Root Finding

Marcin Chrzaszcz, Danny van Dyk
mchrzasz@cern.ch, danny.van.dyk@gmail.com

University of Zurich UZH

Numerical Methods,
26. September, 2016

Plan for today

- General problem of finding the root of a function
  ξ is a root of f iff $f(\xi) = 0$.

- Begin with $D = 1$
  What properties of $f(x)$ can we use to find one, any, or all roots of f? What are the requirements on $f(x)$?

- How about $D > 1$?
  Can we generalize root finding from $D = 1$ do arbitrary D?

General iterative procedure

Let's assume that

- $f(x)$ is a our function of interest,
- ξ is the only root of f,
- we have a point x_0 close to ξ

We now want to find a sequence $\{x_0, x_1, \dots\}$ that

1. converges toward ξ: $\lim_{k\to\infty} x_k = \xi$,
2. is iterative: $x_{k+1} = \Phi[f](x_k)$

We can attempt to Taylor expand f around x_0 to order N, in order to obtain the generator $\Phi[f]$ for the iteration:

Expansion around x_0

$$f(\xi) = 0 = \sum_{n=0}^{N} \frac{(\xi - x_0)^n}{n!} f^{(n)}(x_0)$$

where $f^{(n)}(x_0)$ is the nth derivative of f at the position x_0.

Newton-Raphson method $(N = 1)$

Expansion for $N = 1$

$$f(\xi) = 0 = f(x_0) + (\xi - x_0) \cdot f'(x_0) + \mathcal{O}((\xi - x_0)^2) \qquad (*)$$

Algorithm:

1. start with index $k = 0$
2. solve equation (*), assuming a vanshing approximation error:

$$x_{k+1} \leftarrow x_k - \frac{f(x_k)}{f'(x_k)} \approx \xi + \mathcal{O}((\xi - x_k)^2)$$

$$\Phi[f](x) \equiv x - \frac{f(x)}{f'(x)}$$

3. if $f(x_{k+1}) \leq t$, where t is an a-prior threshold, then stop; otherwise jump back to step #2.

Illustration of Newton-Raphson

$\boldsymbol{k = 0}$:

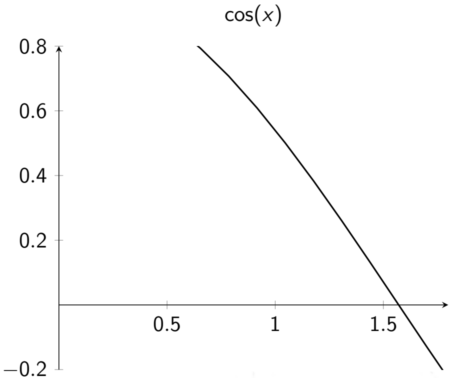


Illustration of Newton-Raphson

$\boldsymbol{k = 0}$:

$$x_0 := 1$$

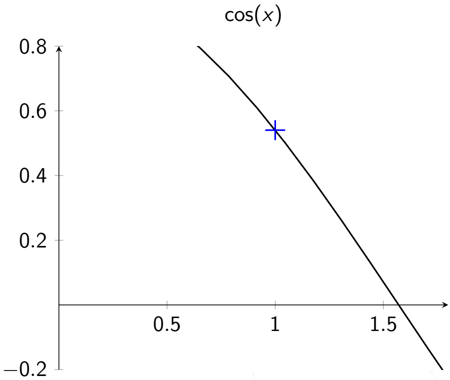


Illustration of Newton-Raphson

$\boldsymbol{k = 0}$:

$$x_0 := 1$$

$\boldsymbol{k = 1}$:

$$x_0 = 1 \qquad f(x_0) = +0.54$$
$$f'(x_0) = -0.84$$

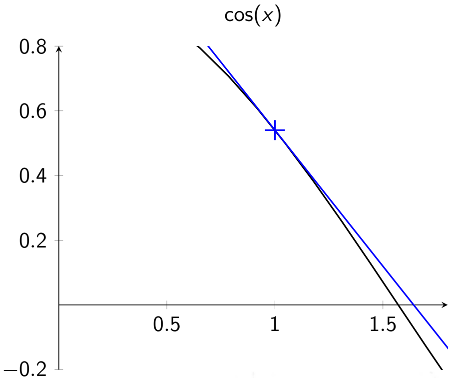


Illustration of Newton-Raphson

$\boldsymbol{k = 0}$:

$$x_0 := 1$$

$\boldsymbol{k = 1}$:

$$x_0 = 1 \qquad f(x_0) = +0.54$$
$$f'(x_0) = -0.84$$

$$x_1 \leftarrow 1.64$$

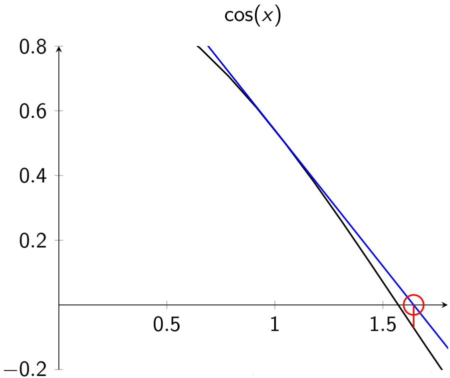


Illustration of Newton-Raphson

$\boldsymbol{k = 0}$:

$$x_0 := 1$$

$\boldsymbol{k = 1}$:

$$x_0 = 1 \qquad f(x_0) = +0.54$$
$$f'(x_0) = -0.84$$

$$x_1 \leftarrow 1.64$$

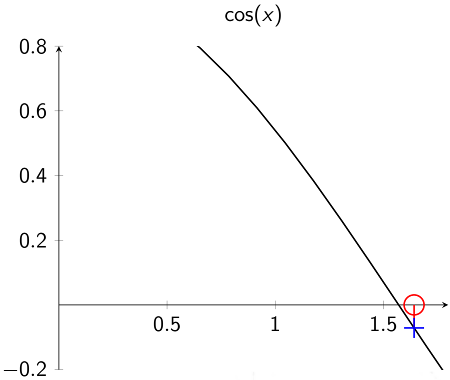


Illustration of Newton-Raphson

$\boldsymbol{k = 0}$:

$$x_0 := 1$$

$\boldsymbol{k = 1}$:

$$x_0 = 1 \qquad f(x_0) = +0.54$$
$$f'(x_0) = -0.84$$

$$x_1 \leftarrow 1.64$$

$\boldsymbol{k = 2}$:

$$x_1 = 1.64 \qquad f(x_0) = -0.07$$
$$f'(x_0) = -0.997$$

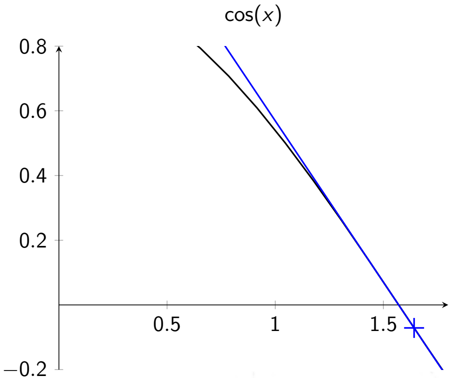


Illustration of Newton-Raphson

$\boldsymbol{k = 0}$:

$$x_0 := 1$$

$\boldsymbol{k = 1}$:

$$x_0 = 1 \qquad f(x_0) = +0.54$$
$$f'(x_0) = -0.84$$

$$x_1 \leftarrow 1.64$$

$\boldsymbol{k = 2}$:

$$x_1 = 1.64 \qquad f(x_0) = -0.07$$
$$f'(x_0) = -0.997$$

$$x_2 \leftarrow 1.57$$

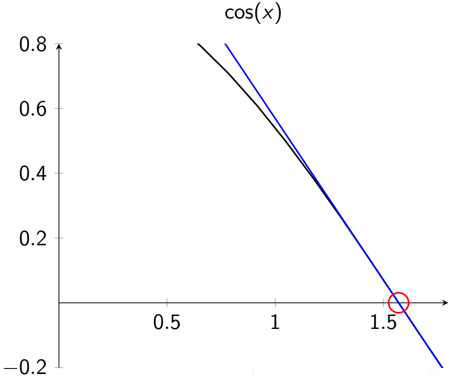


Illustration of Newton-Raphson

$\boldsymbol{k = 0}$:

$$x_0 := 1$$

$\boldsymbol{k = 1}$:

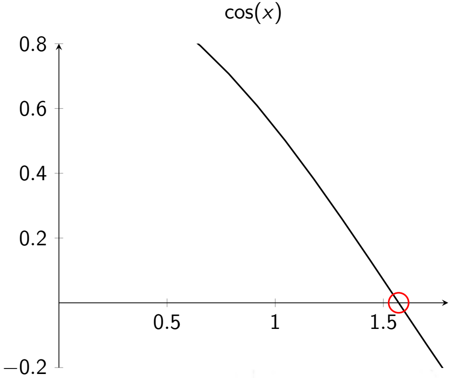
$$x_0 = 1 \qquad f(x_0) = +0.54$$
$$f'(x_0) = -0.84$$

$$x_1 \leftarrow 1.64$$

$\boldsymbol{k = 2}$:

$$x_1 = 1.64 \qquad f(x_0) = -0.07$$
$$f'(x_0) = -0.997$$

$$x_2 \leftarrow 1.57$$



k	x_k	$f(x_k)$
0	1.00	+0.54
1	1.64	−0.07
2	1.57	+0.00
ξ	1.5708	0

Modified Newton-Raphson method $(N = 2)$

Expansion for $N = 2$

$$f(\xi) = 0 = f(x_0) + (\xi - x_0) \cdot f'(x_0) + \frac{1}{2}(\xi - x_0)^2 f''(x_0) + \mathcal{O}((\xi - x_0)^3) \quad (*)$$

Algorithm:

1. start with index $k = 0$
2. solve equation (*), assuming a vanshing approximation error:

$$\begin{aligned} a_\pm &\leftarrow x_k - \frac{f'(x_k) \pm \sqrt{[f'(x_k)]^2 - 2f(x_k)f''(x_k)}}{f''(x_k)} \\ &\approx \xi + \mathcal{O}((\xi - x_k)^3) \end{aligned}$$

3. if $|f(a_+)| < |f(a_-)|$, then $x_{k+1} \leftarrow a_+$; otherwise $x_{k+1} \leftarrow a_-$
4. if $f(x_{k+1}) \leq t$, where t is an a-priori threshold, then stop; otherwise jump back to step #2.

Rate of convergence

Does this iterative procedure converge? If yes, how fast? How can we quantify the rate of convergence?

- we have local convergence of order $p \geq 1$, if for all $x \in U(\xi)$

$$||\Phi[f](x) - \xi|| \leq C \cdot ||x - \xi||^p \,, \quad \text{with} \quad C \geq 0\,.$$

  Note: if $p = 1$ then we must have $C < 1$.
- we have global convergence if $U(\xi) = \mathbb{R}$
- for $D = 1$ we can calculate p if $\Phi[f](x)$ is differentiable to sufficient degree:

$$\Phi(x) - \xi = \Phi(x) - \Phi(\xi) = \frac{(x-\xi)^p}{p!} + o(||x - \xi||^p)$$

Rate of convergence for Newton-Raphson

We have $\Phi(x) \equiv \Phi[f](x)$:

- $\Phi(\xi) = \xi$ by construction
- $\Phi'(\xi) = \frac{f(\xi) \cdot f''(\xi)}{[f'(\xi)]^2} = 0$ (*) by construction
- $\Phi''(\xi) = \frac{f''(\xi)}{f'(\xi)}$

We can therefore write

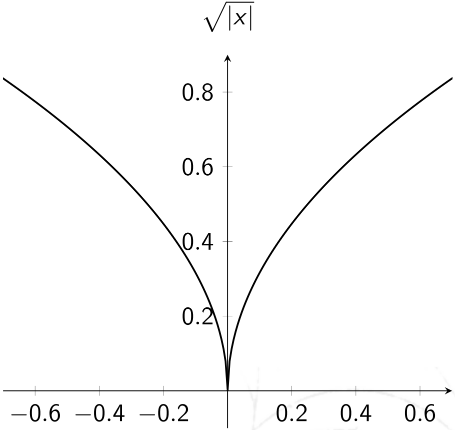
$$\Phi(x) - \Phi(\xi) = \frac{(x-\xi)^2}{2!} \Phi''(\xi) + o(||x-\xi||^2) \,. \tag{1}$$

The Newton-Raphson method converges therefore *at least quadratically*, i.e.: it is a second-order method.

(*): only if $f'(\xi) \neq 0$, which is equivalent to ξ is a simple root of f.

