



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：线性迭代方法的收敛性

我们看到了迭代法解线性方程组的例子

解 $Ax = b$ 时：Jacobi 和 Gauss-Seidel 迭代都通过写出不动点方程，使用不动点迭代

解 $Ax = \lambda x$ ：同时有两个变量： λ 和 x 。严格来说，它们的乘积并不是线性关系，但（线性）迭代法依然适用

Power method: 收敛到最大特征值的方向



Inverse Power method

幂迭代告诉我们如何找到 λ_1 与 v_1
其它特征值和特征向量呢？

给定方阵 A . 假设 A 有 n 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 和线性无关的特征向量为 v_1, v_2, \dots, v_n

A^{-1} 的特征值？

$$\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_n}$$

特征向量？

$$v_1, v_2, \dots, v_n$$

迭代需要使用到 A^{-1} ？ 只需要解线性方程：LU分解



Inverse Power method

幂迭代告诉我们如何找到 λ_1 与 v_1
其它特征值和特征向量呢？

给定方阵 A . 假设 A 有 n 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 和线性无关的特征向量为 v_1, v_2, \dots, v_n

$(A - qI)^{-1}$ 的特征值？

$$\frac{1}{\lambda_1 - q}, \frac{1}{\lambda_2 - q}, \dots, \frac{1}{\lambda_n - q}$$

特征向量？

$$v_1, v_2, \dots, v_n$$



Inverse Power method

给定方阵 A . 假设 A 有 n 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 和线性无关的特征向量为 v_1, v_2, \dots, v_n

$(A - qI)^{-1}$ 的特征值?

$$\frac{1}{\lambda_1 - q}, \frac{1}{\lambda_2 - q}, \dots, \frac{1}{\lambda_n - q}$$

特征向量?

$$v_1, v_2, \dots, v_n$$

如果对 $(A - qI)^{-1}$ 使用幂迭代呢?

$$\text{即 } x_k = (A - qI)^{-1}x_{k-1}$$

注意: 可以通过高斯消元, 先求出矩阵 $B = (A - qI)^{-1}$ 的LU分解, 再进行迭代; 对于迭代次数远远比 n 小的稀疏矩阵, 也可以尝试每一次迭代都解一次线性方程组

$$\text{收敛到 } \frac{1}{\lambda_k - q} = \max_{1 \leq j \leq n} \frac{1}{|\lambda_j - q|}$$

收敛速度? 取决于第二大特征值与最大特征值的比值——如果 q 越接近一个特征值, 越好

如何选择 q ? 可以结合Rayleigh quotient \Rightarrow Rayleigh quotient iteration



Rayleigh Quotient Iteration

$$\sigma_k = \frac{\mathbf{x}_{k-1}^\top A \mathbf{x}_{k-1}}{\mathbf{x}_{k-1}^\top \mathbf{x}_{k-1}},$$
$$\mathbf{w}_k = (A - \sigma_k I)^{-1} \mathbf{x}_{k-1},$$
$$\mathbf{x}_k = \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$$

注意 $(A - \sigma_k I)$ 不再是固定的
迭代次数更少，但是每一次迭代的计算代价更多

思考：当 σ_k 很接近一个特征值的时候， $A - \sigma_k I$ 也非常接近不可逆 (ill conditioning)。这会带来问题吗？



Power method

幂迭代告诉我们如何找到 λ_1 与 v_1
其它特征值和特征向量呢？

$$x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

如果 $\alpha_1 \neq 0$ ，随着 $k \rightarrow \infty$, $A^k x \rightarrow \lambda_1^k \alpha_1 v_1$

如果 $\alpha_1 = 0$ 呢？

• 如果 $\alpha_2 \neq 0$ ，随着 $k \rightarrow \infty$, $A^k x \rightarrow \lambda_2^k \alpha_2 v_2$

• 如果 $\alpha_2 = 0, \alpha_3 \neq 0$ ，随着 $k \rightarrow \infty$, $A^k x \rightarrow \lambda_3^k \alpha_3 v_3$



Power method

幂迭代告诉我们如何找到 λ_1 与 v_1
其它特征值和特征向量呢？

如果 $\alpha_1 \neq 0$ ，随着 $k \rightarrow \infty$, $A^k x \rightarrow \lambda_1^k \alpha_1 v_1$

