Gradient Method: Part 2

Lecture 4-2 - CMSE 382

Prof. Firas Khasawneh

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

Wed, Feb 4, 2026

This Lecture

Topics:

- Condition number
- Gradient descent solution sensitivity
- Diagonal scaling

Announcements:

- Quiz 2 today, end of class
- No office hours Friday

Section 1

Condition number, Sensitivity, and Diagonal Scaling

Condition number

Recall: positive definiteness. Singular matrix.

An $n \times n$ real symmetric matrix A is called:

- **Positive definite** if $\boldsymbol{x}^T A \boldsymbol{x} > 0$ for every non-zero choice of $\boldsymbol{x}$.
  - The diagonal entries of M are positive.
  - There is an *invertible* matrix B such that $A = B^T B$
- **A** is **positive definite** if and only if all its eigenvalues are positive.

A **singular matrix** is a square matrix that does not have an inverse.

- A matrix is singular if and only if its determinant is 0.
- A matrix has a zero eigenvalue if and only if its determinant is 0.

Condition number

Definition

Let $\mathbf{A}$ be a positive definite matrix. Then the **condition number** of $\mathbf{A}$ is defined by $\kappa(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})}$

- $\kappa \geq 1$
- Matrices with large condition number are **ill-conditioned** and matrices with small condition number are **well-conditioned**.
- A singular matrix has an infinite condition number.

- For a linear system $A\mathbf{x} = b$, the condition number measures the sensitivity of the solution $\mathbf{x}$ to fluctuations in the the observed data $\mathbf{b}$:
$$\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \kappa(A) \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|}$$
- **Gradient descent rate of convergence** of $\mathbf{x}_k$ to a given stationary point $\mathbf{x}^*$ depends on $\kappa(\nabla^2 f(\mathbf{x}^*))$.

Example: Solve the system

$$\begin{bmatrix} 1+10^{-5} & 1 \\ 1 & 1+10^{-5} \end{bmatrix} \mathbf{x} = \mathbf{b}$$

- If $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 0.4999975 \\ 0.4999975 \end{bmatrix}$
- If $\mathbf{b} = \begin{bmatrix} 1.01 \\ 1 \end{bmatrix}$ $\mathbf{x} = \begin{bmatrix} 500.50249748 \\ -499.49750251 \end{bmatrix}$

Diagonal scaling
Motivating example

Assume we want to minimize the quadratic function

$$f(x, y) = 1000x^2 + 40xy + y^2 = \mathbf{x}^T A\mathbf{x}, \text{ where } A = \begin{bmatrix} 1000 & 20 \\ 20 & 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

We can write this problem as $\min_{\mathbf{x}} \{\mathbf{x}^\top \mathbf{A}\mathbf{x}\}$

- the condition number of **A** is $\kappa(A) = 1668.001$.
  - Using gradient descent, we would expect slow convergence.

Diagonal Scaling

Main idea

We have an ill-conditioned $\mathbf{A}$ but we want to minimize $\mathbf{x}^\top \mathbf{A}\mathbf{x}$.
Instead,

- pick a non-singular matrix $\mathbf{S}$
- set $\mathbf{x} = \mathbf{S}\mathbf{y}$
- Then instead minimize the transformation

$$\min_{\mathbf{y}} \{(\mathbf{S}\mathbf{y})^\top \mathbf{A}(\mathbf{S}\mathbf{y})\}$$

- The transformation aims to speed up convergence while still giving a good answer.
- The gradient method for the transformed problem becomes

$$\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \mathbf{S}\mathbf{S}^\top \nabla f(\mathbf{x}_k)$$

- Instead of picking $\mathbf{S}$ directly, set $\mathbf{D} = \mathbf{S}\mathbf{S}^\top$ and pick $\mathbf{D}$ instead.

Input: tolerance parameter $\varepsilon > 0$.

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

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

1. Pick a scaling matrix $\mathbf{D}_k \succ 0$
2. Pick a stepsize t_k
   - For example, using exact line search on the function $g(t) = f(\mathbf{x}_k - t\mathbf{D}_k \nabla f(\mathbf{x}_k))$.
3. Set $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \mathbf{D}_k \nabla f(\mathbf{x}_k)$.
4. If $\|\nabla f(\mathbf{x}_{k+1})\| \leq \varepsilon$, then STOP and $\mathbf{x}_{k+1}$ is the output.

How to choose $\mathbf{D}_k$?

- Newton's method: $\mathbf{D}_k = (\nabla^2 f(\mathbf{x}_k))^{-1}$
- Diagonal method: $\mathbf{D}_k = (\nabla^2 f(\mathbf{x}_k))_{ii}^{-1}$
- Often chosen because they result in fast convergence.

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,