華東師範大學
EAST CHINA NORMAL UNIVERSITY

数据科学与工程算法基础

Algorithm Foundations of Data Science and Engineering

第八章 特征值计算

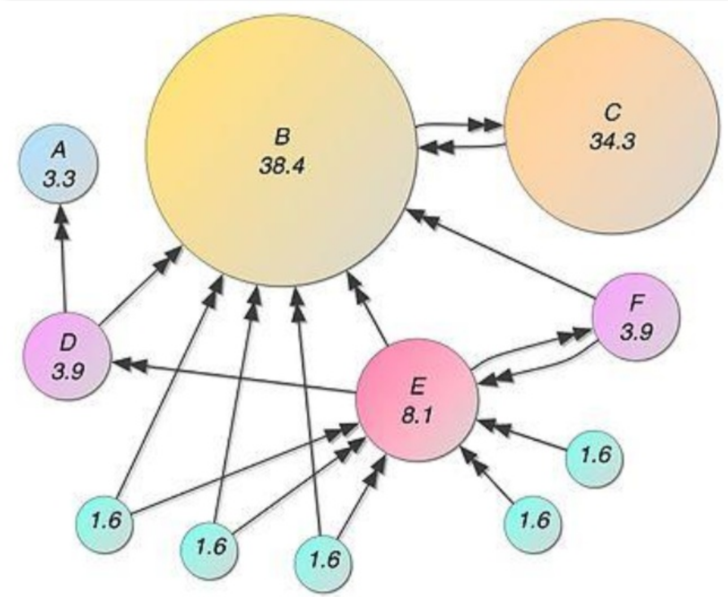
课程提纲

Content

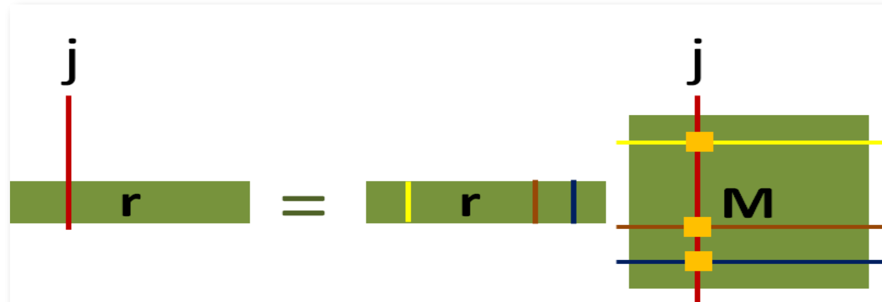
课程提纲

Content

PageRank 与特征值

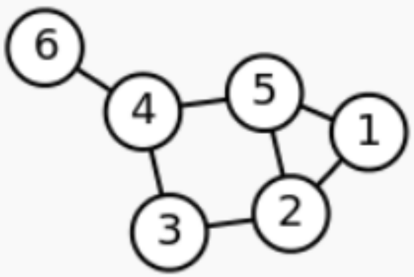


Set $\mathbf{r}^{(0)} = (\frac{1}{n}, \cdots, \frac{1}{n})$;
For $k = 1, 2, \cdots$;
Let $\mathbf{r}^{(k)} = \mathbf{r}^{(k-1)} \cdot M$;



- 一个概率转移矩阵为 P 的齐次有限随机游走的平稳分布为 π，则满足 $\pi P = \pi$
  - π 为转移概率矩阵 P 最大特征值 1 对应的左特征向量
  - 如何有效计算随机游走的平稳分布?
  - 矩阵形式的 PageRank 算法为什么是正确的?