如果 $\alpha_1 = 0$ 呢？

每一次迭代，通过投影确保与已知的特征向量
都是正交的



QR分解与同时迭代

如何同时找出所有的特征值与特征向量？

假设 A 是实数对称矩阵。同时进行幂迭代，并保持向量两两正交。

回忆：正交化，即QR分解

最开始

$$\left[A \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \mid A \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \mid \cdots \mid A \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right] = \left[\bar{q}_1^1 \mid \cdots \mid \bar{q}_m^1 \right] \begin{bmatrix} r_{11}^1 & r_{12}^1 & \cdots & r_{1m}^1 \\ & r_{22}^1 & & \vdots \\ & & \ddots & \vdots \\ & & & r_{mm}^1 \end{bmatrix}$$

$$\begin{aligned} A\bar{Q}_1 &= \left[A\bar{q}_1^1 \mid A\bar{q}_2^1 \mid \cdots \mid A\bar{q}_m^1 \right] \\ &= \left[\bar{q}_1^2 \mid \bar{q}_2^2 \mid \cdots \mid \bar{q}_m^2 \right] \begin{bmatrix} r_{11}^2 & r_{12}^2 & \cdots & r_{1m}^2 \\ & r_{22}^2 & & \vdots \\ & & \ddots & \vdots \\ & & & r_{mm}^2 \end{bmatrix} \\ &= \bar{Q}_2 R_2. \end{aligned}$$



QR分解与同时迭代

如何同时找出所有的特征值与特征向量？

假设 A 是实数对称矩阵。同时进行幂迭代，并保持向量两两正交。

$$\begin{aligned} A\bar{Q}_1 &= [A\bar{q}_1 | A\bar{q}_2 | \cdots | A\bar{q}_m] \\ &= [\bar{q}_1^2 | \bar{q}_2^2 | \cdots | \bar{q}_m^2] \begin{bmatrix} r_{11}^2 & r_{12}^2 & \cdots & r_{1m}^2 \\ & r_{22}^2 & & \vdots \\ & & \ddots & \vdots \\ & & & r_{mm}^2 \end{bmatrix} \\ &= \bar{Q}_2 R_2. \end{aligned}$$

算法：

```
Set  $\bar{Q}_0 = I$   
for  $j = 1, 2, 3, \dots$   
     $A\bar{Q}_j = \bar{Q}_{j+1} R_{j+1}$   
end
```



相似变换

定义：若矩阵 A 与矩阵 B 满足 $B = T^{-1}AT$ ，则说它们是相似的

相似矩阵有着相同的特征值

特征向量也能互相表示

QR分解给出一个特殊的相似变换

$$Q^{-1}AQ = Q^T AQ = Q^T QRQ = RQ$$



QR分解与同时迭代

A 和 $Q^{-1}AQ$ 有相同的特征值

$$A = QR,$$
$$Q^{-1}A Q = RQ$$

QR迭代的等价形式:

- $A_1 = A$
- 分解 $A_k = Q_k R_k$
- $A_{k+1} = R_k Q_k$

如果收敛, 则 $A_\infty = Q_\infty R_\infty = R_\infty Q_\infty$

因此 $A_\infty Q_\infty = Q_\infty A_\infty$



QR分解与同时迭代

假设矩阵 A 的特征向量张满 \mathbb{R}^n ，而且特征值互不相同。
则矩阵 A 和 B 有相同的特征向量当且仅当

$$AB = BA$$

证明: (\Rightarrow)

$$BAy = BA \sum_i \alpha_i x_i = B \sum_i \lambda_i^A \alpha_i x_i = \sum_i \lambda_i^A \lambda_i^B x_i$$

$$AB y = AB \sum_i \alpha_i x_i = A \sum_i \lambda_i^B \alpha_i x_i = \sum_i \lambda_i^A \lambda_i^B x_i$$

由于 y 是任意的，因此 $BA = AB$



QR分解与同时迭代

假设矩阵 A 的特征向量张满 \mathbb{R}^n ，而且特征值互不相同。则矩阵 A 和 B 有相同的特征向量当且仅当

$$AB = BA$$

证明: (\Leftarrow)

令 x 为 A 的一个特征向量，满足 $Ax = \lambda x$ ，则有

$$A(Bx) = (AB)x = (BA)x = \lambda(Bx)$$

- 如果 $Bx \neq 0$ ，则 Bx 也为 A 的一个特征向量，特征值为 λ 。由于 A 的特征值互不相同，则存在实数 c 满足 $Bx = cx$ ，因此 x 为 B 的一个特征向量
- 如果 $Bx = 0$ ，则 x 为 B 的一个特征向量，特征值为 0

因此 A 的特征向量均为 B 的特征向量，而 A 的特征向量张满 \mathbb{R}^n ，因此它们有着相同的特征向量。



奇异值 (singular values)

对于一般的 $m \times n$ 实数矩阵 A , 可以有奇异值分解:

$$A = U S V^T$$

A 的形状: $m \times n$

U 的形状: $m \times m$

S 的形状: $m \times n$

V 的形状: $n \times n$

$$U^T U = I_m, \quad V^T V = I_n$$

观察:

$$A A^T = U S S^T U^T$$

$$A^T A = V S^T S V^T$$

$A A^T$ 和 $A^T A$ 都是实数对称阵, 可做特征值分解!



