Numerical Optimization

Instructor : Sung Chan Jun

Week #2 : September 04, 2023 (Monday 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 – Last Week

- Optimization
  - Optimization problem is to find the best solution under the various constraints such that a given cost function is optimized (minimized or maximized).
  - Applications
    - data analysis and model fitting, structural design problems
    - planning, scheduling, computer tomography
    - too many others …

Recall – Last Week

- Optimization-Categories
  - Based on number of variables : Univariate, Multivariate
  - Based on constraints : Unconstrained, Constrained
  - Based on numerical techniques : Local, Global
  - Based on linearity：Linear, Nonlinear
- Mathematical modeling
  - Problem formulation

Define the optimization problem and model parameters

Define the decision variables

Define the objective (cost function)

Define the constraints

Recall – Last Week

- Optimization Formulation
  - General Formulation
    - Decision variables (design variables) : $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$
    - minimize $f(\mathbf{x})$
      - subject to $c_i(\mathbf{x}) = 0,\ i = 1,\dots,m;\ \ c_i(\mathbf{x}) \geq 0,\ \ i = m+1,\dots,p$
    - Feasibility of $\mathbf{x}$
      - $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ is said to be feasible or a feasible solution, if it satisfies all the constraints in the above optimization problem.
    - Feasible sets
      - Discrete Optimization Problem : Feasible set is discrete
      - Optimization control theory, calculus of variations : Feasible set is a function space (usually infinite dimension)
  - Maximization and minimization : f is maximized ⇔ -f is minimized

Recall – Last Week

- Minimum
  - The point x^* is a strong local minimum of $F(x)$ if there exists $\delta > 0$ such that
    - $F(x)$ is defined on $Nbh(x^*, \delta)$, and $F(x^*) < F(y)$ for all $y \in Nbh(x^*, \delta)$, $y \neq x^*$.
  - The point x^* is a weak local minimum of $F(x)$ if there exists $\delta > 0$ such that
    - $F(x)$ is defined on $Nbh(x^*, \delta)$, and $F(x^*) \leq F(y)$ for all $y \in Nbh(x^*, \delta)$.
  - The point x^* is a global minimum of $F(x)$ if $F(x^*)$ is the least value.

Recall – Last Week

- Optimality Condition : Unconstrained Univariate
  - Let univariate function f(x) be at least twice-continuously differentiable.
  - x* be a local minimum of f(x). What would happen around x*?

<u>Taylor expansion of f(x) at x*</u>

There exists a scalar t ($0 \le t \le 1$) such that $f(x^* + \varepsilon) = f(x^*) + \varepsilon f'(x^*) + \frac{1}{2}\varepsilon^2 f''(x + t\varepsilon)$.

<u>Continuity of f' and f'' at x*</u>

When we assume $f'(x^*) < 0$, there exists positive ε^* such that

$$\varepsilon f'(x^*) + \tfrac{1}{2}\varepsilon^2 f''(x^* + t\varepsilon) < 0 \text{ for all } 0 < \varepsilon \le \varepsilon^*.$$

<u>$f(x^* + \varepsilon) < f(x^*)$ for all such ε</u>

It contradicts the optimality (local minimum) of x*. So, it should be $f'(x^*) \ge 0$.

Recall – Last Week

- (Continued) x^* be a local minimum of $f(x)$. What would happen around x^*?

Continuity of f' and f'' at x*

When we assume $f'(x^*) > 0$, there exists negative ε^ such that

$\varepsilon f'(x^*) + ½\varepsilon 2 f''(x^* + t\varepsilon) < 0$ for all ε^ $\le \varepsilon < 0$.

Then $f(x^* + \varepsilon) < f(x^*)$ for all such ε. It contradicts local minimum of x^*.

Finally, it should be $f'(x^*) = 0$ ('First order optimality')

Taylor Expansion of f(x) at x* by first order optimality

$f(x^* + \varepsilon) = f(x^*) + ½\varepsilon 2 f''(x^* + t\varepsilon)$ for some $0 \le t \le 1$.

Continuity of f'' at x*

Assuming $f''(x^*) < 0$, $f''(x) < 0$ in some Nbd(x^*), that is, $f''(x^* + t\varepsilon) < 0$ for sufficiently small ε. So, $f(x^* + \varepsilon) < f(x^*)$ in that Nbd(x^*).

This contradicts local minimum of x^*.

So, it should be $f''(x^*) \ge 0$. ('second order optimality')

Recall – Last Week

- Optimality Condition : Unconstrained Univariate

  Let objective univariate function f(x) be at least twice-continuously differentiable.

  - When do we make sure that x^* is a local minimum of f(x)?
    - How about $f'(x^*) = 0$ and $f''(x^*) > 0$?

Taylor Expansion of f(x) at x^*

$$f(x^* + \varepsilon) = f(x^*) + \varepsilon f'(x^*) + \tfrac{1}{2}\varepsilon^2 f''(x^* + t\varepsilon) \text{ (due to } f'(x^*) = 0)$$

$$= f(x^*) + \tfrac{1}{2}\varepsilon^2 f''(x^* + t\varepsilon) \text{ for some } 0 \le t \le 1.$$

Continuity of f'' at x^*

There exists $\varepsilon_0 > 0$ such that $f''(x^* + t\varepsilon) > 0$, $\forall\ 0 < |\varepsilon| \le \varepsilon_0$.

Then $f(x^* + \varepsilon) > f(x^*)$ $\forall\ 0 < |\varepsilon| \le \varepsilon_0$. So, x^* is a local minimum.

