Newton's Method

Daniel P. Robinson
Department of Applied Mathematics and Statistics
Johns Hopkins University

September 8, 2020

Notes

Outline

Notes

Problem of interest

Given a function $F : \mathbb{R}^n \to \mathbb{R}^n$, find a vector $x \in \mathbb{R}^n$ such that

$$F(x) = 0$$

- some methods only require evaluating F
  - bisection method ($n = 1$)
  - fixed-point iteration
  - secant method
- some methods require evaluating F and $\nabla F(x)$
  - inverse interpolation
  - Newton's Method
- examples
  - calculating $\sqrt{z}$ by finding a zero of $F(x) = x^2 - z$
  - most nonlinear optimization methods in some way reduce to or are based on applying Newton's Method to some function
- linear $F(x) = Ax - b$ always has either 0, 1, or infinitely many solutions
- nonlinear F may have any number of solutions
  - $F(x) = e^x + 1 = 0$ (no real solutions)
  - $F(x) = x^2 + a = 0$ (0, 1, or 2 solutions)
  - $F(x) = \cos(x)$ (infinitely many solutions)

Notes

What can we expect?

- Zeros usually cannot be computed analytically
- Solutions to even simple equations can be irrational, e.g., $F(x) = x^2 - 2 = 0$
- We cannot expect to solve a system of equations exactly even if it has rational zeros
- We seek algorithms that produce approximate numerical solutions quickly, accurately, and reliably

Notes

Newton's Method in one variable

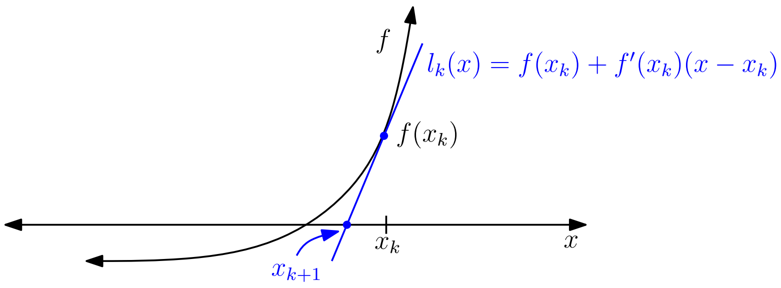
Given x_k consider the first-order Taylor expansion

$$l_k(x) \stackrel{\text{def}}{=} f(x_k) + f'(x_k)(x - x_k) \approx f(x) \quad \text{for } x \text{ near } x_k$$

Define x_{k+1} as the zero of l_k

$$0 = l_k(x_{k+1}) = f(x_k) + f'(x_k)(x_{k+1} - x_k)$$

which yields $$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad \text{(this is the Newton iteration)}$$



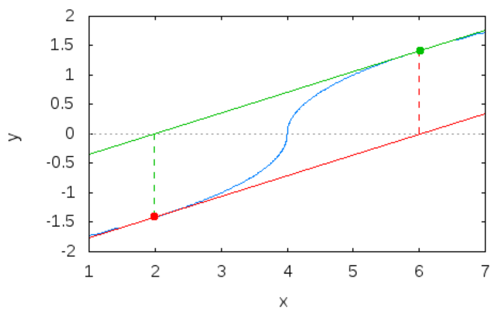
Notes

Newton's Method in one variable

Newton's Method for finding a zero of the function $f(x)$ generates a sequence of iterates $\{x_k\}_{k\geq 0}$ from the formula

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

- each iteration requires both a function and derivative evaluation
- does it work?
  - iterates can cycle, diverge, converge, or simply not converge
  - what if $f'(x_k) = 0$?
  - what if $f(x_k)$ is undefined or equal to an imaginary number?



