r

    When there is a constant c > 0 and integer N such that

    $$\left\| x_{k+1} - x^* \right\| \le c \left\| x_k - x^* \right\|^r \text{ for } k > N, \quad \text{or} \quad 0 \le \lim_{k\to\infty} \frac{\left\| x_{k+1} - x^* \right\|}{\left\| x_k - x^* \right\|^r} < \infty \text{ for } k > N.$$

  - Convergence rate
    - r = 1, linear convergence
    - r = 2, quadratic convergence
    - r > 1, superlinear convergence
    - As r is bigger (in positive), we say that convergence speed is faster.

Univariate Optimization:
Root Finding - Method of Bisection

- Generalization of method of bisection

  - Interval [a, b] is given such that f(a)f(b) < 0.
  - Step 1. compute f(x) at the midpoint x = (a + b)/2.
  - Step 2. if f(x) = 0 or (b – a) < TOL, then terminate.

    if f(x)f(a) < 0, then b := x,

    else a: = x.
  - Step 3. Go to Step 1.

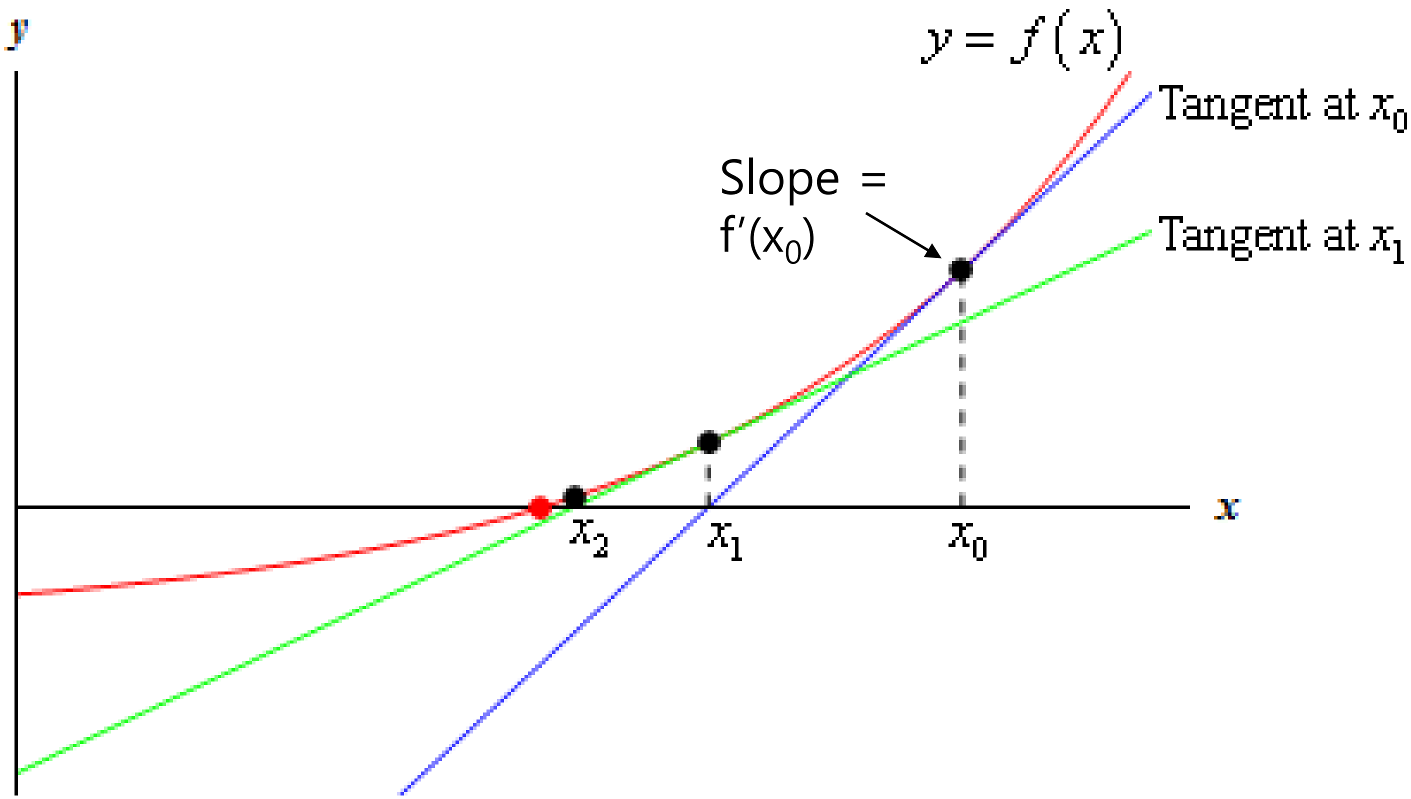
Midpoint x is one of choices. Any interior point In the interval [a, b] is possible to choose as a new point.