Recall – Last Week

- Summary of Optimality Conditions in Univariate Function
  - Assume objective univariate function f(x) is at least twice-continuously differentiable.

(**NC**) Necessary condition for a local minimum

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

i.e. x^* is a local minimum $\Rightarrow f'(x^*) = 0$ & $f''(x^*) \geq 0$

(**SC**) Sufficient condition for a local minimum

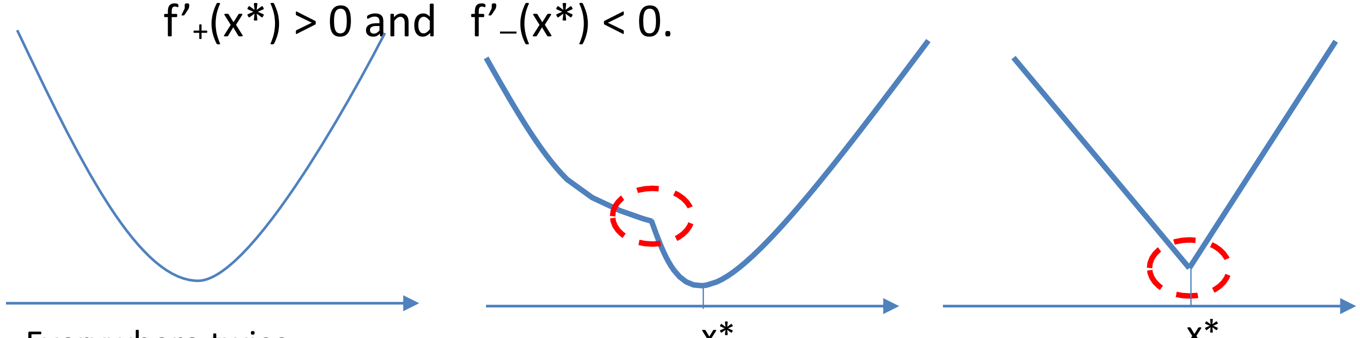
$f'(x^*) = 0$ & $f''(x^*) > 0$

i.e. $f'(x^*) = 0$ & $f''(x^*) > 0 \Rightarrow x^*$ is a local minimum

Recall – Last Week

- Optimality Condition : Unconstrained Univariate
  - When f(x*) or f'(x*) is not continuous
    - When x* is not a point of discontinuity and f(x) is twice-continuously differentiable in the neighborhood of x*
      - The previous optimal condition is OK.
    - When f(x) is continuous at x*, but x* is a point of discontinuity in f'(x)
      - Sufficient conditions for x* to be a strong local minimum are that

$$f'_{+}(x^*) > 0 \text{ and } f'_{-}(x^*) < 0.$$



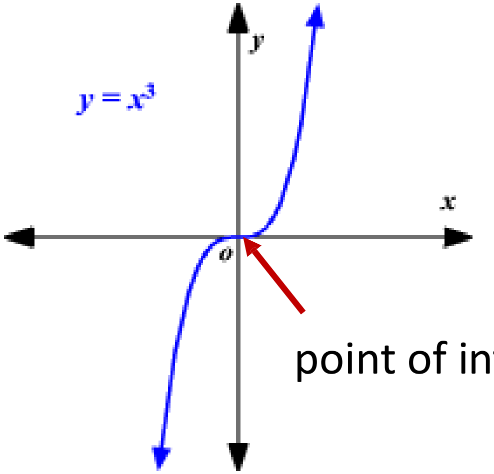
Everywhere twice differentiable

f' is discontinuous in R, but x* is not at a point of discontinuity.

x* is at a point of d iscontinuity of f'.

Optimality Condition : Unconstrained Univariate

- Question?
  - Does there exist a stationary point x* (when f'(x*) = 0) that is neither a minimum nor a maximum?
    - Yes! For example, f(x) = x3 or f(x) = x5 or f(x) = xn (n is odd integer) around 0.



point of inflection

https://www.varsitytutors.com/

- Point of inflection
  - When stationary point x* is neither a minimum nor a maximum

Optimality Condition :
Unconstrained Univariate

(Generalization of optimal conditions)

Assume objective univariate function f(x) is at least n times continuously differentiable.

- 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.
  - What happen when n is odd?

Multivariate Calculus

- Calculus of functions depending on two or more variables
  - Basic concept is the same to the one with one variable, but it is more tricky.
- Differentiation of function $z = f(x, y) : R^2 \rightarrow R^1$
  - partial differentiation (in general)
    - notation : $\partial f(x, y)/\partial x$, $\partial f(x, y)/\partial y$
    - definition : $\partial f(x, y)/\partial x = \lim_{h_x \rightarrow 0} [f(x + h_x, y) - f(x, y)]/h_x$

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

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
    - DF satisfies

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x)$$
$$\Leftrightarrow \lim_{h \to 0} \frac{f(x+h) - f(x) - f'(x)h}{h} = 0$$

