

# Lecture 16

Last time

- ▷ Analysis continued
- ▷ Convex guarantees
- ▷ Extensions

Today

- ▷ What's to come
- ▷ One-dimensional Newton's method.
- ▷ Newton's in  $\mathbb{R}^d$ .

What's to come?

Second-order Methods

- ▷ Newton's Method / Solving Systems of equations.
- ▷ Quasi-Newton Methods.
- ▷ Conjugate gradient.
- ▷ Trust Region Methods.

Newton's Method

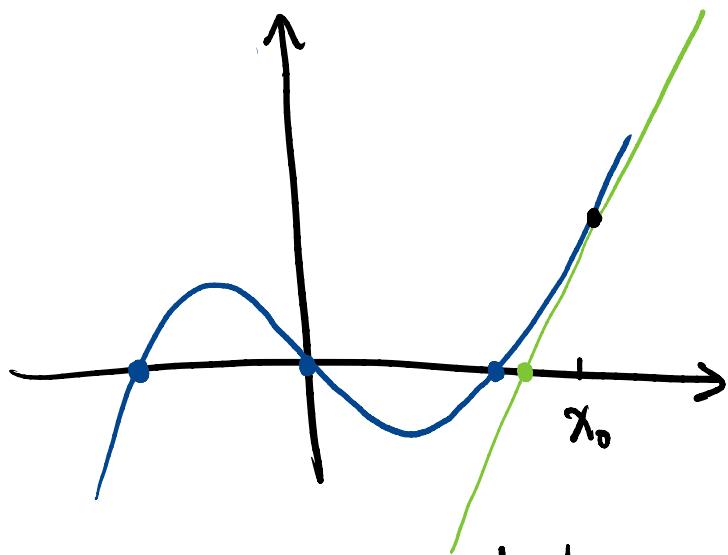
Imagine we had a system of nonlinear equations

$$F(x) = 0$$

with  $F: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and we want to solve for  $x$ . This recovers finding stationary points if  $F = \nabla f$ .

## One-dimensional setup

Assume  $F: \mathbb{R} \rightarrow \mathbb{R}$  is smooth.



The idea of Newton's method is to linearize and then look for a root (zero).

Thus, we update

$$\text{Pick } x_{k+1} \text{ s.t. } F(x_k) + F'(x_k)(x_{k+1} - x_k) = 0$$

Note that reordering this amounts to

$$x_{k+1} = x_k - \frac{F(x_k)}{F'(x_k)}.$$

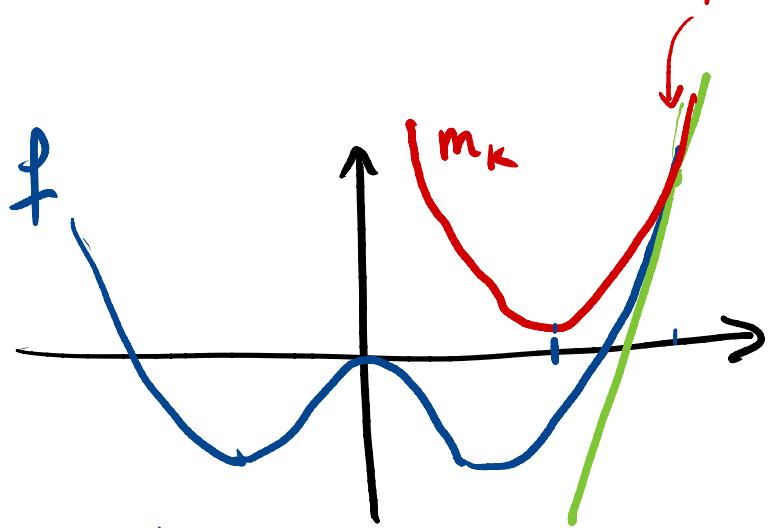
If  $F = f'$ , then this is

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

We need second  
order information

If  $f''(x_k) > 0$  this also corresponds

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \left\{ f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k) (x - x_k)^2 \right\}$$



When  $f''(x_k) < 0$   
we don't have a  
convex model.

This method is **really fast**. As an example: Consider  $F(x) = x^2 - a$ , then

$$F(x) = 0 \Leftrightarrow x = \pm \sqrt{a}.$$

In this case, Newton's method reduces to

$$x_{k+1} = x_k - \frac{F(x_k)}{F'(x_k)} = x_k - \frac{x_k^2 - a}{2x_k} = \frac{1}{2} \left( x_k - \frac{a}{x_k} \right).$$

For  $a=2$  and  $x=1$ , we obtain

$$x_0 = 1 \dots$$

$$x_1 = 1.5 \dots$$

$$x_2 = 1.41 \dots$$

$$x_3 = 1.41421 \dots$$

$$x_4 = 1.41421356237 \dots$$

# correct digits  $\approx 2^k$   
 $x_7 \approx 60$  correct

Aside: This algorithm was used in the video game Quake 3 (1999) to find  $1/\sqrt{x}$ .

