



计算方法

刘景铨

计算机软件新技术国家重点实验室
南京大学



回顾线性迭代方法

- 矩阵多项式的特征空间: $p(A)v = p(\lambda)v$
 - $A \rightarrow p(A)$
 - 特征空间 $\{\lambda_1, \lambda_2, \dots, \lambda_n\} \rightarrow \{p(\lambda_1), p(\lambda_2), \dots, p(\lambda_n)\}$
- 应用之 Power method: 收敛到最大特征值对应的方向
- 这节课:
 - Cayley-Hamilton: 存在矩阵多项式使得 $A^{-1} = p(A)$
 - 通过迭代法, 找出矩阵多项式使得 $A^{-1} \approx p(A)$
 - 解线性方程组的 Krylov 子空间迭代法



一个神奇的多项式

要解 $Ax = b$, 可否求出多项式 $p(A)$ 使得 $p(A)A = I$?

记 $q(x) = 1 - xp(x)$, 则 $q(A)$ 为全零矩阵。

这样的多项式 $q(x)$ 一定存在!

考虑由 $\{I, A, A^2, A^3, \dots, A^{n^2}\}$ 张成的向量空间, 维度是多少? 练习: 证明它们是线性相关的。

Cayley-Hamilton: 只需要 n 次多项式即可



Cayley-Hamilton

考虑关于 x 的 n 次多项式 $q_A(x) = \det(A - xI)$

假设 $q_A(x) = c_0 + c_1x + \cdots + c_nx^n$

类似地定义 $q_A(B) = c_0I + c_1B + \cdots + c_nB^n$

定理(Cayley-Hamilton): $q_A(A) = \mathbf{0}$ 为全零矩阵

非平凡的: 注意 $\det(A - xI)$ 仅仅为一个数



Cayley-Hamilton (选讲)

考虑关于 x 的 n 次多项式 $q_A(x) = \det(A - xI)$

定理(Cayley-Hamilton): $q_A(A) = \mathbf{0}$ 为全零矩阵

证明(sketch): 先考虑 A 是可以对角化的。注意 $q_A(A)$ 也一定是可以对角化的。

取 A 的任一特征值 $\lambda \in R$ 和特征向量 v 满足 $Av = \lambda v$ 。
由前面的观察可知, v 也是 $q_A(A)$ 的特征向量, 而且特征值满足 $q_A(A)v = q_A(\lambda)v = \mathbf{0}$

可见 $q_A(A)$ 的每个特征向量对应的特征值均为 $\mathbf{0}$

又因为 $q_A(A)$ 可以对角化, 此时 $q_A(A)$ 只能是全零矩阵



Cayley-Hamilton (选讲)

考虑关于 x 的 n 次多项式 $q_A(x) = \det(A - xI)$

定理(Cayley-Hamilton): $q_A(A) = \mathbf{0}$ 为全零矩阵

证明(cont'd, sketch): 对于 A 不可以对角化的情况, 可以使用Jordan标准形分解; 或者利用可对角化矩阵在复数域上的稠密性(任意一个矩阵, 均可表示为可对角化矩阵的极限)



Cayley-Hamilton

考虑关于 x 的 n 次多项式 $q_A(x) = \det(A - xI)$

定理(Cayley-Hamilton): $q_A(A) = \mathbf{0}$ 为全零矩阵

注意 $q_A(0) = \det(A) \neq \mathbf{0}$

$$A^{-1} = \frac{(-1)^{n-1}}{\det(A)} (A^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1I)$$

$$A^{-1} = p(A)$$

如何计算 $q(x)$? 高斯消元?



