Numerical Methods II

SSCM 3423

Chapter 2

This chapter solves system of nonlinear equations using Newton method and steepest descent method.

Dr. Yeak Su Hoe
Department of Mathematical Sciences
Faculty of Science, Universiti Teknologi Malaysia
81300 UTM Johor Bahru, Malaysia
s.h.yeak@utm.my

June 2016

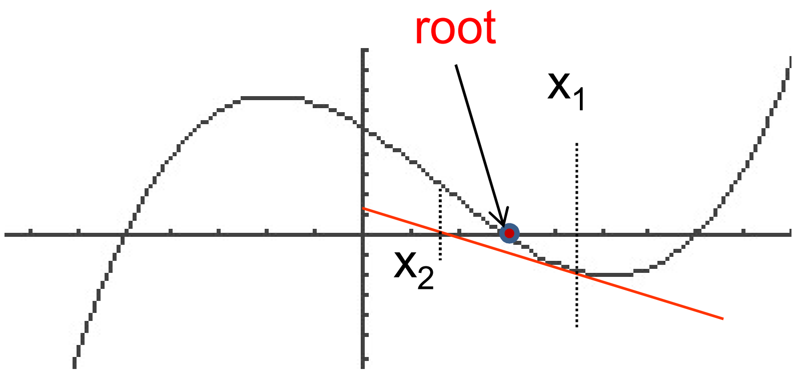
Nonlinear equation - Rootfinding

Newton's Method for Approximating Roots

Given: $x_i \rightarrow$ an initial guess of the root of $f(x)=0$
Assumption : x_1 better than x_0, x_2 better than x_1, etc.

Taylor Theorem : $f(x+h) \approx f(x) + h f'(x)$
Find h such that $f(x+h)=0$.

$$\Longrightarrow \quad h \approx -\frac{f(x)}{f'(x)}$$



A new guess of the root:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

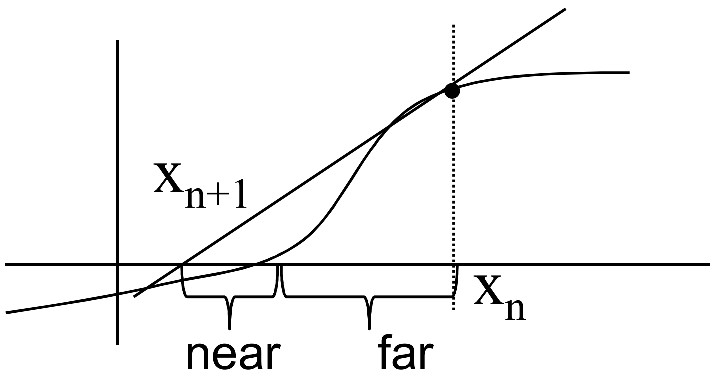
Newton's Method for Approximating Roots

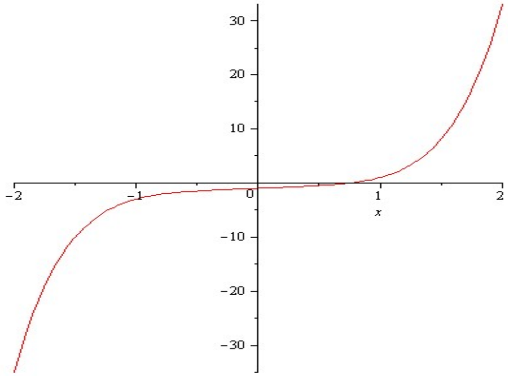
- Given $f(x)$ we seek a root of $f(x)=0$.
- If x_n is an approximation for the root

Then we claim that x_{n+1}

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

is a better approximation





Try !
example

$$f(x) = x^5 + x - 1 = 0$$

Answer:
0.7548776667

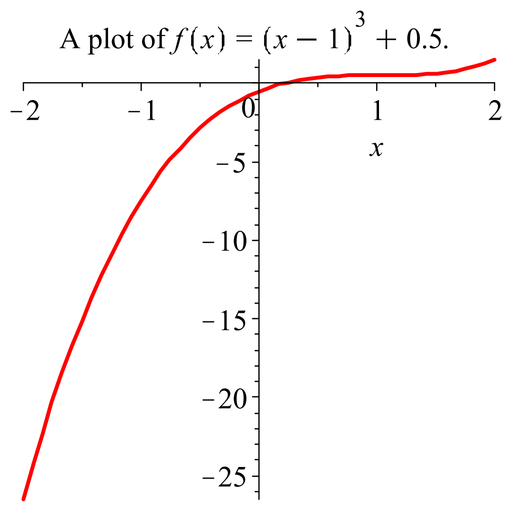
Assumptions:

f(x) is continuous and the first derivative is known

An initial guess x_0 such that $f'(x_0) \neq 0$ is given

Advantages

- Converges fast (quadratic convergence), if it converges.
- Requires only one guess



Drawbacks

1. <u>Divergence at inflection points</u>
   Selection of the initial guess or an iteration value of the root that is close to the inflection point of the function may start diverging away from the root in using Newton-Raphson method.

For example, to find the root of the equation $f(x) = (x-1)^3 + 0.5 = 0$

The Newton-Raphson method reduces to $x_{i+1} = x_i - \dfrac{(x_i - 1)^3 + 0.5}{3(x_i - 1)^2}$

The root starts to diverge at certain Iteration because the previous estimate of 0.9 is close to the inflection point of x=1.

Eventually after many iterations the root converges to the exact value of x=0.206

Drawbacks – Oscillations near local maximum and minimum

seek a root of $f(x)=0$

Table A Oscillations near local maxima and mimima in Newton-Raphson method.

Iteration Number	x_i	$f(x_i)$
0	–1.0000	3.00
1	0.5	2.25
2	–1.75	5.063
3	–0.30357	2.092
4	3.1423	11.874
5	1.2529	3.570
6	–0.17166	2.029
7	5.7395	34.942
8	2.6955	9.266
9	0.97678	2.954

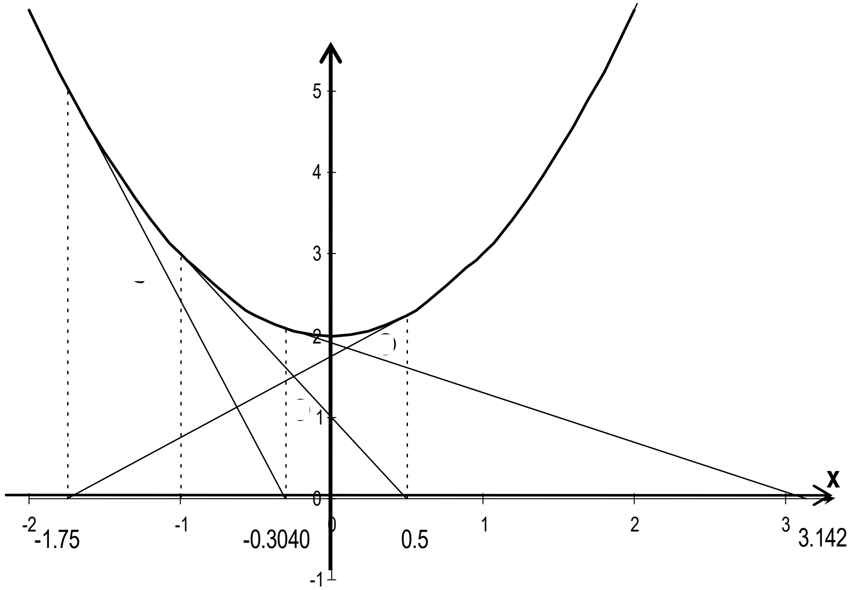


Figure A Oscillations around local minima for $f(x)=x^2+2$

Application of Newton method
Finding a square-root

- Example: √2 = 1.4142135623730950488016887242097
- Let x_0 be one and apply Newton's method.