图的鲁棒性与特征值

Labeled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2&0&0&0&0&0\\0&3&0&0&0&0\\0&0&2&0&0&0\\0&0&0&3&0&0\\0&0&0&0&3&0\\0&0&0&0&0&1 \end{pmatrix}$	$\begin{pmatrix} 0&1&0&0&1&0\\1&0&1&0&1&0\\0&1&0&1&0&0\\0&0&1&0&1&1\\1&1&0&1&0&0\\0&0&0&1&0&0 \end{pmatrix}$	$\begin{pmatrix} 2&-1&0&0&-1&0\\-1&3&-1&0&-1&0\\0&-1&2&-1&0&0\\0&0&-1&3&-1&-1\\-1&-1&0&-1&3&0\\0&0&0&-1&0&1 \end{pmatrix}$

- 图的拉普拉斯矩阵 $L = D - A$，其中 D 为对角的度矩阵，A 为邻接矩阵
- 图鲁棒性衡量了图结构受到攻击时保持结构的能力
  - 代数连通性：Laplacian 矩阵第二小特征值
  - 如何计算?

特征值与特征向量

- 对给定矩阵 $A \in \mathbb{R}^{n\times n}$ 和非零向量 v，若 $Av = \lambda v$，则 λ 是 A 的一个特征值，v 是 A 对应于特征值 λ 的特征向量
- 线性变换
  - 线性变换 = 拉伸 + 旋转
  - 矩阵 A 作用于向量 v 像是一个函数，在 A 的作用下输入向量 v 变换为输出 λv
  - 也就是说，矩阵 A 对其特征向量只进行拉伸
  - $Av = \lambda v$ 可以改写为 $Av - \lambda v = 0$ 或者 $(A - \lambda I)v = 0$
  - 齐次线性方程组具有非零解，所以 $|A - \lambda I| = 0$

计算特征值和特征向量

- 给定矩阵 $A = \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix}$
- 特征方程为 $|A - \lambda I| = \lambda^2 - 3\lambda + 2 = 0$，因此特征值为 $\lambda_1 = 1$，$\lambda_2 = 2$
- 通过解线性方程得到特征向量 $(A - \lambda_i I)v = 0$
  - 得对应于特征值 $\lambda_1 = 1$ 的特征向量为 $(1,1)^\top$
  - 得对应于特征值 $\lambda_2 = 2$ 的特征向量为 $(2,1)^\top$
- 但这种方法（高斯消元法）是不可扩展的，很难在大规模矩阵上使用

课程提纲

Content

幂法

- 设矩阵 A 有唯一的主特征值 λ_1 （模最大）
- 可以运用幂法求解主特征值及其特征向量

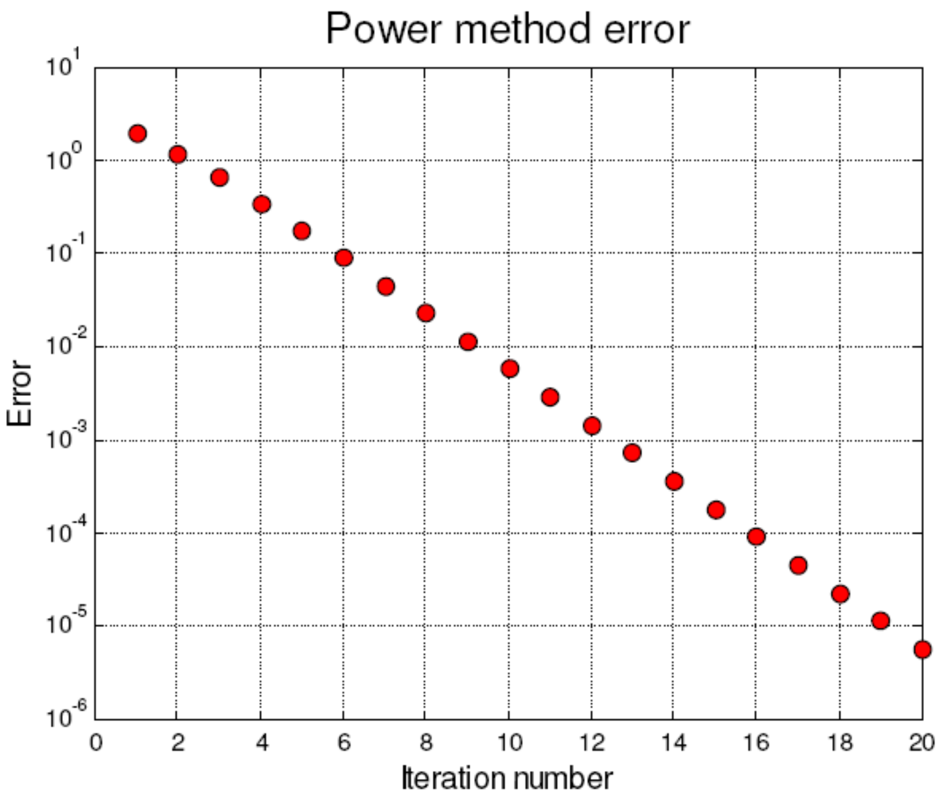
1: Pick a starting vector $\mathbf{x}^{(0)} \in \mathbb{R}^n$, such that $\|\mathbf{x}^{(0)}\| = 1$;
2: For $k = 1, 2, \cdots$;
3: Let $\mathbf{v} = A\mathbf{x}^{(k-1)}$;
4: Let $\mathbf{x}^{(k)} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$;