<Matlab demo 1>

Notes

Consider the zero-finding problem $F(x) = 0$ where

$$F(x) = \begin{pmatrix} F_1(x) \\ \vdots \\ F_n(x) \end{pmatrix} \in \mathbb{R}^n$$

and the Jacobian matrix is given by

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

Newton's Method in higher dimensions

Iteratively compute zeros of

$$l_k(x) \stackrel{\text{def}}{=} F(x_k) + \nabla F(x_k)(x - x_k)$$

by

$$x_{k+1} = x_k + s_k \quad \text{where } s_k \text{ solves } \quad \nabla F(x_k)s = -F(x_k)$$

- assumes that $\nabla F(x_k)$ is nonsingular
- computing $\nabla F(x_k)$ (and solving with it) can be very expensive!
- can be viewed as finding a zero of n individual affine models simultaneously

$$F_i(x_k) + \nabla F_i(x_k)^T s = 0 \quad \Longleftrightarrow \quad \nabla F_i(x_k)^T s = -F_i(x_k) \quad \text{for all } i = 1, \ldots, n$$

<Matlab demo 2>

Notes

- A sequence of iterates $\{x_k\}$ is said to converge to x^* if

$$\lim_{k \to \infty} \|x_k - x^*\| = 0$$

- It would be great if Newton's Method for solving

$$F(x) = 0$$

was guaranteed to satisfy $\lim_{k\to\infty} x_k = x^*$ and $F(x^*) = 0$

- If iterates $\{x_k\}$ generated by an algorithm converge from any initial point x_0, then we say the algorithm is globally convergent
- Newton's Method is not globally convergent
- If a sequence of iterates $\{x_k\}$ generated by an algorithm converges once x_k is "close enough" to a zero x^*, then we say it is locally convergent
- When the iterates do converge, we are also interested in how fast they do so, i.e., the rate of convergence

Notes

Definition (q-linearly convergent)

We say that the sequence $\{x_k\}$ is q-linearly convergent to x^* if there exists a constant $c \in [0,1)$ such that

$$\|x_{k+1} - x^*\| \le c\|x_k - x^*\|$$

- $\{1 + 2^{-k}\}$ converges linearly to $x^* = 1$

Definition (q-superlinearly convergent)

We say that the sequence $\{x_k\}$ is q-superlinearly convergent to x^* if $\{x_k\}$ converges to x^* and there exists a sequence $\{c_k\}$ satisfying

$$\lim_{k\to\infty} c_k = 0$$

and

$$\|x_{k+1} - x^*\| \le c_k\|x_k - x^*\|$$

Definition (q-quadratically convergent)

We say that the sequence $\{x_k\}$ is q-quadratically convergent to x^* if $\{x_k\}$ converges to x^* and there exists a constant $c \ge 0$ such that

$$\|x_{k+1} - x^*\| \le c\|x_k - x^*\|^2$$

- $\{1 + 2^{-2^k}\}$ converges quadratically to $x^* = 1$

Notes

Theorem (Local convergence of Newton's method)

Let $F : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function on an open convex set D, and assume that $F(x^*) = 0$ for some $x^* \in \mathcal{D}$ and that $\nabla F(x^*)$ is nonsingular. Then there exists an open neighborhood $\mathcal{S}$ containing x^* such that, for any x_0 in $\mathcal{S}$, the Newton iterates are well defined, remain in $\mathcal{S}$ and converge to x^*.

- if any Newton iterate gets "close enough" to x^*, then the Newton iterates will converge to x^*
- this does not say anything about how fast they will converge!

Notes

With stronger assumptions, we show that Newton's Method is quadratically convergent

Theorem (Quadratic-convergence of Newton's Method)

Let $x^* \in \mathbb{R}^n$ satisfy $F(x^*) = 0$ and assume that the following hold:

- F *is continuously differentiable in an open convex set* $\mathcal{X} \subseteq \mathbb{R}^n$ *containing* x^*
- $\nabla F(x^*)$ *is nonsingular*
- *there exists* $r > 0$ *such that* $\mathcal{B}(x^*, r) \subset \mathcal{X}$ *and the Jacobian* ∇F *is Lipschitz continuous with constant* L *in* $\mathcal{B}(x^*, r)$

It follows that there exists an $\varepsilon > 0$ *such that for all* $x_0 \in \mathcal{B}(x^*, \varepsilon)$ *the sequence of iterates generated by Newton's Method is well-defined, converges to* x^**, and satisfies*

$$\|x_{k+1} - x^*\| \leq c\|x_k - x^*\|^2 \quad \text{for some } c > 0 \qquad \text{(quadratic convergence)}$$

