Mathematics Methods for Computer Science

Instructor: Xubo Yang

SJTU-SE DALAB

Lecture

Conjugate Gradients I: Setup

Time for Gaussian Elimination

$$A \in \mathbb{R}^{n \times n} \Rightarrow$$

Time for Gaussian Elimination

$$A \in \mathbb{R}^{n \times n} \Rightarrow O(n^3)$$

Common Case

"Easy to apply, hard to invert"

- Sparse matrices
- Special structure

New Philosophy

Iteratively improve approximation rather than solve in closed form.

For Today

$$A\vec{x} = \vec{b}$$

- Square
- Symmetric
- Positive Definite

Variational Viewpoint

$$A\vec{x} = \vec{b}$$

$$\Updownarrow$$

$$\min_{\vec{x}} \left[\frac{1}{2}\vec{x}^\top A\vec{x} - \vec{b}^\top\vec{x} + c\right]$$

Gradient Descent Strategy

1. Compute search direction

Gradient Descent Strategy

(梯度下降策略)

1. Compute search direction

$$\vec{d}_k \equiv -\nabla f(\vec{x}_{k-1}) = \vec{b} - A\vec{x}_{k-1}$$

Gradient Descent Strategy

1. Compute search direction

$$\vec{d}_k \equiv -\nabla f(\vec{x}_{k-1}) = \vec{b} - A\vec{x}_{k-1}$$

2. Do line search to find

$$\vec{x} \equiv \vec{x}_{k-1} + \alpha_k \vec{d}_k$$

(fixed point iteration)

Line Search Along $\vec{d}$ from $\vec{x}$

$$\min_{\alpha} g(\alpha) \equiv f(\vec{x} + \alpha\vec{d})$$

Line Search Along $\vec{d}$ from $\vec{x}$

$$\min_{\alpha} g(\alpha) \equiv f(\vec{x} + \alpha \vec{d})$$

$$\alpha = \frac{\vec{d}^\top (\vec{b} - A\vec{x})}{\vec{d}^\top A \vec{d}}$$

Gradient Descent with Closed-Form Line Search

$$\vec{d}_k = \vec{b} - A\vec{x}_{k-1}$$

$$\alpha = \frac{\vec{d}^\top \vec{d}}{\vec{d}^\top A\vec{d}}$$

$$\vec{x} = \vec{x}_{k-1} + \alpha_k \vec{d}_k$$

Convergence

Change of backward errors in iteration k (See book)

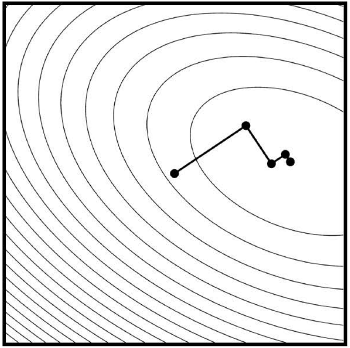
$$R_k \equiv \frac{f(\vec{x}_k) - f(\vec{x}^*)}{f(\vec{x}_{k-1}) - f(\vec{x}^*)} \leq 1 - \frac{1}{\text{cond}\, A}$$

后向误差R_k对于速度、质量的影响：
条件数越大，越poorly conditioned，从而收敛越慢，需要的迭代次数越多，并且误差较大。

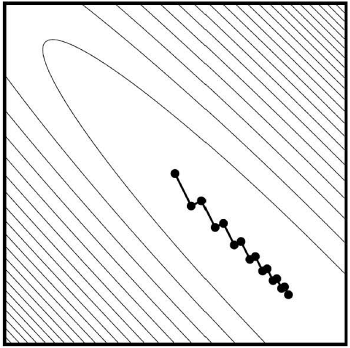
Conclusions:

- Conditioning affects speed and quality
- Unconditional convergence ($\text{cond}\, A \geq 1$)

Visualization



Well conditioned A



Poorly conditioned A

此图说明了条件数对于梯度下降法找根的算法的性能影响很大。

Can We Do Better?

- **Can iterate forever:**
  Should stop after $O(n)$ iterations!
- **Lots of repeated work**
  when poorly conditioned

Observation: "A-Norm"

$$f(\vec{x}) = \frac{1}{2}(\vec{x} - \vec{x}^*)^\top A(\vec{x} - \vec{x}^*) + \text{const}.$$

Observation: "A-Norm"

$$f(\vec{x}) = \frac{1}{2}(\vec{x} - \vec{x}^*)^\top A (\vec{x} - \vec{x}^*) + \text{ const.}$$

$$A = LL^T$$

Observation: “A-Norm”

$$f(\vec{x}) = \frac{1}{2}(\vec{x} - \vec{x}^*)^\top A(\vec{x} - \vec{x}^*) + \text{ const.}$$

$$A = LL^T$$

$$\Longrightarrow f(\vec{x}) = \frac{1}{2}\left\|L^\top(\vec{x} - \vec{x}^*)\right\|_2^2 + \text{ const.}$$

Substitution

$$\vec{y} \equiv L^\top \vec{x}, \quad \vec{y}^* \equiv L^\top \vec{x}^*$$
$$\implies \bar{f}(\vec{y}) = \|\vec{y} - \vec{y}^*\|_2^2$$

