

## 12. Newton's method

- sets of nonlinear equations
- the derivative matrix and linearization
- Newton's method
- examples

# Set of nonlinear equations

$n$  nonlinear equations in  $n$  variables

$$\begin{aligned} f_1(x_1, \dots, x_n) &= 0 \\ f_2(x_1, \dots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, \dots, x_n) &= 0 \end{aligned}$$

in vector notation:

$$f(x) = 0$$

where  $x \in \mathbf{R}^n$  and  $f : \mathbf{R}^n \rightarrow \mathbf{R}^n$  are defined as

$$f(x) = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ f_2(x_1, \dots, x_n) \\ \vdots \\ f_n(x_1, \dots, x_n) \end{bmatrix}$$

# Derivative matrix

$f : \mathbf{R}^n \rightarrow \mathbf{R}^n$ , differentiable

$$Df(\hat{x}) = \begin{bmatrix} \frac{\partial f_1(\hat{x})}{\partial x_1} & \frac{\partial f_1(\hat{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\hat{x})}{\partial x_n} \\ \frac{\partial f_2(\hat{x})}{\partial x_1} & \frac{\partial f_2(\hat{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\hat{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(\hat{x})}{\partial x_1} & \frac{\partial f_n(\hat{x})}{\partial x_2} & \cdots & \frac{\partial f_n(\hat{x})}{\partial x_n} \end{bmatrix}$$

is the *derivative matrix* (or *Jacobian matrix*) evaluated at  $\hat{x}$

**example:**  $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$

$$f(x) = \begin{bmatrix} e^{2x_1+x_2} - x_1 \\ x_1^2 - x_2 \end{bmatrix}$$

$$Df(\hat{x}) = \begin{bmatrix} 2e^{2\hat{x}_1+\hat{x}_2} - 1 & e^{2\hat{x}_1+\hat{x}_2} \\ 2\hat{x}_1 & -1 \end{bmatrix}$$

# Linearization

if  $f : \mathbf{R}^n \rightarrow \mathbf{R}^n$  is differentiable at  $\hat{x} \in \mathbf{R}^n$ , and  $x$  is near  $\hat{x}$ , then

$$f_i(x) \approx f_i(\hat{x}) + \frac{\partial f_i(\hat{x})}{\partial x_1}(x_1 - \hat{x}_1) + \frac{\partial f_i(\hat{x})}{\partial x_2}(x_2 - \hat{x}_2) + \cdots + \frac{\partial f_i(\hat{x})}{\partial x_n}(x_n - \hat{x}_n)$$

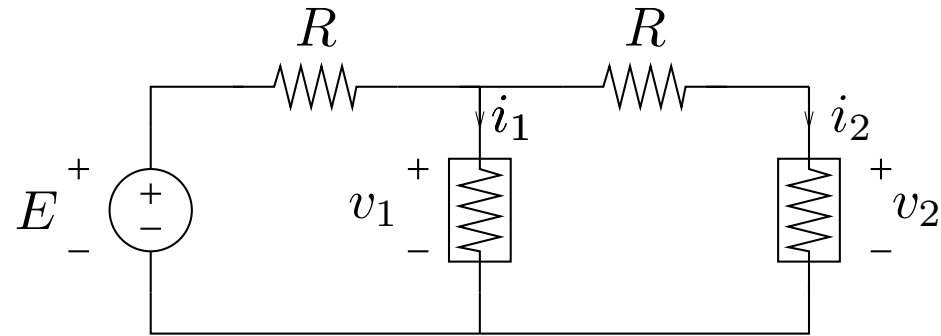
in matrix notation:

$$f(x) \approx f(\hat{x}) + Df(\hat{x})(x - \hat{x})$$

**example** (previous page):  $f$  linearized around  $\hat{x} = 0$  is

$$f(x) \approx \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

## Example: nonlinear static circuit



two nonlinear resistors with  $i$ - $v$  characteristics  $i_1 = g_1(v_1)$ ,  $i_2 = g_2(v_2)$

circuit equations

$$f_1(v_1, v_2) = g_1(v_1) + \frac{v_1 - E}{R} + \frac{v_1 - v_2}{R} = 0$$

$$f_2(v_1, v_2) = g_2(v_2) + \frac{v_2 - v_1}{R} = 0$$

two nonlinear equations in two variables  $v_1$ ,  $v_2$

derivative matrix evaluated at  $\hat{v}_1, \hat{v}_2$ :

$$Df(\hat{v}) = \begin{bmatrix} g'_1(\hat{v}_1) + 2/R & -1/R \\ -1/R & g'_2(\hat{v}_2) + 1/R \end{bmatrix}$$

linearized equations around  $\hat{v}_1, \hat{v}_2$ :