Richardson iteration

一个特殊情形: 给定实数 $0 < a < 1$, 如何用 a 表示出 $a^{-1}b$?

$$b + (1 - a)b + (1 - a)^2b + \dots = \frac{b}{1 - (1 - a)} = a^{-1}b$$

等价于 $x_0 = \mathbf{0}$, $x_{k+1} = (1 - a)x_k + b$

Richardson iteration (对于对称正定的矩阵 A):

$$\begin{aligned} x_0 &= \mathbf{0} \\ \overrightarrow{x_{k+1}} &= (I - \alpha A)\overrightarrow{x_k} + \alpha \overrightarrow{b} = \overrightarrow{x_k} - \alpha(A\overrightarrow{x_k} - \overrightarrow{b}) \end{aligned}$$



Richardson iteration

一个特殊情形: 给定实数 $0 < a < 1$, 如何用 a 表示出 $a^{-1}b$?

$$b + (1 - a)b + (1 - a)^2b + \dots = \frac{b}{1 - (1 - a)} = a^{-1}b$$

等价于 $x_0 = \mathbf{0}$, $x_{k+1} = (1 - a)x_k + b$

Richardson iteration:

$$\begin{aligned} x_0 &= \mathbf{0} \\ \vec{x}_{k+1} &= (I - \alpha A)\vec{x}_k + \alpha \vec{b} = \vec{x}_k - \alpha(A\vec{x}_k - \vec{b}) \end{aligned}$$

矩阵多项式的视角: 特征空间: $p(A)v = p(\lambda)v$

- $A \rightarrow p(A)$
- 特征空间 $\{\lambda_1, \lambda_2, \dots, \lambda_n\} \rightarrow \{p(\lambda_1), p(\lambda_2), \dots, p(\lambda_n)\}$

最优化的视角: 对于对称的矩阵 A , 这正好是目标函数 $\frac{1}{2}x^T Ax - b^T x$ 的梯度下降方法!

$$\nabla \left(\frac{1}{2}x^T Ax - b^T x \right) = \frac{1}{2}(A + A^T)x - b$$



Detour: An Optimization Perspective

给定连续可导 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 最小化 f

例子:

最小二乘法: $f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2$

向量投影: $f(c) = \|c\vec{a} - \vec{b}\|_2^2$

求伪逆(Pseudoinverse): $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$, subject to $\mathbf{Ax} = \mathbf{b}$

对称矩阵的特征值: $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x}$, subject to $\mathbf{x}^\top \mathbf{x} = 1$

线性规划: $f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x}$, subject to $\mathbf{Ax} \geq \mathbf{b}$

主成分分析 (Principal component analysis):

$$f(\mathbf{C}) = \|\mathbf{X} - \mathbf{CC}^\top \mathbf{X}\|_F, \text{ subject to } \mathbf{CC}^\top = \mathbf{I}_{d \times d}$$

离散组合优化: 最小生成树, 最小割、最大流 (网络流), 最短路径, 最大匹配; 最大独立集、团、支配集, 最小覆盖, 背包问题, 最大割, max-SAT, 旅行商问题, 最长路径 (哈密顿路径)



Optimality notion

给定连续可导 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 最小化 f

x_* 是一个全局最优解(Global minimum), 如果

$$f(x) \geq f(x_*), \forall x$$

x_* 是一个局部最优解(Local minimum), 如果

$$\exists \delta > 0, \forall x: |x - x_*| < \delta, \quad \text{均有 } f(x) \geq f(x_*)$$



Optimality condition from calculus

给定连续可导 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 和一个点 x_0

$$\nabla f(x) := \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

$$f(x) \approx f(x_0) + \nabla f(x_0) \cdot (x - x_0), \quad \forall x$$

特别地, 可以选取 $x - x_0 = \alpha (\nabla f(x_0))^\top$

$$f\left(x_0 + \alpha (\nabla f(x_0))^\top\right) - f(x_0) \approx \alpha \|\nabla f(x_0)\|_2^2$$

当 α 足够小的时候, α 的符号决定 f 局部的增减性

当 $\nabla f(x_0) = 0$, 则点 x_0 为驻点(stationary point)

