HW4 posted tonight

Second-Order Methods (Topic for the rest of the semester)

1st Newton's Method / Solving Nonlinear Equations

1. One-Dimensional Newton
2. $\mathbb{R}^d$ Newton
3. Convergence Analysis
4. Problems with Newton

2nd Modified / Quasi-Newton Methods

3rd Trust-Region Methods

4th Conjugate Gradient Methods

Newton's Method

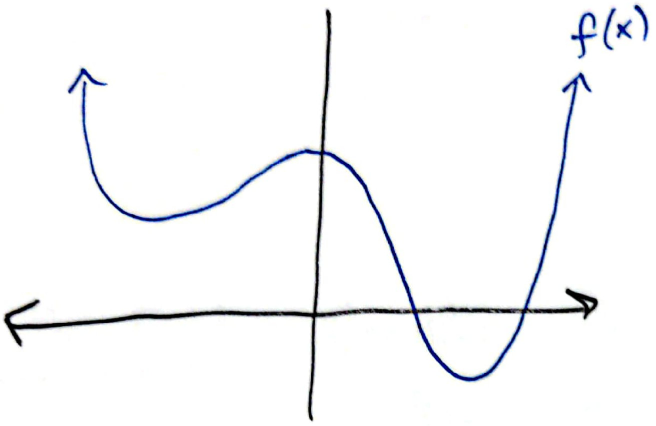
The classic setup of Newton seeks solutions to a system of nonlinear equations, $x \in \mathbb{R}^d$

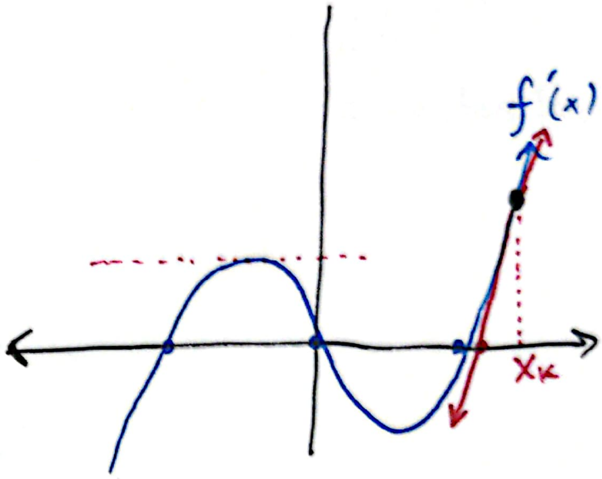
$F(x) = 0, \; F: \mathbb{R}^d \to \mathbb{R}^d$ — "Nonlinear Equation Solving".

$\underbrace{\nabla f(x)}_{F} = 0, \; f: \mathbb{R}^d \to \mathbb{R}$ — "Nonlinear Optimization"

1. One-Dimensional Setup

Give $f : \mathbb{R} \to \mathbb{R}$, then





$$\min_{x \in \mathbb{R}^d} f(x)$$

Find $f'(x) = 0$
"root finding"

Take $F(x) = 0$, $(F = f')$, linearize $F(x)$ at current x_k

Then move to root of that linear equation.

Formally, $F(x+s) \approx F(x_K) + F'(x_K) \cdot s$

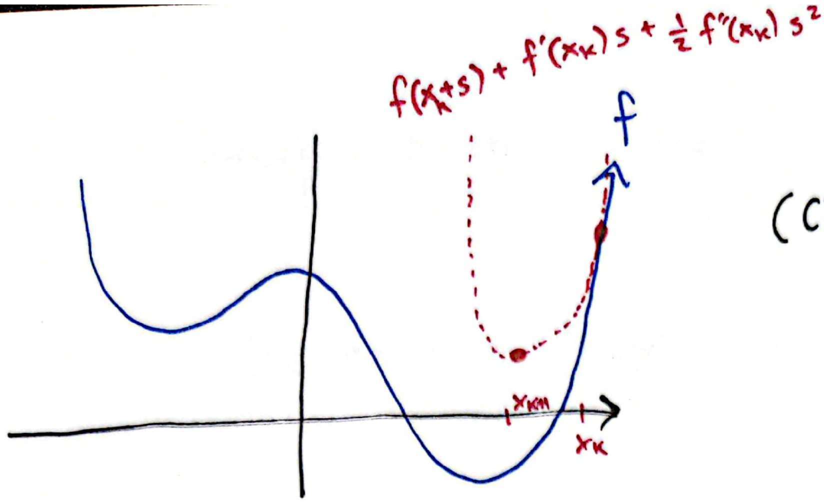
Pick x_{K+1} s.t. $F(x_K) + F'(x_K)(x_{K+1} - x_K) = 0$

$\quad \llcorner x_K + s$

$$\Leftrightarrow x_{K+1} = x_K - \frac{F(x_K)}{F'(x_K)}$$ (provided $F'(x_K) \neq 0$)

$$\left(\text{with } s = -\frac{F(x_K)}{F'(x_K)}\right)$$

For $F = f'$, $x_{K+1} = x_K - \dfrac{f'(x_K)}{f''(x_K)}$.



