Non-Linear Equations in MD

Vasily Arzhanov

KTH

Overview

- Multivariate functions
- Gradient and its geometical meaning
- Linearization in 2D and MD
- Vector-valued functions
- Jacobian matrix
- Newton's method in 2D and MD
- Quasi-Newton Method

Multivariate Functions

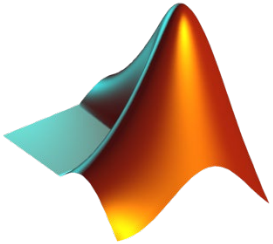
$$f = f(x, y)$$

$$f = f(x, y, z)$$

$$f = f(x_1, \ldots, x_n)$$

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

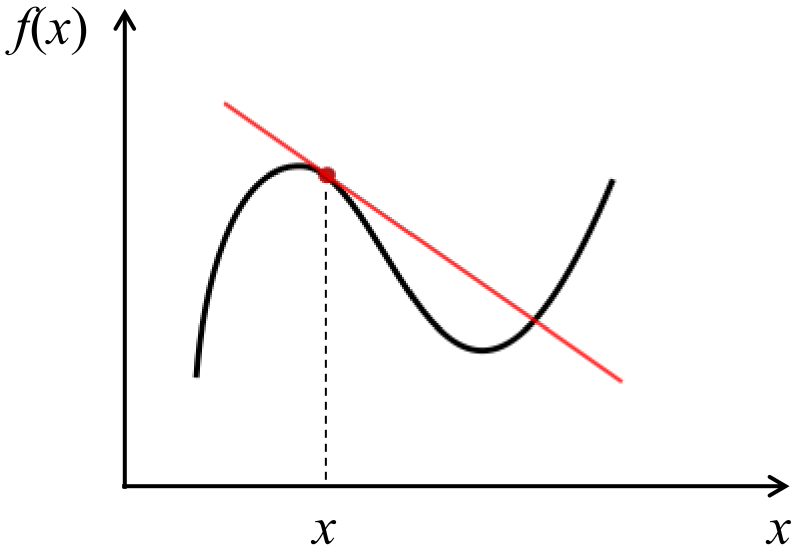
Graph of a bivariate function



Derivatives in 1D

$$\frac{\mathrm{d}f}{\mathrm{d}x} \equiv \frac{df}{dx} \equiv f'(x) \equiv \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Change rate at x.



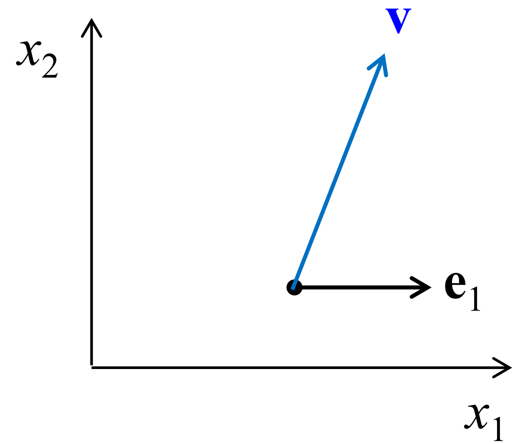
Derivatives in 2D

$$\mathbf{x} \equiv \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad z = f(x_1, x_2) = f(\mathbf{x}) \qquad \mathbf{e}_1 \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\partial_1 f \equiv \frac{\partial f}{\partial x_1} \equiv \lim_{h \to 0} \frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h} = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{e}_1) - f(\mathbf{x})}{h}$$

$$\partial_{\mathbf{v}} f \equiv \frac{\partial f}{\partial \mathbf{v}} \equiv \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}$$

Change rate at **x** in direction **v**



Gradient in 2D

$$\frac{\partial f}{\partial \mathbf{v}} \equiv \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} = \lim_{h \to 0} \frac{f(x_1 + hv_1, x_2 + hv_2) - f(x_1, x_2)}{h}$$

$$\lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} = \partial_1 f \cdot v_1 + \partial_2 f \cdot v_2$$

$$\operatorname{grad} f \equiv \nabla f \equiv \left[\partial_1 f, \partial_2 f\right] \qquad \frac{\partial f}{\partial \mathbf{v}} = \nabla f \cdot \mathbf{v}$$

Gradient Meaning

$$\frac{\partial f}{\partial \mathbf{v}} = \nabla f \cdot \mathbf{v} \qquad \left|\nabla f \cdot \mathbf{v}\right| \le \left\|\nabla f\right\|_2 \cdot \left\|\mathbf{v}\right\|_2$$