Square root of a equivalent to find root of $f(x) = x^2 - a = 0$

$$f'(x) = 2x$$

$$x_{i+1} = x_i - \frac{x_i^2 - 2}{2x_i} = \frac{1}{2}\left(x_i + \frac{2}{x_i}\right)$$

$$x_0 = 1$$

$$x_1 = \frac{1}{2}\left(1 + \frac{2}{1}\right) = \frac{3}{2} = 1.5000000000$$

$$x_2 = \frac{1}{2}\left(\frac{3}{2} + \frac{4}{3}\right) = \frac{17}{12} \approx 1.4166666667$$

Note the rapid convergence

$$x_3 = \frac{1}{2}\left(\frac{17}{12} + \frac{24}{17}\right) = \frac{577}{408} \approx 1.414215686$$

$$x_4 = 1.4142135623\underbrace{7}\underbrace{46}$$

$$x_5 = 1.41421356237309504880168\underbrace{9}\underbrace{6}$$

$$x_6 = 1.4142135623730950488016887242097$$

Convergence Notation

Let $x_1, x_2, \ldots,$ converge to x (x_2 better than x_1, x_3 better than x_2 and so on), then

$$\text{Linear Convergence:} \quad \frac{|x_{n+1} - x|}{|x_n - x|} \le C$$

$$\text{Quadratic Convergence:} \quad \frac{|x_{n+1} - x|}{|x_n - x|^2} \le C$$

$$\text{Convergence of order } P: \quad \frac{|x_{n+1} - x|}{|x_n - x|^p} \le C$$

where C is a constant independent of x.

- A method with convergence order ***q*** converges faster than a method with convergence order ***p*** if ***q>p***.
- Quadratic convergence is faster than linear convergence.
- Methods of convergence order ***p>1*** are said to have super linear convergence.

Order of Convergence for Fixed Point Iteration Scheme

Fixed Point Iteration Scheme

Theorem. Let g be a continuous function on closed interval $[a,b]$ with $\alpha>1$ continuous derivatives on open interval (a,b). Let $p\in(a,b)$ be a fixed point of g.

If $\mathbf{g'(p)=g''(p)=\ldots=g^{(\alpha-1)}(p)=0}$,

But $\mathbf{g^{(\alpha)}(p)\neq 0}$, then there exists a $\delta>0$ such that for any $p_0\in[p-\delta, p+\delta]$, the sequence $\mathbf{p_n=g(p_{n-1})}$ converges to fixed point p of order α with asymptotic error constant

$$\lim_{n\to\infty}\frac{|e_{n+1}|}{|e_n|^{\alpha}}=\frac{|g^{(\alpha)}(p)|}{\alpha!}$$

Let, $x_n=r-\varepsilon_n$, $x_{n+1}=r-\varepsilon_{n+1}$, Where $r=G(r)$

$x_{n+1}=G(x_n) \rightarrow r-\varepsilon_{n+1}=G(r-\varepsilon_n)$ Using Taylor series, we get

$$r-\varepsilon_{n+1}=G(r)-\varepsilon_n G'(r)+\varepsilon_n^2\frac{G''(r)}{2!}+\cdots=r-\varepsilon_n G'(r)+\varepsilon_n^2\frac{G''(r)}{2!}+\cdots$$

The leading term in Taylor series gives $\varepsilon_{n+1}\approx G'(r)\,\varepsilon_n$ $\quad\varepsilon_1\approx G'(r)\,\varepsilon_0$ $\quad\varepsilon_n\approx [G'(r)]^n\varepsilon_0$

So, the fixed-point iteration is a first order scheme, provided $G'(r)\neq 0$

The scheme converges if $|G'(r)|<1$, diverges if $|G'(r)|>1$

e.g. $x-e^{-x}=0$

The error decreases monotonically if $0\leq G'(r)<1$,
The error decreases oscillatory if $-1\leq G'(r)<0$.

$x_{n+1}=e^{-xn}$ Or $x_{n+1}=-\ln(x_n)$

$x^2-2x-3=0$, roots are -1, 3. $\quad x_{n+1}=\sqrt{2x_n+3}$ $\quad x_{n+1}=\dfrac{3}{x_n-2}$

Newton's method for nonlinear systems

$$\begin{bmatrix} g_1 \\ \vdots \\ g_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

The system of equations $g_i(y_1,y_2,\ldots,y_n)=0 \quad (1\le i\le n)$
can be expressed simply as $\mathbf{G}(\mathbf{Y})=\mathbf{0}$
by letting $\mathbf{Y}=(y_1,y_2,\ldots,y_n)^T$ and $\mathbf{G}=(g_1,g_2,\ldots,g_n)^T$. Using the Taylor's series expansion, we get

$$\mathbf{0}=\mathbf{G}(\mathbf{Y}+\mathbf{H}) \approx \mathbf{G}(\mathbf{Y})+\mathbf{G}'(\mathbf{Y})\mathbf{H}\ , \quad \textit{(where } \mathbf{Y}+\mathbf{H} \textit{ is more accurate solution)}$$

where $\mathbf{H}=(h_1,h_2,\ldots,h_n)^T$ and $\mathbf{G}'(\mathbf{Y})$ is the $n\times n$ Jacobian matrix $\mathbf{J}(\mathbf{Y})$:

$$\mathbf{J}(\mathbf{Y})=\mathbf{G}'(\mathbf{Y})=\begin{bmatrix} \partial g_1/\partial y_1 & \partial g_1/\partial y_2 & \cdots & \partial g_1/\partial y_n \\ \partial g_2/\partial y_1 & \partial g_2/\partial y_2 & \cdots & \partial g_2/\partial y_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial g_n/\partial y_1 & \partial g_n/\partial y_2 & \cdots & \partial g_n/\partial y_n \end{bmatrix}$$

$$d\mathbf{F}=\mathbf{F}'\left(\mathbf{x}^{(n)}\right)\Delta\mathbf{x}$$

$$df=\frac{\partial f}{\partial x_1}\Delta x_1+\frac{\partial f}{\partial x_2}\Delta x_2+\cdots+\frac{\partial f}{\partial x_n}\Delta x_n$$

