ROB 101 - Fall 2021

Solutions of Nonlinear Equations (Vector-valued Functions)

November 8, 2021

UNIVERSITY OF MICHIGAN

Learning Objectives

- Extend our horizons from linear equations to nonlinear equations.
- Appreciate the power of using algorithms to iteratively construct approximate solutions to a problem.
- Accomplish all of this without assuming a background in Calculus.

Outcomes

- Linear approximations of nonlinear functions.

- Extensions of these ideas to vector-valued functions of several variables, that is $f : \mathbb{R}^m \to \mathbb{R}^n$, with key notions being the gradient and Jacobian of a function.

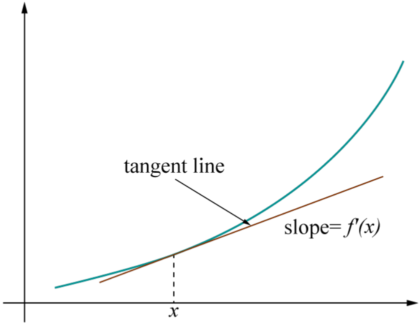
Essence Newton's Method

- Recall Newton's method is a root-finding algorithm which produces successively better approximations to the roots (or zeroes) of a real-valued function.

- The idea is to start with an initial guess (reasonably close to the true root), then to *approximate the function by its tangent line*, and finally to compute the x-intercept of this tangent line by elementary algebra.

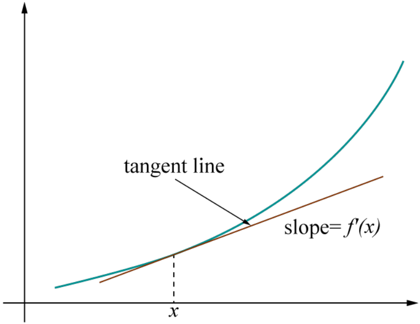
Core concept: approximate the nonlinear function by its tangent line at the current operating point.

Q. What is the best linear approximation of a nonlinear function?



- The linear function $y(x)$ that passes through the point (x_0,y_0) with slope a can be written as

$$y(x) = y_0 + a\,(x - x_0)\,.$$



Linear Approximation at a Point

- We define the *linear approximation of a function at a point* x_0 by taking $y_0 := f(x_0)$ and $a := \frac{df(x_0)}{dx} = f'(x_0)$.
- This gives us

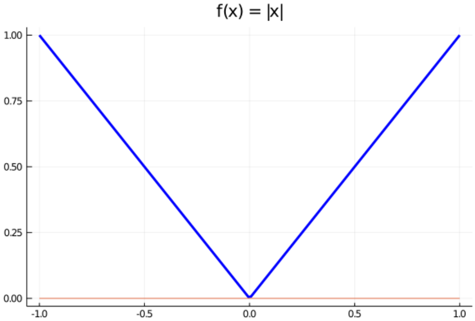
$$\boxed{f(x) \approx f(x_0) + \frac{df(x_0)}{dx}\left(x - x_0\right).}$$

Numerical Approximations of a Derivative

For a smooth (continuous and differentiable) function and sufficiently small h, the following *finite difference* approximations of the derivatives are possible.

- Forward difference $\frac{df(x_0)}{dx} \approx \frac{f(x_0+h)-f(x_0)}{h}$.
- Backward difference $\frac{df(x_0)}{dx} \approx \frac{f(x_0)-f(x_0-h)}{h}$.
- Symmetric or central difference $\frac{df(x_0)}{dx} \approx \frac{f(x_0+h)-f(x_0-h)}{2h}$.

Explain why the function $f(x) = |x|$ is not differentiable at $x_0 = 0$.



We compute

- forward difference:
$$\frac{df(0)}{dx} \approx \frac{f(0+h) - f(0)}{h} = \frac{|h| - 0}{h} = \frac{h}{h} = \boxed{+1},$$
- backward difference:
$$\frac{df(0)}{dx} \approx \frac{f(0) - f(0-h)}{h} = \frac{0 - |-h|}{h} = \frac{-h}{h} = \boxed{-1},$$
- and central difference:
$$\frac{df(0)}{dx} \approx \frac{f(0+h) - f(0-h)}{2h} = \frac{|h| - |-h|}{2h} = \frac{h-h}{2h} = \boxed{0}.$$

Remark

These three methods giving very different approximations to the "slope" at the origin is a strong hint that the function is not differentiable at the origin.
By following different paths as we approach x_0, approaching x_0 from the left versus the right for example, gives different answers for the "slope" of the function at x_0. In Calculus, you'll learn that this means the function is not differentiable at x_0.

Consider a function $f : \mathbb{R}^m \to \mathbb{R}^n$. We seek a means to build a linear approximation of the function near a given point $x_0 \in \mathbb{R}^m$.

- When $m = n = 1$, we can approximate a function by

$$f(x) \approx f(x_0) + \frac{df(x_0)}{dx}(x - x_0) =: f(x_0) + a(x - x_0).$$

Consider a function $f : \mathbb{R}^m \to \mathbb{R}^n$. We seek a means to build a linear approximation of the function near a given point $x_0 \in \mathbb{R}^m$.