- 幂法是一种迭代算法
  - 达到收敛条件算法终止
  - 向量单位化的目的为了防止计算溢出
  - 可以利用幂法求解随机游走的平稳分布
  - 为什么幂法是收敛的?

幂法分析

- 假设矩阵 A 标准化的特征向量构成一组标准正交基，满足
$$v_i^\top v_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$
- 令 $x^{(0)}$ 为特征向量生成子空间中任意非零向量，则
$$x^{(0)} = \sum_{i=1}^{n} c_i v_i \text{ 且 } c_1 \neq 0$$
- 则 $x^{(k)}$ 收敛到主特征向量 v_1，因为
$$\lim_{k\to\infty} A^k x^{(0)} = \lim_{k\to\infty} \sum_{i=1}^{n} c_i A^k v_i = \lim_{k\to\infty} \sum_{i=1}^{n} c_i \lambda_i^k v_i$$
$$= \lim_{k\to\infty} c_1 \lambda_1^k \left(v_1 + \sum_{i=1}^{n} \frac{c_i}{c_1} \left(\frac{\lambda_i}{\lambda_1}\right)^k v_i \right) = c_1 \lambda_1^k v_1$$

幂法分析（续）



- 假设 $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$
- 随着迭代次数的增加，幂法收敛速度取决于 $\frac{\lambda_2}{\lambda_1}$，即幂法是线性收敛的
- 如果主特征值具有重数 r，计算得到的特征向量将是 r 个特征向量的线性组合

幂法扩展

- 反幂法
  - 使用 A^{-1} 而不是 A，因为 A^{-1} 的特征值是 $1/\lambda^i$
  - 给出了一种求解矩阵绝对值最小的非零特征值的方法
- 平移法
  - $A-\mu I$ 的特征值为 $\lambda_i-\mu$
  - 运用幂法可以找到 $|\lambda_i-\mu|$ 最大的特征值
  - 运用反幂法可以找到 $|1/(\lambda_i-\mu)|$ 最大的特征值，即 A 最接近 μ 的特征值
  - 因为无法事先知道 λ_i，只能尝试不同的 μ
  - 难以找到矩阵 A 所有的特征值

反幂法

- 直接计算 A^{-1} 复杂度高
  - $A^{-1}x^{(k-1)} = v$ 等价于 $x^{(k-1)} = Av$
  - 求解线性方程组方法多如逐次超松弛迭代法效率更高

1: Pick a starting vector $\mathbf{x}^{(0)} \in \mathbb{R}^n$, such that $\|\mathbf{x}^{(0)}\| = 1$;
2: For $k = 1, 2, \cdots$;
3: Solve $A\mathbf{v} = \mathbf{x}^{(k-1)}$ for $\mathbf{v}$;
4: Let $\mathbf{x}^{(k)} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$;

  - 其收敛性分析与幂法相同
  - 对奇异矩阵，不能使用反幂法

平移法

- $A-\mu I$ 的特征值为 $\lambda_i-\mu$
- 给定 μ，计算矩阵 A 最接近 μ 的特征值

1: Pick some μ close the desired eigenvalue;
2: Pick a starting vector $\mathbf{x}^{(0)} \in \mathbb{R}^n$, such that $\|\mathbf{x}^{(0)}\| = 1$;
3: For $k = 1, 2, \cdots$;
4:     Solve $(A-\mu I)\mathbf{v} = \mathbf{x}^{(k-1)}$ for $\mathbf{v}$;
5:     Let $\mathbf{x}^{(k)} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$;