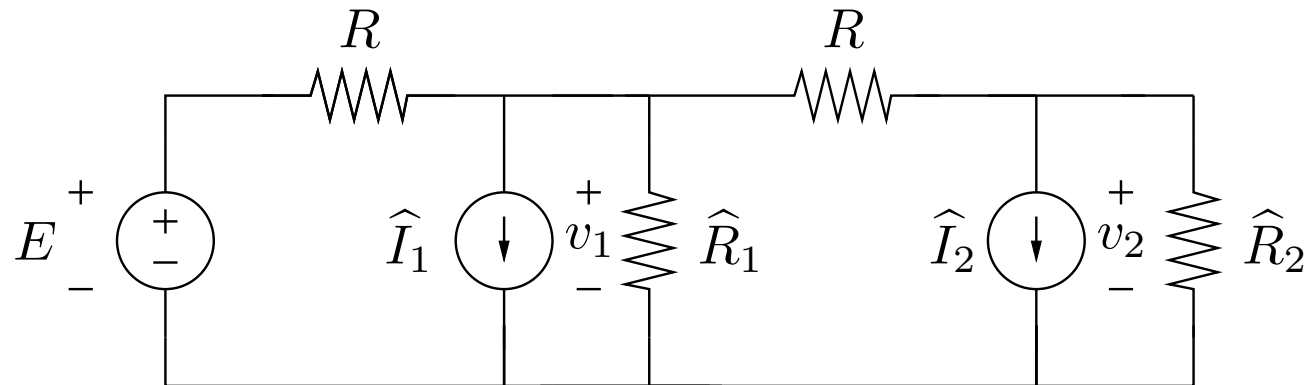
$$\begin{bmatrix} f_1(\hat{v}_1, \hat{v}_2) \\ f_2(\hat{v}_1, \hat{v}_2) \end{bmatrix} + \begin{bmatrix} g'_1(\hat{v}_1) + 2/R & -1/R \\ -1/R & g'_2(\hat{v}_2) + 1/R \end{bmatrix} \begin{bmatrix} v_1 - \hat{v}_1 \\ v_2 - \hat{v}_2 \end{bmatrix} = 0,$$

*i.e.*,

$$g_1(\hat{v}_1) - g'_1(\hat{v}_1)\hat{v}_1 + g'_1(\hat{v}_1)v_1 + \frac{v_1 - E}{R} + \frac{v_1 - v_2}{R} = 0$$
$$g_2(\hat{v}_2) - g'_2(\hat{v}_2)\hat{v}_2 + g'_2(\hat{v}_2)v_2 + \frac{v_2 - v_1}{R} = 0$$

two linear equations in two variables  $v_1, v_2$

**interpretation:** linearized equations describe a linear circuit



$$\hat{I}_1 = g_1(\hat{v}_1) - g'_1(\hat{v}_1)\hat{v}_1, \quad \hat{R}_1 = 1/g'_1(\hat{v}_1)$$

$$\hat{I}_2 = g_2(\hat{v}_2) - g'_2(\hat{v}_2)\hat{v}_2, \quad \hat{R}_2 = 1/g'_2(\hat{v}_2)$$

# Newton's method

$f : \mathbf{R}^n \rightarrow \mathbf{R}^n$ , differentiable

---

**given** an initial  $x$ , a required tolerance  $\epsilon > 0$

**repeat**

1. evaluate  $g = f(x)$  and  $H = Df(x)$ .
2. **if**  $\|g\| \leq \epsilon$ , **return**  $x$ .
3. Solve  $Hv = -g$ .
4.  $x := x + v$ .

**until** maximum number of iterations is exceeded.

---

- each iteration requires one evaluation of  $f(x)$  (*i.e.*,  $n$  scalar function evaluations) and  $Df(x)$  (*i.e.*,  $n^2$  derivatives)
- we assume  $Df(x)$  is nonsingular