- Possible choices of next point in the interval [a, b]
  - midpoint – conventional method of bisection
  - internally dividing point of interval [a, b] AB in the ratio 1:2 or 2:1
  - internally dividing point of interval [a, b] AB in the ratio n:m
  - random choice

Univariate Optimization:
Newton's Method

- Newton's method
  - Approximate f(x) by tangent line at the given point.
  - Assume f(x) is differentiable.
  - $x_{k+1} = x_k - f(x_k)/f'(x_k)$
  - Pros
    - Very fast converging (convergence rate 2)
  - Cons
    - Convergence depending on initial guess
    - not working when $f'(x_k)$ is small
    - Derivative is required.

Newton's Method



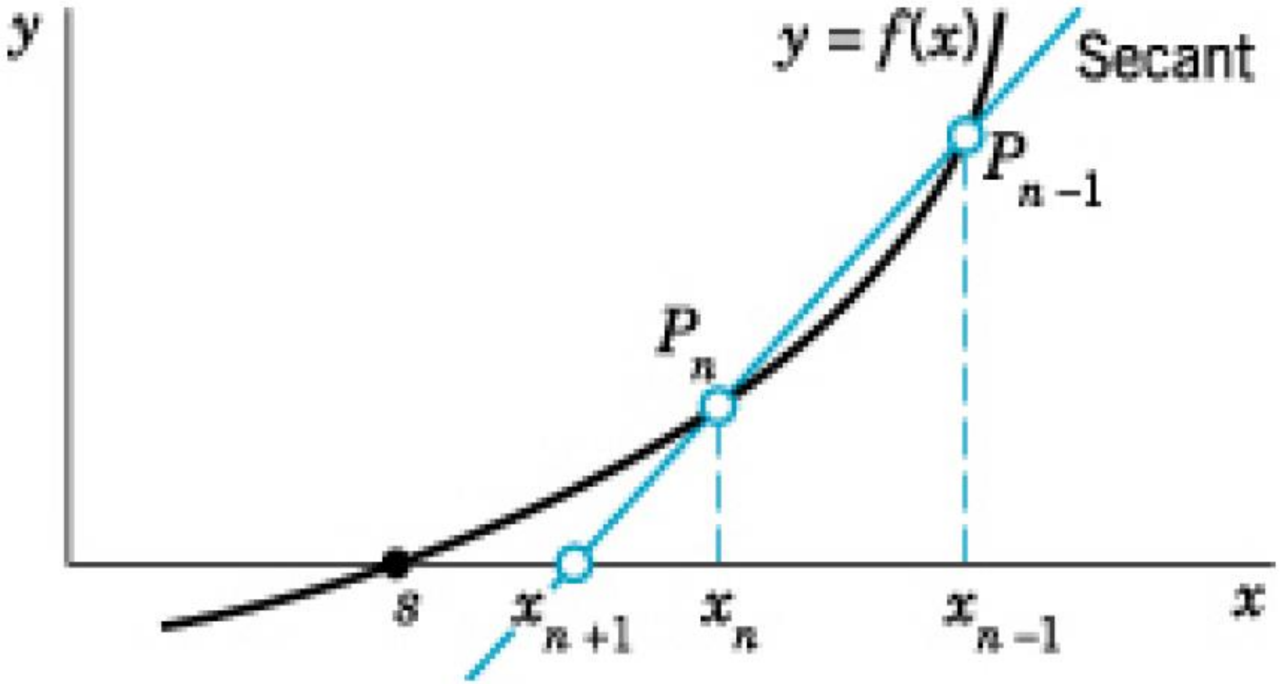
http://tutorial.math.lamar.edu/

Univariate Optimization:
Secant Method

- Secant method (method of linear interpolation)
  - Computing f'(x) is very expensive and impossible to compute in some cases.
  - Approximating tangent line by straight line passing two recent iterates (many variants exist)
  - $x_{k+1} = x_k - [(x_k - x_{k-1})/(f(x_k) - f(x_{k-1})]f(x_k)$
  - Pros : rapid convergent (roughly rate 1.6180)
  - Cons : may be divergent if straight line approximation is extrapolation