The correction vector **H** is obtained by solving linear system

$$\mathbf{J}(\mathbf{Y})\mathbf{H}= -\mathbf{G}(\mathbf{Y})$$

If Jacobian matrix is tridiagonal matrix, then **H** can be solved using Thomas algorithm. If the matrix size is 2×2, then just use the inverse of matrix **J**, $\mathbf{H}=\mathbf{J}^{-1}(-\mathbf{G})$.
Finally, Newton's iteration for n nonlinear equations in n variables is given by

$$\mathbf{Y}^{(k+1)}=\mathbf{Y}^{(k)}+\mathbf{H}^{(k)} \;\rightarrow\; \mathbf{Y}^{(k+1)}=\mathbf{Y}^{(k)}-\mathbf{J}^{-1}\mathbf{G}$$

where the Jacobian system is

$$\mathbf{J}(\mathbf{Y}^{(k)})\mathbf{H}^{(k)}= -\mathbf{G}(\mathbf{Y}^{(k)}).$$

Preliminaries – Taylor's series and derivative w.r.t. vectors

Taylor's series

$$f(x+h) = f(x) + h\frac{f'(x)}{1!} + h^2\frac{f''(x)}{2!} + \cdots + h^n\frac{f^{(n)}(x)}{n!} + R_n^*.$$

$$R_n = \frac{f^{(n+1)}(\theta x)}{(n+1)!}(x-a)^{n+1}, \quad a < \theta x < x.$$

$$X(t+h) = X(t) + hX'(t) + \frac{h^2}{2!}X''(t) + \frac{h^3}{3!}X^{(3)}(t) + \cdots$$

$$X''(t) = \tfrac{d}{dt}F(t, X)$$
$$x_1(t),\ x_2(t), \cdots$$

y (scalar or a vector)	$\partial y/\partial \mathbf{x}$
A$\mathbf{x}$	A
$\mathbf{x}^T$A	A^T
$\mathbf{x}^T\mathbf{x}$	$2\mathbf{x}^T$
$\mathbf{x}^T$A$\mathbf{x}$	A$\mathbf{x}$+A$^T\mathbf{x}$

x is column vector

$$f(x+a, y+b) = f(x,y) + \frac{1}{1!}D_1[f(x,y)] + \frac{1}{2!}D_2[f(x,y)] + \cdots + \frac{1}{n!}D_n[f(x,y)] + R_n.$$

$$D_n = \left(a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y}\right)^n \quad R_n = \frac{1}{(n+1)!}D_{n+1}[f(x+\theta_1 a, y+\theta_2 b)]$$

$$0 < \theta_1 < 1,\ 0 < \theta_2 < 1.$$

Newton's method for nonlinear systems

Example
$$y + x^2 - 0.5 - x = 0$$
$$x^2 - 5xy - y = 0$$
Initial guess $x = 1, \quad y = 0$

$$\mathbf{J}(\mathbf{Y}) = \mathbf{G}'(\mathbf{Y}) = \begin{bmatrix} \partial g_1/\partial y_1 & \partial g_1/\partial y_2 & \cdots & \partial g_1/\partial y_n \\ \partial g_2/\partial y_1 & \partial g_2/\partial y_2 & \cdots & \partial g_2/\partial y_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial g_n/\partial y_1 & \partial g_n/\partial y_2 & \cdots & \partial g_n/\partial y_n \end{bmatrix}$$

$$F\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y + x^2 - 0.5 - x \\ x^2 - 5xy - y \end{bmatrix}, \quad F' = \begin{bmatrix} 2x-1 & 1 \\ 2x-5y & -5x-1 \end{bmatrix}, \quad X_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Exact solution:
[x = 1.233317793, y = .2122450145],
[x = -.1781281996, y = .2901421450],
[x = -.4551895934, y = -.1623871594].

Iteration 1:

$$F\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} y + x^2 - 0.5 - x \\ x^2 - 5xy - y \end{bmatrix} = \begin{bmatrix} -0.5 \\ 1 \end{bmatrix} =, \quad F' = \begin{bmatrix} 2x-1 & 1 \\ 2x-5y & -5x-1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & -6 \end{bmatrix}$$

$$X_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 2 & -6 \end{bmatrix}^{-1} \begin{bmatrix} -0.5 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.25 \\ 0.25 \end{bmatrix}$$

Iteration 2:

$$F\begin{bmatrix} 1.25 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 0.0625 \\ -0.25 \end{bmatrix} =, \quad F' = \begin{bmatrix} 1.5 & 1 \\ 1.25 & -7.25 \end{bmatrix}$$

$$X_2 = \begin{bmatrix} 1.25 \\ 0.25 \end{bmatrix} - \begin{bmatrix} 1.5 & 1 \\ 1.25 & -7.25 \end{bmatrix}^{-1} \begin{bmatrix} 0.0625 \\ -0.25 \end{bmatrix} = \begin{bmatrix} 1.2332 \\ 0.2126 \end{bmatrix}$$

Try initial guess:
$x = -0.5, 0.5$
$x = -0.5, -0.5$

$$\mathbf{Y}^{(k+1)} = \mathbf{Y}^{(k)} + \mathbf{H}^{(k)} \rightarrow \mathbf{Y}^{(k+1)} = \mathbf{Y}^{(k)} - \mathbf{J}^{-1}\mathbf{G}$$

Note: **F = G**

Newton's method for nonlinear systems

Example

$$x_1^3 - 2x_2 - 2 = 0$$
$$x_1^3 - 5x_3^2 + 7 = 0$$
$$x_2 x_3^2 - 1 = 0$$

Initial guess $\mathbf{x} = [1 \quad 1 \quad 1]^T$

Exact solution:

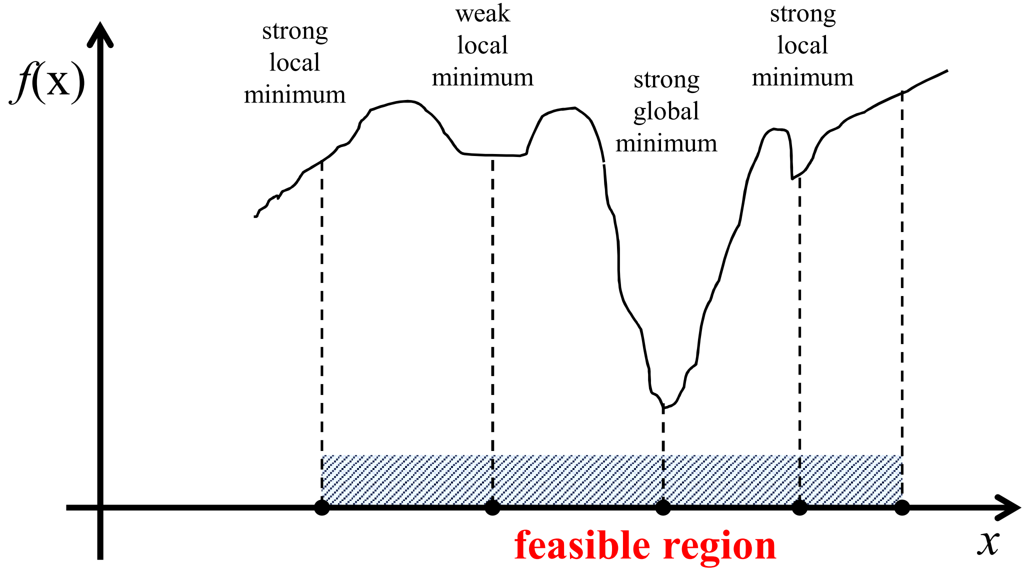
$$\mathbf{x} = [3^{1/3} \quad 0.5 \quad \sqrt{2}]^T$$