- When $m = n = 1$, we can approximate a function by

$$f(x) \approx f(x_0) + \frac{df(x_0)}{dx}(x - x_0) =: f(x_0) + a(x - x_0).$$

- For the general case of $f : \mathbb{R}^m \to \mathbb{R}^n$, can we find an $n \times m$ matrix A such that

$$f(x) \approx f(x_0) + A(x - x_0).$$

Linear Approximation of $f : \mathbb{R}^m \to \mathbb{R}^n$

- A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is called a vector-valued function.
- We compute its best linear approximation via the generalization of the derivative $A := \frac{\partial f(x)}{\partial x}\Big|_{x=x_0}$ such that

$$f(x) \approx f(x_0) + A(x - x_0).$$

- The input is a vector such as $x \in \mathbb{R}^m$. If we use the standard basis of $\mathbb{R}^m$, we have

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_m = \sum_{i=1}^{m} x_i e_i.$$

- The input is a vector such as $x \in \mathbb{R}^m$. If we use the standard basis of $\mathbb{R}^m$, we have

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_m = \sum_{i=1}^{m} x_i e_i.$$

- Then we can use a finite difference approximation to compute each column of $A = \begin{bmatrix} a_1^{\mathrm{col}} & \dots & a_m^{\mathrm{col}} \end{bmatrix}$ as

$$\boxed{a_i^{\mathrm{col}} = \frac{\partial f(x_0)}{\partial x_i} = \frac{f(x_0 + he_i) - f(x_0 - he_i)}{2h}}$$

- For the special case of $n = 1$, $f : \mathbb{R}^m \to \mathbb{R}$, the matrix A is a row vector and called the *gradient* of f.
- The following notations are all common

$$A = \nabla f = \text{grad } f.$$

- The symbol ∇ (nabla or del) is a common notation to refer to the gradient of f.

- For the general case of $f : \mathbb{R}^m \to \mathbb{R}^n$, A is an $n \times m$ matrix and called the *Jacobian* of f, i.e.,

$$A_{n\times m} = \begin{bmatrix} a_1^{\text{col}} & \dots & a_m^{\text{col}} \end{bmatrix} = \frac{\partial f(x)}{\partial x}.$$

- Each column of the Jacobian $a_i^{\text{col}} = \frac{\partial f(x)}{\partial x_i} \in \mathbb{R}^n$ shows the rate of change of f along e_i.

For the function

$$f(x_1,x_2,x_3) := \left[\begin{array}{c} x_1x_2x_3 \\ \log(2+\cos(x_1)) + x_2^{x_1} \\ \frac{x_1x_3}{1+x_2^2} \end{array}\right],$$

compute its Jacobian at the point

$$x_0 = \left[\begin{array}{c} \pi \\ 1.0 \\ 2.0 \end{array}\right]$$

and evaluate the "accuracy" of its linear approximation.

Using $h = 0.001$ and central differences we get

$$a_1^{\text{col}} = \frac{\partial f(x_0)}{\partial x_1} = \begin{bmatrix} 2.0 \\ 0.0 \\ 1.0 \end{bmatrix},$$

$$a_2^{\text{col}} = \frac{\partial f(x_0)}{\partial x_2} = \begin{bmatrix} 6.2832 \\ 3.1416 \\ -3.1416 \end{bmatrix},$$

$$a_3^{\text{col}} = \frac{\partial f(x_0)}{\partial x_3} = \begin{bmatrix} 3.1416 \\ 0.0000 \\ 1.5708 \end{bmatrix}.$$

The Jacobian at x_0 is

$$A := \frac{\partial f(x_0)}{\partial x} = \begin{bmatrix} 2.0000 & 6.2832 & 3.1416 \\ 0.0000 & 3.1416 & 0.0000 \\ 1.0000 & -3.1416 & 1.5708 \end{bmatrix},$$

and the linear approximation is

$$f(x) \approx f(x_0) + A(x - x_0) =$$
$$\begin{bmatrix} 6.2832 \\ 1.0000 \\ 3.1416 \end{bmatrix} + \begin{bmatrix} 2.0000 & 6.2832 & 3.1416 \\ 0.0000 & 3.1416 & 0.0000 \\ 1.0000 & -3.1416 & 1.5708 \end{bmatrix} \begin{bmatrix} x_1 - \pi \\ x_2 - 1.0 \\ x_3 - 2.0 \end{bmatrix}.$$

- To assess the quality of the linear approximation, we measure the error defined as

$$e(x) := \|f(x) - f_{\text{lin}}(x)\|,$$

where $f_{\text{lin}}(x) := f(x_0) + A(x - x_0)$.

- To assess the quality of the linear approximation, we measure the error defined as

$$e(x) := \|f(x) - f_{\mathrm{lin}}(x)\|,$$

where $f_{\mathrm{lin}}(x) := f(x_0) + A(x - x_0)$.

- We will seek to estimate the maximum value of $e(x)$ over a region containing the point x_0. Define

$$S(x_0) := \{x \in \mathbb{R}^3 \mid |x_i - x_{0i}| \le d, i = 1{,}2{,}3\}$$

and

$$\mathrm{Max\ Error} := \max_{x \in S(x_0)} e(x) = \max_{x \in S(x_0)} \|f(x) - f_{\mathrm{lin}}(x)\|.$$

For $d = 0.25$, we used a "random search" routine and estimated that

$$\text{Max Error} = 0.12.$$

To put this into context,

$$\max_{x \in S(x_0)} \|f(x)\| = 8.47,$$

and thus the relative error is about 1.5%.

Let's switch to the Julia notebook for some more examples.

Next Time

- Newton-Raphson Method
- Read Chapter 11 of ROB 101 Book