

Lecture 14 矩阵函数、非负矩阵 — 2025.12.16

教授：邵美悦

Scribe: 路人甲

目录

1 矩阵函数	1
1.1 定义	1
1.2 性质	4
1.3 常见的矩阵函数	6
2 非负矩阵	10
2.1 正矩阵	10
2.2 非负矩阵	14
2.3 Perron 根	15
2.4 随机矩阵	17

1 矩阵函数

1.1 定义

矩阵函数是我们在第一节课就接触的概念。它有非常多的定义方式。下面我们先展示几种比较差的定义方式。

定义 1.1 (幂级数定义). 设 $f(z) = \sum_{k=0}^{\infty} c_k z^k$ 的收敛半径为 r . 若 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 满足 $\rho(\mathbf{A}) < r$, 则定义

$$f(\mathbf{A}) = \sum_{k=0}^{\infty} c_k \mathbf{A}^k.$$

这个定义非常自然：我们只需要把标量 z 替换为矩阵 \mathbf{A} 即可。然而，这个定义有两个问题：

1. 如果函数是 $p(\lambda) = \det(\lambda \mathbf{I} - \mathbf{A})$ 的话，我们要怎么替换？直接替换会得到 $\det(\mathbf{A} \mathbf{I} - \mathbf{A}) = 0$ ，这样替换显然是不对的；正确的做法是先把 $p(\lambda)$ 写成多项式的形式，然后再替换。
2. 同一个函数在不同区域内有不同的幂级数展开式，如果我用不同的幂级数展开来定义矩阵函数，如图 1，我们可以在 A 点展开获得幂级数，也可以在 C 点展开获得幂级数，这两种方式得到的矩阵一定是一样的吗？这个论证起来就比较麻烦了。这暴露出一个严重的问题：矩阵函数不应该依赖于“你怎么写这个函数”。
3. 这个定义的适用范围有限。定义要求谱落在某个圆盘内，要求函数在该圆盘内解析。然而，很多我们关心的函数并不满足这个条件，比如矩阵的对数函数、矩阵的平方根函数等——它们根本没法用幂级数在全谱上统一定义，如果我们想要强行用幂级数来定义这些函数，就势必要分块定义；分块定义一方面会带来计算上的麻烦，另一方面也会导致上面提到的“依赖于怎么写函数”的问题更加严重。

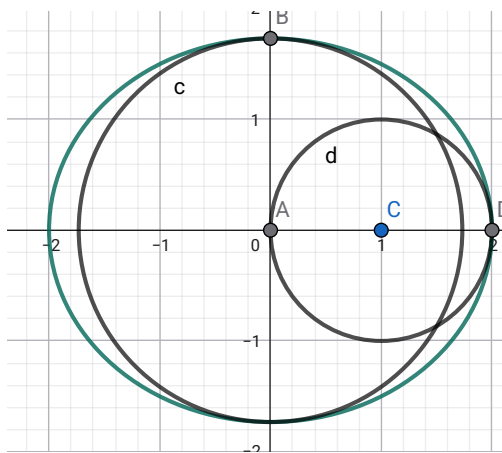


图 1: 不同的幂级数展开式

第二个坏定义是通过 Jordan 标准型来定义矩阵函数。

定义 1.2 (Jordan 标准型定义). 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 且 $\mathbf{A} = \mathbf{P} \mathbf{J} \mathbf{P}^{-1}$, 其中 $\mathbf{J} = \text{diag}\{\mathbf{J}_1, \dots, \mathbf{J}_m\}$ 是 \mathbf{A} 的 Jordan 标准型, 则定义 \mathbf{A} 的矩阵函数为

$$f(\mathbf{A}) = \mathbf{P} \cdot \text{diag}\{\mathbf{F}_1, \dots, \mathbf{F}_m\} \cdot \mathbf{P}^{-1},$$