Newton's Method
$x_{k+1} = x_k - f(x_k)/f'(x_k)$

Secant Method



http://ocw.snu.ac.kr

Secant Method

- How to approximate tangent line
  - There are many ways to do it
  - Two point approximation
    - $f'(x_n) \approx (f(x_n) - f(x_{n-1}))/(x_n - x_{n-1})$
    - $f'(x_n) \approx (f(x_n) - f(x_{n-2}))/(x_n - x_{n-2})$
  - Three point approximation
    - $f'(x_n) \approx \alpha(f(x_n) - f(x_{n-1}))/(x_n - x_{n-1}) + (1-\alpha)(f(x_n) - f(x_{n-2}))/(x_n - x_{n-2})$

Secant Method

- How to approximate tangent line (Richardson's extrapolation)

  Assumption : all points on axis are even spaced, that is, h is fixed

  - 3-point approximation
    - Forward difference $$f'(t_i) \approx \frac{-f(t_i+2h)+4f(t_i+h)-3f(t_i)}{2h}$$
    - Backward difference $$f'(t_i) \approx \frac{3f(t_i)-4f(t_i-h)+f(t_i-2h)}{2h}$$
  - 4-point approximation (Central difference)

$$f'(t_i) \approx \frac{-f(t_i+2h)+8f(t_i+h)-8f(t_i-h)+f(t_i-2h)}{12h}$$

  - 6-point approximation (Central difference)

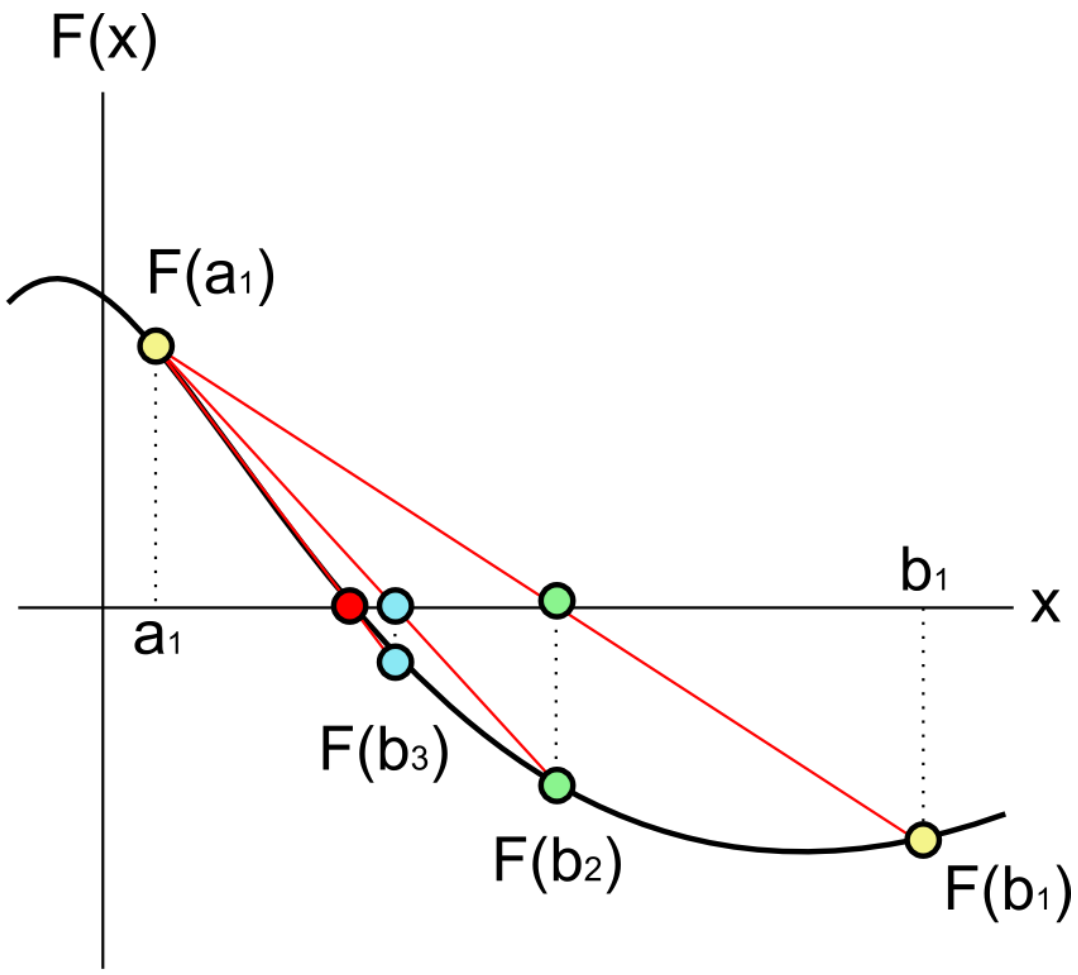
$$f'(t_i) \approx \frac{f(t_i+3h)-9f(t_i+2h)+45f(t_i+h)-45f(t_i-h)+9f(t_i-2h)-f(t_i-3h)}{60h}$$