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

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

Multivariate Calculus

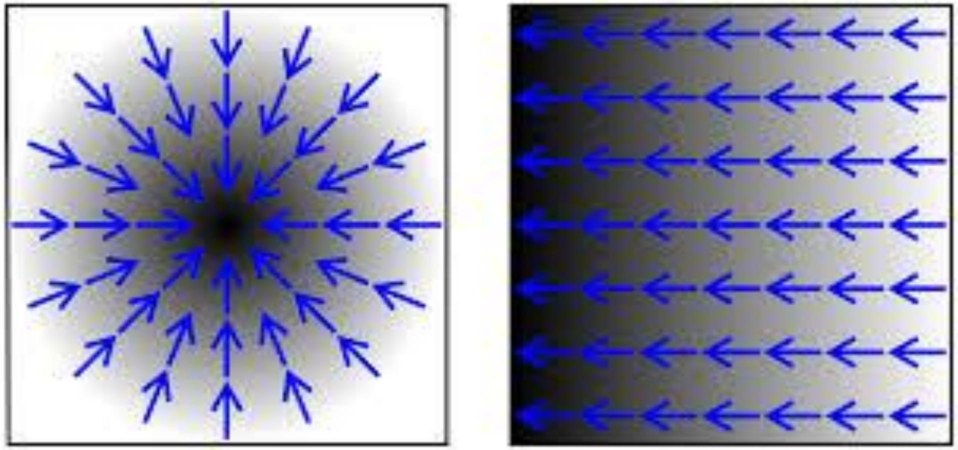
Let $f(\mathbf{x})$ be a scalar valued function $R^n \to R$.

- Gradient (grad f, ∇f)
  - $\nabla f(\mathbf{x}) = \nabla f(x_1, x_2, \ldots, x_n) = (\partial f/\partial x_1 \;\; \partial f/\partial x_2 \;\; \ldots \;\; \partial f/\partial x_n)^T$
  - Physical meaning : steepest increasing direction

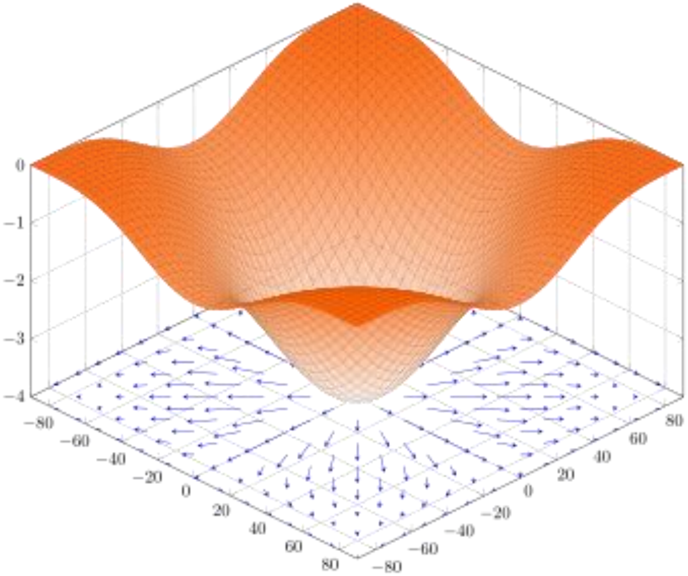
Let F be a vector valued function $R^n \to R^n$. F is called vector field.

- Divergence (div F, $\nabla \cdot F$): Let $F = (f_1, f_2, \ldots, f_n)$
  - $\nabla \cdot F = \partial f_1/\partial x_1 + \partial f_2/\partial x_2 + \ldots + \partial f_n/\partial x_n$
  - Physical meaning : rate of volume change per unit volume

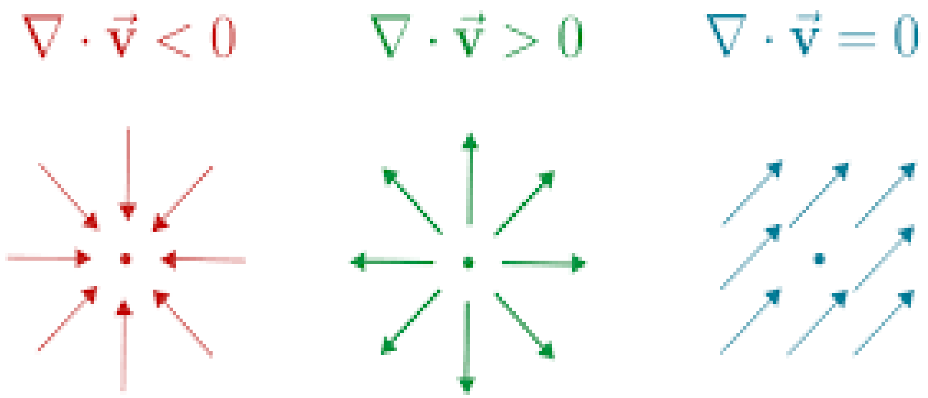
Gradient & Divergence



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



$f(x, y) = -(\cos^2 x + \cos^2 y)^2$



https://www.khanacademy.org

The divergence of a vector field at a given point measures how much it is flowing out of, or into, that point.

Multivariate Calculus

- 2^{nd} derivative for univariate f(x)
  - notation : f'' or $\partial^2 f(x)/\partial x^2$

- 2^{nd} derivative for multivariate $f(\mathbf{x}) = f(x_1, x_2): R^2 \rightarrow R$
  - notation : $H(\mathbf{x}) = H(x_1, x_2)$ called 'Hessian'
  - definition : matrix $H(x_1, x_2)$ defined by $\begin{pmatrix} \partial^2 f / \partial x_1^2 & \partial^2 f / \partial x_1 \partial x_2 \\ \partial^2 f / \partial x_2 \partial x_1 & \partial^2 f / \partial x_2^2 \end{pmatrix}$