$$F = \begin{bmatrix} x_1^3 - 2x_2 - 2 \\ x_1^3 - 5x_3^2 + 7 \\ x_2 x_3^2 - 1 \end{bmatrix}, \quad F' = \begin{bmatrix} 3x_1^2 & -2 & 0 \\ & ? & \\ & ? & \end{bmatrix}, \quad X_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{Y}^{(k+1)} = \mathbf{Y}^{(k)} + \mathbf{H}^{(k)} \rightarrow \mathbf{Y}^{(k+1)} = \mathbf{Y}^{(k)} - \mathbf{J}^{-1}\mathbf{F}^{(k)}$$

$$\mathbf{J}(\mathbf{Y}) = \mathbf{G}'(\mathbf{Y}) = \begin{bmatrix} \partial g_1/\partial y_1 & \partial g_1/\partial y_2 & \cdots & \partial g_1/\partial y_n \\ \partial g_2/\partial y_1 & \partial g_2/\partial y_2 & \cdots & \partial g_2/\partial y_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial g_n/\partial y_1 & \partial g_n/\partial y_2 & \cdots & \partial g_n/\partial y_n \end{bmatrix}$$

Types of minima



- **Global minimum/maximum** is the minimum/maximum value in the feasible region.
- **Local minimum/maximum** is the minimum/maximum value in the local region only.

First-order optimality condition

Minimize $f(x)$

- For function of one variable, we can find extremum by differentiating function and setting derivative to zero
- For function of n variables, we need to find critical point, i.e. solution of nonlinear system

$$\nabla \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$$

where $\nabla f(x)$ is gradient vector of f, whose i^{th} component is $\partial f(x) / \partial x_i$

- For continuously differentiable f: $S \subseteq R^n \rightarrow R$, any interior point x^* of S at which f has local minimum must be critical point of f
- **Not** all critical points are minima: they can be **maxima or saddle point**.

Second-order optimality condition

Minimize $f(x)$

- For twice continuously differentiable f: $S \subseteq R^n \to R$, we can distinguish among the critical points by using **Hessian matrix $H_f(x)$** defined by

$$\left(H_f(x)\right)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

which is symmetric

- At critical point x^*, if $H_f(x^*)$ is
  - Positive definite, then x^* is minimum of f
  - Negative definite, then x^* is maximum of f
  - Indefinite, then x^* is saddle point of f
  - Singular, then various behaviour are possible

$$\nabla f(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

$$\mathbf{H} = \nabla^2 f(\mathbf{x}) = \frac{\partial(\nabla f(\mathbf{x}))}{\partial \mathbf{x}} = \nabla(\nabla f(\mathbf{x})) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix}$$

A is positive definite if
$\mathbf{x}^T A \mathbf{x} > 0$, for all $\mathbf{x}$ (all eigenvalue>0)
A is negative definite if
$\mathbf{x}^T A \mathbf{x} < 0$, for all $\mathbf{x}$ (all eigenvalue<–0)

A Hermitian matrix which is neither positive definite, negative definite, positive-semidefinite, nor negative-semidefinite is called *indefinite*. (having both positive and negative eigenvalues).

Steepest descent

- Basic principle is to minimize the N-dimensional function by a series of 1D line-minimizations:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$$

- The steepest descent method chooses $\mathbf{p}_k$ to be parallel to the gradient

$$\mathbf{p}_k = -\nabla f(\mathbf{x}_k)$$

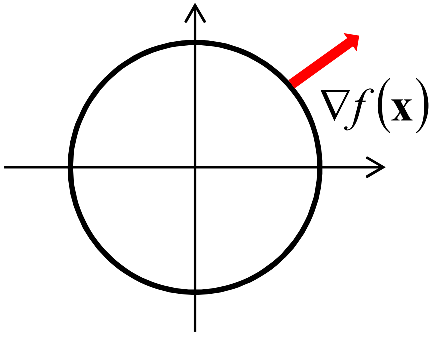
- Step-size α_k is chosen to minimize $f(\mathbf{x}_k + \alpha_k \mathbf{p}_k)$.

  For quadratic forms there is a closed form solution:

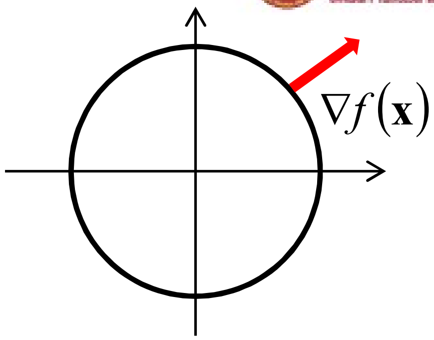
$$\alpha_k = \frac{\mathbf{p}_k^T \mathbf{p}_k}{\mathbf{p}_k^T \mathbf{H} \mathbf{p}_k}$$

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H \mathbf{x} - \mathbf{b}^T \mathbf{x}$$

Try with $f = x^2 + y^2$



Steepest descent



Let $\mathbf{x}_k$=[1 0]^T, ∇f=2x**i**+2y**j**, $\mathbf{z}_k$= $-\nabla f/|\nabla f|$,
Minimize $f(\mathbf{x}_k+\alpha\mathbf{z}_k)$

Try with
$f=x^2+y^2$

Let α_0=0, α_1=0.2, α_2=0.4, we get

$$f_1 = f(\mathbf{x}_k - \alpha_1 \mathbf{z}_k) = f\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = 1,$$