其中

$$\mathbf{F}_i = f(\mathbf{J}_i) = \begin{bmatrix} f(\lambda_i) & f^{(1)}(\lambda_i) & \cdots & \cdots & \frac{f^{(n_i-1)}(\lambda_i)}{(n_i-1)!} \\ 0 & f(\lambda_i) & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & f^{(1)}(\lambda_i) \\ 0 & 0 & \cdots & \cdots & f(\lambda_i) \end{bmatrix} \in \mathbb{C}^{n_i \times n_i}, \quad i = 1, \dots, m.$$

此处我们假设 f 在 λ_i 处有 $n_i - 1$ 阶导数。

这个定义看起来也很自然：我们先把矩阵化简成 Jordan 标准型，然后对每个 Jordan 块分别定义矩阵函数，最后再把它们拼接起来。然而，这是一个“存在性定理型定义”，而不是可用的定义：

1. 如果考虑排列，Jordan 标准型就不是唯一的了。假设我们有两个不同的 Jordan 标准型分解 $A = \tilde{P}\tilde{J}\tilde{P}^{-1}$ 和 $A = \hat{P}\hat{J}\hat{P}^{-1}$ ，那么我们得到的矩阵函数分别是 $\tilde{P}f(\tilde{J})\tilde{P}^{-1}$ 和 $\hat{P}f(\hat{J})\hat{P}^{-1}$ ，这两个矩阵函数是否相等？这个问题的答案是肯定的，但证明起来比较麻烦。换言之，**这个定义依赖于基的选择**，矩阵函数应该是坐标自由的。
2. Jordan 标准型的计算非常不稳定，数值计算中很少使用 Jordan 标准型分解。因此，这个定义在数值计算中并不实用。**任何一个对 Jordan 形式敏感的定义都不可能是一个好的定义。**
3. 这个定义要求函数在每个特征值处都有足够阶数的导数，否则无法定义矩阵函数。这限制了函数的选择范围。

当然，这个定义也并非一无是处：它告诉我们矩阵函数只依赖于函数在谱上的行为，并且对相似变换不变。

那么，有没有一种定义，它完美地继承了以上两种定义的优点，但同时又规避了它们的缺点呢？或者，我们究竟想要什么样的定义呢？我们的设计指标是：

1. 只依赖于矩阵本身，而与坐标无关，即与基的选取无关。
2. 只依赖于函数在谱附近的行为。
3. 与多项式函数在形式上尽可能相似。
4. 对相似变换不变： $f(VAV^{-1}) = Vf(A)V^{-1}$ 。

接下来，我们就介绍一种满足以上设计指标的定义方式：

定义 1.3 (Cauchy 积分定义). 设 Ω 是复平面上的区域，且包含 $A \in \mathbb{C}^{n \times n}$ 的所有特征值。对于 Ω 上的解析函数 $f(\lambda)$ ，我们定义

$$f(A) = \frac{1}{2\pi i} \int_{\partial\Omega} f(z)(zI - A)^{-1} dz.$$

这也是我们在第一节课中提到的矩阵函数的定义方式。这个定义方式有以下优点：

1. 完全内在，与坐标无关。

2. 谱在定义中出现得恰到好处：我们只要求曲线只要包围谱，函数只要在谱附近解析即可，而 $(z\mathbf{I} - \mathbf{A})^{-1}$ 自动编码了谱信息。
3. 完全不依赖于怎么写函数：这个定义告诉我们只要 $f(\lambda) = g(\lambda)$ ，那么就一定有 $f(\mathbf{A}) = g(\mathbf{A})$ 。
4. 对相似变换不变：这是因为积分算子是线性的，与相似变换可交换。
5. 对于解析函数，它与幂级数定义、Jordan 标准型定义是一致的。前面的定义不是错，而不是本质。

或许唯一的缺点是：我们需要验证对不同的区域 Ω 积分结果是否相同。不过，这个验证并不难，因为如果我们有二个区域 Ω_1 和 Ω_2 ，将它们连接起来，直接利用 Cauchy 积分定理即可证明积分结果相同。

结合以上的讨论，我们可以看到：矩阵函数的本质并不是把数换成矩阵，而是通过谱，把解析函数作用于算子。

1.2 性质

矩阵函数有很多重要的性质，下面我们列举几个：

定理 1.4 (相似变换不变性). 设 $\mathbf{A}, \mathbf{V} \in \mathbb{C}^{n \times n}$ ，且 \mathbf{V} 可逆，则对于解析函数 f ，有

$$f(\mathbf{V}\mathbf{A}\mathbf{V}^{-1}) = \mathbf{V}f(\mathbf{A})\mathbf{V}^{-1}.$$