- 其收敛性与幂法相同
- 通过选择不同的 μ，可以找到矩阵 A 更多的特征值
- 如何确定 μ? （二分查找）

课程提纲

Content

牛顿迭代法

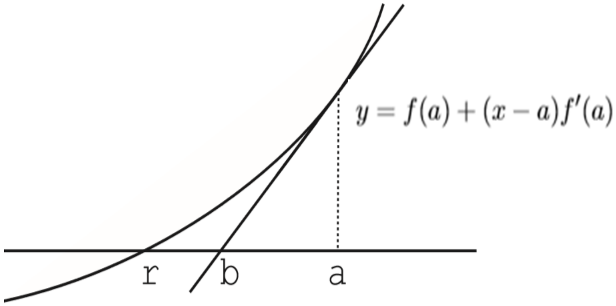
- 假设 r 是方程 $f(x)=0$ 的一个解，如果 $f(x)$ 是一个可微函数，如何求它的解?
- 牛顿迭代法

1:	Pick a starting point x_0;
2:	Until the convergence condition;
3:	$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$;

  - 假定真实解为 r，$h = x_n - r$ 表明 x_n 距离真实解的距离
  - 当 h 比较小时

$$0 = f(r) = f(x_n + h) \approx f(x_n) + hf'(x_n)$$

$$= f(x_n) + \frac{\mathrm{d}f(x)}{\mathrm{d}x}\Big|_{x=x_n} \Delta(x_n)$$



瑞利商

- 瑞利商 (Rayleigh Quotient)
  - 给定实对称矩阵 A，其瑞利商为 $R_A(\cdot):\mathbb{R}^n\backslash\{0\}\to\mathbb{R}$ 的函数
  - 对 $\forall x\in\mathbb{R}^n\backslash\{0\}$，其定义为 $R_A(x)=\dfrac{\langle Ax,x\rangle}{\langle x,x\rangle}$
- 矩阵 A 标准化的特征向量构成一组标准正交基，满足 $v_i^\top v_j=\begin{cases}1 & \text{if } i=j\\ 0 & \text{otherwise}\end{cases}$
- 因此
  - $\langle x,x\rangle=x^\top x=(\sum_{i=1}^{n}c_iv_i)^\top(\sum_{i=1}^{n}c_iv_i)=\sum_{i=1}^{n}c_i^2$
  - $\langle Ax,x\rangle=x^\top Ax=(\sum_{i=1}^{n}c_iv_i)^\top(\sum_{i=1}^{n}c_iAv_i)=\sum_{i=1}^{n}\lambda_ic_i^2$

瑞利商（续）

- 瑞利商改写为 $R_A(x) = \frac{\langle Ax, x\rangle}{\langle x, x\rangle} = \frac{\sum_{i=1}^n \lambda_i c_i^2}{\sum_{i=1}^n c_i^2}$
- 如果 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$，则
  - $\lambda_1 = \max_{\mathbb{R}^n \setminus \{0\}} R_A(x)$
  - $\lambda_n = \min_{\mathbb{R}^n \setminus \{0\}} R_A(x)$
  - 对实矩阵 A，v_j 为特征值 λ_j 对应的特征向量，则 $R_A(v_j) = \lambda_j$
  - 因此 $\lambda_1 \geq R_A(x) \geq \lambda_n$