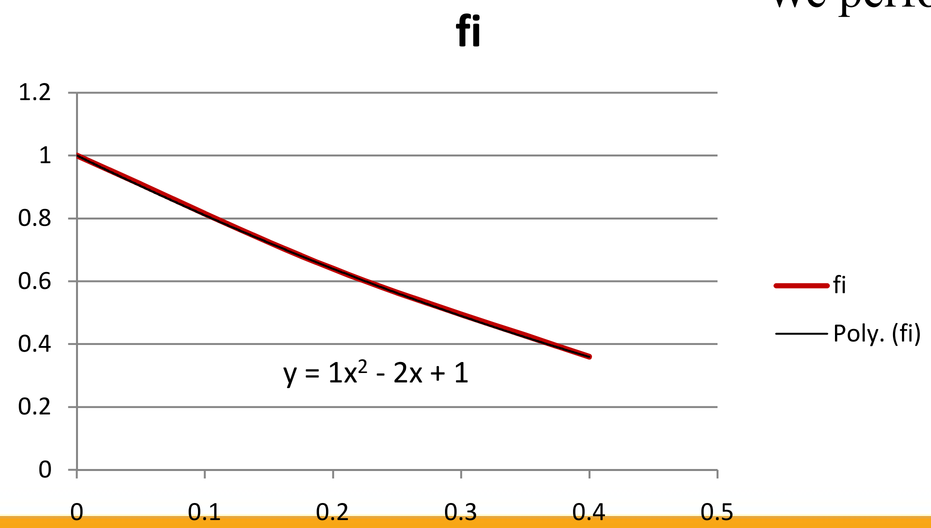
$$f_2 = f(\mathbf{x}_k - \alpha_2 \mathbf{z}_k) = f\left(\begin{bmatrix}0.8\\0\end{bmatrix}\right) = 0.64,$$

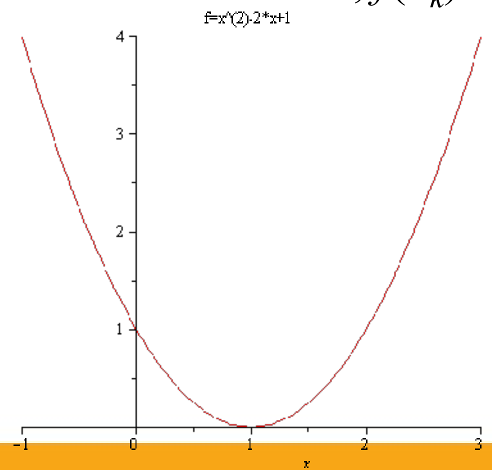
$$f_3 = f(\mathbf{x}_k - \alpha_3 \mathbf{z}_k) = f\left(\begin{bmatrix}0.6\\0\end{bmatrix}\right) = 0.36,$$

i	**1**	**2**	**3**
α_i	0	0.2	0.4
f_i	1	0.64	0.36

We perform interpolation using Ms Excel

So, $f(\mathbf{x}_k)$ is minimum when α is 1





Steepest descent method

- Let f: $R^n \to R$, be a real-valued function of n real variable
- At any point x where gradient vector is nonzero, negative gradient, $-\nabla f(x)$, points downhill toward lower values of f
- In fact, $-\nabla f(x)$ is locally direction of steepest descent: f decreases more rapidly along direction of negative gradient than along any other
- **Steepest descent method**: starting from initial guess x_0, we approximate solution given by

$$\boldsymbol{x_{k+1} = x_k - \alpha_k \nabla f(x_k)}$$

where α_k is **line search** parameter that determines how far to go in given direction.

Try with $f = x^2 + y^2$

Steepest Descent

- **Steepest descent algorithm**:
- Initial guess: $x_0 \in R^n$

Our objective is

$$\min_x f(x)$$

Step A:set $i = 0$

Step B: if $\nabla \boldsymbol{f}(\boldsymbol{x_i})\mathbf{=0}$ then stop,

else, compute **search direction**

$h_i = -\nabla f(x_i)$

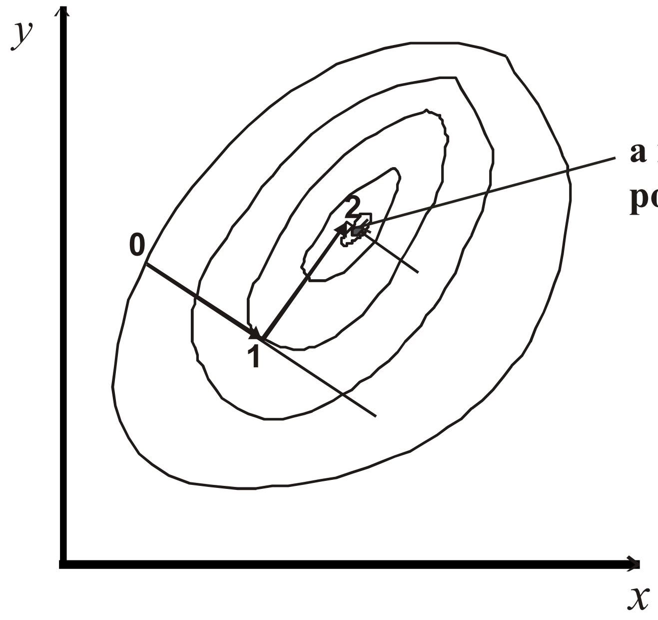
Step C: compute the step size α_k

$$\alpha_i \in \arg\min_{\alpha \geq 0} f\left(x_i + \alpha\, h_i\right)$$

Step D: set $x_{i+1} = x_i + \alpha_i \cdot h_i$, and go to step B.

Steepest Descent

steepest descent method to find the minimum can be applied to solve a system of nonlinear equations $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$



$$f_1(x_1, x_2, \ldots, x_n) = 0,$$
$$f_2(x_1, x_2, \ldots, x_n) = 0,$$
$$\vdots$$
$$f_n(x_1, x_2, \ldots, x_n) = 0.$$

$$g(x_1, \cdots, x_n) = \sum_{i=1}^{n} [f_i(x_1, \cdots, x_n)]^2$$

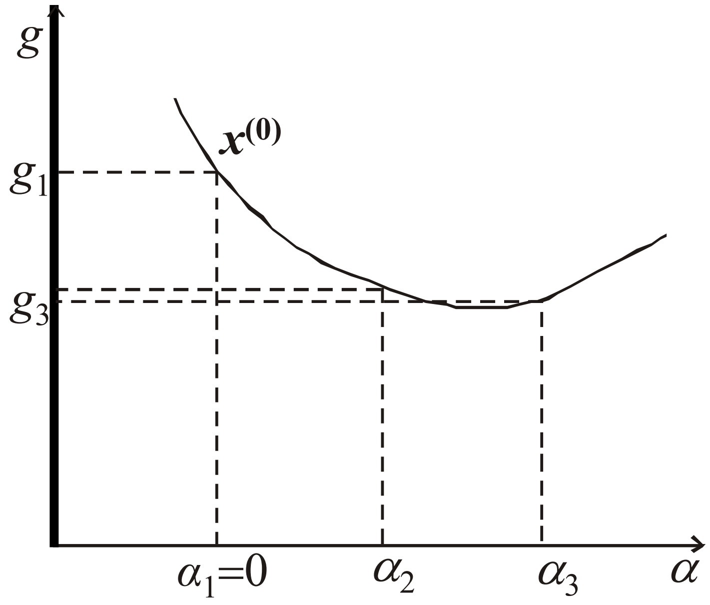
$$g = \boldsymbol{f} \times \boldsymbol{f}^{\mathrm{T}}$$

where $\boldsymbol{f} = [f_1 \; f_2 \ldots f_n]$ and $\boldsymbol{f}^{\mathrm{T}}$ is the transpose of $\boldsymbol{f}$

Where, we get $\min_{\mathbf{x}} g(\mathbf{x}) = 0$

Steepest Descent

1. Evaluate g at an initial approximation $\boldsymbol{x}^{(0)}$.
2. Determine a direction from $\boldsymbol{x}^{(0)}$ using the gradient of g.
3. Move to a new appropriate position $\boldsymbol{x}^{(1)}$ so that $g(\boldsymbol{x}^{(1)}) < g(\boldsymbol{x}^{(0)})$.
4. Repeat steps 2-4 with $\boldsymbol{x}^{(0)}$ replaced by $\boldsymbol{x}^{(1)}$.