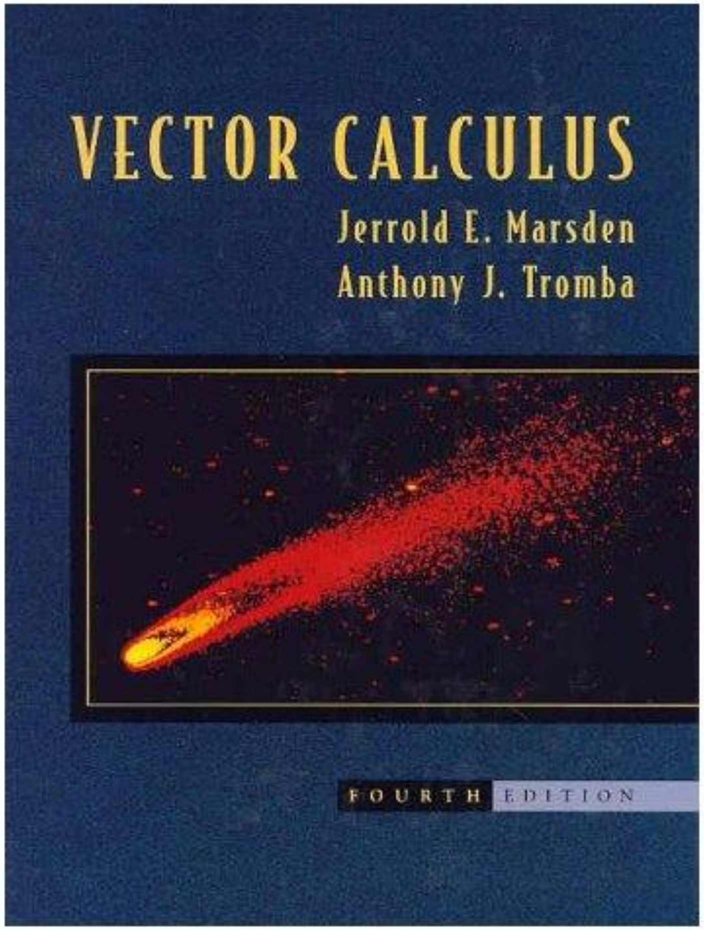
Multivariate Calculus

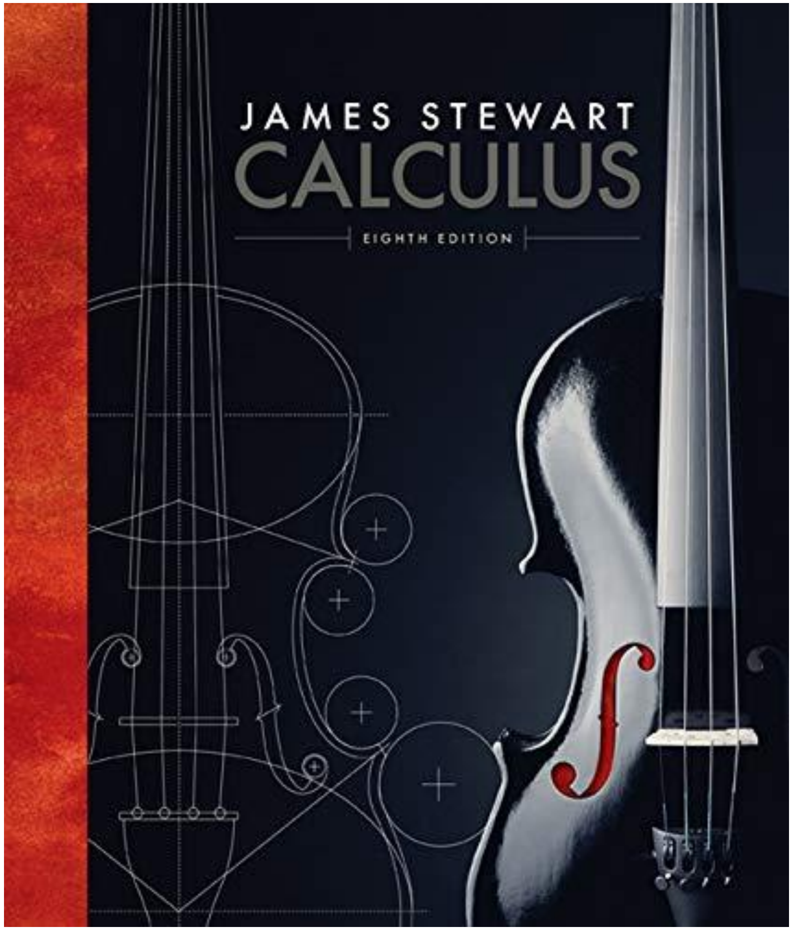
- How about differentiation of $F(\mathbf{x}) = F(x_1, x_2, \ldots, x_n) : R^n \rightarrow R^m$?

  and derivative DF? or Hessian?

  - Please refer to calculus book.

Multivariate Calculus





Optimality Conditions
Unconstrained Multivariate

- Assume objective function f(**x**) is at least twice-continuously differentiable.

Let objective univariate function f(x) be at least twice-continuously differentiable.

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

- (**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.

Positive Definiteness of Matrix

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

$$M = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

is positive definite since for any non-zero column vector z with entries a, b and c, we have

$$\begin{aligned} z^{\mathsf{T}} M z = (z^{\mathsf{T}} M)\, z &= \begin{bmatrix} (2a-b) & (-a+2b-c) & (-b+2c) \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \\ &= (2a-b)a + (-a+2b-c)b + (-b+2c)c \\ &= 2a^2 - ba - ab + 2b^2 - cb - bc + 2c^2 \\ &= 2a^2 - 2ab + 2b^2 - 2bc + 2c^2 \\ &= a^2 + a^2 - 2ab + b^2 + b^2 - 2bc + c^2 + c^2 \\ &= a^2 + (a-b)^2 + (b-c)^2 + c^2 \end{aligned}$$