Cauchy-Schwartz

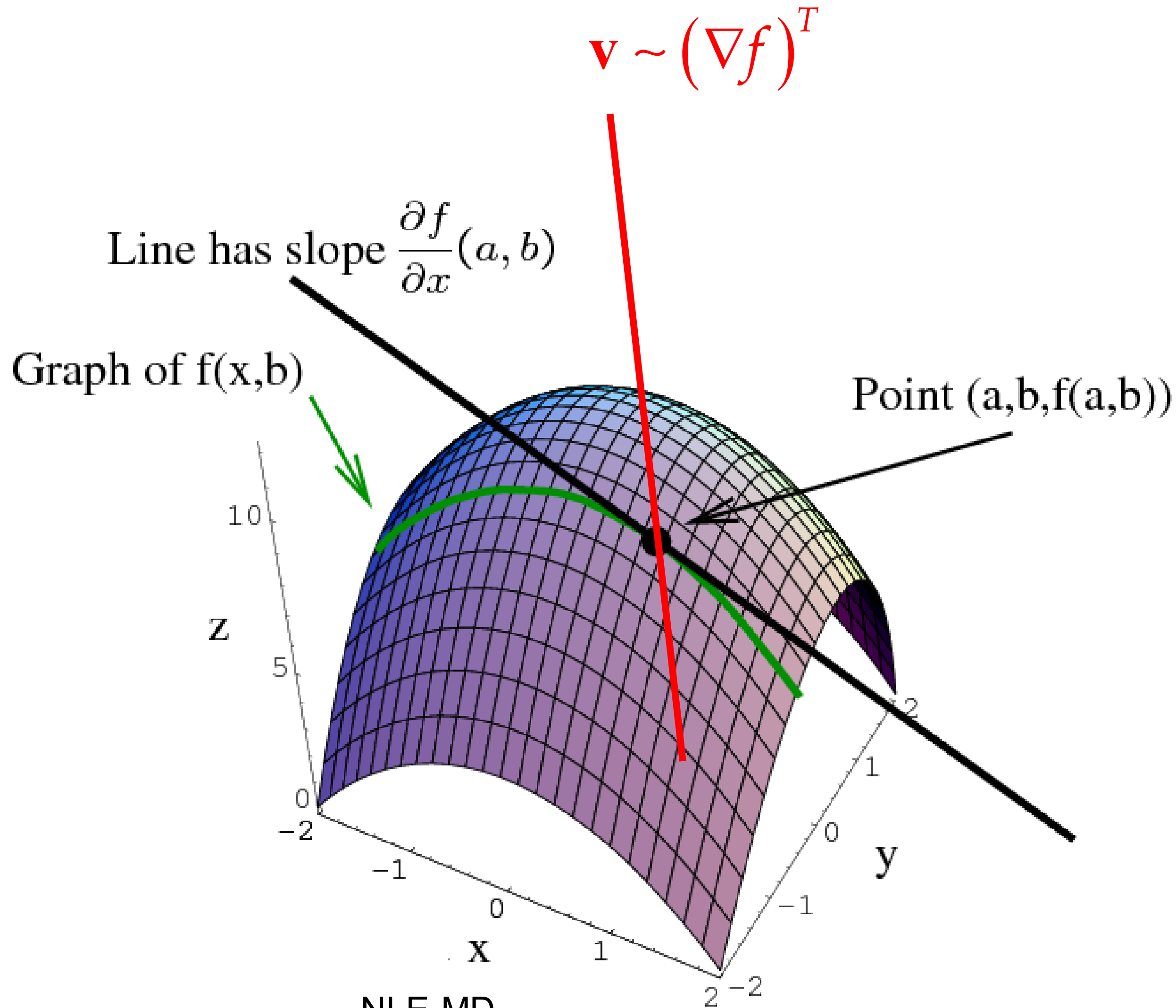
$$\left\|\mathbf{v}\right\|_2 = 1 \qquad \left|\nabla f \cdot \mathbf{v}\right| \le \left\|\nabla f\right\|_2$$

$$\mathbf{v} \sim \left(\nabla f\right)^T \qquad \left|\nabla f \cdot \mathbf{v}\right| = \left\|\nabla f\right\|_2$$

$$\max_{\mathbf{v}} \left|\frac{\partial f}{\partial \mathbf{v}}\right| = \max_{\mathbf{v}} \left|\nabla f \cdot \mathbf{v}\right|$$