瑞利商迭代法

- 在平移法中，放松固定平移值 μ 的假设
  - 每轮迭代中平移值随着迭代向量的变化而变化
  - 瑞利商迭代法

1: Pick a starting vector $\mathbf{x}^{(0)} \in \mathbb{R}^n$, such that $\|\mathbf{x}^{(0)}\| = 1$;
2: Let $\lambda^{(0)} = r(\mathbf{x}^{(0)}) = (\mathbf{x}^{(0)})^T A \mathbf{x}^{(0)}$;
3: For $k = 1, 2, \cdots$;
4: Solve $(A - \lambda^{(k-1)} I)\mathbf{v} = \mathbf{x}^{(k-1)}$ for $\mathbf{v}$;
5: Let $\mathbf{x}^{(k)} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$;
6: Let $\lambda^{(k)} = r(\mathbf{x}^{(k)}) = (\mathbf{x}^{(k)})^T A \mathbf{x}^{(k)}$;

- 当实矩阵 A 是对称矩阵时，瑞利商迭代法一定收敛
  - 瑞利商迭代法为什么收敛?
  - 收敛到什么?

瑞利商迭代法分析

- 定义函数 $f(v, \lambda) = Av - \lambda v$
- 计算函数 $f(v, \lambda)$ 的全微分（$\Delta z = \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y$），得到

  $\Delta f(v, \lambda) = (A - \lambda I)\Delta v - (\Delta\lambda)v$
- 牛顿迭代法第 $k+1$ 轮给出的近似解可以计算为

$$
\begin{aligned}
0 &= f(v_k, \lambda_k) + \Delta f(v_k, \lambda_k) \\
&= (A - \lambda_k I)(v_k + \Delta v_k) - (\Delta\lambda_k)v_k \\
&= (A - \lambda_k I)v_{k+1} - (\Delta\lambda_k)v_k
\end{aligned}
$$