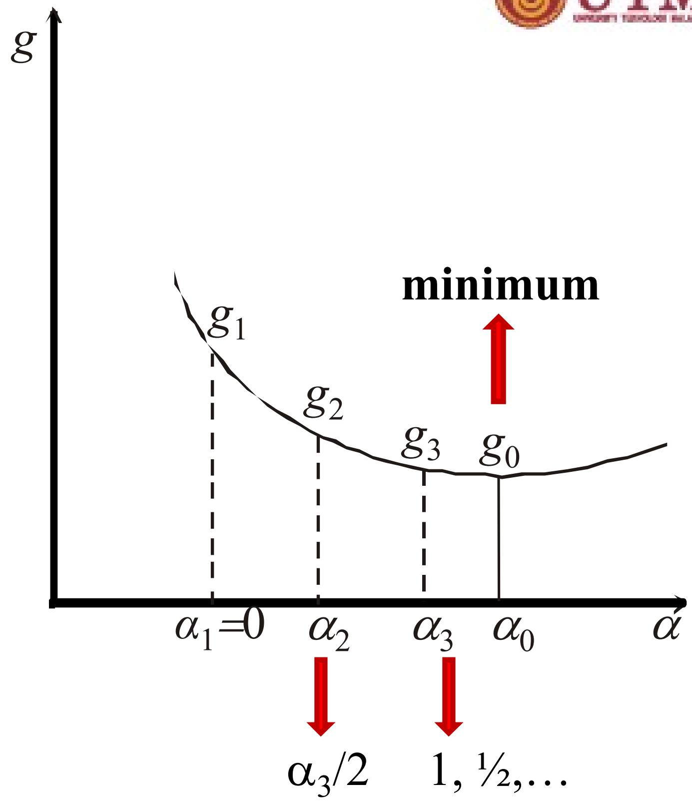


The direction of greatest decrease in the value of g at $\boldsymbol{x}$ is the direction given by $-\nabla g(\boldsymbol{x})$, therefore $\boldsymbol{x}^{(1)}$ is given by

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - \alpha \nabla g(\boldsymbol{x})$$

where α will be determined

Steepest Descent



The procedure is to fit a quadratic to three points (α_1, g_1), (α_2, g_2), and (α_3, g_3) then choose α so that g is a minimum in the direction of steepest descent. The steps are

a) Let $\alpha_1 = 0$ at $\boldsymbol{x}^{(0)}$, therefore $g_1 = g(\boldsymbol{x}^{(0)})$.
b) Let $\alpha_3 = 1$ and evaluate $g_3 = g(\boldsymbol{x}^{(0)} - \alpha_3 \boldsymbol{z})$
c) If $g_3 > g_1$, let $\alpha_3 = \alpha_3/2$ and repeat step (b)
d) Let $\alpha_2 = \alpha_3/2$ and evaluate $g_2 = g(\boldsymbol{x}^{(0)} - \alpha_2 \boldsymbol{z})$

The quadratic through three points (α_1, g_1), (α_2, g_2), and (α_3, g_3) has the form

$$P(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2) = g_1 + g^1{}_1\alpha + g^2{}_1\,\alpha(\alpha - \alpha_2)$$

where

$$h_1 = \frac{g_2 - g_1}{\alpha_2 - \alpha_1},\quad h_2 = \frac{g_3 - g_2}{\alpha_3 - \alpha_2},\quad h_3 = \frac{h_2 - h_1}{\alpha_3 - \alpha_1}$$

$$g_1^1 = \frac{g_2 - g_1}{\alpha_2 - \alpha_1},\quad g_2^1 = \frac{g_3 - g_2}{\alpha_3 - \alpha_2},\quad g_1^2 = \frac{g_2^1 - g_1^1}{\alpha_3 - \alpha_1}$$

$$z_0 = \left\|\nabla g\left(\mathbf{x}^{(0)}\right)\right\|_2$$

$$\mathbf{z} = \frac{1}{z_0}\nabla g\left(\mathbf{x}^{(0)}\right) = \nabla g\left(\mathbf{x}^{(0)}\right) \Big/ \left\|\nabla g\left(\mathbf{x}^{(0)}\right)\right\|_2$$

At the location where

$$\frac{dP}{d\alpha} = h_1 + 2h_3\alpha - h_3\alpha_2 = 0 \;\rightarrow \alpha_0 = 0.5\left(\alpha_2 - \frac{h_1}{h_3}\right) = 0.5\left(\alpha_2 - \frac{g_1^1}{g_1^2}\right)$$

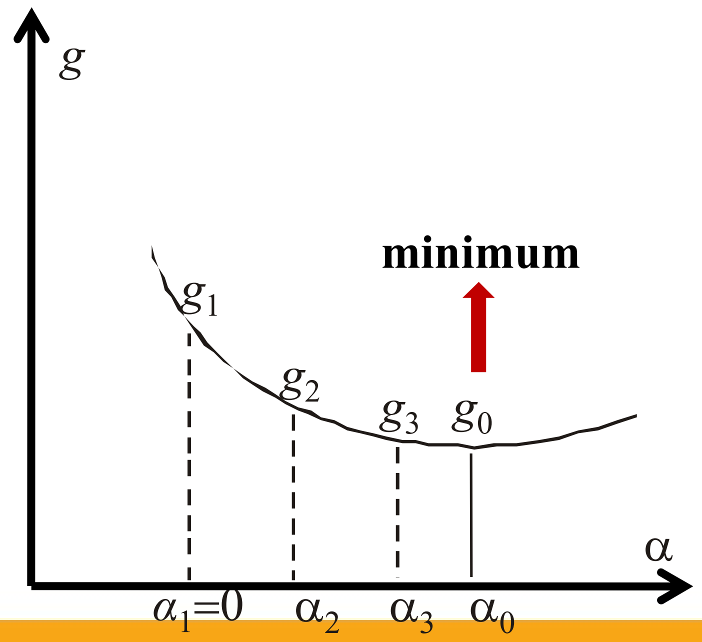
Steepest Descent

e) Evaluate $g_0 = g(\boldsymbol{x}^{(0)} - \alpha_0 \boldsymbol{z})$

Since a quadratic through three points can have a minimum or a maximum as shown in Fig below, α is chosen so that g is the lowest value between g_0 and g_3 as follows

f) If $g_0 < g_3$ then $\alpha = \alpha_0$ else $\alpha = \alpha_3$

4. Repeat steps 2-4 with $\boldsymbol{x}^{(0)}$ replaced by $\boldsymbol{x}^{(1)}$.



a) Let $\alpha_1 = 0$ at $\boldsymbol{x}^{(0)}$, therefore $g_1 = g(\boldsymbol{x}^{(0)})$.
b) Let $\alpha_3 = 1$ and evaluate $g_3 = g(\boldsymbol{x}^{(0)} - \alpha_3 \boldsymbol{z})$
c) If $g_3 > g_1$, let $\alpha_3 = \alpha_3/2$ and repeat step (b)
d) Let $\alpha_2 = \alpha_3/2$ and evaluate $g_2 = g(\boldsymbol{x}^{(0)} - \alpha_2 \boldsymbol{z})$