Is achieved when $\mathbf{v} \sim \left(\nabla f\right)^T$

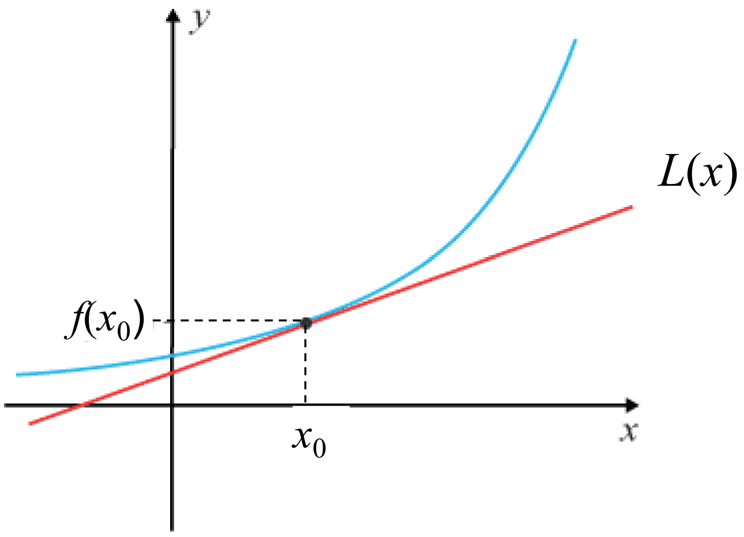
Greatest Increase



Linearization in 1D

In 1D: $\nabla f(x_0) \equiv f'(x_0)$

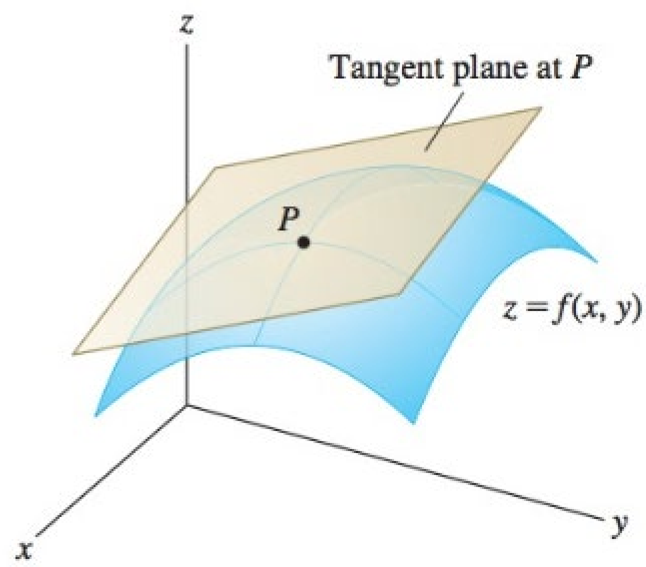
$$L(x) = f(x_0) + f'(x_0)\left(x - x_0\right) = f(x_0) + \nabla f(x_0)\left(x - x_0\right)$$



$$f(x) \approx L(x)$$

Tangent Plane

$$L(x,y) = f(x_0,y_0) + f_x(x_0,y_0)\left(x - x_0\right) + f_y(x_0,y_0)\left(y - y_0\right)$$



$$f(x,y) \approx L(x,y)$$

Linearization in 2D

$$L(x) = f(x_0) + f'(x_0)\left(x - x_0\right) = f(x_0) + \nabla f(x_0)\left(x - x_0\right)$$

$$L(x,y) = f(x_0, y_0) + f_x(x_0, y_0)\left(x - x_0\right) + f_y(x_0, y_0)\left(y - y_0\right)$$

$$\mathbf{x} \equiv \begin{bmatrix} x \\ y \end{bmatrix} \qquad \mathbf{x}_0 \equiv \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \qquad \nabla f(\mathbf{x}_0) = \left[f_x(x_0, y_0), f_y(x_0, y_0) \right]$$

$$\begin{bmatrix} r_1 & r_2 \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \equiv r_1 c_1 + r_2 c_2 \qquad f(\mathbf{x}) \approx L(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \left(\mathbf{x} - \mathbf{x}_0\right)$$