Substitution

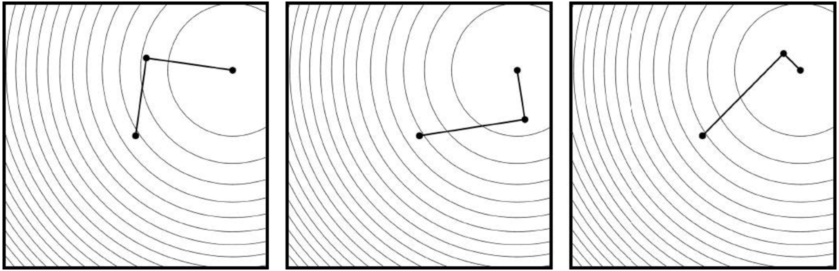
$$\vec{y} \equiv L^\top \vec{x}, \quad \vec{y}^* \equiv L^\top \vec{x}^*$$
$$\implies \bar{f}(\vec{y}) = \|\vec{y} - \vec{y}^*\|_2^2$$

Proposition

Suppose $\{\vec{w}_1, \ldots, \vec{w}_n\}$ are orthogonal in $\mathbb{R}^n$. Then, $\vec{f}$ is minimized in at most n steps by line searching in direction $\vec{w}_1$, then direction $\vec{w}_2$; and so on.

可以如此理解：取正交基x，y，z，则相当于在三个方向上面分别找到最小点，则组合起来就是空间上的最小点。

Substitution



Any two orthogonal directions suffice!

Undoing Change of Coordinates

Line search on $\bar{f}$ along $\vec{w}$ is the same as line search on f along $(L^\top)^{-1}\vec{w}$.

Undoing Change of Coordinates

Line search on $\bar{f}$ along $\vec{w}$ is the same as line search on f along $(L^\top)^{-1}\vec{w}$.

$$\begin{aligned} 0 = \vec{w}_i \cdot \vec{w}_j &= \left(L^\top \vec{v}_i\right)^\top \left(L^\top \vec{v}_j\right) \\ &= \vec{v}_i^\top \left(LL^\top\right) \vec{v}_j = \vec{v}_i^\top A \vec{v}_j \end{aligned}$$

Conjugate Directions

(A-共轭向量)

A-Conjugate Vectors

Two vectors $\vec{v}$ and $\vec{w}$ are A-conjugate if $\vec{v}A\vec{w} = 0$.

Conjugate Directions

A-Conjugate Vectors

Two vectors $\vec{v}$ and $\vec{w}$ are A-conjugate if $\vec{v}A\vec{w} = 0$.

Corollary

Suppose $\{\vec{v}_1, \ldots, \vec{v}_n\}$ are A-conjugate. Then, f is minimized in at most n steps by line searching in direction $\vec{v}_1$; then direction $\vec{v}_2$, and so on.

High-Level Ideas So Far

- Steepest descent may not be fastest descent (surprising!)

High-Level Ideas So Far

- Steepest descent may not be fastest descent (surprising!)
- Two inner products:

$$\vec{v} \cdot \vec{w}$$

$$\langle \vec{v}, \vec{w} \rangle_A \equiv \vec{v}^\top A \vec{w} \quad \text{(A内积)}$$

New Problem

Find n A-conjugate directions.

Gram-Schmidt?

Gram-Schmidt?

- Potentially unstable

Gram-Schmidt?

- Potentially unstable
- Storage increases with each iteration

Another Clue

$$\vec{r}_k \equiv \vec{b} - A\vec{x}_k$$

Another Clue

$$\vec{r}_k \equiv \vec{b} - A\vec{x}_k$$
$$\vec{r}_{k+1} = \vec{r}_k - \alpha_{k+1} A\vec{v}_{k+1}$$

Another Clue

$$\vec{r}_k \equiv \vec{b} - A\vec{x}_k$$
$$\vec{r}_{k+1} = \vec{r}_k - \alpha_{k+1} A\vec{v}_{k+1}$$

Proposition

When performing gradient descent on f,
$$\operatorname{span}\{\vec{r}_0, \ldots, \vec{r}_k\} = \operatorname{span}\{\vec{r}_0, A\vec{r}_0, \ldots, A^k\vec{r}_0\}$$

Another Clue

$$\vec{r}_k \equiv \vec{b} - A\vec{x}_k$$
$$\vec{r}_{k+1} = \vec{r}_k - \alpha_{k+1} A\vec{v}_{k+1}$$

Proposition

When performing gradient descent on f,
$$\operatorname{span}\{\vec{r}_0, \ldots, \vec{r}_k\} = \operatorname{span}\{\vec{r}_0, A\vec{r}_0, \ldots, A^k\vec{r}_0\}$$

Krylov space?!

Gradient Descent: Issue

$$\vec{x}_k - \vec{x}_0 \neq \underset{\vec{v} \in \operatorname{span}\left\{\vec{r}_0, A\vec{r}_0, \ldots, A^{k-1}\vec{r}_0\right\}}{\operatorname{argmin}} f\left(\vec{x}_0 + \vec{v}\right)$$

Gradient Descent: Issue

$$\vec{x}_k - \vec{x}_0 \neq \underset{\vec{v} \in \operatorname{span}\left\{\vec{r}_0, A\vec{r}_0, \ldots, A^{k-1}\vec{r}_0\right\}}{\operatorname{argmin}} f\left(\vec{x}_0 + \vec{v}\right)$$

But if this did hold ...

Convergence in n steps!