# Eigen Vector Note

FanmingL

2018 年 12 月 5 日

## 1 特征值问题

考虑一个特征值问题，

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{1}$$

这个问题的解，我们很容易由特征值分解的常用方法，先求特征值，再求特征向量的方式求出来，那么我们是否能够直接求出 $\mathbf{x}$ 呢，或者说把这个问题转化成一个优化问题去按照我们的一些常用的优化方法比 (梯度下降/上升法、遗传算法、模拟退火法等) 去求解。

其实我们很容易想到，利用特征向量 $\mathbf{x}$ 和 $\mathbf{A}\mathbf{x}$ 线性相关这一特点，从几何上理解即，$\mathbf{x}$ 和 $\mathbf{A}\mathbf{x}$ 两个向量共线，所以这个问题可以很直观的转化成如下优化问题

$$\min \quad \|\mathbf{x} \times (\mathbf{A}\mathbf{x})\|_2^2 \tag{2}$$

其中 $\times$ 代表外积，当 $\mathbf{x}$ 和 $\mathbf{A}\mathbf{x}$ 共线的话，$\mathbf{x} \times (\mathbf{A}\mathbf{x})$ 的值为零向量。因为叉乘的结果的是与 $\mathbf{x}$ 和 $\mathbf{A}\mathbf{x}$ 的夹角 $\theta$ 的正弦值 $\sin\theta$ 成正比的。当其共线时，$\theta = 0$ 故 $\|\mathbf{x} \times (\mathbf{A}\mathbf{x})\|_2^2$ 取到最小值 0。

这个优化目标很直观，但是，不好求导，但是可以用非梯度的方法去优化。

我们可以看到当 $\mathbf{x}$ 和 $\mathbf{A}\mathbf{x}$ 共线时 $\theta$ 为零，$\cos\theta$ 取到最大值 1，而 $\cos\theta$ 可以表示为 $\mathbf{x}$ 和 $\mathbf{A}\mathbf{x}$ 向量单位话之后的内积。所以我们可以写出以下优化目标。

$$\max \quad \frac{(\mathbf{x}^T\mathbf{A}\mathbf{x})^2}{\mathbf{x}^T\mathbf{x}} \quad s.t. \quad \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = 1 \tag{3}$$

令 $J(\mathbf{x})$，那么我们求的特征向量即

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \quad \frac{(\mathbf{x}^T\mathbf{A}\mathbf{x})^2}{\mathbf{x}^T\mathbf{x}} \quad s.t. \quad \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = 1$$

我们可以验证在 $\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = 1$ 的约束下，当且仅当 $\mathbf{x}$ 是 $\mathbf{A}$ 的特征向量时，$J(\mathbf{x})$ 取到最大值。假设 $\mathbf{A}\mathbf{x} = \mathbf{a}$，所以 $\mathbf{a}^T\mathbf{a} = 1$。

$$\begin{aligned} J(\mathbf{x}) &= \frac{(\mathbf{x}^T\mathbf{a})^2}{\mathbf{x}^T\mathbf{x}} \\ &\leq \frac{\mathbf{x}^T\mathbf{x}\mathbf{a}^T\mathbf{a}}{\mathbf{x}^T\mathbf{x}} \\ &= \mathbf{a}^T\mathbf{a} \\ &= 1 \end{aligned} \tag{4}$$

上述推导中，用到了柯西施瓦茨不等式，即

$$< \mathbf{a}, \mathbf{b} >^2 \quad \leq \quad < \mathbf{a}, \mathbf{a} >< \mathbf{b}, \mathbf{b} >$$

其中 $< *, * >$ 表示二者的内积。该不等式当且仅当 $\mathbf{a}$ 和 $\mathbf{b}$ 线性相关时，等号成立，即当且仅当存在常数 $\lambda$ 使得 $\mathbf{a} = \lambda\mathbf{b}$ 等号成立。

所以，式4中

$$\frac{(\mathbf{x}^T\mathbf{a})^2}{\mathbf{x}^T\mathbf{x}} \leq \frac{\mathbf{x}^T\mathbf{x}\mathbf{a}^T\mathbf{a}}{\mathbf{x}^T\mathbf{x}}$$

当且仅当存在常数 $\lambda$ 使得 $\mathbf{x} = \lambda\mathbf{a} = \lambda\mathbf{A}\mathbf{x}$ 等号成立。

对于 $J(\mathbf{x}) \leq 1$ 这一不等式，同样当且仅当存在常数 $\lambda$ 使得 $\mathbf{x} = \lambda\mathbf{A}\mathbf{x}$ 时，等式成立，$J(\mathbf{x})$ 取到最大值 1，而我们知道 $\lambda$ 就是 $\mathbf{A}$ 的特征值。

所以当且仅当 $\mathbf{x}$ 为 $\mathbf{A}$ 的特征向量时，$J(\mathbf{x})$ 取到最大值 1。

实际上，当 $\mathbf{x}$ 为 $\mathbf{A}$ 的任意特征向量时，$J(\mathbf{x})$ 都能取到最大值 1。因为，当 $\mathbf{a} = \lambda\mathbf{x}$ 且 $\mathbf{a}^T\mathbf{a} = \lambda^2\mathbf{x}^T\mathbf{x} = 1$ 时，

$$J(\mathbf{x}) = \frac{\lambda^2(\mathbf{x}^T\mathbf{x})^2}{\mathbf{x}^T\mathbf{x}} = \lambda^2\mathbf{x}^T\mathbf{x} = 1$$