$$f(x) \approx L(x) = f(x_0) + f'(x_0)\left(x - x_0\right)$$

Vector-Valued Functions

$$\begin{cases} f(x,y)=0 \\ g(x,y)=0 \end{cases} \qquad \mathbf{F}(x,y)=\mathbf{F}(\mathbf{x})=\begin{bmatrix} f(x,y) \\ g(x,y) \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}=\mathbf{0}$$

$$f(x,y)\approx f(x_0,y_0)+f_x(x_0,y_0)\left(x-x_0\right)+f_y(x_0,y_0)\left(y-y_0\right)$$
$$g(x,y)\approx g(x_0,y_0)+g_x(x_0,y_0)\left(x-x_0\right)+g_y(x_0,y_0)\left(y-y_0\right)$$

$$\begin{bmatrix} f(x,y) \\ g(x,y) \end{bmatrix}\approx\begin{bmatrix} f(x_0,y_0) \\ g(x_0,y_0) \end{bmatrix}+\begin{bmatrix} f_x(x_0,y_0)\left(x-x_0\right)+f_y(x_0,y_0)\left(y-x_0\right) \\ g_x(x_0,y_0)\left(x-x_0\right)+g_y(x_0,y_0)\left(y-y_0\right) \end{bmatrix}$$

Jacobian Matrix

$$\begin{bmatrix} f(x,y) \\ g(x,y) \end{bmatrix} \approx \begin{bmatrix} f(x_0,y_0) \\ g(x_0,y_0) \end{bmatrix} + \begin{bmatrix} f_x(x_0,y_0)\left(x-x_0\right)+f_y(x_0,y_0)\left(y-x_0\right) \\ g_x(x_0,y_0)\left(x-x_0\right)+g_y(x_0,y_0)\left(y-y_0\right) \end{bmatrix}$$

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}_0) + \begin{bmatrix} f_x(x_0,y_0) & f_y(x_0,y_0) \\ g_x(x_0,y_0) & g_y(x_0,y_0) \end{bmatrix} \cdot \begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix}$$

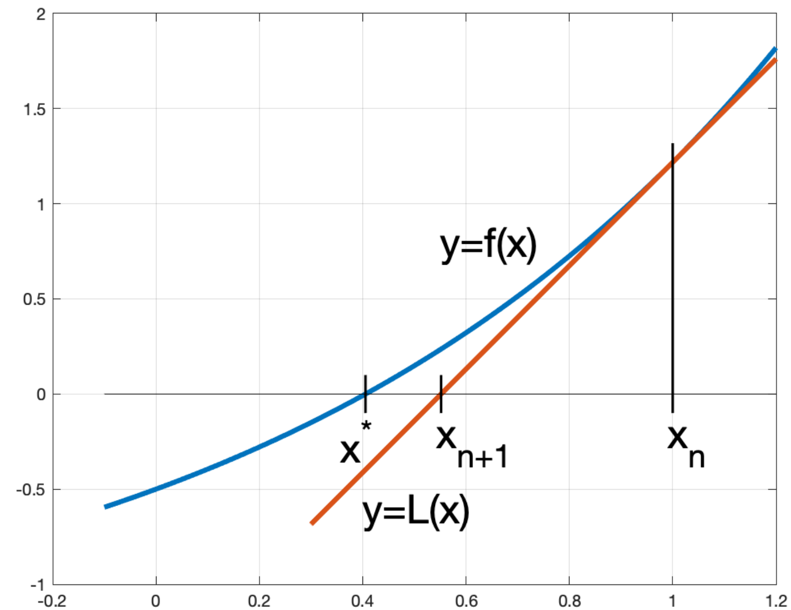
$$\mathbf{J}(\mathbf{x}_0) = \mathbf{J}(x_0,y_0) \equiv \begin{bmatrix} f_x(x_0,y_0) & f_y(x_0,y_0) \\ g_x(x_0,y_0) & g_y(x_0,y_0) \end{bmatrix}$$

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{L}(\mathbf{x}) \equiv \mathbf{F}(\mathbf{x}_0) + \mathbf{J}(\mathbf{x}_0)\cdot\left(\mathbf{x}-\mathbf{x}_0\right)$$

$$f(x) \approx L(x) = f(x_0) + f'(x_0)\left(x-x_0\right)$$

Newton's Method in 1D

$f(x) = 0 \qquad f(x),\ x \in \mathbb{R}$



Let root be x^* $\quad f(x^*) = 0$

Linearization about x_n

$$f(x) \approx L(x) = f(x_n) + f'(x_n)\left(x - x_n\right)$$

Let x_{n+1} be solution to $\quad 0 = L(x_{n+1}) = f(x_n) + f'(x_n)\left(x_{n+1} - x_n\right)$