Proof:
Since $\nabla F(x^*)$ is nonsingular by assumption, we may define

$$M \overset{\text{def}}{=} \|[\nabla F(x^*)]^{-1}\| < \infty$$

For radius of differentiability r, Lipschitz constant L for the Jacobian ∇F, and constant M, we may define

$$\varepsilon \overset{\text{def}}{=} \min\left(r, \frac{1}{2ML}\right)$$

Notes

We first show that if $\|x_0 - x^*\| \leq \varepsilon$ then $\nabla F(x_0)$ is nonsingular and bounded by using

Lemma

Let $\|\cdot\|$ *be a matrix norm that satisfies* $\|AB\| \leq \|A\|\|B\|$ *and* $\|I\| = 1$*. If* A *is nonsingular and* $\|A^{-1}(B - A)\| < 1$*, then* B *is nonsingular and*

$$\|B^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}(B - A)\|}$$

Norm inequalities, definition of M, Lipschitz continuity of ∇F, and definition of ε yield

$$\begin{aligned}\|[\nabla F(x^*)^{-1}(\nabla F(x_0) - \nabla F(x^*))\| &\leq \|[\nabla F(x^*)^{-1}\|\|\nabla F(x_0) - \nabla F(x^*)\| \\ &\leq ML\|x_0 - x^*\| \leq ML\varepsilon \leq \tfrac{1}{2}\end{aligned}$$

so that the previous lemma ensures that $\nabla F(x_0)$ is nonsingular and

$$\|[\nabla F(x_0)]^{-1}\| \leq \frac{\|[\nabla F(x^*)]^{-1}\|}{1 - \|[\nabla F(x^*)]^{-1}(\nabla F(x_0) - \nabla F(x^*))\|} \leq 2\|[\nabla F(x^*)]^{-1}\| = 2M$$

as desired.

Notes

Using simple algebra and the fact that $F(x^*) = \mathbf{0}$, we have

$$\begin{aligned}
x_1 - x^* &= x_0 - x^* - [\nabla F(x_0)]^{-1} F(x_0) \\
&= x_0 - x^* - [\nabla F(x_0)]^{-1} (F(x_0) - F(x^*)) \\
&= [\nabla F(x_0)]^{-1} \Big[F(x^*) - \underbrace{\big(F(x_0) + \nabla F(x_0)(x^* - x_0)\big)}_{\text{affine model of } F \text{ at } x_0} \Big]
\end{aligned}$$

and then using the bound on the previous slide and Taylor expansion we have

$$\begin{aligned}
\|x_1 - x^*\| &\le \|[\nabla F(x_0)]^{-1}\| \|F(x^*) - \big(F(x_0) + \nabla F(x_0)(x^* - x_0)\big)\| \\
&\le (2M) \left(\tfrac{1}{2} L \|x_0 - x^*\|^2 \right) \\
&= ML\|x_0 - x^*\|^2 && (1) \\
&= c\|x_0 - x^*\|^2 \quad \text{for } c \stackrel{\text{def}}{=} ML && (2)
\end{aligned}$$

which is the condition required to prove quadratic convergence if $\{x_k\}$ converges to x^*. However, we have from (1) and the definition of ε that

$$\|x_1 - x^*\| \le ML\|x_0 - x^*\|^2 = ML\|x_0 - x^*\| \|x_0 - x^*\| \le \tfrac{1}{2} \|x_0 - x^*\|.$$

Since $\|x_1 - x^*\| < \varepsilon$, we can repeat the argument and thus $\{x_k\}$ does in fact converge to x^*. It now follows from (2) that the iterates $\{x_k\}$ converge quadratically to x^*. ∎

Notes

Summary

- We need x_k to be close enough to x^* so that $\nabla F(x_k)$ is nonsingular
- The region of guaranteed convergence is the ball centered at x^* with radius

$$\varepsilon = \min\left(r, \frac{1}{2ML}\right)$$

- Large M and/or L implies a small ball of guaranteed convergence
- Newton's Method may converge even outside of this ball
- Large L means the gradient changes rapidly near x^*
- Large M means Newton step may send us far away

Notes

Relevance for optimization

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; f(x)$$

First order necessary condition:

$$\nabla f(x^*) = \mathbf{0}$$

So we are searching for zeros of the gradient map $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$. Ready for applying Newton's method!

Initialize x^0 to some starting point. Compute iterates

$$x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

Called the Newton step and $-(\nabla^2 f(x^k))^{-1} \nabla f(x^k)$ is called the Newton direction.

Notes

Notes