R. DE Levie, J. Chem. Sci., 2009

Univariate Optimization:
Regular Falsi Method

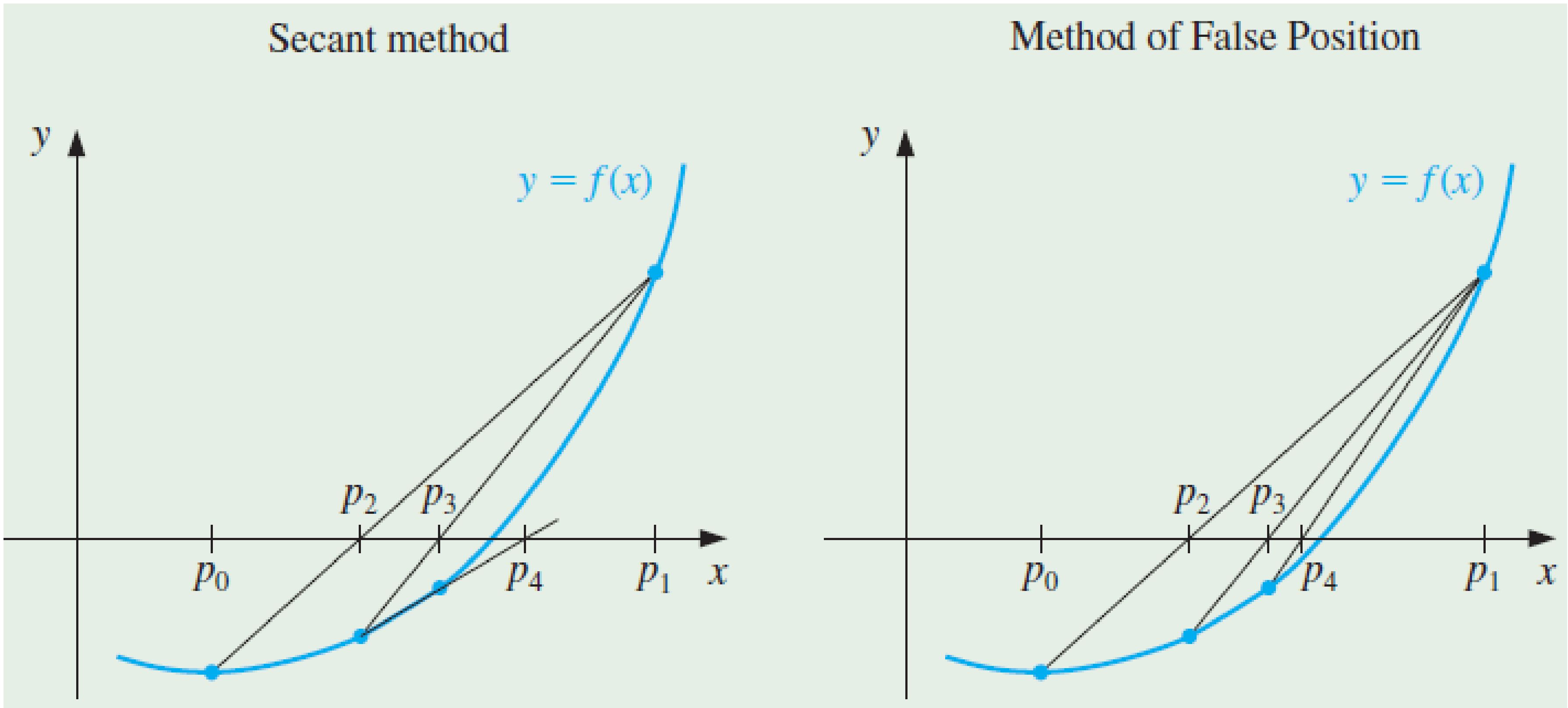
- Regular falsi method (method of false position)
  - Modified version of secant method & bisection method
  - Consider the given interval I_k = [a, b] such that f(a)f(b) < 0.
  - Apply a secant method with two initial points a & b. Find a point x_{k+1} intersecting with x-axis and a secant.
  - Choose updated interval as follows:
    - $I_{k+1} = [a, x_{k+1}]$, if f(a) and $f(x_{k+1})$ have different signs,
    - $I_{k+1} = [x_{k+1}, b]$, if f(a) and $f(x_{k+1})$ have the same signs.
    - This removes danger of extrapolation.
  - Keep doing in the same manner until termination criterion is satisfied.