## interpretation of one Newton iteration

$$x^+ = x - Df(x)^{-1}f(x) \quad (x = x^{(k)}, x^+ = x^{(k+1)})$$

- linearize  $f$  around current iterate  $x$

$$f_{\text{aff}}(y) = f(x) + Df(x)(y - x)$$

- solve linearized equation  $f_{\text{aff}}(y) = 0$

$$f(x) + Df(x)(y - x) = 0$$

for  $y$ , *i.e.*,

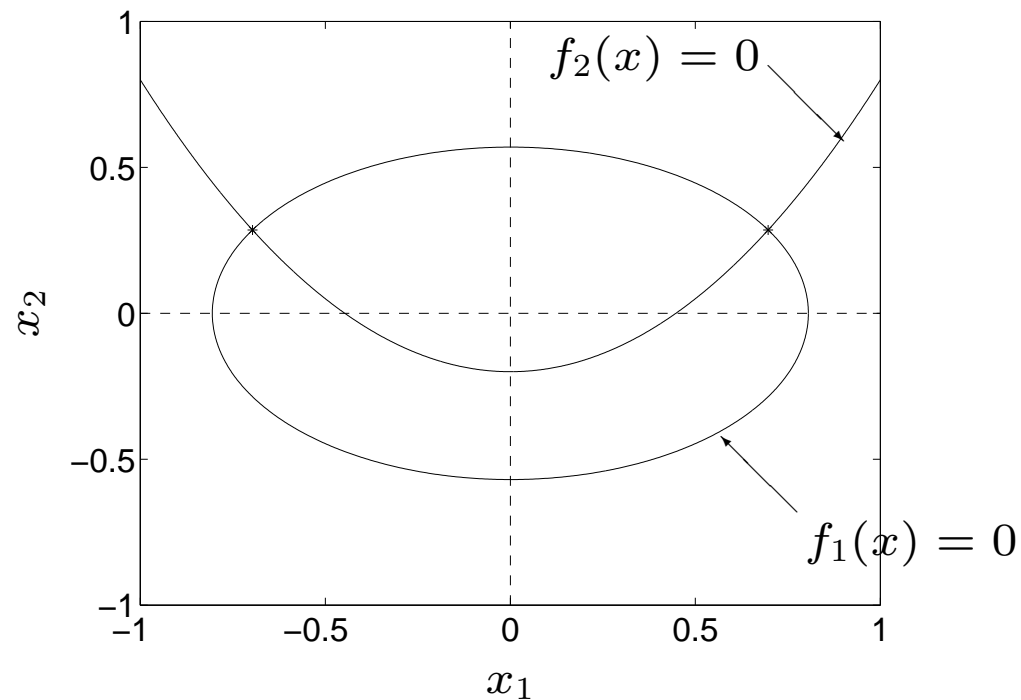
$$y = x - Df(x)^{-1}f(x)$$

- take  $y$  as new iterate  $x^+$

## Example

$$f_1(x_1, x_2) = \log(x_1^2 + 2x_2^2 + 1) - 0.5 = 0$$

$$f_2(x_1, x_2) = x_2 - x_1^2 + 0.2 = 0$$



two equations in two variables; two solutions  $(0.70, 0.29)$ ,  $(-0.70, 0.29)$

**derivative matrix of  $f$  at  $x$ :**

$$Df(x) = \begin{bmatrix} 2x_1/(x_1^2 + 2x_2^2 + 1) & 4x_2/(x_1^2 + 2x_2^2 + 1) \\ -2x_1 & 1 \end{bmatrix}$$