此时需要看二阶导数



Optimality condition from calculus

给定连续可导 $f: \mathbb{R}^n \rightarrow \mathbb{R}$

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

通常也把Hessian矩阵 H_f 记作 $\nabla^2 f$



Optimality condition from calculus

给定连续可导 $f: \mathbb{R}^n \rightarrow \mathbb{R}, \forall x$

$$f(x) \approx f(x_0) + \nabla f(x_0) \cdot (x - x_0) + \frac{1}{2} (x - x_0)^\top H_f(x_0) (x - x_0)$$

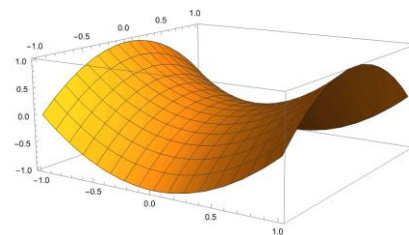
如果 $\nabla f(x_0) = 0$

$H_f(x_0) > 0$: 局部最小

$H_f(x_0) < 0$: 局部最大

$H_f(x_0)$ 同时有正的和负的特征值: saddle point

$H_f(x_0)$ 不可逆: 不一定是局部极值 (Morse theory)

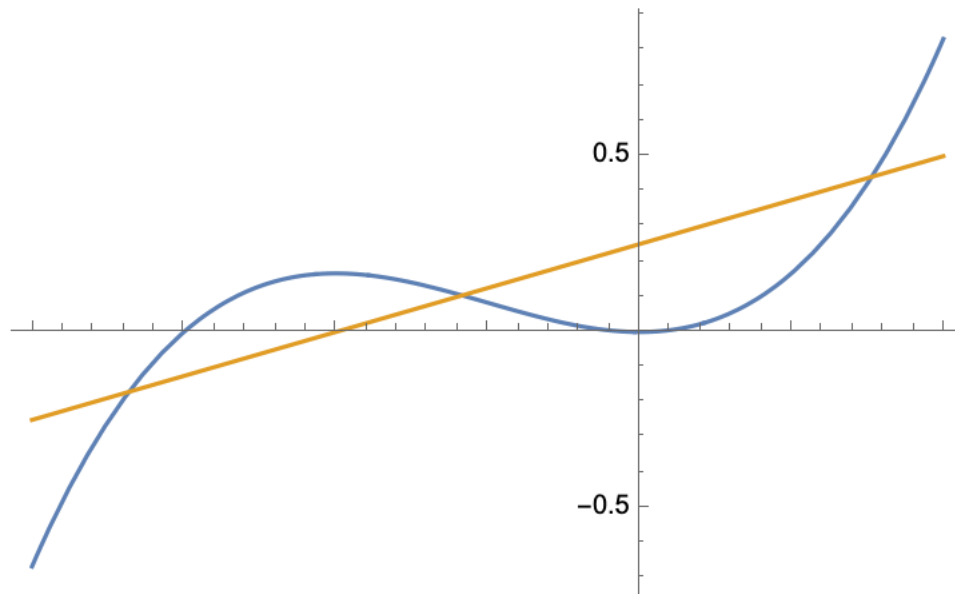
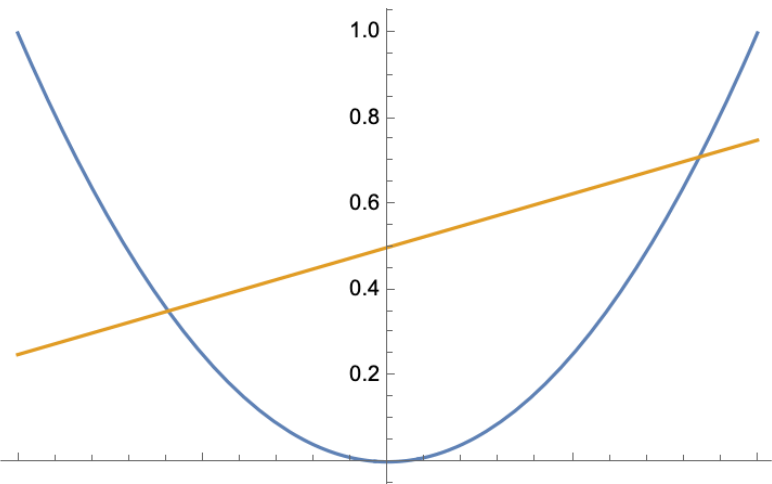