SVD

定理： AA^T 和 $A^T A$ 有着相同的非零特征值。

证明： 考虑 AA^T 的特征值 $\lambda \neq 0$ 和特征向量 u : $AA^T u = \lambda u$

$$A^T A(A^T u) = A^T (AA^T u) = A^T (\lambda u) = \lambda (A^T u)$$

可见，只要 $A^T u \neq 0$ ，则 $A^T u$ 为 $A^T A$ 的一个特征向量，特征值也是 λ 。不妨假设 $A^T u = 0$ ，则 $AA^T u = A(A^T u) = 0$ ，与 $\lambda \neq 0$ 矛盾。

类似地，考虑 $A^T A$ 的特征值 $\lambda \neq 0$ 和特征向量 v : $A^T A v = \lambda v$ ，也有 $AA^T(Av) = A(A^T A v) = A(\lambda v) = \lambda(Av)$ 。

同理有 $Av \neq 0$ ，因此 Av 为 AA^T 的一个特征向量，特征值也是 λ 。



SVD

定理： AA^T 和 $A^T A$ 有着相同的非零特征值

把它们记作 $\{\lambda_i\}$ ，另外对 AA^T 和 $A^T A$ 进行谱分解，可得到正规正交的向量 $\{\mathbf{u}_i\}, \{\mathbf{v}_i\}$ 作为它们的特征向量：

$$AA^T \mathbf{u}_i = \lambda_i \mathbf{u}_i$$

$$A^T A \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

把 $\{\mathbf{u}_i\}$ 组合成矩阵 \mathbf{U} ， $\{\mathbf{v}_i\}$ 组合成矩阵 \mathbf{V} ，则有 $\mathbf{U}^T \mathbf{U} = \mathbf{I}_m$ ， $\mathbf{V}^T \mathbf{V} = \mathbf{I}_n$

回顾之前的证明：

- 若 \mathbf{u} 为 AA^T 的一个特征向量，则 $A^T \mathbf{u}$ 为 $A^T A$ 的一个特征向量，且有着相同的特征值；
- 若 \mathbf{v} 为 $A^T A$ 的一个特征向量，则 $A \mathbf{v}$ 为 AA^T 的一个特征向量，且有着相同的特征值



SVD

回顾之前的证明:

- 若 \mathbf{u} 为 \mathbf{AA}^\top 的一个特征向量, 则 $\mathbf{A}^\top \mathbf{u}$ 为 $\mathbf{A}^\top \mathbf{A}$ 的一个特征向量, 且有着相同的特征值;
- 若 \mathbf{v} 为 $\mathbf{A}^\top \mathbf{A}$ 的一个特征向量, 则 $\mathbf{A}\mathbf{v}$ 为 \mathbf{AA}^\top 的一个特征向量, 且有着相同的特征值

$$\begin{aligned}\mathbf{AA}^\top \mathbf{u}_i &= \lambda_i \mathbf{u}_i \\ \mathbf{A}^\top \mathbf{A} \mathbf{v}_i &= \lambda_i \mathbf{v}_i\end{aligned}$$

利用上述观察可得: 存在常数 c_i, d_i 使得

$$\begin{aligned}\mathbf{A}^\top \mathbf{u}_i &= c_i \mathbf{v}_i \\ \mathbf{A} \mathbf{v}_i &= d_i \mathbf{u}_i\end{aligned}$$

接下来, 利用 \mathbf{v}_i 和 \mathbf{u}_i 都是单位长度的性质, 可以求出 c_i, d_i

注意到 $\mathbf{u}_i^\top \mathbf{AA}^\top \mathbf{u}_i$ 可以有两种写法:

$$\begin{aligned}\mathbf{u}_i^\top \mathbf{AA}^\top \mathbf{u}_i &= \lambda_i \mathbf{u}_i^\top \mathbf{u}_i = \lambda_i \\ \mathbf{u}_i^\top \mathbf{AA}^\top \mathbf{u}_i &= (\mathbf{A}^\top \mathbf{u}_i)^\top \mathbf{A}^\top \mathbf{u}_i = c_i^2 \mathbf{v}_i^\top \mathbf{v}_i = c_i^2\end{aligned}$$

因此 $c_i^2 = \lambda_i$, 类似可得 $d_i^2 = \lambda_i$, 因此有:

$$\begin{aligned}\mathbf{A}^\top \mathbf{u}_i &= \sqrt{\lambda_i} \mathbf{v}_i \\ \mathbf{A} \mathbf{v}_i &= \sqrt{\lambda_i} \mathbf{u}_i\end{aligned}$$