核心思想是利用积分算子的线性性：

证明.

$$\begin{aligned} f(\mathbf{V}\mathbf{A}\mathbf{V}^{-1}) &= \frac{1}{2\pi i} \int_{\partial\Omega} f(z)(z\mathbf{I} - \mathbf{V}\mathbf{A}\mathbf{V}^{-1})^{-1} dz \\ &= \frac{1}{2\pi i} \int_{\partial\Omega} f(z)\mathbf{V}(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{V}^{-1} dz \\ &= \mathbf{V} \left(\frac{1}{2\pi i} \int_{\partial\Omega} f(z)(z\mathbf{I} - \mathbf{A})^{-1} dz \right) \mathbf{V}^{-1} \\ &= \mathbf{V}f(\mathbf{A})\mathbf{V}^{-1}. \end{aligned}$$

□

一个重要推论是：矩阵函数的特征值是函数作用于矩阵的特征值上。

推论 1.5 (谱映射定理). 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ 且 $\lambda_1, \dots, \lambda_n$ 是 \mathbf{A} 的特征值 (按重数计数). 则对于解析函数 f , 有

$$\text{spec}(f(\mathbf{A})) = \{f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)\} = f(\text{spec}(\mathbf{A})).$$

证明. 设 $\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}$ 是 \mathbf{A} 的 Jordan 标准型分解, 则

$$f(\mathbf{A}) = \mathbf{V}f(\mathbf{J})\mathbf{V}^{-1}.$$

因为 \mathbf{J} 是 Jordan 块的直和, 而每个 Jordan 块的特征值都是对应的特征值, 所以 $f(\mathbf{J})$ 的特征值就是 f 作用于每个特征值上. 因此, $f(\mathbf{A})$ 的特征值也就是 f 作用于每个特征值上. \square

谱映射定理也验证了我们之前的直觉: 矩阵函数的本质是通过谱, 把解析函数作用于算子. 这个定理在泛函分析中非常重要.

在 Cauchy 积分定义中, 我们要求 Ω 包含矩阵的所有特征值. 那么, 假如 Ω 只包含部分特征值会怎样呢? 或者, 假如我们想要提取出矩阵的某些特征值对应的特征子空间, 把不需要的特征子空间过滤掉, 又该怎么办呢? 这时, 为简化讨论, 设 \mathbf{A} 是正规矩阵, 并假设 Ω 只包含部分特征值, 我们可以定义

$$f(z) = \mathbf{1}_\Omega = \begin{cases} 1, & z \in \Omega, \\ 0, & z \notin \Omega. \end{cases}$$

则我们有

$$\mathbf{P}(\mathbf{A}) = \frac{1}{2\pi i} \int_{\partial\Omega} f(z)(z\mathbf{I} - \mathbf{A})^{-1} dz = \frac{1}{2\pi i} \int_{\partial\Omega} (z\mathbf{I} - \mathbf{A})^{-1} dz.$$

那么 $\mathbf{P}(\mathbf{A})$ 会等于什么呢? 实际上, 它是一个投影算子:

$$\mathbf{P}(\mathbf{A}) = \sum_{\lambda_i \in \Omega} \mathbf{q}_i \mathbf{p}_i^*,$$

也就是说, 它把所有特征值在 Ω 内的特征子空间投影出来, 而把其他特征子空间过滤掉了. 这是因为, 考虑谱分解 $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^*$, 则

$$\mathbf{P}(\mathbf{A}) = \mathbf{Q} \begin{bmatrix} \mathbf{P}(\lambda_1) & & \\ & \ddots & \\ & & \mathbf{P}(\lambda_n) \end{bmatrix} \mathbf{Q}^*.$$

于是我们自然就能够得到上面的结论.