1. Evaluate g at an initial approximation $\boldsymbol{x}^{(0)}$.
2. Determine a direction from $\boldsymbol{x}^{(0)}$ using the gradient of g.
3. Move to a new appropriate position $\boldsymbol{x}^{(1)}$ so that $g(\boldsymbol{x}^{(1)}) < g(\boldsymbol{x}^{(0)})$.
4. Repeat steps 2-4 with $\boldsymbol{x}^{(0)}$ replaced by $\boldsymbol{x}^{(1)}$.

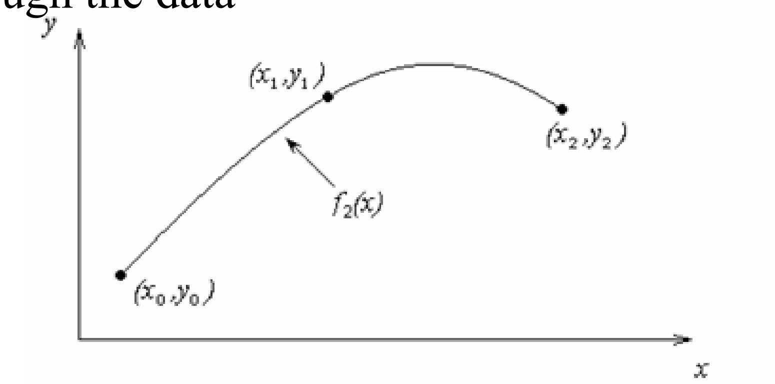
$$\mathbf{z} = \frac{1}{z_0}\nabla g\left(\mathbf{x}^{(0)}\right) = \nabla g\left(\mathbf{x}^{(0)}\right) \Big/ \left\|\nabla g\left(\mathbf{x}^{(0)}\right)\right\|_2$$

Steepest Descent

Newton's forward divided-difference interpolation

Given (x_0,y_0), (x_1,y_1) and (x_2,y_2), fit a quadratic interpolant through the data

$$f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1)$$



$$b_0 = f(x_0)$$

$$b_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

$$b_2 = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}$$

$$b_0 = f[x_0] = f(x_0)$$

$$b_1 = f[x_1, x_0] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

General form

$$b_2 = f[x_2, x_1, x_0] = \frac{f[x_2, x_1] - f[x_1, x_0]}{x_2 - x_0} = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}$$

$$f(x) = f[x_0] + f[x_1, x_0](x - x_0) + f[x_2, x_1, x_0](x - x_0)(x - x_1)$$

alternative form

$$f(x) = f_0 + f_0^{[1]}(x - x_0) + f_0^{[2]}(x - x_0)(x - x_1)$$

where

$$f(x_i) = f_i, \quad f_i^{[0]} = f_i, \quad f_i^{[j]} = \frac{f_{i+1}^{[j-1]} - f_i^{[j-1]}}{x_{i+j} - x_i}$$

Steepest Descent

Example

Use the method of steepest descent with the initial guess $\boldsymbol{x}$ = [0 0 0] to obtain the solutions to the following equations

$$f_1(x_1, x_2, x_3) = 3x_1 - \cos(x_2\, x_3) - ½ = 0$$
$$f_2(x_1, x_2, x_3) = x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 = 0$$
$$f_3(x_1, x_2, x_3) = e^{-x_1 x_2} + 20x_3 + (10\pi - 3)/3 = 0$$

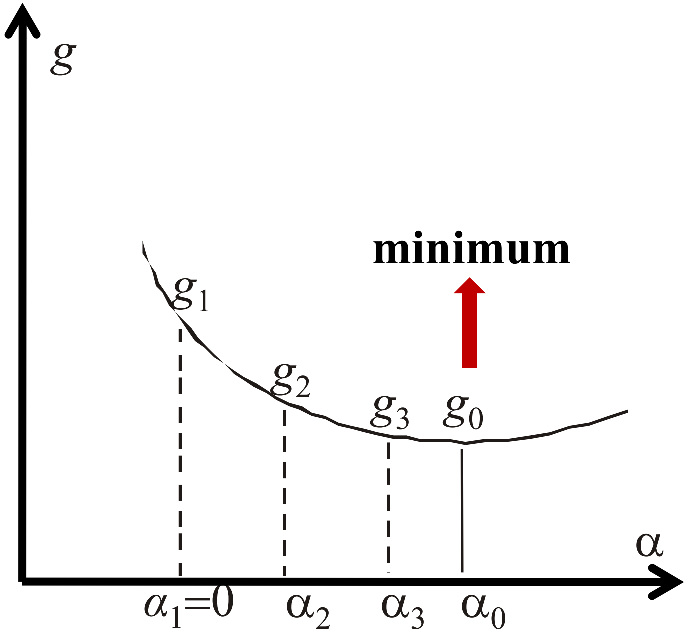
Exact, $\mathbf{x}=(0.5,0,-0.5235988)^T$

$$\boldsymbol{x}^{(0)} = [0\ 0\ 0];\ \boldsymbol{f} = [f_1\ f_2\ f_3]$$
$$g = \boldsymbol{f}\times\boldsymbol{f}^{T} = f_1^2 + f_2^2 + f_3^2 = 111.975$$

$$\nabla g(\mathbf{x}) = \frac{\partial g}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial g(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial g(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

$$\mathbf{H} = \nabla^2 g(\mathbf{x}) = \frac{\partial(\nabla g(\mathbf{x}))}{\partial \mathbf{x}} = \nabla(\nabla g(\mathbf{x})) = \begin{bmatrix} \frac{\partial^2 g(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 g(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 g(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 g(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix}$$

Steepest Descent



$f_1(x_1, x_2, x_3) = 3x_1 - \cos(x_2 x_3) - ½ = 0$

$f_2(x_1, x_2, x_3) = x_1^{\,2} - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 = 0$