所以寻找 $\mathbf{A}$ 的特征向量是与式3表示的优化问题是等价。由于存在多个特征向量，所以 $J(\mathbf{x})$ 有多个极值点，即 $J(\mathbf{x})$ 有多个极值点，是非凸的，同样是个很复杂的函数。我们可以用梯度上升法，求解问题。

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \frac{\alpha}{2}\nabla_{\mathbf{x}} J(\mathbf{x}_n)$$

但是梯度上升法对每个初始值只能收敛到一个极值点，若我们希望得到所有的 $\mathbf{A}$ 的特征向量，需要尝试不同的初始值，颇为麻烦。但我们可以用非梯度的方法，如模拟退火法，和遗传算法，获得尽可能多的解。

下面计算 $J(\mathbf{x})$ 的梯度上升法的迭代公式。考虑到约束条件，引入拉格朗日乘子 $\lambda$

$$J(\mathbf{x}) = \frac{(\mathbf{x}^T\mathbf{A}\mathbf{x})^2}{\mathbf{x}^T\mathbf{x}} - \lambda(\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - 1)$$

令 $J(\mathbf{x})$ 对 $\mathbf{x}$ 求偏导

$$\begin{aligned}
\frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} &= \frac{1}{(\mathbf{x}^T\mathbf{x})^2}\left(\frac{\partial(\mathbf{x}^T\mathbf{A}\mathbf{x})^2}{\partial\mathbf{x}}\mathbf{x}^T\mathbf{x} - (\mathbf{x}^T\mathbf{A}\mathbf{x})^2\frac{\partial(\mathbf{x}^T\mathbf{x})}{\partial\mathbf{x}}\right) - 2\lambda\mathbf{A}^T\mathbf{A}\mathbf{x} \\
&= \frac{1}{(\mathbf{x}^T\mathbf{x})^2}\left(2(\mathbf{x}^T\mathbf{A}\mathbf{x})\frac{\partial(\mathbf{x}^T\mathbf{A}\mathbf{x})}{\partial\mathbf{x}}\mathbf{x}^T\mathbf{x} - 2(\mathbf{x}^T\mathbf{A}\mathbf{x})^2\mathbf{x}\right) - 2\lambda\mathbf{A}^T\mathbf{A}\mathbf{x} \\
&= \frac{1}{(\mathbf{x}^T\mathbf{x})^2}\left(4(\mathbf{x}^T\mathbf{A}\mathbf{x})\mathbf{A}\mathbf{x}\mathbf{x}^T\mathbf{x} - 2(\mathbf{x}^T\mathbf{A}\mathbf{x})^2\mathbf{x}\right) - 2\lambda\mathbf{A}^T\mathbf{A}\mathbf{x} \\
&= \frac{2}{(\mathbf{x}^T\mathbf{x})^2}\left(2\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{x}^T\mathbf{x}\mathbf{A}\mathbf{x} - (\mathbf{x}^T\mathbf{A}\mathbf{x})^2\mathbf{x}\right) - \lambda\mathbf{A}^T\mathbf{A}\mathbf{x}
\end{aligned}$$

于是我们得到 $\mathbf{x}$ 的迭代公式

$$\mathbf{x}_{new} = \mathbf{x} + \alpha\left[\frac{2\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{x}^T\mathbf{x}\mathbf{A}\mathbf{x} - (\mathbf{x}^T\mathbf{A}\mathbf{x})^2\mathbf{x}}{(\mathbf{x}^T\mathbf{x})^2} - \lambda\mathbf{A}^T\mathbf{A}\mathbf{x}\right] \tag{5}$$

将 $\mathbf{x}_{new}$ 代入至约束条件，即

$$\mathbf{x}_{new}^T\mathbf{A}^T\mathbf{A}\mathbf{x}_{new} = 1 \tag{6}$$

可得到关于 $\lambda$ 的一元二次方程，令 $C = \frac{2\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{x}^T\mathbf{x}\mathbf{A}\mathbf{x} - (\mathbf{x}^T\mathbf{A}\mathbf{x})^2\mathbf{x}}{(\mathbf{x}^T\mathbf{x})^2}, D = \mathbf{A}^T\mathbf{A}\mathbf{x}$

$$\begin{aligned}
(\alpha^2C^T\mathbf{A}^T\mathbf{A}C + 2\alpha\mathbf{x}^T\mathbf{A}^T\mathbf{A}C) - 2\lambda(\alpha\mathbf{x}^T\mathbf{A}^T\mathbf{A}D + \alpha^2C^T\mathbf{A}^T\mathbf{A}D) + \alpha^2\lambda^2D^T\mathbf{A}^T\mathbf{A}D = 0 \\
\alpha^2D^T\mathbf{A}^T\mathbf{A}D\lambda^2 - 2(\alpha\mathbf{x}^T + \alpha^2C^T)\mathbf{A}^T\mathbf{A}D\lambda + (\alpha^2C^T + 2\alpha\mathbf{x}^T)\mathbf{A}^T\mathbf{A}C = 0
\end{aligned} \tag{7}$$

于是可以用上式解出 $\lambda$，将解出的 $\lambda$ 代入至迭代公式5，即可得到完整的，不含拉格朗日乘数的迭代公式了。