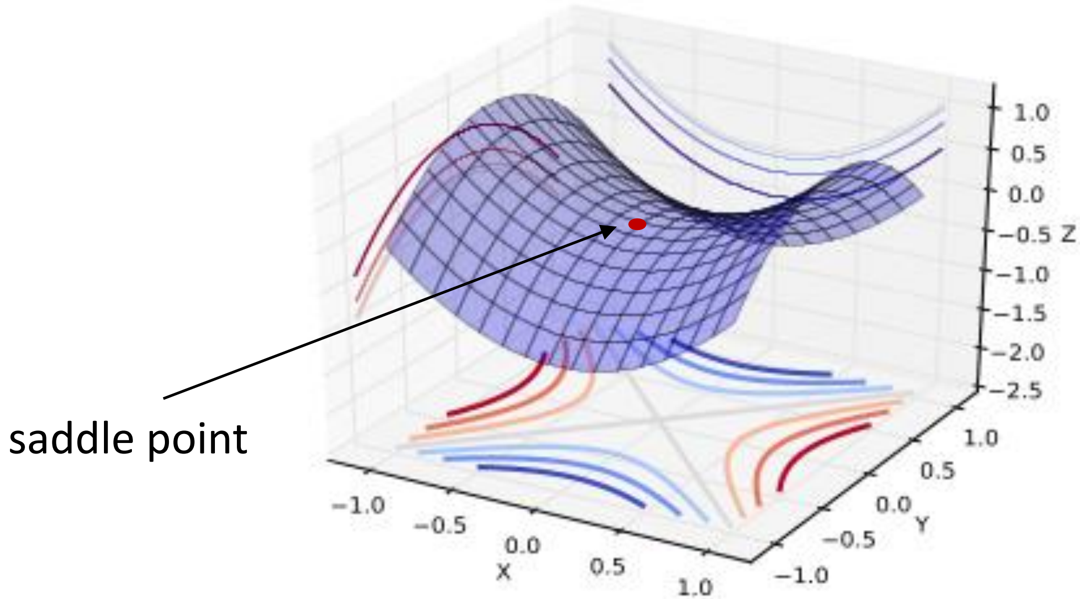
https://en.wikipedia.org

Positive Definiteness of Matrix

- Notes
  - A matrix in which some elements are negative may still be positive definite.
  - Conversely, a matrix whose entries are all positive is not necessarily positive definite.
- Induced partial ordering by positive definiteness
  - For arbitrary square matrices M and N, we write that $M > (\text{or} \geq) N$ when $M - N$ is positive definite (or positive semi-definite). This defines a partial ordering on the set of all square matrices.

Optimality Conditions
Unconstrained Multivariate

- If grad(f(x)) = 0 and x is neither a minimum nor a maximum, it is called a ‘saddle point’.



https://www.kdnuggets.com

- Recall that when univariate, it is called a ‘point of inflection’.

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 $\boxed{f(a)f(b) < 0}$.
  - Step 1. compute $f(x)$ at the midpoint $x = (a + b)/2$.
  - Step 2. if $f(x) = 0$ or $(b - a) < TOL$, then terminate.