### **one Newton iteration**

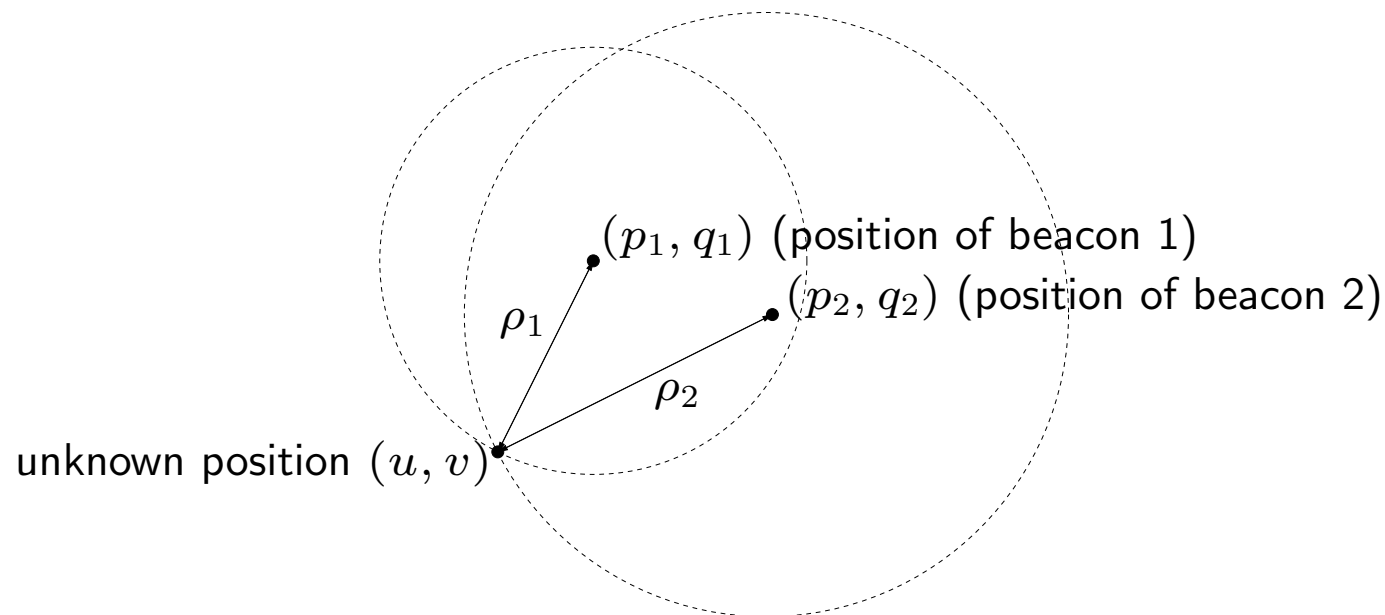
- evaluate  $H = Df(x)$  and  $g = f(x)$
- solve  $H\Delta x = -g$  (two linear equations in two variables)
- set  $x := x + \Delta x$

### **results**

- $x^{(0)} = (1, 1)$ : converges to  $x^* = (0.70, 0.29)$  in about 4 iterations
- $x^{(0)} = (-1, 1)$ : converges to  $x^* = (-0.70, 0.29)$  in about 4 iterations
- $x^{(0)} = (1, -1)$  or  $x^{(0)} = (-1, -1)$ : does not converge

## Example: navigation by range measurements

determine position by measuring the distances  $\rho_1, \rho_2$  to two beacons



two equations in two variables  $u, v$ :

$$f_1(u, v) = \sqrt{(p_1 - u)^2 + (q_1 - v)^2} - \rho_1 = 0$$

$$f_2(u, v) = \sqrt{(p_2 - u)^2 + (q_2 - v)^2} - \rho_2 = 0$$

**derivative matrix** at  $(u, v)$  (assuming  $(u, v) \neq (p_1, q_1)$ ,  $(u, v) \neq (p_2, q_2)$ ):

$$Df(u, v) = - \begin{bmatrix} \frac{p_1 - u}{\sqrt{(p_1 - u)^2 + (q_1 - v)^2}} & \frac{q_1 - v}{\sqrt{(p_1 - u)^2 + (q_1 - v)^2}} \\ \frac{p_2 - u}{\sqrt{(p_2 - u)^2 + (q_2 - v)^2}} & \frac{q_2 - v}{\sqrt{(p_2 - u)^2 + (q_2 - v)^2}} \end{bmatrix}$$

## one Newton iteration

- evaluate  $g = f(u, v)$  and  $H = Df(u, v)$
- solve

$$H \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = -g$$

(two linear equations in two variables)

- $u := u + \Delta u$ ,  $v := v + \Delta v$

## numerical example

$$(p_1, q_1) = (10, 10), \quad (p_2, q_2) = (10, -10), \quad \rho_1 = 14, \quad \rho_2 = 16$$

- start Newton's method at  $(u^{(0)}, v^{(0)}) = (0, 0)$
- converges in three iterations to  $(u^*, v^*) = (-1.12, 1.5)$ :

$k$	$((u^{(k)} - u^*)^2 + (v^{(k)} - v^*)^2)^{1/2}$
0	1.87
1	0.12
2	$4.80 \cdot 10^{-4}$
3	$8.27 \cdot 10^{-9}$

# Convergence of Newton's method

## convergence result

if  $Df(x^*)$  is nonsingular and  $x^{(0)}$  is sufficiently close to  $x^*$ , then Newton's method converges and there exists a  $c > 0$  such that

$$\|x^{(k+1)} - x^*\| \leq c \|x^{(k)} - x^*\|^2$$

(proof in course reader)

- quadratic convergence
- explains very fast convergence of Newton's method when started near a solution
- in practice, we don't know what  $c$  is, or how close  $x^{(0)}$  has to be