当然, 如果你想要把这些特征子空间的特征值也提取出来, 你也可以定义

$$\mathbf{T}(\mathbf{A}) = \frac{1}{2\pi i} \int_{\partial\Omega} z(z\mathbf{I} - \mathbf{A})^{-1} dz.$$

原理是类似的, 不作赘述. 对于非正规矩阵, 这两个公式也可以提取出这些特征值对应的特征子空间, 只不过此时投影算子不再是正交投影了, 而是斜投影. 这也印证了“矩阵函数的本质是谱的作用”这句话.

1.3 常见的矩阵函数

矩阵指数函数 由幂级数定义我们知道

$$e^{\mathbf{A}} = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!}.$$

由定义，我们可以得到一系列性质：

$$1. e^{\mathbf{0}} = \mathbf{I}.$$

$$2. \frac{d}{dt} e^{t\mathbf{A}} = \mathbf{A} e^{t\mathbf{A}}.$$

3. 若 $\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}$ ，则有 $e^{\mathbf{A}} = \mathbf{V}e^{\mathbf{J}}\mathbf{V}^{-1}$ ，其中

$$e^{\mathbf{J}_i} = \begin{bmatrix} e^{\lambda_i} & e^{\lambda_i} & \frac{e^{\lambda_i}}{2!} & \cdots & \frac{e^{\lambda_i}}{(n_i-1)!} \\ 0 & e^{\lambda_i} & e^{\lambda_i} & \cdots & \frac{e^{\lambda_i}}{(n_i-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & e^{\lambda_i} \\ 0 & 0 & 0 & \cdots & e^{\lambda_i} \end{bmatrix} \in \mathbb{C}^{n_i \times n_i}, \quad i = 1, \dots, m.$$

或许有一些同学会自然地认为

$$e^{\mathbf{A}+\mathbf{B}} = e^{\mathbf{A}}e^{\mathbf{B}}.$$

实际上这个等式并不成立。考虑

$$\begin{aligned} e^{\mathbf{A}+\mathbf{B}} &= \mathbf{I} + (\mathbf{A} + \mathbf{B}) + \frac{(\mathbf{A} + \mathbf{B})^2}{2!} + \frac{(\mathbf{A} + \mathbf{B})^3}{3!} + \cdots, \\ e^{\mathbf{A}}e^{\mathbf{B}} &= \left(\mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \cdots \right) \left(\mathbf{I} + \mathbf{B} + \frac{\mathbf{B}^2}{2!} + \frac{\mathbf{B}^3}{3!} + \cdots \right). \end{aligned}$$

展开后我们会发现，只有当 \mathbf{A} 和 \mathbf{B} 乘法可交换时，这两个式子才相等。

矩阵指数函数在微分方程中有着重要的应用。我们知道，标量微分方程

$$\frac{dx}{dt}(t) = \lambda x(t)$$

可以通过两边同时乘上 $e^{-\lambda t}$ 来求解。最终的解为

$$x(t) = e^{\lambda t} x(0).$$

类似地，考虑矩阵微分方程

$$\frac{d\mathbf{x}}{dt}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

两边同时乘上 $e^{-t\mathbf{A}}$ ，有

$$e^{-t\mathbf{A}}\mathbf{x}' = e^{-t\mathbf{A}}\mathbf{A}\mathbf{x} + e^{-t\mathbf{A}}\mathbf{B}\mathbf{u}.$$

亦即

$$\frac{d}{dt} (e^{-tA} \mathbf{x}) = e^{-tA} \mathbf{B} \mathbf{u}.$$

积分后得到

$$e^{-tA} \mathbf{x}(t) - \mathbf{x}(0) = \int_0^t e^{-sA} \mathbf{B} \mathbf{u}(s) ds.$$

因此，矩阵微分方程的解为

$$\mathbf{x}(t) = e^{tA} \mathbf{x}(0) + \int_0^t e^{(t-s)A} \mathbf{B} \mathbf{u}(s) ds. \quad (1)$$