(Check stationary point of 2^{nd} order model = root of (2^{nd} order model)$'(x)$ = Newton Step)

<u>Really Fast</u> Consider $F(x) = x^2 - a$.

Roots $x = \pm\sqrt{a}$

Easy to run Newton $F'(x) = 2x$

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

Fix $a=2$, $x_0 = 1$

x_0 1

x_1 1.5...

x_2 1.41 ....

x_3 1.41421 ...

x_4 1.41421356237 ....

correct digits $\approx 2^k$ (x_7 ~60 correct)

Aside, Quake 3 (1999), $\underbrace{1/\sqrt{x}}$ approximate, the 2 Newton steps.
Algorithm bit

Formally, suppose $\delta_k \to 0$ (think δ_k objective gap, distance to optimal, gradient zero).

δ_k converges "linearly" if $\exists c \in (0,1)$, $N \geq 0$ s.t.

$$\forall k \geq N \quad \delta_{k+1} \leq c \cdot \delta_k .$$

δ_k converges "sublinearly" if no such c exists.

(for example $1/k$)

δ_k converges "superlinearly" if $\exists \{c_k\} \subseteq [0,1)$, $N \geq 0$ s.t.

$$\forall k \geq N \quad \delta_{k+1} \leq c_k \delta_k$$

$$\lim c_k = 0.$$

δ_k converges "quadratically" if $\exists c \in (0,1)$, $N \geq 0$ s.t.

$$\forall k \geq N \quad \delta_{k+1} \leq c\, \delta_k^2 .$$

(superlinear $\delta_{k+1} \leq c\, \delta_k^2 = \underbrace{(c\, \delta_k)}_{\text{"}c_k \to 0} \delta_k$).

Sometimes we might not know $F'(x_k)$ $(= f''(x_k)$ for optimization$)$

Reasonable fix the Secant Method.

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

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

Under modest regularity conditions, this works.

$x_k \to x^*$ (stationary point $f'(x) = 0$)

(superlinearly $|x_k - x^*| \to 0$).

$$|x_{k+1} - x^*| \leq c \cdot |x_k - x^*|^{\varphi}, \quad \varphi = \frac{\sqrt{5}-1}{2}$$

$= 1,618 \ldots$

$\Rightarrow$ Superlinear but not quadratic.

2. Newton in $\mathbb{R}^d$

$$\min_{x \in \mathbb{R}^d} f(x) \quad \Rightarrow \quad \underbrace{\nabla f}_{\hookrightarrow F : \mathbb{R}^d \to \mathbb{R}^d} (x) = 0$$

"Nonlinear system of equation solving"

Idea: Linearize $F(x)$, then solve that linear system to get closer to solving $F(x)=0$.

Recall, the Jacobian of $F(x)$

$$\nabla F(x) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x) & \cdots & \frac{\partial F_1}{\partial x_d}(x) \\ \vdots & & \\ \frac{\partial F_d}{\partial x_1}(x) & \cdots & \frac{\partial F_d}{\partial x_d}(x) \end{pmatrix}$$

The Jacobian is the unique linear operator s.t.

$$\forall s \in \mathbb{R}^d \quad \lim_{t \to 0} \frac{F(x+ts) - (F(x) + \nabla F(x)\, ts)}{t} = 0.$$

(The Hessian $\nabla^2 f$ is the Jacobian of ∇f.)

Newton's Method

Linearize $F(x) \approx F(x_k) + \nabla F(x_k) \cdot (x - x_k)$

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

(only works if $\nabla F(x_k)$ is full rank $\Leftrightarrow$ nonsingular

$\Leftrightarrow$ System has unique sol.)

$\Leftrightarrow$ invertible

$$\Rightarrow x_{k+1} = x_k - \nabla F(x_k)^{-1} F(x_k)$$

"Newton direction" $= -\nabla F(x_k)^{-1} F(x_k)$

$\left(= -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)\right.$ for optimization)

•

3D pictures

Linearize $\nabla f(x)$, go to the zero

$\Leftrightarrow$ Move to stationary point of 2nd order model

$$f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T \nabla^2 f(x_k) s$$

Convex quad when $\nabla^2 f \succ 0$

"descent direction"

saddle

$\nabla^2 f$ indefinite

Concave quad

$\nabla^2 f \prec 0$

"ascent direction"

3. Convergence of Newton's Method

We won't prove the following classic theorem:

Theorem (Local Convergence)

Let $F: \mathbb{R}^d \to \mathbb{R}^d$ be cont diff and assume $F(x^*) = 0$ for some x^*. If $\nabla F(x^*)$ is nonsingular, then some neighborhood S of x^* has any $x_0 \in S$ produce Newton steps

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