Convexity

如果给定的不仅仅是一个点上的导数，而是一个邻域上导数的信息，那么判断标准可以有所简化。

例子：凸函数(convex function)





Convexity

如果给定的不仅仅是一个点上的导数，而是一个邻域上导数的信息，那么判断标准可以有所简化。

例子：凸函数(convex function)的等价刻画

- Jensen's inequality:

$$\forall t \in [0,1], \forall x, y \in \mathbb{R}^n, f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

- 一阶条件（对可导的 f ）：

$$\forall x, y \in \mathbb{R}^n, f(x) \geq f(y) + \nabla f(y) \cdot (x - y)$$

- 二阶条件（对二阶可导的 f ）：

$$\forall x \in \mathbb{R}^n, H_f(x) \succeq 0$$



Steepest descent

给定连续可导 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 最小化 f

使用函数求根的方法, 找极值点

- $\nabla f: \mathbb{R}^n \rightarrow \mathbb{R}^n$, 需要找到 $\nabla f = 0$

梯度下降方法:

- 在当前的点, 找到“最速下降”的方向 (steepest descent)
- $f(x) \approx f(x_0) + \nabla f(x_0) \cdot (x - x_0)$
- 假设 $f(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0)$
- “最速下降”的方向是什么?
 - 找到单位长度的向量 y , 使得 $\nabla f(x_0) \cdot y$ 长度最大化



Steepest descent

给定向量 \vec{a} ，求单位长度的向量 \vec{b} ，使得 $|\langle \vec{a}, \vec{b} \rangle|$ 最大化。

Cauchy-Schwarz不等式：

$$|\langle \vec{a}, \vec{b} \rangle| \leq \|\vec{a}\|_2 \|\vec{b}\|_2,$$

当且仅当存在乘数(scalar)使得 $\vec{a} = c \vec{b}$ 时，等号可以取到

- 假设 $f(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0)$
- “最速下降”的方向是什么？
 - 找到单位长度的向量 y ，使得 $\nabla f(x_0) \cdot y$ 长度最大化
 - $y = \alpha \nabla f(x_0)$



回到Richardson iteration

一个特殊情形: 给定实数 $0 < a < 1$, 如何用 a 表示出 $a^{-1}b$?

$$b + (1 - a)b + (1 - a)^2b + \dots = \frac{b}{1 - (1 - a)} = a^{-1}b$$

等价于 $x_0 = \mathbf{0}$, $x_{k+1} = (1 - a)x_k + b$

Richardson iteration:

$$\begin{aligned} x_0 &= \mathbf{0} \\ \overrightarrow{x_{k+1}} &= (I - \alpha A)\overrightarrow{x_k} + \alpha \overrightarrow{b} = \overrightarrow{x_k} - \alpha(A\overrightarrow{x_k} - \overrightarrow{b}) \end{aligned}$$

矩阵多项式的视角: 特征空间: $p(A)v = p(\lambda)v$

- $A \rightarrow p(A)$
- 特征空间 $\{\lambda_1, \lambda_2, \dots, \lambda_n\} \rightarrow \{p(\lambda_1), p(\lambda_2), \dots, p(\lambda_n)\}$

另一个视角: 对于对称的 A , 这正好是目标函数 $\frac{1}{2}x^\top Ax - b^\top x$ 的梯度下降方法!

$$\nabla \left(\frac{1}{2}x^\top Ax - b^\top x \right) = \frac{1}{2}(A + A^\top)x - b = A\overrightarrow{x_k} - \overrightarrow{b}$$



Richardson iteration

$$A^{-1} = p(A)$$

可否近似 $p(A)$?

