Newton Method: Part 1

Lecture 5-1 - CMSE 382

Prof. Elizabeth Munch

Michigan State University
::
Dept of Computational Mathematics, Science & Engineering

Mon, Feb 9, 2026

This Lecture

Topics:

- Linear approximation theorem
- Quadratic approximation theorem
- Pure Newton's method
- Newton method quadratic local convergence

Announcements:

- Homework 2 posted, due Thursday, Feb 12 at 11:59pm
- Midterm 1 on Wednesday, Feb 18.

Section 1

Review: Linear and Quadratic Approximation

Linear approximation theorem

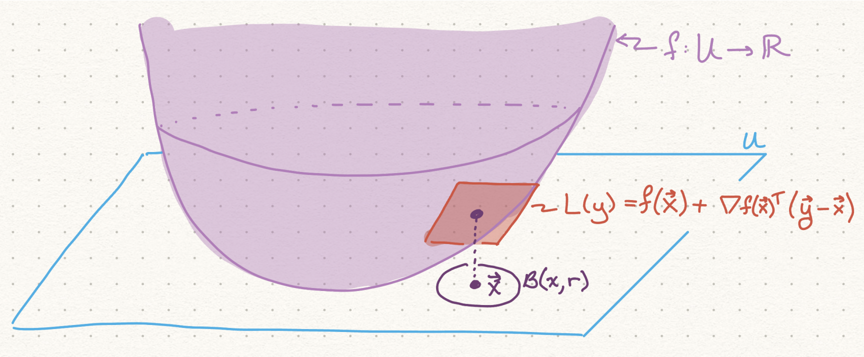
Theorem (Linear Approximation Theorem)

- *Let* $f : U \to \mathbb{R}$ *be a twice continuously differentiable function over an open set* $U \subseteq \mathbb{R}^n$*,*
- *Let* $\mathbf{x} \in U$*,* $r > 0$ *satisfy* $B(\mathbf{x}, r) \subseteq U$*.*

Then for any $\mathbf{y} \in B(\mathbf{x}, r)$ *there exists* $\boldsymbol{\xi} \in [\mathbf{x}, \mathbf{y}]$ *such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\boldsymbol{\xi})(\mathbf{y} - \mathbf{x}).$$

Linear approximation visual



$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + o(\|\mathbf{x} - \mathbf{y}\|).$$

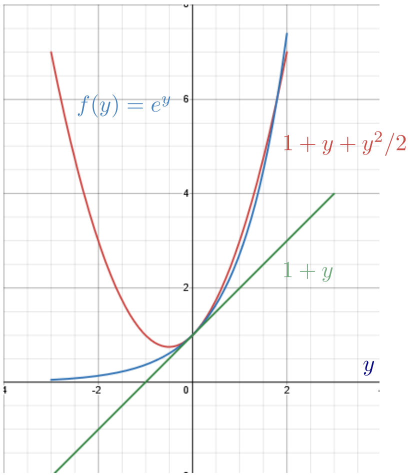
Quadratic approximation theorem

Theorem (Quadratic Approximation Theorem)

- *Let $f : U \to \mathbb{R}$ be a twice continuously differentiable function over an open set $U \subseteq \mathbb{R}^n$.*
- *Let $\mathbf{x} \in U$, $r > 0$ satisfy $B(\mathbf{x}, r) \subseteq U$.*

Then for any $\mathbf{y} \in B(\mathbf{x}, r)$

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + o(\|\mathbf{y} - \mathbf{x}\|^2).$$



Section 2

Newton's Method

Newton's method

Newton's method focuses on optimizing

$$\min\{f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\}$$

where f is twice continuously differentiable.

- Newton's method is a second order method
  - Uses information from the Hessian
  - Contrast with first order methods like gradient descent which only uses gradient information
- Requires that $\nabla^2 f(\mathbf{x})$ is positive definite for every $\mathbf{x} \in \mathbf{R}^n$
  - A unique optimal solution $\mathbf{x}^*$ exists.

Main idea

Minimize the quadratic approximation around $\mathbf{x}_k$:

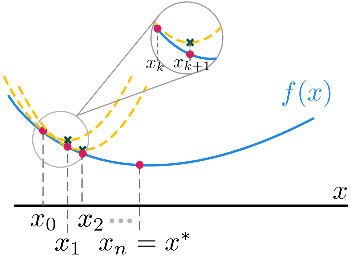
$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \left\{ f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^\top(\mathbf{x}-\mathbf{x}_k) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_k)^\top \nabla^2 f(\mathbf{x}_k)(\mathbf{x}-\mathbf{x}_k) \right\}$$

- This function is not well-defined unless $\nabla^2 f(\mathbf{x}_k) \succ 0$.
- The unique minimizer is the unique stationary point

$$\nabla f(\mathbf{x}_k) + \nabla^2 f(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = \mathbf{0}$$

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

- $-(\nabla^2 f(\mathbf{x}_k))^{-1}\nabla f(\mathbf{x}_k)$ is called the **Newton direction**.



