

## Newton's Method

Motivation:  $n=1$  single variable case  $f: I \rightarrow \mathbb{R}$ ,  $I = [a, b]$

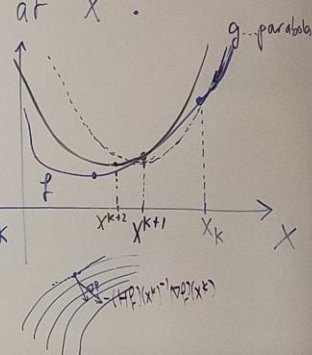
Idea is to approximate  $f$  using a quadratic function  $g$  at  $x^k$ :

$$g(x) = f(x^k) + f'(x^k)(x - x^k) + \frac{1}{2} f''(x^k)(x - x^k)^2$$

Instead of  $\min f(x)$ , at each step  $k$  we solve  $\min g(x)$ , i.e.  $(\nabla g)(x) = 0$

$$\text{Then } (\nabla f)(x^k) + (H_f)(x^k)(x - x^k) = 0 \Rightarrow x - x^k = -(H_f(x^k))^{-1}(\nabla f)(x^k) = d^k$$

Newton direction



This gives  $x^{k+1} = x^k - (H_f(x^k))^{-1}(\nabla f)(x^k)$  as the iteration

## Algorithm: Newton's Method

Step 0: Initialization:  $x^0$ ,  $k=0$

TOLERANCE chosen by user

Step 1:  $d^k = -(H_f(x^k))^{-1}(\nabla f)(x^k)$  If  $\|d^k\| \leq \epsilon$ , stop (since  $x^{k+1} = x^k + d^k$ )

Step 2: Set  $x^{k+1} = x^k + d^k$  for iteration  $k \rightarrow k+1$ . Go to Step 1.

Example: Consider minimizing  $f(x_1, x_2) = 2x_1^4 + x_2^2 - 4x_1x_2 + 5x_2$ .

$$(\nabla f)(x) = \begin{pmatrix} 8x_1^3 - 4x_2 \\ 2x_2 - 4x_1 + 5 \end{pmatrix}, \quad (H_f)(x) = \begin{pmatrix} 24x_1^2 & -4 \\ -4 & 2 \end{pmatrix}$$

Starting from  $x^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ , we then have

$$(\nabla f)(x^0) = \begin{pmatrix} 0 \\ 5 \end{pmatrix}, \quad (H_f)(x^0) = \begin{pmatrix} 0 & -4 \\ -4 & 2 \end{pmatrix}$$

$$\leadsto x^1 = x^0 - (H_f)(x^0)^{-1}(\nabla f)(x^0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \underbrace{\begin{pmatrix} 0 & -4 \\ -4 & 2 \end{pmatrix}^{-1}}_{-\frac{1}{16} \begin{pmatrix} 2 & 4 \\ 4 & 0 \end{pmatrix}} \begin{pmatrix} 0 \\ 5 \end{pmatrix} = \begin{pmatrix} \frac{20}{16} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{5}{4} \\ 0 \end{pmatrix}.$$

Step 2:  $(\nabla f)(x^1) = \begin{pmatrix} \frac{125}{8} \\ 0 \end{pmatrix}, \quad (H_f)(x^1) = \begin{pmatrix} \frac{75}{2} & -4 \\ -4 & 2 \end{pmatrix} > 0$

$$\leadsto x^2 = x^1 - (H_f)(x^1)^{-1}(\nabla f)(x^1) = \begin{pmatrix} \frac{5}{4} \\ 0 \end{pmatrix} - \begin{pmatrix} \frac{75}{2} & -4 \\ -4 & 2 \end{pmatrix}^{-1} \begin{pmatrix} \frac{125}{8} \\ 0 \end{pmatrix} = \dots = \begin{pmatrix} \frac{5}{4} \\ 0 \end{pmatrix} - \frac{1}{59} \begin{pmatrix} \frac{125}{4} \\ \frac{125}{2} \end{pmatrix} \text{ etc.}$$

Remarks: • We need  $(H_f)(x^k)$  to be invertible at each iteration!

Directional derivative There is no guarantee that  $f(x^{k+1}) \leq f(x^k)$ . We only know that

$$\boxed{\begin{aligned} (D_{d^k} f)(x^k) &= (\nabla f)(x^k)^T d^k \\ &= (\nabla f)(x^k)^T \left( - (H_f)(x^k)^{-1} (\nabla f)(x^k) \right) = - (\nabla f)(x^k)^T (H_f)(x^k)^{-1} (\nabla f)(x^k) < 0 \end{aligned}}$$

IF:  $(H_f)(x^k)^{-1} > 0$ , i.e.,  $(H_f)(x^k) > 0$ .

Proof of Theorem

Definition: (Descent direction) (vector)

Let  $f$  be differentiable at  $x$ , then  $g \in \mathbb{R}^n$  is called a descent direction if

$$(D_g f)(x) = (\nabla f)(x)^T g < 0.$$

Example:  $-\nabla f(x)$  is a descent direction...

$$\lim_{t \rightarrow 0} \frac{f(x+tg) - f(x)}{t}$$

Theorem: If  $(H_f)(x^k) > 0$ , then  $d^k$  is a descent direction.