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

Newton's Method in MD

$$\mathbf{F}(\mathbf{x}) = \mathbf{0} \qquad \mathbf{F}(\mathbf{x}),\ \ \mathbf{x} \in \mathbb{R}^m \qquad \text{Let root be } \mathbf{x}^* \quad \mathbf{F}(\mathbf{x}^*) = \mathbf{0}$$

Linearization about $\mathbf{x}_n$ $\quad \mathbf{F}(\mathbf{x}) \approx \mathbf{L}(\mathbf{x}) = \mathbf{F}(\mathbf{x}_n) + \mathbf{J}(\mathbf{x}_n)\left(\mathbf{x} - \mathbf{x}_n\right)$

Let $\mathbf{x}_{n+1}$ be a solution to $\quad \mathbf{0} = \mathbf{L}(\mathbf{x}_{n+1}) = \mathbf{F}(\mathbf{x}_n) + \mathbf{J}(\mathbf{x}_n)\left(\mathbf{x}_{n+1} - \mathbf{x}_n\right)$

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{J}(\mathbf{x}_n)^{-1}\mathbf{F}(\mathbf{x}_n)$$

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

Alternative Formulation

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{J}(\mathbf{x}_n)^{-1}\mathbf{F}(\mathbf{x}_n)$$

$$\mathbf{J}(\mathbf{x}_n)\boldsymbol{\delta}_n = -\mathbf{F}(\mathbf{x}_n)$$

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \boldsymbol{\delta}_n$$

Remarks on Newton's

- Local convergence
- Quadratic convergence if **J** is non-singular
- Convergence is fast
- Requires all partial derivatives

Euclidean length $$\left\| \mathbf{x}_{n+1} - \mathbf{x}^* \right\| \approx C \left\| \mathbf{x}_n - \mathbf{x}^* \right\|^2$$

Example in 2D

$$\begin{cases} y - x^3 = 0 \\ x^2 + y^2 = 1 \end{cases}$$

$$\mathbf{F}(x,y) = \begin{bmatrix} y - x^3 \\ x^2 + y^2 - 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$f(x,y) = y - x^3$$
$$g(x,y) = x^2 + y^2 - 1$$

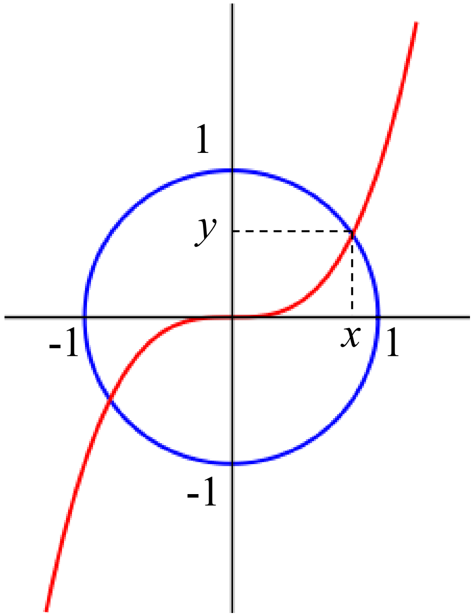
$$\mathbf{J}(x,y) = \begin{bmatrix} -3x^2 & 1 \\ 2x & 2y \end{bmatrix}$$

