Newton's Method: Part 2

Lecture 5-2 - CMSE 382

Prof. Elizabeth Munch

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

Wed, Feb 11, 2026

This Lecture

Topics:

- Damped Newton's method
- Cholesky factorization
- Hybrid gradient-Newton method

Announcements:

- Homework 2 posted, due Thursday, Feb 12 at 11:59pm
- Midterm 1 on Wednesday, Feb 18.
- No office hours Friday, Feb 13.

Section 1

Damped Newton's Method

Damped Newton method

Motivating example: Consider the function $f(x,y) = \sqrt{x^2+1} + \sqrt{y^2+1}$ whose optimal solution is $(0,0)$.

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} (1+x^2)^{-3/2} & 0 \\ 0 & (1+y^2)^{-3/2} \end{bmatrix}. \quad \nabla^2 f(\mathbf{x}) \succ 0.$$

- There is **NO** $m > 0$ such that $\lambda_{\min}(\nabla^2 f(\mathbf{x})) \geq m$ for any $\mathbf{x} \in \mathbb{R}^n$

Slow convergence

```
>> pure_newton(f,g,h,[1;1],1e-8)
iter=  1 f(x)=2.8284271247
iter=  2 f(x)=2.8284271247
   :                    :
iter= 30 f(x)=2.8105247315
iter= 31 f(x)=2.7757389625
iter= 32 f(x)=2.6791717153
iter= 33 f(x)=2.4507092918
iter= 34 f(x)=2.1223796622
iter= 35 f(x)=2.0020052756
iter= 36 f(x)=2.0000000081
iter= 37 f(x)=2.0000000000
```

Divergence

```
>> pure_newton(f,g,h,[10;10],1e-8);
iter=  1 f(x)=2000.0009999997
iter=  2 f(x)=1999999999.9999990000
iter=  3 f(x)=1999999999999997300000000000.0000000
iter=  4 f(x)=1999999999999999230000000000000000000....
iter=  5 f(x)=       Inf
```

- Descent of the generated sequence is not guaranteed even when $\nabla^2 f(\mathbf{x}) \succ 0$.
- Rectify by introducting a step size t_k (Damped Newton's method).

Damped Newton's Method

Input: $\varepsilon > 0$ tolerance parameter
Initialization: Pick $\mathbf{x}_0 \in \mathbb{R}^n$ arbitrarily
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 $\nabla^2 f(\mathbf{x}_k)\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$

(2) Pick t_k using constant stepsize, exact line search, or backtracking

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

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

Damped Newton method

Revisiting the previous example: Consider the function $f(x,y) = \sqrt{x^2+1} + \sqrt{y^2+1}$ whose optimal solution is $(0,0)$.

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} (1+x^2)^{-3/2} & 0 \\ 0 & (1+y^2)^{-3/2} \end{bmatrix}. \quad \nabla^2 f(\mathbf{x}) \succ 0.$$

- There is **NO** $m > 0$ such that $\lambda_{\min}(\nabla^2 f(\mathbf{x})) \geq m$ for any $\mathbf{x} \in \mathbb{R}^n$

Pure Newton's Divergence

```
>> pure_newton(f,g,h,[10;10],1e-8);
iter=   1 f(x)=2000.0009999997
iter=   2 f(x)=1999999999.9999990000
iter=   3 f(x)=1999999999999999730000000000.0000000
iter=   4 f(x)=19999999999999999230000000000000000000....
iter=   5 f(x)=         Inf
```

Damped Newton's-Convergence

```
>> newton_backtracking(f,g,h,[10;10],0.5,0.5,1e-8);
iter=   1 f(x)=4.6688169339
iter=   2 f(x)=2.4101973721
iter=   3 f(x)=2.0336386321
    :                :
iter=  16 f(x)=2.0000000005
iter=  17 f(x)=2.0000000000
```

Section 2

Cholesky decomposition

Cholesky decomposition

- When A is symmetric and positive definite, it has a *Cholesky* decomposition given by

$$A = LL^T,$$

where L is a lower triangular matrix (a matrix with zeros everywhere above the diagonal).

- If A is diagonal, the Cholesky decomposition represents the matrix square root:

$$L = A^{\frac{1}{2}} = \begin{bmatrix} \sqrt{A_{11}} & 0 & \ldots 0 \\ 0 & \sqrt{A_{22}} & \ldots 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \ldots \sqrt{A_{nn}} \end{bmatrix}$$

- Example: $A = \begin{bmatrix} 9 & 0 \\ 0 & 25 \end{bmatrix}$

$$L = A^{\frac{1}{2}} = \begin{bmatrix} \sqrt{9} & 0 \\ 0 & \sqrt{25} \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$$

Verify: $LL^T = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 9 & 0 \\ 0 & 25 \end{bmatrix}$

Section 3

Hybrid Gradient-Newton Method

Hybrid Gradient-Newton Method

Motivation:

- Newton's method assumes that $\nabla^2 f(\mathbf{x}) > 0$.
- Gradient Descent does not ue the Hessian.

We avoid the assumption $\nabla^2 f(\mathbf{x}) > 0$ by constructing a hybrid method.

The hybrid gradient-Newton method:

- Does not require the Hessian to be positive definite
- Is likely to converge faster than the gradient method
- **Approach**: At each iteration, determine if $\nabla^2 f(\mathbf{x}_k) \succ 0$
  - If $\nabla^2 f(\mathbf{x}_k) \succ 0$, use a Newton step
  - Otherwise, use a gradient descent step

Hybrid Gradient-Newton Method

Input: $\varepsilon > 0$ tolerance parameter, method for finding stepsize t (constant step size, exact line search, or backtracking

Initialization: Pick $\mathbf{x}_0 \in \mathbb{R}^n$ arbitrarily

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

- If $\nabla^2 f(\mathbf{x}_k) \succ 0$, then take $\mathbf{d}_k$ as the solution to the system $\nabla^2 f(\mathbf{x}_k)\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$. Else, set $\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$.
- Pick stepsize t according to the input method
- Set $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$
- If $\|\nabla f(\mathbf{x}_{k+1})\| \leq \varepsilon$, then stop and $\mathbf{x}_{k+1}$ is the output.

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,