这是一个非常重要的结果。它在控制理论中有着广泛的应用。

现在，带着这个结果，我们来考虑一个扰动分析问题： $e^{A+\Delta A}$ 和 e^A 会差多少？当 $A\Delta A \neq \Delta A A$ 时，我们无法直接使用 $e^{A+\Delta A} = e^A e^{\Delta A}$ 来分析。那么，我们该如何分析呢？我们可以考虑引入一个参数 t ，定义

$$\mathbf{X}(t) = e^{t(A+\Delta A)}.$$

则有 $\mathbf{X}(1) = e^{A+\Delta A}$. 对 $\mathbf{X}(t)$ 求导，有

$$\frac{d\mathbf{X}}{dt}(t) = (A + \Delta A) \mathbf{X}(t) = A \mathbf{X}(t) + \Delta A \mathbf{X}(t).$$

我们利用 (1) 的结果，得到

$$\mathbf{X}(t) = e^{tA} + \int_0^t e^{(t-s)A} \Delta A \mathbf{X}(s) ds.$$

令 $t = 1$ ，有

$$e^{A+\Delta A} = e^A + \int_0^1 e^{(1-s)A} \Delta A e^{s(A+\Delta A)} ds.$$

如果我们还有其他条件，比如 $A^* = -A$ 及 $\Delta A^* = -\Delta A$ ，我们则可以进行进一步的估计。由于反 Hermite 矩阵的特征值是纯虚数，它的指数函数满足

$$e^A = Q e^\Lambda Q^*, \quad e^\Lambda = \begin{bmatrix} e^{i\theta_1} & & \\ & \ddots & \\ & & e^{i\theta_n} \end{bmatrix}$$

于是

$$\exp(A)^* \exp(A) = Q \exp(\Lambda)^* \exp(\Lambda) Q^* = I.$$

因此， e^A 是酉矩阵。因此

$$\|e^{A+\Delta A} - e^A\|_{\text{ui}} \leq \int_0^1 \|e^{(1-s)A} \Delta A e^{s(A+\Delta A)}\|_{\text{ui}} ds = \|\Delta A\|_{\text{ui}},$$

其中 $\|\cdot\|_{\text{ui}}$ 是任意酉不变范数。

接下来，我们来看看矩阵指数函数在矩阵方程中的应用。下面这个方程被称为 Lyapunov 方程：

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^* = \mathbf{B}.$$

假如我们要求 $\operatorname{Re}(\lambda(\mathbf{A})) < 0$ ，则 Lyapunov 方程有唯一解，且解可以表示为

$$\mathbf{X} = - \int_0^{+\infty} \mathbf{e}^{t\mathbf{A}} \mathbf{B} \mathbf{e}^{t\mathbf{A}^*} dt.$$

遗憾的是，凭我们现有的知识，我们还无法证明这个结果。不过我们可以验证这个结果的正确性。首先，我们得先弄清楚 $\operatorname{Re}(\lambda(\mathbf{A})) > 0$ 这个条件到底是什么意思。根据谱映射定理，若 \mathbf{A} 有特征值 $\alpha + \beta i$ ，那么 $\mathbf{e}^{\mathbf{A}}$ 就有特征值 $\mathbf{e}^{\alpha + \beta i}$ 。 $\alpha < 0$ 意味着， $|\mathbf{e}^{\alpha + \beta i}| = |\mathbf{e}^\alpha| \cdot |\mathbf{e}^{i\beta}| = |\mathbf{e}^\alpha| < 1$ 。这也就是说， $\rho(\mathbf{e}^{\mathbf{A}}) < 1$ 。于是

$$\lim_{t \rightarrow +\infty} \mathbf{e}^{t\mathbf{A}} = \mathbf{0}.$$

这是因为当 $t \rightarrow +\infty$ 时， $\mathbf{e}^{t\mathbf{A}}$ 的每个特征值的模都趋近于零。现在，我们来验证 Lyapunov 方程的解：