Pure Newton's Method

Input: $\varepsilon > 0$ tolerance parameter
Initialization: Pick $\mathbf{x}_0 \in \mathbb{R}^n$ arbitrarily.

- $\mathbf{x}_0$ too far from $\mathbf{x}^*$ can cause divergence.

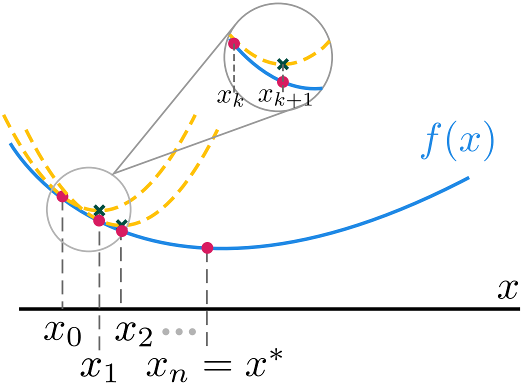
General step: For any $k = 0, 1, 2, \ldots,$ do:

(1) Compute the Newton direction $\mathbf{d}_k$, which is the solution to the linear system $\nabla^2 f(\mathbf{x}_k)\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$

- More efficient than inverting $\nabla^2 f(\mathbf{x}_k)$ in $\mathbf{d}_k = -(\nabla^2 f(\mathbf{x}_k))^{-1}\nabla f(\mathbf{x}_k)$

(2) Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$

(3) If $\|\nabla f(\mathbf{x}_{k+1})\| < \varepsilon$, stop and output $\mathbf{x}_{k+1}$.



Newton's Method

Example $f(x, y) = \frac{1}{2}(10x^2 + y^2) + 5\log(1 + e^{-x-y})$. (Source)

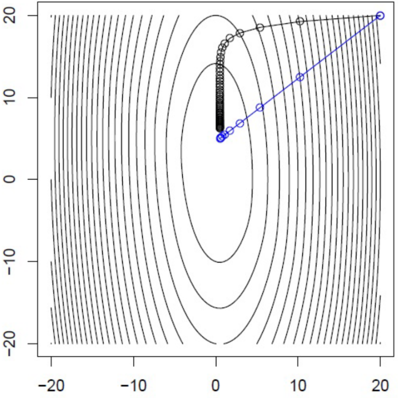


Figure: *

Comparison of Newton's Method (blue) with Gradient Descent (black)

Section 3

Convergence

Quadratic Local Convergence of Newton's Method

Theorem: Let f be a twice continuously differentiable function defined over $\mathbb{R}^n$. Let $\{\mathbf{x}_k\}_{k\geq 0}$ be the sequence generated by Newton's method, and let $\mathbf{x}^*$ be the unique minimizer of f over $\mathbb{R}^n$.

Assume that

- there exists $m > 0$ such that $\lambda_{\min}(\nabla^2 f(\mathbf{x})) \geq m$ for any $\mathbf{x} \in \mathbb{R}^n$
- there exists an L such that $\nabla^2 f(\mathbf{x})$ is Lipschitz with constant L

Then for any $k = 0, 1, 2, \ldots$, the following inequality holds

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\| \leq \frac{L}{2m}\|\mathbf{x}_k - \mathbf{x}^*\|^2$$

Moreover, if $\|\mathbf{x}_0 - \mathbf{x}^*\| \leq \frac{m}{L}$, then for $k = 0, 1, 2, \ldots$,

$$\|\mathbf{x}_k - \mathbf{x}^*\| \leq \frac{2m}{L}\left(\frac{1}{2}\right)^{2^k}$$

- Rapid convergence! (no. of accuracy digits is doubled at each k)
- Can still converge if the assumptions are not met

Groups - Round 2

Group 1
Abigail, Shanze, Jack, Quang Minh,

Group 2
Igor, Atticus, K M Tausif, Long,

Group 3
Yousif, Zheng, Jake, Purvi,

Group 4
Maye, Alice, Arjun, Kyle,

Group 5
Monirul Amin, Jay, Brandon, Luis,

Group 6
Scott, Ha, Lora, Tianjian,

Group 7
Braedon, Sai, Joseph, Noah,

Group 8
Michal, Aidan, Jonid, Dev,

Group 9
Vinod, Saitej, Anthony, Breena,

Group 10
Karen, Dori, Lowell, Aaron,

Group 11
Jamie, Sanskaar, Dominic, Lauryn,

Group 12
Andrew, Arya, Daniel, Morgan,