Quick review of convergence naming

Suppose  $\delta_k \rightarrow 0$  (This could be the objective gap, the distance to a solution or  $\|\nabla f(x_k)\|$ ).

We say that

- ▷  $\delta_k$  converges linearly if  $\exists c \in (0, 1)$ ,  $N \geq 0$  s.t.  $\forall k \geq N \quad \delta_{k+1} \leq c\delta_k$
- ▷  $\delta_k$  converges sublinearly if no such  $c$  exists.
- ▷  $\delta_k$  converges superlinearly if  $\exists \{c_n\} \subseteq [0, 1]$ ,  $N \geq 0$  s.t.  $c_n \rightarrow 0$  and  $\forall k \geq N \quad \delta_{k+1} \leq c_k \delta_k$ .
- ▷  $\delta_k$  converges quadratically if  $\exists c \in (0, 1)$ ,  $N \geq 0$  s.t.  $\forall k \geq N \quad \delta_{k+1} \leq c\delta_k^2$ .

This is super linear since  $c\delta_k \rightarrow 0$ .

Secant Method

If we don't know  $F'(x_k)$  it is reasonable to approximate it with

$$F'(x_k) \approx \frac{F(x_k) - F(x_{k-1})}{x_k - x_{k-1}}$$

and thus

$$x_{k+1} \leftarrow x_k - \left( \frac{x_k - x_{k-1}}{\frac{F(x_k) - F(x_{k-1})}{F'(x_k)}} \right) F(x_k).$$

Under modest regularity conditions, we have  $x_k \rightarrow x^*$  with  $F(x^*) = 0$ .

Moreover

$$|x_{k+1} - x^*| \leq c \cdot |x_k - x^*|^{\varphi}$$

where  $\varphi = \frac{\sqrt{5} + 1}{2} = 1.618 \dots$  is the golden ratio. Thus convergence is superlinear, but not quadratic.

## Newton in $\mathbb{R}^d$

In a bunch of applications we want to solve systems of equations

$$F(x) = 0 \quad \text{with} \quad F: \mathbb{R}^d \rightarrow \mathbb{R}^d \text{ smooth.}$$

For example:

▷ Optimization  $\nabla f(x) = 0$ .

- ▷ Computer graphics
- ▷ Physics (Equilibrium states thermodynamics)
- ▷ Robotics (inverse kinematics)
- ▷ ...

Key idea: Linearize  $F(x)$ , then solve linear system.

Recall that the Jacobian of  $F(x)$  is

$$\nabla F(x) = \begin{bmatrix} \frac{\partial F_1(x)}{\partial x_1} & \dots & \frac{\partial F_1(x)}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_d(x)}{\partial x_1} & \dots & \frac{\partial F_d(x)}{\partial x_d} \end{bmatrix}.$$

When  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  is  $C^2$ , The Hessian  $\nabla^2 f$  is the Jacobian of  $\nabla f$ .

Then, Newton's method updates by

Finding  $x_{k+1}$  s.t.  $F(x_k) + \nabla F(x_k)(x_{k+1} - x_k) = 0$ .

If  $\nabla F(x_{k+1})$  is full rank, then the system has a unique solution and

$$x_{k+1} = x_k - \underbrace{\nabla F(x_k)^{-1}}_{\text{Newton's direction}} F(x_k).$$

In optimization land this is equivalent to

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k).$$

Notice that this is equivalent to constructing a second-order approximation of  $f$  at  $x_k$ :

$$f_k(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2} (x - x_k)^T \nabla^2 f(x_k) (x - x_k)$$

and finding a critical point of  $f_k$ .

Unlike before we don't have that  $f_k$  is convex:

- ▷ If  $\nabla^2 f(x_k) < 0 \Rightarrow f_k$  is concave  
*Ascent direction.*
- ▷ If  $\nabla^2 f(x_k) > 0 \Rightarrow f_k$  is convex  
*Descent direction*
- ▷ If  $\nabla^2 f(x_k)$  is indefinite  $\Rightarrow f_k$  has a saddle.

### Convergence of Newton's method

We state the following without a proof, but we'll get back to a

non asymptotic version next class

### Theorem (Local convergence)

Let  $F: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be continuously differentiable and assume  $F(x^*) = 0$  for some  $x^*$ . If  $\nabla F(x^*)$  is nonsingular, then some neighborhood  $S$  of  $x^*$  we have that if  $x_0 \in S$ , the iterates of Newton's method satisfy

$$x_k \in S, \quad x_k \rightarrow x^*, \quad \nabla F(x_k) \text{ nonsingular.}$$

### Warnings

- ▷ Global convergence is not granted.
- ▷ If  $\nabla F(x_k)$  is singular the method is not well-defined.
- ▷ Even if  $\nabla F(x_k)$  is nonsingular, we can have numerical issues