$f_3(x_1, x_2, x_3) = e^{-x_1 x_2} + 20x_3 + (10\pi - 3)/3 = 0$

Exact, $\mathbf{x}=(0.5,0,-0.5235988)^T$

$$\nabla g(\mathbf{x}) = \frac{\partial g}{\partial \mathbf{x}} = \begin{pmatrix} \dfrac{\partial g(\mathbf{x})}{\partial x_1} \\ \vdots \\ \dfrac{\partial g(\mathbf{x})}{\partial x_n} \end{pmatrix} = \begin{bmatrix} 2f_1 \cdot 3 + 2f_2 \cdot 2x_1 + 2f_3 \cdot e^{-x_1 x_2}(-x_2) \\ 2f_1 \cdot (\sin(x_2 x_3) x_3) + 2f_2 \cdot (-162(x_2 + 0.1)) + 2f_3 \cdot e^{-x_1 x_2}(-x_1) \\ 2f_1 \cdot (\sin(x_2 x_3) x_2) + 2f_2 \cdot (\cos x_3) + 2f_3 \cdot 20 \end{bmatrix}$$

$\mathbf{x}^{(0)} = [0\ 0\ 0];$

$g(\mathbf{x}^{(0)}) = f_1^{\,2} + f_2^{\,2} + f_3^{\,2} = 111.975$ $\quad z_0 = \left\| \nabla g(\mathbf{x}^{(0)}) \right\|_2 = 419.554$

Let $\mathbf{z} = \frac{1}{z_0} \nabla g(\mathbf{x}^{(0)}) = (-0.0214514, -0.0193062, 0.999583)^T$

With α_1=0, $g_1 = g(\mathbf{x}^{(0} - \alpha_1 \mathbf{z}) = g(\mathbf{x}^{(0}) = 111.975$ Now, let α_3=1,

$g_3 = g(\mathbf{x}^{(0} - \alpha_3 \mathbf{z}) = 93.5649$

Since, $g_3 < g_1$, we accept α_3, set $\alpha_2 = \alpha_3/2 = 0.5$

$g_2 = g(\mathbf{x}^{(0} - \alpha_2 \mathbf{z}) = 2.53557$

$$\frac{dP}{d\alpha} = h_1 + 2h_3\alpha - h_3\alpha_2 = 0 \rightarrow \alpha_0 = 0.5\left(\alpha_2 - \frac{h_1}{h_3}\right) = 0.5\left(\alpha_2 - \frac{g_1^1}{g_1^2}\right)$$

$$f_2(x) = f[x_0] + f[x_1, x_0](x - x_0) + f[x_2, x_1, x_0](x - x_0)(x - x_1)$$

$b_0 = f[x_0] = f(x_0),$

$b_1 = f[x_1, x_0] = \dfrac{f(x_1) - f(x_0)}{x_1 - x_0},$

$b_2 = f[x_2, x_1, x_0] = \dfrac{f[x_2, x_1] - f[x_1, x_0]}{x_2 - x_0}.$

Find quadratic polynomial that interpolate data (0,111.975), (0.5, 2.53557) &(1,93.5649). We use Newton's forward divided-difference interpolation

$$P(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2) = g_1 + g^1{}_1\alpha + g^2{}_1\alpha(\alpha - \alpha_2)$$

Steepest Descent

$P(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2) = g_1 + g^1{}_1\alpha + g^2{}_1\,\alpha(\alpha - \alpha_2)$

interpolate data (0,111.975), (0.5, 2.53557) & (1,93.5649).

$\alpha_1 = 0, \quad g_1 = 111.975,$

$\alpha_2 = 0.5, \quad g_2 = 2.53557, h_1 = \dfrac{g_2 - g_1}{\alpha_2 - \alpha_1} = -218.878,$

$\alpha_3 = 1, \quad g_3 = 93.5649, h_2 = \dfrac{g_3 - g_2}{\alpha_3 - \alpha_2} = 182.059, \quad h_3 = \dfrac{h_2 - h_1}{\alpha_3 - \alpha_1} = 400.937$

$$P(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2) = 111.975 - 218.878\alpha + 400.937\alpha(\alpha - 0.5)$$

x =[0.011218 0.010096 -0.522741], g = 2.327617
x =[0.137860 -0.205453 -0.522059], g = 1.274058
x =[0.266959 0.005511 -0.558494], g = 1.068131
x =[0.272734 -0.008118 -0.522006], g = 0.468309
x =[0.308689 -0.020403 -0.533112], g = 0.381087
x =[0.314308 -0.014705 -0.520923], g = 0.318837
x =[0.324267 -0.008525 -0.528431], g = 0.287024
x =[0.330809 -0.009678 -0.520662], g = 0.261579
x =[0.339809 -0.008592 -0.528080], g = 0.238486
x =[0.345746 -0.009034 -0.520941], g = 0.217440

We have $P'(\alpha)=0$ when $\alpha = \alpha_0 = 0.522959$. Since $g_0 = g(\mathbf{x}^{(0)} - \alpha_0\mathbf{z}) = 2.32762$ is smaller than g_1 and g_3, we set

$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \alpha_0\mathbf{z} = \mathbf{x}^{(0)} - 0.522959\mathbf{z} = (0.0112182, 0.0100964, -0.522741)^T$ We get, $g(\mathbf{x}^{(1)}) = 2.32762$

$$\frac{dP}{d\alpha} = h_1 + 2h_3\alpha - h_3\alpha_2 = 0 \;\rightarrow \alpha_0 = 0.5\left(\alpha_2 - \frac{h_1}{h_3}\right) = 0.5\left(\alpha_2 - \frac{g_1^1}{g_1^2}\right)$$

Exact, $\mathbf{x} = (0.5, 0, -0.5235988)^T$

$$f_2(x) = f[x_0] + f[x_1, x_0](x - x_0) + f[x_2, x_1, x_0](x - x_0)(x - x_1)$$

$$f[x_0] = f(x_0), \quad f[x_1, x_0] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}, \quad f[x_2, x_1, x_0] = \frac{f[x_2, x_1] - f[x_1, x_0]}{x_2 - x_0}.$$

$$f(x) = f_0 + f_0^{[1]}(x - x_0) + f_0^{[2]}(x - x_0)(x - x_1)$$

$$f(x_i) = f_i, \quad f_i^{[0]} = f_i, \quad f_i^{[j]} = \frac{f_{i+1}^{[j-1]} - f_i^{[j-1]}}{x_{i+j} - x_i}$$

References

- Brian Bradie, A Friendly Introduction to Numerical Analysis, Prentice Hall, New Jersey, 2006.
- Fausett L.V, Numerical Methods; algorithm and applications, Prentice Hall, New Jersey,2003
- Rao S.S, Applied Numerical Methods for Engineers and Scientist, Prentice Hall, New Jersey,2002
- Faires J.D. Burden R, Numerical Methods, 2nd Edition, Thomson Brooks/Cole, Australia,1998
- Burden R.L, Faires J.D & Reynolds A.C, Numerical Analysis, 5th edition, PWS-KENT Pub, Boston, 1993