等价地, 记 $q(x) = 1 - xp(x)$; 则需要寻找多项式 q 满足 $q(0) = 1, q(x) \approx 0, \forall x > 0$

要解 $Ax = b$, 即找出 $x_* = p(A)b \in \text{span}\{b, Ab, A^2b, \dots\}$

Richardson iteration: 设 $x_0 = 0$,

$$x_{k+1} = (I - \alpha A)x_k + \alpha b$$

误差 $e_k = x_k - x_*$ 满足 $e_k = (I - \alpha A)e_{k-1} = (I - \alpha A)^k e_0$

收敛性: 当且仅当 $\rho(I - \alpha A) = \max\{|1 - \alpha\lambda_1|, |1 - \alpha\lambda_n|\} < 1$

可以令 $1 - \alpha\lambda_1 = -(1 - \alpha\lambda_n)$ 即 $\alpha = \frac{2}{\lambda_1 + \lambda_n}$

此时 $\rho(I - \alpha A) = \frac{\lambda_n - \lambda_1}{\lambda_1 + \lambda_n} = 1 - \frac{2}{\frac{\lambda_n + \lambda_1}{\lambda_1}}$, 收敛需要 $O\left(\left(1 + \frac{\lambda_n}{\lambda_1}\right) \cdot \log \frac{1}{\epsilon}\right)$ 步

回顾: $\frac{\lambda_n}{\lambda_1}$ 正是条件数!



回到Richardson iteration: Steepest descent

Richardson iteration:

$$\begin{aligned} \mathbf{x}_0 &= \mathbf{0} \\ \overrightarrow{\mathbf{x}}_{k+1} &= (I - \alpha A)\overrightarrow{\mathbf{x}}_k + \alpha \overrightarrow{\mathbf{b}} = \overrightarrow{\mathbf{x}}_k - \alpha(A\overrightarrow{\mathbf{x}}_k - \overrightarrow{\mathbf{b}}) \end{aligned}$$

对于对称的 A , 这正好是目标函数 $\frac{1}{2} \mathbf{x}^\top A \mathbf{x} - \mathbf{b}^\top \mathbf{x}$ 的梯度下降方法!

$$\nabla \left(\frac{1}{2} \mathbf{x}^\top A \mathbf{x} - \mathbf{b}^\top \mathbf{x} \right) = \frac{1}{2} (A + A^\top) \mathbf{x} - \mathbf{b} = A\overrightarrow{\mathbf{x}}_k - \overrightarrow{\mathbf{b}}$$

目标函数 $\frac{1}{2} \mathbf{x}^\top A \mathbf{x} - \mathbf{b}^\top \mathbf{x}$ 是convex的: A 是正定的

利用strong convexity, 也可以证明梯度下降的收敛时间正比于条件数 $\frac{\lambda_n}{\lambda_1}$

参考: Convex Optimization by Stephen Boyd and Lieven Vandenberghe



Richardson iteration

$$A^{-1} = p(A)$$

可否近似 $p(A)$?

等价地, 记 $q(x) = 1 - xp(x)$; 需要寻找 $q(0) = 1, q(x) \approx 0, \forall x > 0$

要解 $Ax = b$, 即找出 $x_* = p(A)b \in \text{span}\{b, Ab, A^2b, \dots\}$

Richardson iteration: 设 $x_0 = 0$,

$$x_{k+1} = (I - \alpha A)x_k + \alpha b$$

注意, 这里的 $x_k = p_k(A)b \in \text{span}\{b, Ab, A^2b, \dots, A^{k-1}b\}$; Krylov子空间