Pathological Example

$k = 0$:

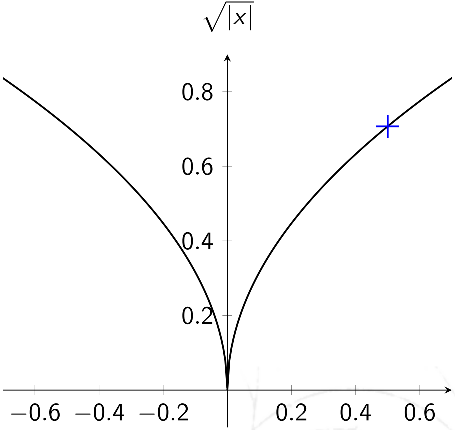


endless loop

Pathological Example

$\boldsymbol{k = 0}$:

$$x_0 := 0.5$$



endless loop

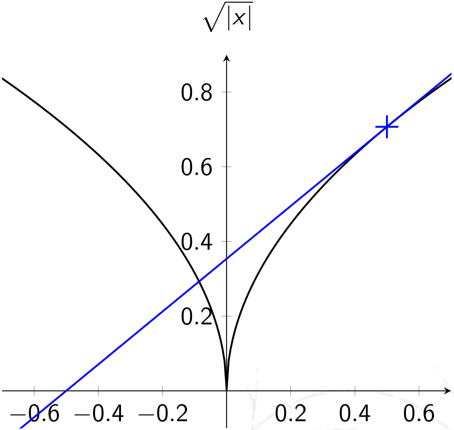
Pathological Example

$k = 0$:

$$x_0 := 0.5$$

$k = 1$:

$$x_0 = 0.5 \qquad f(x_0) = +0.707$$
$$f'(x_0) = -0.707$$



endless loop

Pathological Example

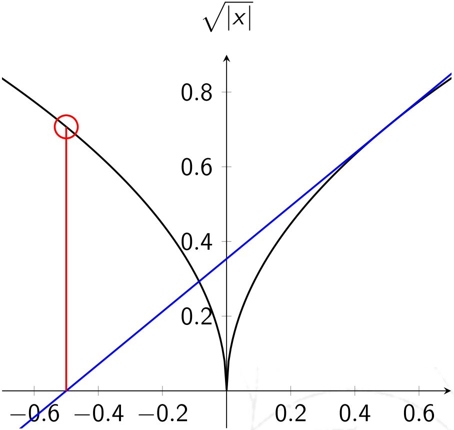
$\boldsymbol{k = 0}$:

$$x_0 := 0.5$$

$\boldsymbol{k = 1}$:

$$x_0 = 0.5 \qquad f(x_0) = +0.707$$
$$f'(x_0) = -0.707$$

$$\color{red}{x_1 \leftarrow -0.5}$$



endless loop

Pathological Example

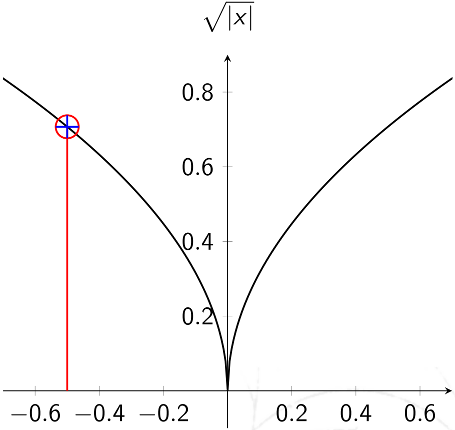
$\boldsymbol{k = 0}$:

$$x_0 := 0.5$$

$\boldsymbol{k = 1}$:

$$x_0 = 0.5 \qquad f(x_0) = +0.707$$
$$f'(x_0) = -0.707$$

$$x_1 \leftarrow -0.5$$



endless loop

Pathological Example

$\boldsymbol{k = 0}$:

$$x_0 := 0.5$$

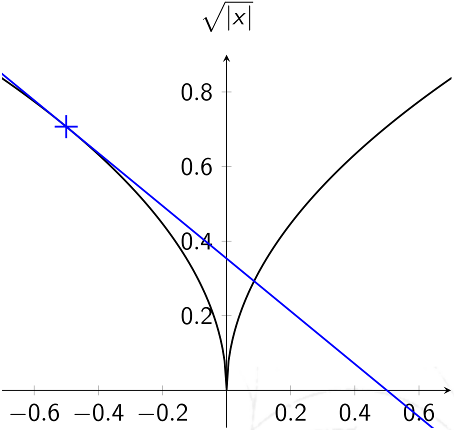
$\boldsymbol{k = 1}$:

$$x_0 = 0.5 \qquad f(x_0) = +0.707$$
$$f'(x_0) = -0.707$$

$$x_1 \leftarrow -0.5$$

$\boldsymbol{k = 2}$:

$$x_1 = -0.5 \qquad f(x_0) = +0.707$$
$$f'(x_0) = -0.707$$



endless loop

Pathological Example

$\boldsymbol{k = 0}$:

$$x_0 := 0.5$$

$\boldsymbol{k = 1}$:

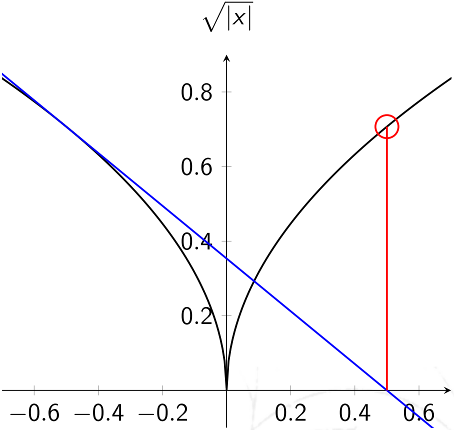
$$x_0 = 0.5 \qquad f(x_0) = +0.707$$
$$f'(x_0) = -0.707$$

$$x_1 \leftarrow -0.5$$

$\boldsymbol{k = 2}$:

$$x_1 = -0.5 \qquad f(x_0) = +0.707$$
$$f'(x_0) = -0.707$$

$$x_2 \leftarrow +0.5$$



endless loop

Pitfalls of Newton-Raphson

We had previously discussed that Newton-Raphson is guaranteed to converge in an environment $U(\xi)$ of the root ξ.

However, it may fail to converge or converge only very slowly...

- if $U(\xi)$ is a small interval and the initial point lies outside of U,
- if the algorithm encounters a stationary point χ of f (i.e. $f'(\chi) = 0$),
- if ξ is a m-multiple root of f (i.e. $f(x) \sim (x - \xi)^m \dots$)

The last point can be accomodated for: If you know the mulitplicity m of the root, modify the Newton-Raphson step to read:

$$x_{k+1} \leftarrow x_k - m \frac{f(x_k)}{f'(x_k)}$$

Basically, this rescales the derivative from $f'(x) \to f'(x)/m$.

Honorable mention: Horner Scheme

If $f(x) \equiv p_n(x)$ is a polynomial of degree n in x, there might be multiple roots of f, i.e. roots ξ_n with $f'(\xi_n) = 0$.

The Horner Scheme allows to efficiently calculate all the (multiple) roots of the polynomial $p_n(x)$.

However, this is of limited use: The polynomials dominantly arise in numerical computations as the characteric polynomial χ_M of a matrix M. In these cases, the roots of χ_M correspond to eigenvalue of M. However, the are better/more stables ways to numerically compute all eigenvalues of M, instead of finding all root of χ_M.

Intermission: Derivatives

Both the original and the modified Newton-Raphson methods involve taking the derivative f' of our target function f at arbitrary points. Of course, if we can analytically calculate the derivative, there is no problem.

What about functions that we can only evaluate numerically? (Think: Experiments, numerical calculations)

Enter the field of *numerical differentiation*!

Intermission: Derivatives

How can we obtain the derivative $f'(x)$ numerically? Let's start from an expansion of $f(x)$ close to the point of interest x_0:

$$f(x_0 + 2h) = f(x_0) + f'(x_0) \cdot 2h + \frac{f''(x_0)}{2} \cdot 4h^2 + \dots$$

$$f(x_0 + h) = f(x_0) + f'(x_0) \cdot h + \frac{f''(x_0)}{2} \cdot h^2 + \dots$$

$$f(x_0) = f(x_0)$$

$$f(x_0 - h) = f(x_0) - f'(x_0) \cdot h + \frac{f''(x_0)}{2} \cdot h^2 + \dots$$

$$f(x_0 - 2h) = f(x_0) - f'(x_0) \cdot 2h + \frac{f''(x_0)}{2} \cdot 4h^2 + \dots$$

Variant #1: Forward difference quotient

$$\frac{f(x_0+h) - f(x_0)}{h} = f'(x_0) + \frac{f''(x_0)}{2} \cdot h = f'(x_0) + o(h)$$