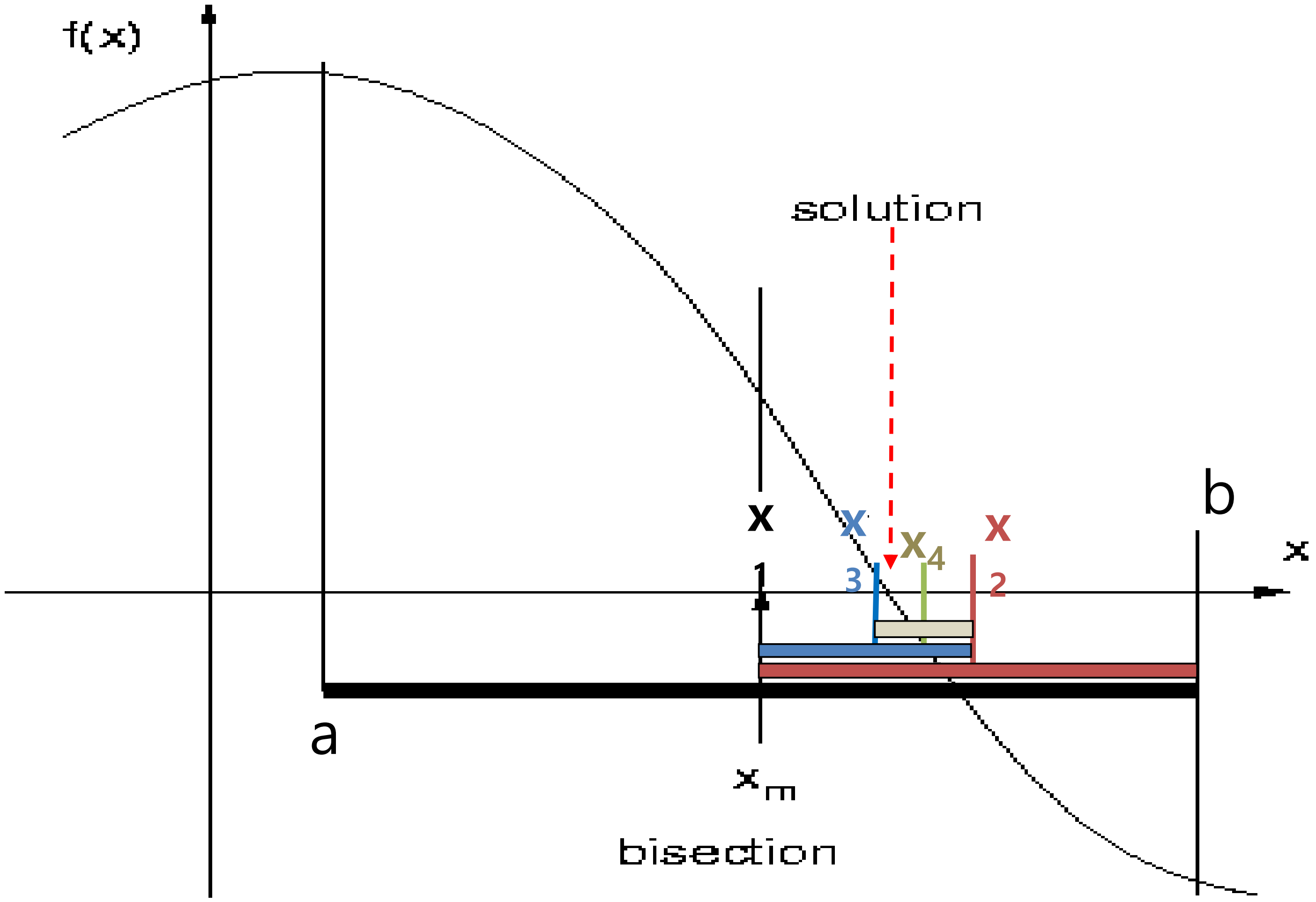
    if $f(x)f(a) < 0$, then $b := x$,

    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.

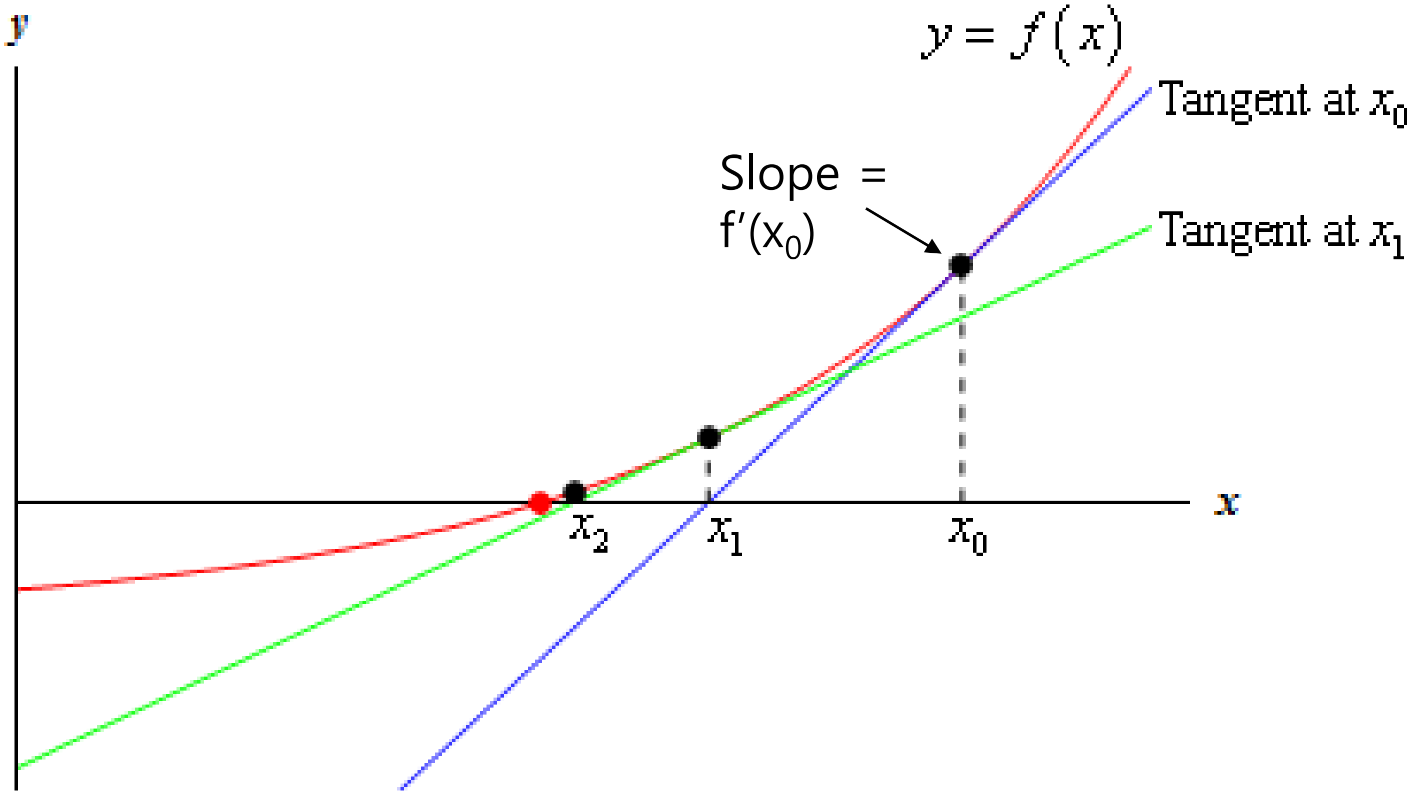
Midpoint x is one of choices. Any interior point In the interval [a, b] is possible to choose as a new point.