可以证明: 相当于考虑 $q(x) = 1 - xp(x) = (1 - \alpha x)^k$

$$-Ae_k = b - Ax_k = (I - Ap_k(A))b = q_k(A)b$$

$$-e_k = x_* - x_k = (I - p_k(A)A)x_* = q_k(A)x_*$$



Chebyshev iteration (选讲)

$$A^{-1} = p(A)$$

目标: 近似 $p(A)$

等价地, 记 $q(x) = 1 - xp(x)$; 需要寻找 $q(0) = 1, q(x) \approx 0, \forall x > 0$

要解 $Ax = b$, 即找出 $x_* = p(A)b \in \text{span}\{b, Ab, A^2b, \dots\}$

Chebyshev iteration:

$$x_{k+1} = (I - \alpha_k A)x_k + \alpha_k b$$

误差 $e_k = x_k - x_*$ 满足 $e_k = \prod_i (I - \alpha_i A) e_0$

$$\|e_k\| \leq \left\| \prod_i (I - \alpha_i A) \right\| \|e_0\|$$

要最小化 $\|\prod_i (I - \alpha_i A)\|$: Chebyshev 多项式!

注意, 这里的 $x_k \in \text{span}\{b, Ab, A^2b, \dots\}$



Conjugate gradients

给定正定矩阵 A ，可定义内积：

$$\langle x, y \rangle_A := x^\top A y$$

- 线性
- 对称性
- 正定性 $\langle x, x \rangle_A > 0 \quad \forall x \neq 0$

如果 $\langle x, y \rangle_A = 0$ ，则称 x, y 关于 A 共轭

内积自带的范数： $\|x\|_A^2 := x^\top A x$

令 x_* 满足 $Ax_* = b$ ，考虑关于 x 的函数：

$$\frac{1}{2} x^\top A x - b^\top x = \frac{1}{2} \|x - x_*\|_A^2 + \text{constant}$$



Conjugate gradients

给定正定矩阵 A ，可定义内积：

$$\langle x, y \rangle_A := x^\top A y$$

性质：共轭向量一定是线性无关的。因此可以组成一组基。

证明：假设 p_1, p_2, \dots, p_k 互相共轭，并且是线性相关的：存在 $\alpha_1, \dots, \alpha_k$ 使得：

$$\alpha_1 p_1 + \dots + \alpha_k p_k = 0$$

另一方面

$$\begin{aligned} & (\alpha_1 p_1 + \dots + \alpha_k p_k)^\top A (\alpha_1 p_1 + \dots + \alpha_k p_k) \\ &= \alpha_1^2 \|p_1\|_A^2 + \alpha_2^2 \|p_2\|_A^2 + \dots + \alpha_k^2 \|p_k\|_A^2 > 0 \end{aligned}$$

与上式矛盾



Conjugate gradients

$$A^{-1} = p(A)$$

目标：近似 $p(A)$

等价地，记 $q(x) = 1 - xp(x)$ ；需要寻找 $q(0) = 1$ ， $q(x) \approx 0, \forall x > 0$

要解 $Ax = b$ ，即找出 $x_* = p(A)b \in \text{span}\{b, Ab, A^2b, \dots\}$

考虑所有这样的迭代算法：

记 Krylov 子空间 $K_0 = \{0\}$, $K_i = \text{span}\{b, Ab, \dots, A^{i-1}b\}$

设 $x_i = \operatorname{argmin}_{x \in K_i} \|x - x_*\|_A^2$ ，其中 x_* 满足 $Ax_* = b$

引理：记 $v_i := x_i - x_{i-1}$ ，则 $\{v_i\}$ 两两共轭： $v_i^\top Av_j = 0, \forall i \neq j$



Conjugate gradients

记Krylov子空间 $K_0 = \{0\}$, $K_i = \text{span}\{b, Ab, \dots, A^{i-1}b\}$

设 $x_i = \operatorname{argmin}_{x \in K_i} \|x - x_*\|_A^2$, 其中 x_* 满足 $Ax_* = b$

引理: 记 $v_i := x_i - x_{i-1}$, 则 $\{v_i\}$ 两两共轭: $v_i^\top Av_j = 0, \forall i \neq j$