$$\begin{aligned} \mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^* &= - \int_0^{+\infty} \left(\mathbf{A} \mathbf{e}^{t\mathbf{A}} \mathbf{B} \mathbf{e}^{t\mathbf{A}^*} + \mathbf{e}^{t\mathbf{A}} \mathbf{B} \mathbf{e}^{t\mathbf{A}^*} \mathbf{A}^* \right) dt \\ &= - \int_0^{+\infty} \left(\frac{d}{dt} \mathbf{e}^{t\mathbf{A}} \mathbf{B} \mathbf{e}^{t\mathbf{A}^*} + \mathbf{e}^{t\mathbf{A}} \mathbf{B} \frac{d}{dt} \mathbf{e}^{t\mathbf{A}^*} \right) dt \\ &= - \int_0^{+\infty} \frac{d}{dt} \left(\mathbf{e}^{t\mathbf{A}} \mathbf{B} \mathbf{e}^{t\mathbf{A}^*} \right) dt \\ &= - \lim_{t \rightarrow +\infty} \mathbf{e}^{t\mathbf{A}} \mathbf{B} \mathbf{e}^{t\mathbf{A}^*} + \mathbf{e}^{0 \cdot \mathbf{A}} \mathbf{B} \mathbf{e}^{0 \cdot \mathbf{A}^*} = \mathbf{B}. \end{aligned}$$

Lyapunov 方程是 Sylvester 方程的一个特例。对于 Sylvester 方程 $\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{B} = \mathbf{C}$ ，如果

$$\begin{cases} \operatorname{Re}(\lambda(\mathbf{A})) < 0, \\ \operatorname{Re}(\lambda(\mathbf{B})) > 0, \end{cases}$$

则 Sylvester 方程有唯一解，且解可以表示为

$$\mathbf{X} = - \int_0^{+\infty} \mathbf{e}^{t\mathbf{A}} \mathbf{C} \mathbf{e}^{t\mathbf{B}} dt.$$

幂函数 幂函数是最简单的矩阵函数之一。不过，我们需要注意幂函数的定义域问题。对于正整数 $p \in \mathbb{N}$ ， \mathbf{A}^p 对任意矩阵 \mathbf{A} 都有定义；但若 $p \in \mathbb{Z}$ ，此时 \mathbf{A} 必须是可逆矩阵；若 $p \in \mathbb{R}$ ，我们需要进一步要求 \mathbf{A} 是正定矩阵——具体地，若 $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^*$ ，则有

$$\mathbf{A}^p = \mathbf{Q}\mathbf{\Lambda}^p\mathbf{Q}^*, \quad \mathbf{\Lambda}^p = \begin{bmatrix} \lambda_1^p & & \\ & \ddots & \\ & & \lambda_n^p \end{bmatrix}.$$

从另一个角度来看，对于正定矩阵，由于它的特征值都是正实数，在 Cauchy 积分定义中我们可以选择包含正实数轴的区域来定义矩阵幂函数——这暗示着，正定矩阵的幂函数似乎可以通过正实数轴上的积分来定义，因为虚轴上并没有奇点。事实上，若 $\alpha \in (0, 1)$ ，即正分数幂，我们有

$$\mathbf{A}^\alpha = \frac{\sin(\pi\alpha)}{\pi\alpha} \int_0^{+\infty} (\mathbf{I} + t^{1/\alpha} \mathbf{A}^{-1})^{-1} dt.$$

这个矩阵恒等式来自于标量恒等式

$$x^\alpha = \frac{\sin(\pi\alpha)}{\pi\alpha} \int_0^{+\infty} \frac{1}{1 + t^{1/\alpha} x^{-1}} dt, \quad x > 0, \alpha \in (0, 1).$$

读者可自行验证。利用这个结论可以解决 10 月 21 日作业的第 6 题，限于篇幅，此处不作展开。

矩阵对数函数 对数函数只在正实数上有定义，因此我们要求矩阵必须是正定矩阵。而对于正定矩阵的对数函数，我们也可以有类似的实积分表示：

$$\log(\mathbf{A}) = \int_0^{+\infty} \left(\frac{1}{1+t} \mathbf{I} - (\mathbf{A} + t\mathbf{I})^{-1} \right)^{-1} dt.$$

这些积分表示的推导都比较复杂，但它们的思想是类似的：都是利用矩阵谱的特殊性质，把积分路径限制在实轴上。

矩阵符号函数 对复数 z ，我们定义它的符号函数为

$$\text{sign}(z) = \begin{cases} 1, & \text{Re}(z) > 0, \\ -1, & \text{Re}(z) < 0 \\ \text{not defined} & \text{Re}(z) = 0. \end{cases}$$