Analytic Solution

$$x^6 + x^2 = 1$$

$$t^3 + t = 1$$

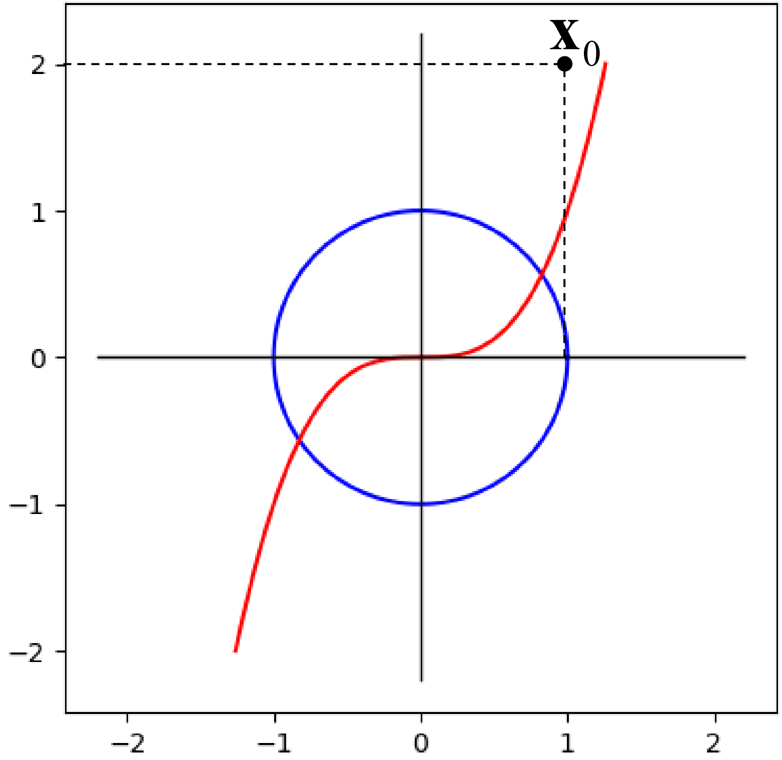
$$t^3 + mt = n$$



$$t = \sqrt[3]{\frac{n}{2} + \sqrt{\left(\frac{n}{2}\right)^2 + \left(\frac{m}{3}\right)^3}} - \sqrt[3]{-\frac{n}{2} + \sqrt{\left(\frac{n}{2}\right)^2 + \left(\frac{m}{3}\right)^3}}$$

$$x = 0.826031357654187 \quad y = 0.563624162161258$$

Initial Guess



$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad x_0 = 1 \qquad y_0 = 2$$

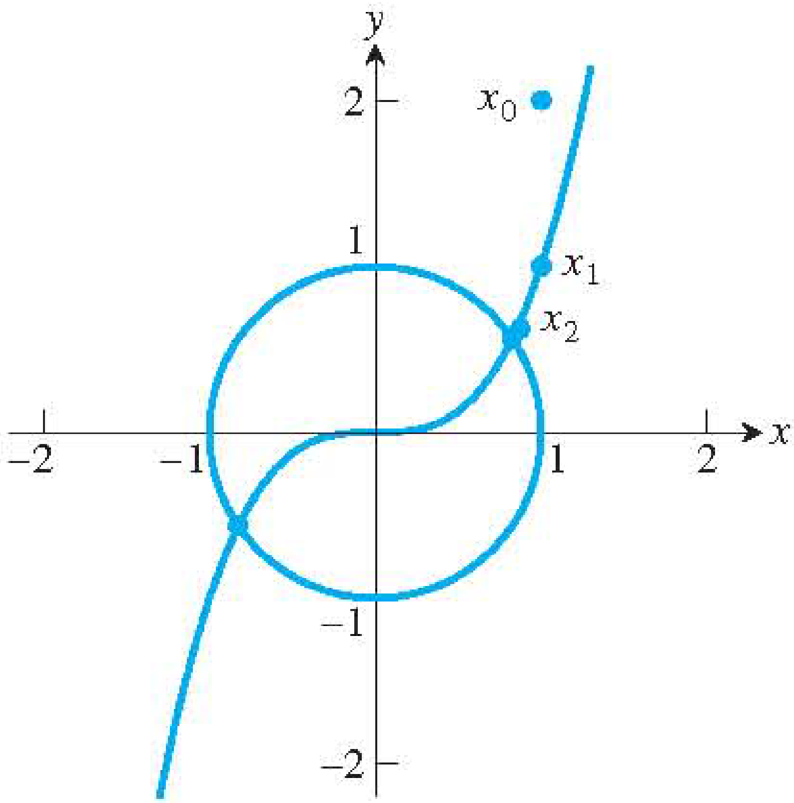
First Iteration

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \begin{aligned} x_0 &= 1 \\ y_0 &= 2 \end{aligned} \qquad \mathbf{J}(x,y) = \begin{bmatrix} -3x^2 & 1 \\ 2x & 2y \end{bmatrix} \qquad \mathbf{J}(x_0, y_0) = \begin{bmatrix} -3 & 1 \\ 2 & 4 \end{bmatrix}$$

$$\begin{bmatrix} -3 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \mathbf{J}(\mathbf{x}_0)\boldsymbol{\delta}_0 = -\mathbf{F}(\mathbf{x}_0) = -\begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$$\mathbf{x}_1 = \mathbf{x}_0 + \boldsymbol{\delta}_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Second Iteration