Regular Falsi Method



https://commons.wikimedia.org/

Regula Falsi Method (False Position)



http://www.uobabylon.edu.iq

Univariate Optimization:
Root-finding Methods

- Bracketing methods
  - General approaches choosing any nested interval of the previous interval.
    - Given interval $\mathbf{I}_0$ such that $x \in \mathbf{I}_0$ where $f(x) = 0$.
    - Find $\{\mathbf{I}_j\}$ such that $\mathbf{I}_j \subset \mathbf{I}_{j-1}$ and $x \in \mathbf{I}_j$. (make sure that length of interval $\mathbf{I}_j$ should be sufficiently reduced)
    - It generates a set of nested intervals, which is guaranteed to converge.
  - Example : the method of bisection.

Univariate Optimization:
Root-finding Methods

- Root finding techniques
    - Method of bisection : bracketing, that is, interval is used.
    - Newton's method : straight line is used.
    - Secant method : straight line is used.
    - Regula falsi method : straight line and bracketing are used.
- Why are straight lines mainly used in root finding techniques?
    - Straight line (1^{st} order polynomial) is the simplest shape in approximation.
    - Finding root (intersecting point with x-axis) of straight line is very easy.

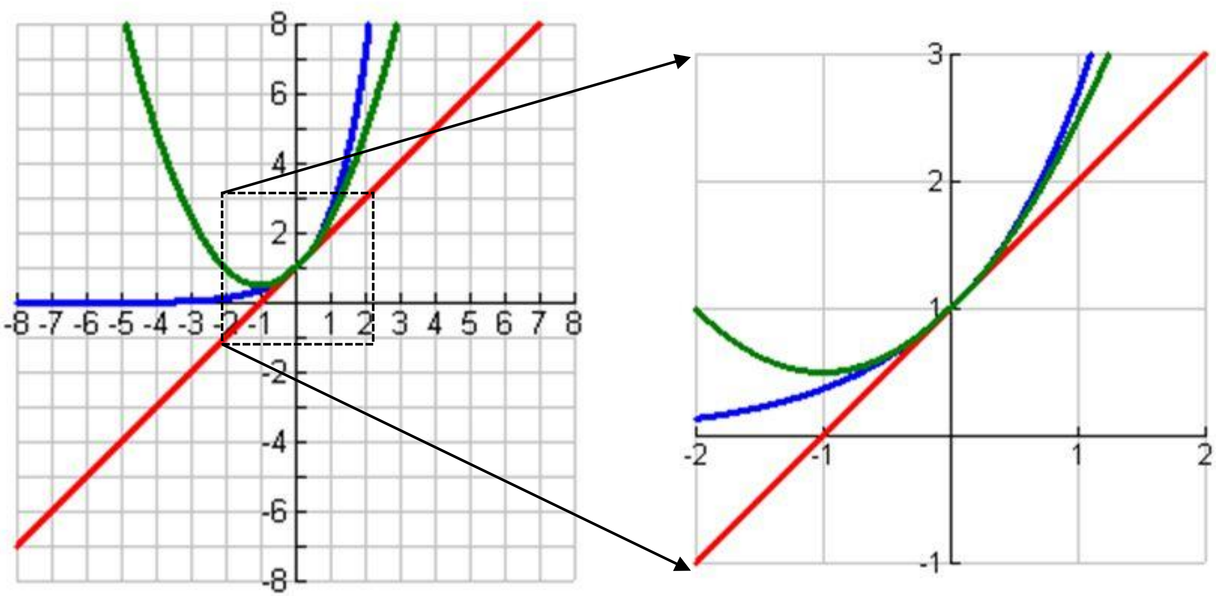
Univariate Optimization:
Root-finding Methods

$$f(x) = e^x$$

$$L(x) = 1 + x$$

$$Q(x) = 1 + x + \frac{1}{2}x^2$$

Q: In terms of approximation sense, which one is better, a straight line (1^{st} order polynomial) or quadratic function (2^{nd} order polynomial)?



http://slideplayer.com/slide/7717334/

Univariate Optimization:
Root-finding Methods

- How about other approaches in place of a straight line?
  - Straight line (1st order polynomial) is the simplest shape in approximation, so more complex shape may be applicable in the same context.
  - Curve (2nd order or higher order polynomial) may be possible.
- More advanced root-finding approaches
  - Higher order polynomial or other function approximation
    - In place of a straight line, higher order polynomials (quadratic, cubic…) or other functions are possible to approximate original function f(x).