SVD

由于

$$A^T \mathbf{u}_i = \sqrt{\lambda_i} \mathbf{v}_i$$

$$A \mathbf{v}_i = \sqrt{\lambda_i} \mathbf{u}_i$$

不失一般性地假设 $m > n$ ，则有

$$AV = U \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$$

$$U^T AV = \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$$

$$\Rightarrow A = USV^T$$

S 的对角线元素被称作奇异值(singular values)，通常记作 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$

从几何上看，矩阵 A 的作用为：

- \mathbb{R}^n 上的旋转变换
- 伸缩变换
- \mathbb{R}^m 上的旋转变换



SVD与Rayleigh quotient

回顾特征值的min-max刻画，记 $R_B(\mathbf{x}) := \frac{\mathbf{x}^\top \mathbf{B} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$
对于对称的矩阵 \mathbf{B} ， $R_B(\mathbf{x})$ 在特征值处取到极值

回顾 $\|A\|_2 := \max_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$ ， $cond_2(A) := \|A\|_2 \cdot \|A^{-1}\|_2$
对于对称正定的矩阵 A ，

$$\|A\|_2 = \lambda_{max}(A)$$
$$cond_2(A) = \lambda_{max}(A) / \lambda_{min}(A)$$

对于一般的矩阵呢？

$$\|A\|_2 = \sqrt{\lambda_{max}(A^\top A)} = \sigma_{max}(A)$$
$$cond_2(A) = \frac{\sqrt{\lambda_{max}(A^\top A)}}{\sqrt{\lambda_{min}(A^\top A)}} = \frac{\sigma_{max}(A)}{\sigma_{min}(A)}$$



拉格朗日乘数法(选讲)

对于一般的矩阵 A , $\max_{\|x\|_2=1} \|Ax\|_2$ 什么时候取到极值?

除了通过特征值的min-max刻画, 还可以通过拉格朗日乘数法, 把带约束的最优化问题转化为不带约束的:

$$L(\lambda) := \sup_x \mathbf{x}^\top A^\top A \mathbf{x} - \lambda(\mathbf{x}^\top \mathbf{x} - 1)$$

对 \mathbf{x} 求梯度也可得 $A^\top A \mathbf{x} = \lambda \mathbf{x}$

因此也可得到 $\max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\lambda_{max}(A^\top A)}$



SVD与low rank approximations

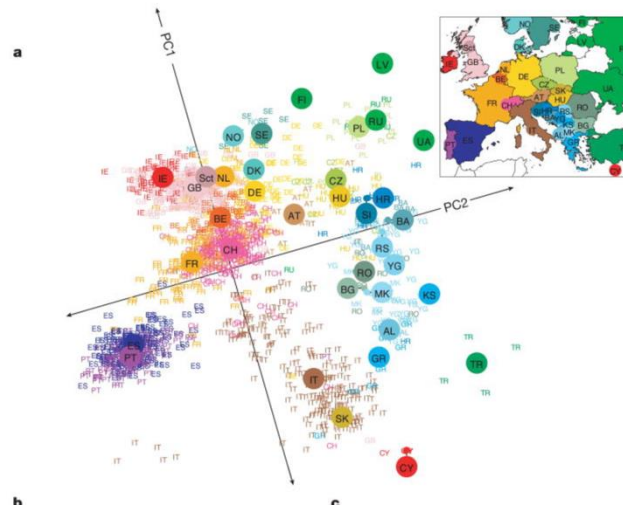
$$A = \sum_{i=1}^{\min(m,n)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

考虑如下的近似 $\tilde{A} := \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

Eckart-Young定理: \tilde{A} 是所有列空间的维度最多为 r 的矩阵中, 能同时最小化 $\|\tilde{A} - A\|_2$ 和 $\|\tilde{A} - A\|_F$ 的矩阵

Many applications:

- Principal component analysis
 - Data visualization
 - [Genomes can encode geography](#)
 - Eigenfaces
- [Matrix completions](#)
- Recommender systems
- Latent semantic analysis
- [Approximation algorithm for Max-cut](#)





SVD与伪逆pseudoinverse (选讲)

$$A = \sum_{i=1}^{\min(m,n)} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\top}$$

$$A^+ = VS^+U^{\top} = \sum_{i:\sigma_i \neq 0} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^{\top}$$

如果 A 可逆, 则 $A^+ = A^{-1}$

如果 A 是overdetermined, 则 $A^+ \mathbf{b}$ 为 $A\mathbf{x} = \mathbf{b}$ 的最小二乘解

如果 A 是underdetermined, 则 $A^+ \mathbf{b}$ 为 $A\mathbf{x} = \mathbf{b}$ 中2范数最小的最小二乘解