证明: 假设 $i < j$. 注意到 x_j 的定义, 是 K_j 中最小化目标函数 $\|x - x_*\|_A^2$ 的。

这意味着 $\nabla \frac{1}{2} \|x_j - x_*\|_A^2 = Ax_j - b$ 必须与 K_j 正交:

否则, 在 K_j 中可以往梯度的投影方向上进行扰动, 令目标函数更小
因此 $Ax_j - b$ 也与 K_{j-1} 正交。

类似地, $Ax_{j-1} - b$ 也必须与 K_{j-1} 正交。因此,

$$Av_j = Ax_j - Ax_{j-1} \in K_{j-1}^\perp$$

而 $x_i, x_{i-1} \in K_i \Rightarrow v_i \in K_i \subseteq K_{j-1}$, 因此 $v_i^\top Av_j = 0$

一个推论: $K_i = \text{span}\{v_1, v_2, \dots, v_i\}$



Conjugate gradients

记 Krylov 子空间 $K_0 = \{0\}$, $K_i = \text{span}\{b, Ab, \dots, A^{i-1}b\}$

设 $x_i = \operatorname{argmin}_{x \in K_i} \|x - x_*\|_A^2$, 其中 x_* 满足 $Ax_* = b$

引理: 记 $v_i := x_i - x_{i-1}$, $r_i := b - Ax_i$, 则

$$v_i = \frac{v_i^\top r_{i-1}}{\|r_{i-1}\|^2} \left(r_{i-1} - \frac{r_{i-1}^\top A v_{i-1}}{v_{i-1}^\top A v_{i-1}} v_{i-1} \right)$$

证明(sketch): 首先注意到 $Ax_{i-1} - b \in K_{i-1}^\perp$, 但是 $Ax_{i-1} - b \in K_i$,

因此 $K_i = \text{span}\{v_1, v_2, \dots, v_{i-1}, r_{i-1}\}$

换言之, 可以写出 $v_i = c_0 r_{i-1} + \sum_{j=1}^{i-1} c_j v_j$

要确定 c_0 , 考虑 $v_i^\top r_{i-1} = c_0 \|r_{i-1}\|^2$

要确定 c_{i-1} , 注意到 $v_j^\top A v_i = 0, \forall j < i - 1$, 令 $v_{i-1}^\top A v_i = 0$ 即可

类似地, 令 $v_j^\top A v_i = 0$ 可得到其它的 $c_j = 0$



Conjugate gradients

记 Krylov 子空间 $K_0 = \{0\}, K_i = \text{span}\{b, Ab, \dots, A^{i-1}b\}$

设 $x_i = \text{argmin}_{x \in K_i} \|x - x_*\|_A^2$, 其中 x_* 满足 $Ax_* = b$

引理: 记 $v_i := x_i - x_{i-1}, r_i := b - Ax_i$, 则

$$v_i = \frac{v_i^\top r_{i-1}}{\|r_{i-1}\|^2} (r_{i-1} - \frac{r_{i-1}^\top A v_{i-1}}{v_{i-1}^\top A v_{i-1}} v_{i-1})$$

简化: 令 $d_i = \frac{\|r_{i-1}\|^2}{v_i^\top r_{i-1}} v_i$,

即有 $x_i = x_{i-1} + \frac{\|r_{i-1}\|^2}{d_i^\top A d_i} d_i, d_i = r_{i-1} + \frac{\|r_{i-1}\|^2}{\|r_{i-2}\|^2} d_{i-1}$

可以证明: conjugate gradients 在运行 k 步之后, 误差最多是正比于从所有 $\deg(q) < k$ 且 $q(0) = 1$ 的多项式中, 最小的 $\|q\|_\infty$

事实上:

$$\|x_k - x_*\|_A^2 \leq \inf_{q(0)=1, \deg(q) \leq k} \max_i q(\lambda_i)^2 \cdot \|b\|_{A^{-1}}^2$$