- 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



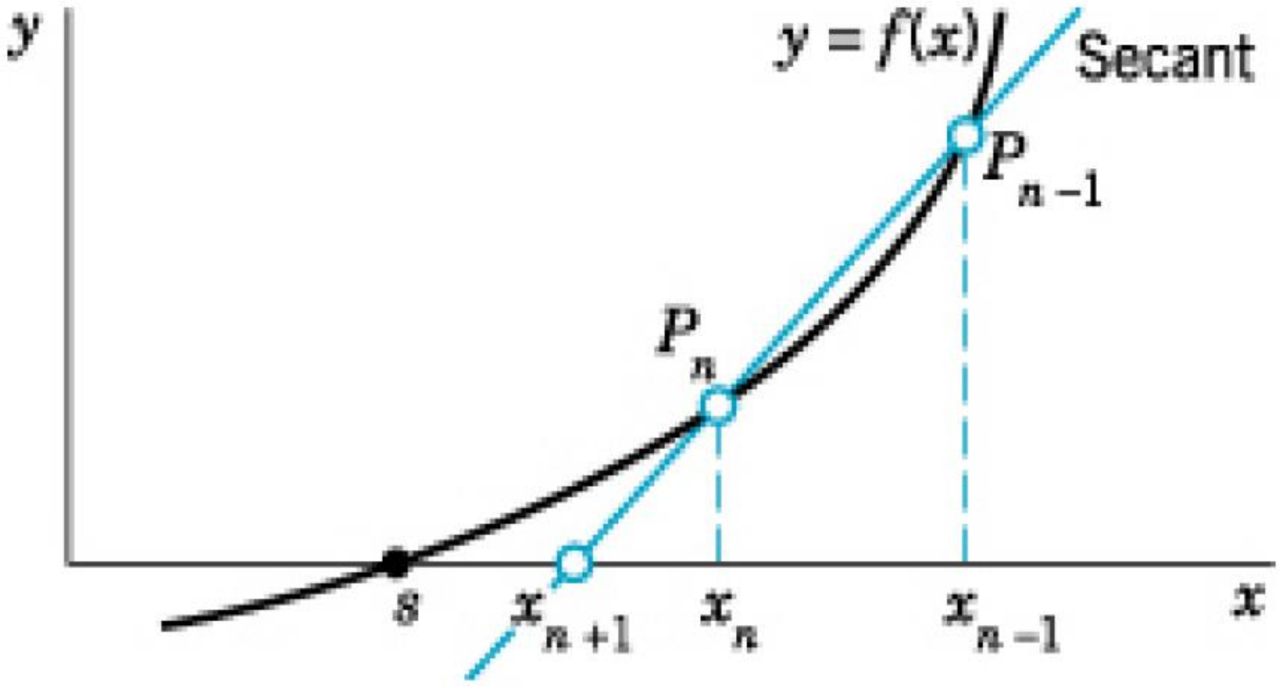
http://tutorial.math.lamar.edu/

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 : 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 $f'(t_i) \approx \dfrac{-f(t_i+2h)+4f(t_i+h)-3f(t_i)}{2h}$
    - Backward difference $f'(t_i) \approx \dfrac{3f(t_i)-4f(t_i-h)+f(t_i-2h)}{2h}$
  - 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)

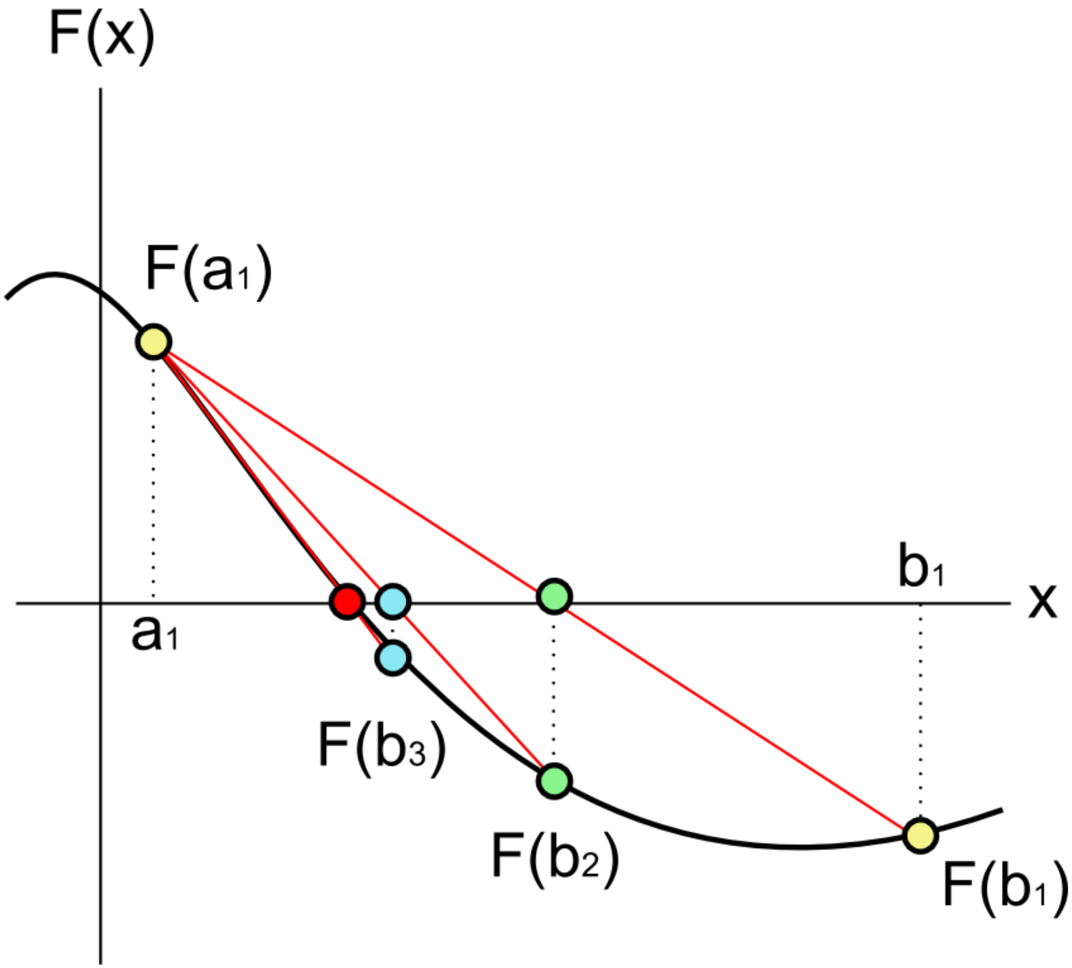
$$f'(t_i) \approx \frac{f(t_i+3h)-9f(t_i+2h)+45f(t_i+h)-45f(t_i-h)+9f(t_i-2h)-f(t_i-3h)}{60h}$$