Univariate Optimization: Root-finding Methods

- More advanced root-finding approaches
  - Higher order polynomial approximation
    - Higher order polynomials (quadratic, cubic…) are used for approximation of original function f(x).
      - That would be much rapidly convergent.
      - One problem for higher polynomial approximation is to seek the zero point of it, which may be more difficult than a straight line.
      - Up to 4-th order polynomials, it is possible to find roots with the given root formulations.
  - Other function approximation than polynomial
    - Key points to consider
      - higher order approximation + easy to find a zero point

Root Formulation of Polynomials

- 1^{st} order polynomial : $y = ax + b$ (when $a \neq 0$)
  - $x = -b/a$
- 2^{nd} order polynomial : $y = ax^2 + bx + c$ (when $a \neq 0$)
  - $x = [-b \pm \sqrt{(b^2 - 4ac)}]/2a$
- 3^{rd} order polynomial : $y = ax^3 + bx^2 + cx + d$ (when $a \neq 0$)
  - $$\begin{aligned} x_1 = &-\frac{b}{3a} \\ &-\frac{1}{3a}\sqrt[3]{\frac{2b^3-9abc+27a^2d+\sqrt{(2b^3-9abc+27a^2d)^2-4(b^2-3ac)^3}}{2}} \\ &-\frac{1}{3a}\sqrt[3]{\frac{2b^3-9abc+27a^2d-\sqrt{(2b^3-9abc+27a^2d)^2-4(b^2-3ac)^3}}{2}} \\ x_2 = &-\frac{b}{3a} \\ &+\frac{1+i\sqrt{3}}{6a}\sqrt[3]{\frac{2b^3-9abc+27a^2d+\sqrt{(2b^3-9abc+27a^2d)^2-4(b^2-3ac)^3}}{2}} \\ &+\frac{1-i\sqrt{3}}{6a}\sqrt[3]{\frac{2b^3-9abc+27a^2d-\sqrt{(2b^3-9abc+27a^2d)^2-4(b^2-3ac)^3}}{2}} \\ x_3 = &-\frac{b}{3a} \\ &+\frac{1-i\sqrt{3}}{6a}\sqrt[3]{\frac{2b^3-9abc+27a^2d+\sqrt{(2b^3-9abc+27a^2d)^2-4(b^2-3ac)^3}}{2}} \\ &+\frac{1+i\sqrt{3}}{6a}\sqrt[3]{\frac{2b^3-9abc+27a^2d-\sqrt{(2b^3-9abc+27a^2d)^2-4(b^2-3ac)^3}}{2}} \end{aligned}$$

http://wiki.mathnt.net

Root Formulation of Polynomials

- 4^{th} order polynomial : $y = ax^4 + bx^3 + cx^2 + dx + e$ (when $a \neq 0$)
  - $x = ?$
  - For its detail, refer to https://en.wikipedia.org/wiki/Quartic_function#General_formula_for_roots
- 5^{th} or higher order polynomials
  - No general root formulation exists.
    - It was proved by Abel (1802–1829).

Univariate Optimization:
Root Finding Methods

- Additional ideas
  - Rational function approximation (rational interpolation)
    - Approximate f(x) by rational function of the form

$$f_{rat}(x) = \frac{x - c}{d_0 + d_1 x + d_2 x^2}$$

    - d_0, d_1, d_2, c are chosen so that the function value and derivatives of $f_{rat}(x)$ agree with those of f(x) at two points.
    - This approximation is easy to find zero point, which is just ‘c’.
  - Other function approximations
    - We can generate any kinds of approximations, which is better approximated and root is easy to find.

Univariate Optimization:
Root Finding Methods

- Safeguarded methods
  - A guaranteed and reliable method : the method of bisection
  - A fast-convergent, but less reliable method : secant method
  - Mixed methods : bisection + secant
    - If f is well-behaved, it gives the rapid convergence (secant). In the worst case, it is not less efficient than the guaranteed method (bisection).