与高斯消元相比: 对于稀疏矩阵, 恰当运用迭代法可以收敛更快
在没有数值误差的情况下, CG 最多迭代 $n + 1$ 次

Conjugate Gradient Method

$x_0 =$ initial guess

$d_0 = r_0 = b - Ax_0$

for $k = 0, 1, 2, \dots, n - 1$

if $r_k = 0$, stop, end

$$\alpha_k = \frac{r_k^\top r_k}{d_k^\top A d_k}$$

$$x_{k+1} = x_k + \alpha_k d_k$$

$$r_{k+1} = r_k - \alpha_k A d_k$$

$$\beta_k = \frac{r_{k+1}^\top r_{k+1}}{r_k^\top r_k}$$

$$d_{k+1} = r_{k+1} + \beta_k d_k$$

end



小结

- **Richardson iteration**

- 两个视角：矩阵多项式、梯度最速下降
- 思考：如果换作其它范数下面的最速下降？
- 收敛性：当且仅当 $\rho(I - \alpha A) = \max\{|1 - \alpha\lambda_1|, |1 - \alpha\lambda_n|\} < 1$
- 记 $x_k = p_k(A)b \in \text{span}\{b, Ab, A^2b, \dots, A^{k-1}b\}$ ；Krylov子空间
- $x_* - x_k = (I - p_k(A)A)x_* = q_k(A)x_*$
- 收敛所需迭代次数正比于条件数 $\frac{\lambda_n}{\lambda_1}$

- **Conjugate gradients**（共轭梯度方法）

- 记Krylov子空间 $K_0 = \{0\}, K_i = \text{span}\{b, Ab, \dots, A^{i-1}b\}$
- $x_i = \text{argmin}_{x \in K_i} \|x - x_*\|_A^2$ ，其中 x_* 满足 $Ax_* = b$
- CG的收敛所需迭代次数正比于 $\sqrt{\frac{\lambda_n}{\lambda_1}}$ ，依然取决于条件数
- 精确数值：最多迭代 $n+1$ 次；有限精度：也需要选择合适的停止条件

- 对稀疏矩阵，或者有好的初始猜测时，迭代比高斯消元更加适合



预条件(Preconditioning) (选讲)

要解 $Ax = b$, 选取矩阵 M , 改为解 $M^{-1}Ax = M^{-1}b$

$$\text{cond}(M^{-1}A) \neq \text{cond}(A)$$

注意: Richardson iteration和CG都需要正定矩阵
 $M^{-1}A$ 可能不再是正定的, 甚至可能不是对称的

假设 M 是对称正定, 则有Cholesky分解: $M = EE^T$, 并且
$$\text{cond}(M^{-1}A) = \text{cond}(E^{-1}AE^{-T})$$



预条件(Preconditioning) (选讲)

假设 M 是对称正定, 则有Cholesky分解: $M = EE^T$, 并且

$$\text{cond}(M^{-1}A) = \text{cond}(E^{-1}AE^{-T})$$

证明: $E^{-1}AE^{-T}$ 是对称正定的, 有完整的特征基, 因此只需要证明 $M^{-1}A$ 和 $E^{-1}AE^{-T}$ 有着一样的特征值即可。任取 $E^{-1}AE^{-T}$ 的特征向量和特征值

$$\begin{aligned} E^{-1}AE^{-T}v &= \lambda v \Rightarrow E^{-T}E^{-1}AE^{-T}v = \lambda E^{-T}v \Rightarrow M^{-1}Ay \\ &= \lambda y \end{aligned}$$

因为它们的奇异值一样, 它们的条件数也是一样的

因此, 可以先解 $E^{-1}AE^{-T}y = E^{-1}b$, 这是对称正定的系统
然后解得 $x = E^{-T}y$



下节课预告

图的邻接矩阵

图上的随机游走

数值分析在图论与组合中的应用