Iteration History

Step	x_n	y_n
0	1.00000000000000	2.00000000000000
1	1.00000000000000	1.00000000000000
2	0.87500000000000	0.62500000000000
3	0.82903634826712	0.56434911242604
4	0.82604010817065	0.56361977350284
5	0.82603135773241	0.56362416213163
6	0.82603135765419	0.56362416216126
7	0.82603135765419	0.56362416216126
	0.82603135765419	0.56362416216126

Approximate Jacobian

$$\mathbf{F}(x,y)=\begin{bmatrix} f(x,y) \\ g(x,y) \end{bmatrix}$$

- No analytic form
- Defined implicitly
- Found iteratively
- Given as black box

$$\frac{\mathbf{F}(x+h,y)-\mathbf{F}(x,y)}{h}=\begin{bmatrix} \big(f(x+h,y)-f(x,y)\big)/h \\ \big(g(x+h,y)-g(x,y)\big)/h \end{bmatrix}=\begin{bmatrix} \tilde{f}_x(x,y) \\ \tilde{g}_x(x,y) \end{bmatrix}$$

$$\mathbf{J}(x,y)\approx\begin{bmatrix} \tilde{f}_x(x,y) & \tilde{f}_y(x,y) \\ \tilde{g}_x(x_0,y) & \tilde{g}_y(x,y) \end{bmatrix}$$

Secant Method in MD

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{J}(\mathbf{x}_n)^{-1}\mathbf{F}(\mathbf{x}_n)$$

$$x_{n+1} = x_n - q_n^{-1} f(x_n) \qquad q_n = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$$

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{Q}_n^{-1}\mathbf{F}(\mathbf{x}_n) \qquad q_n \cdot \Delta x_n = \Delta f_n$$

$$\mathbf{Q}_n \cdot \left(\mathbf{x}_n - \mathbf{x}_{n-1}\right) = \mathbf{F}(\mathbf{x}_n) - \mathbf{F}(\mathbf{x}_{n-1})$$

Broyden's Method

Best appr. $\mathbf{Q}_n \approx \mathbf{J}(\mathbf{x}_n)$ $\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{Q}_n^{-1}\mathbf{F}(\mathbf{x}_n)$

Cond 1: $\mathbf{Q}_{n+1} \approx \mathbf{Q}_n$

Cond 2: $\mathbf{Q}_{n+1} \cdot \left(\mathbf{x}_{n+1} - \mathbf{x}_n\right) = \mathbf{F}(\mathbf{x}_{n+1}) - \mathbf{F}(\mathbf{x}_n)$

$$\mathbf{Q}_{n+1} \cdot \Delta\mathbf{x}_{n+1} = \Delta\mathbf{F}(\mathbf{x}_{n+1})$$

$$\mathbf{Q}_{n+1} = \mathbf{Q}_n + \frac{1}{\left\|\Delta\mathbf{x}_{n+1}\right\|^2}\left[\Delta\mathbf{F}(\mathbf{x}_{n+1}) - \mathbf{Q}_n\Delta\mathbf{x}_{n+1}\right] \cdot \Delta\mathbf{x}_{n+1}^T$$

Column by Row

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \cdot \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} = \begin{bmatrix} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \\ a_3b_1 & a_3b_2 & a_3b_3 \\ a_4b_1 & a_4b_2 & a_4b_3 \end{bmatrix}$$

Jacobian in multi-D

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_1(x_1,\ldots,x_m) \\ f_2(x_1,\ldots,x_m) \\ \vdots \\ f_n(x_1,\ldots,x_m) \end{bmatrix}$$

$$\partial_1 f_1 \equiv f'_{1,x_1}(x_1,\ldots,x_m)$$

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \partial_1 f_1 & \partial_2 f_1 & \cdots & \partial_m f_1 \\ \partial_1 f_2 & \partial_2 f_2 & \cdots & \partial_m f_2 \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_n & \partial_2 f_n & \cdots & \partial_m f_n \end{bmatrix}$$

Important

- Multivariate functions
- Gradient and its geometical meaning
- Linearization in 2D and MD
- Vector-valued functions
- Jacobian matrix
- Newton’s method in 2D and MD
- Quasi-Newton Method