Intermission: Derivatives

How can we obtain the derivative $f'(x)$ numerically? Let's start from an expansion of $f(x)$ close to the point of interest x_0:

$$f(x_0+2h) = f(x_0) + f'(x_0)\cdot 2h + \frac{f''(x_0)}{2}\cdot 4h^2 + \dots$$

$$\boxed{f(x_0+h) = f(x_0) + f'(x_0)\cdot h + \frac{f''(x_0)}{2}\cdot h^2 + \dots}$$

$$f(x_0) = f(x_0)$$

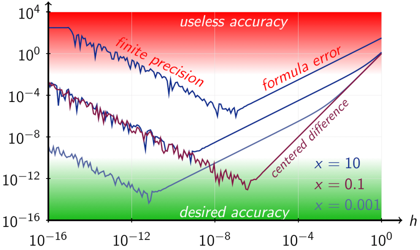
$$\boxed{f(x_0-h) = f(x_0) - f'(x_0)\cdot h + \frac{f''(x_0)}{2}\cdot h^2 + \dots}$$

$$f(x_0-2h) = f(x_0) - f'(x_0)\cdot 2h + \frac{f''(x_0)}{2}\cdot 4h^2 + \dots$$

Variant #2: Central difference quotient

$$\frac{f(x_0+h) - f(x_0-h)}{2h} = f'(x_0) + \frac{f'''(x_0)}{3}h^2 = f'(x_0) + o(h^2)$$

Intermission: How to choose h?



source: Wikipedia

- two sources of numerical errors compete:
  - formula errors, and
  - round-off (finite precision) errors

- formula error can be reduced through "intelligent formulae"
- round-off error can be reduced through (schematically):

$$x_{\text{approx}} \leftarrow x + h$$
$$h_{\text{approx}} \leftarrow x_{\text{approx}} - x$$
$$f' \leftarrow \frac{f(x_{\text{approx}}) - f(x)}{h_{\text{approx}}}$$

- optimal choice of h will be close to

$$h \sim \sqrt{\varepsilon} \cdot x$$

with ε: working precision

Intermission: Higher derivatives

What about $f''(x_0)$, $f^{(3)}(x_0)$ or even higher derivatives?

Central difference quotient for $f^{(n)}$

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{f(x_0+(n-2k)h)}{(2h)^n}=f^{(n)}(x_0)+o(h^2)$$

Intermission: Multiple partial derivatives

What about (multiple) partial derivatives $\partial_x \partial_y f(x, y)$?

Central difference "stencil" for $\partial_x \partial_y f(x, y)$

Combine what we learned previously:

$$\frac{af(+h_x, +h_y) + bf(+h_x, -h_y) + cf(-h_x, +h_y) + df(-h_x, -h_y)}{4h_x h_y}$$

$$= \ldots$$

Intermission: Multiple partial derivatives

What about (multiple) partial derivatives $\partial_x \partial_y f(x, y)$?

Central difference "stencil" for $\partial_x \partial_y f(x, y)$

Combine what we learned previously:

$$\frac{+f(+h_x, +h_y) - f(+h_x, -h_y) - f(-h_x, +h_y) + f(-h_x, -h_y)}{4 h_x h_y} = \partial x \partial_y f(x, y)|_{x=0, y=0} + o(h_x^2) + o(h_y^2)$$

End of intermission.

Newton-Raphson for $D > 1$?

What happens if we generalize to $D > 1$?

Let's consider the case $D = 2$ first, for a polynomial of degree 1

$$f(x_1, x_2) = (1 - x) \cdot (2 - y)$$

f has infinitely many roots:

- $x = 1$ and y arbitrary
- $y = 2$ and x arbitrary

In order to pin-point only one point in $D = 2$, we need to solve 2 equations:

$$f_1(x_1, x_2) = 0, \qquad f_2(x_1, x_2) = 0$$

In D dimensions, we will need to solve D simultaneous equations!

Generalization to $D > 1$

Substitute:

$$
\begin{aligned}
f(x) \in \mathbb{R} \quad &\rightarrow \quad f(x) \in \mathbb{R}^D \\
x \in \mathbb{R} \quad &\rightarrow \quad x \in \mathbb{R}^D \\
f'(x) \in \mathbb{R} \quad &\rightarrow \quad [D_f]_{ij} \equiv \frac{\partial f_i(x)}{\partial x_j} \in \mathbb{R}^{D \times D}
\end{aligned}
$$

Then the Newton-Raphson steps reads

$$
x^{(k+1)} \leftarrow x^{(k)} - \left[D_f(x^{(k)}) \right]^{-1} f(x^{(k)}) \, .
$$

Problems:

- the functional matrix $D_f(x^{(k)})$ might be singular, or numerically close to being singular
- inverting D_f is generally computationally expensive!

In the next two lectures, Marcin will discuss how to solve systems of linear equations. These methods will also introduce you to methods to numerically compute the above $x^{(k+1)}$ from $D_f(x^{(k)})$.

Backup