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

Univariate Optimization:
Regular Falsi Method

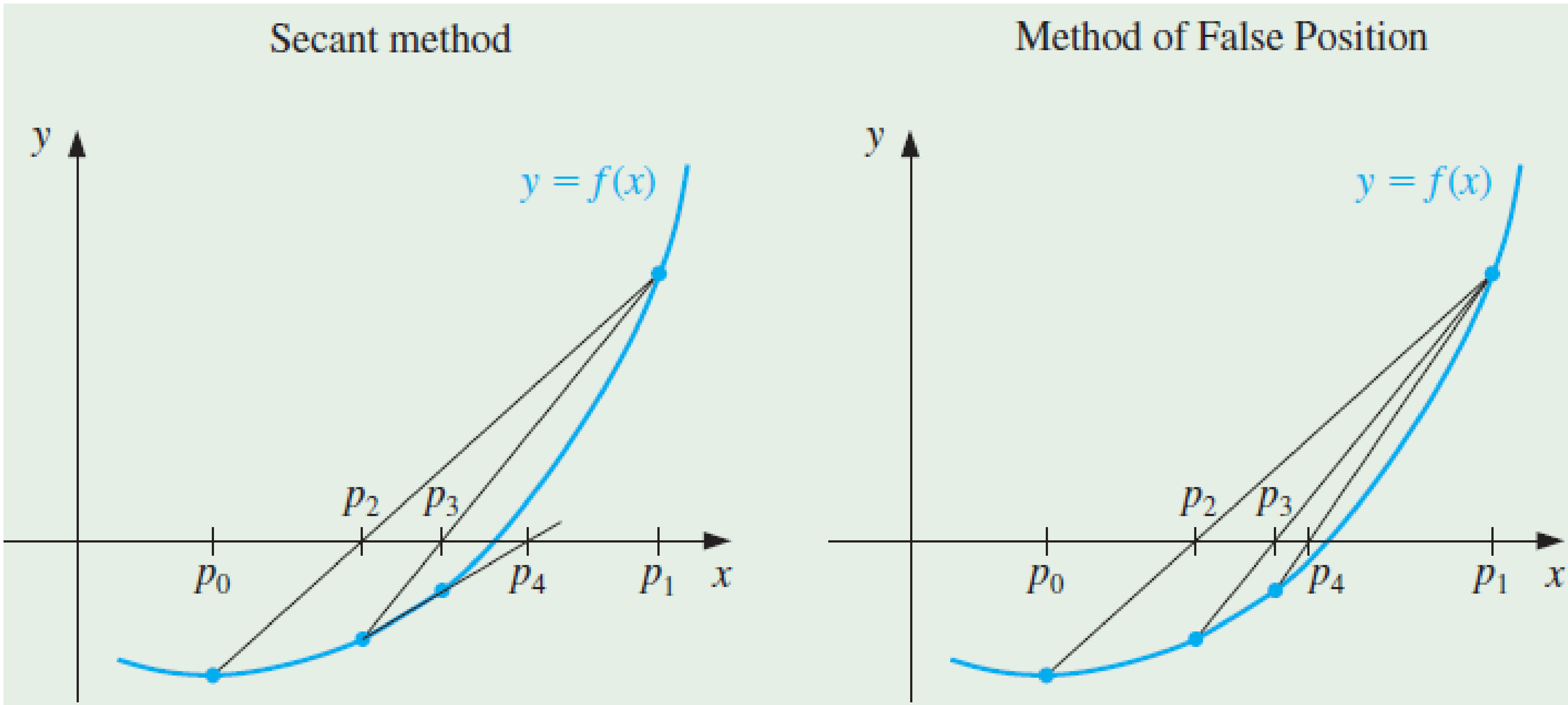
- Regular falsi method (method of false position)
  - Modified version of secant method & bisection method
  - Consider the given interval [a, b] such that $f(a)f(b) < 0$.
  - Apply a secant method with two initial points a & b. Find a point x intersecting with x-axis and a secant.
  - Choose updated interval as [a, x] or [x, b] depending on which corresponding function value agrees in sign with $f(x_{k+1})$. 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

- 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.
- How about other approaches in place of straight line?
  - Straight line (1^{st} order polynomial) is the simplest shape in approximation, so more complex shape may be applicable in the same context.
  - Curve (2^{nd} order or higher order polynomial) may be possible.

Univariate Optimization:
Root Finding Methods

- Additional ideas
  - Higher order polynomial approximation
    - In place of straight line, higher order polynomials (quadratic, cubic…) are possible to approximate 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 straight line.
    - Key points to consider
      - higher order approximation + easy to find a zero point

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’.

Univariate Optimization:
Root Finding Methods

- Bracketing methods
  - 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

- 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).