- 因此，$v_{k+1} = (\Delta\lambda_k)(A - \lambda_k I)^{-1}v_k$

瑞利商迭代法收敛性分析

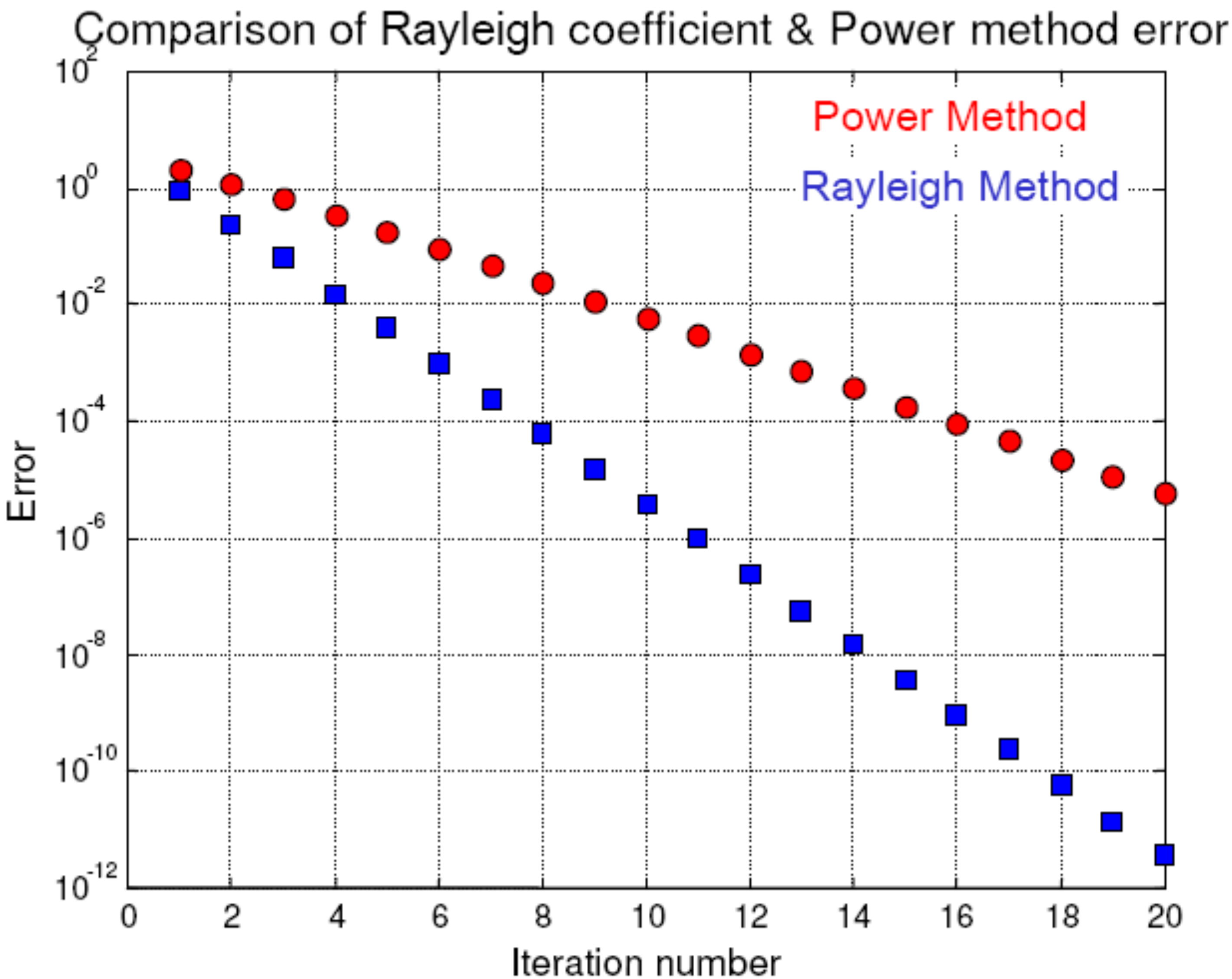
- 为简单起见，仅考虑两个最大的主特征值

$$R_A\left(x^{(k)}\right) \approx \frac{(c_1\lambda_1^k v_1 + c_2\lambda_2^k v_2)^\top (c_1\lambda_1^{k+1} v_1 + c_2\lambda_2^{k+1} v_2)}{(c_1\lambda_1^k v_1 + c_2\lambda_2^k v_2)^\top (c_1\lambda_1^k v_1 + c_2\lambda_2^k v_2)}$$

$$= \frac{c_1^2\lambda_1^{2k+1} + c_2^2\lambda_2^{2k+1}}{c_1^2\lambda_1^{2k} + c_2^2\lambda_2^{2k}} = \lambda_1 + O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\right)$$

- 瑞利商迭代法是收敛的，其极限为最大主特征值 λ_1
- 随着迭代次数的增加，瑞利商收敛速度取决于 $\left(\frac{\lambda_2}{\lambda_1}\right)^2$

瑞利商迭代法 VS 幂法



Comparison of Rayleigh coefficient & Power method error
Power Method
Rayleigh Method
Error
Iteration number
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
0
2
4
6
8
10
12
14
16
18
20

降阶技术

- 矩阵 A 标准化的特征向量构成一组标准正交基，满足 $v_i^\top v_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$
- 通过正交分解可得 $A = \sum_{i=1}^{n} \lambda_i v_i v_i^\top$
  - 求得主特征值 λ_1 后，构造矩阵 $A^{(1)} = \sum_{i=2}^{n} \lambda_i v_i v_i^\top$
  - 因为 $A^{(1)} = \sum_{i=2}^{n} \lambda_i v_i v_i^\top = A - \lambda_1 v_1 v_1^\top$
  - 计算 $A^{(1)} v_i = \begin{cases} 0 & \text{if } i = 1 \\ \lambda_i v_i & \text{otherwise} \end{cases}$

降阶技术（续）

- 通过降阶技术，把原矩阵的特征值变为 $0, \lambda_2, \ldots, \lambda_n$，且满足 $|\lambda_2| \geq \ldots \geq |\lambda_n| \geq 0$
- 运用幂法或者瑞利商迭代法可以进一步求解矩阵 $A^{(1)}$ 的主特征值
- 依次迭代，定义矩阵 $A^{(k+1)} = A^{(k)} - \lambda_k v_k v_k^\top$，求得后续特征值和特征向量

本章小结

- 特征值计算方法
  - 幂法
  - 反幂法
  - 瑞利商迭代法
  - 降阶方法
- 特征值与特征向量的应用场景
  - 图中顶点 PageRank 值
  - 图鲁棒性度量
  - 特征提取
  - 数据降维