于是，若矩阵 \mathbf{A} 没有纯虚特征值，设其 Jordan 标准型分解为 $\mathbf{A} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$ ，我们定义它的符号函数为

$$\text{sign}(\mathbf{A}) = \mathbf{P} \text{diag}\{\text{sign}(\mathbf{J}_1), \dots, \text{sign}(\mathbf{J}_m)\} \mathbf{P}^{-1},$$

其中

$$\text{sign}(\mathbf{J}_i) = \text{sign}(\lambda_i) \mathbf{I}, \quad i = 1, \dots, m.$$

如果我们事先排好序，使得

$$\text{sign}(\mathbf{J}) = \begin{bmatrix} \mathbf{I}_{n_+} & 0 \\ 0 & -\mathbf{I}_{n_-} \end{bmatrix}$$

则有

$$\text{sign}(\mathbf{A}) = \mathbf{P} \begin{bmatrix} \mathbf{I}_{n_+} & 0 \\ 0 & -\mathbf{I}_{n_-} \end{bmatrix} \mathbf{P}^{-1}.$$

我们对 \mathbf{P} 和 \mathbf{P}^{-1} 也进行相应的分块：

$$\mathbf{P} = [\mathbf{X}_1, \mathbf{X}_2], \quad \mathbf{P}^{-1} = \begin{bmatrix} \mathbf{Y}_1^* \\ \mathbf{Y}_2^* \end{bmatrix},$$

则有

$$\text{sign}(\mathbf{A}) = \mathbf{X}_1 \mathbf{Y}_1^* - \mathbf{X}_2 \mathbf{Y}_2^*.$$

另一方面，由于 $\mathbf{X} \mathbf{X}^{-1} = \mathbf{I}$ ，我们有

$$\mathbf{I} = \mathbf{X}_1 \mathbf{Y}_1^* + \mathbf{X}_2 \mathbf{Y}_2^*.$$

结合以上两个式子，我们得到

$$\mathbf{X}_2 \mathbf{Y}_2^* = \frac{1}{2}(\mathbf{I} - \text{sign}(\mathbf{A})).$$

而我们又注意到 $\mathbf{A} \mathbf{X}_2 = \mathbf{X}_2 \mathbf{J}_2$ ，这说明 $\text{span}(\mathbf{X}_2)$ 张成了 \mathbf{A} 的负特征子空间，因此如果我们能够得到 $\text{sign}(\mathbf{A})$ 的近似，通过对 $\mathbf{I} - \text{sign}(\mathbf{A})$ 做满秩分解，我们就能得到负特征子空间的一个近似。类似地， $\text{span}(\mathbf{X}_1)$ 张成了 \mathbf{A} 的正特征子空间，因此我们也可以通过 $\text{sign}(\mathbf{A})$ 得到正特征子空间的近似。由于符号函数可以用 Newton 方法高效地计算出来，因此它在特征值分解中有着重要的应用。读者可参考 Golub 和 Van Loan 的 *Matrix Computations* 的 9.4 节了解更多内容。

2 非负矩阵

请注意，矩阵的复数是 *Matrices*，不是 *Matrixes*！

在第一部分中，我们通过解析函数 $f(z)$ 刻画了矩阵 \mathbf{A} 的行为。现在我们将视角从解析性转回矩阵的元素结构。

2.1 正矩阵

在数值计算中，我们总是希望快速定位矩阵的主导特征信息。对于一般的矩阵来说，这往往很难；但正矩阵由于其特殊的结构，使得我们可以快速地定位主导特征信息。乘幂法之所以在很多问题中都能够快速收敛，正是因为现实问题中的矩阵大多是正矩阵。

如图 2，我们随机生成了一个正矩阵，并将其谱画了出来。我们可以看到，最大特征值明显地脱离了其他特征值，而且这个最大特征值正好落在正实轴上。

Eigenvalue Distribution of a Positive Matrix (Perron Theorem)

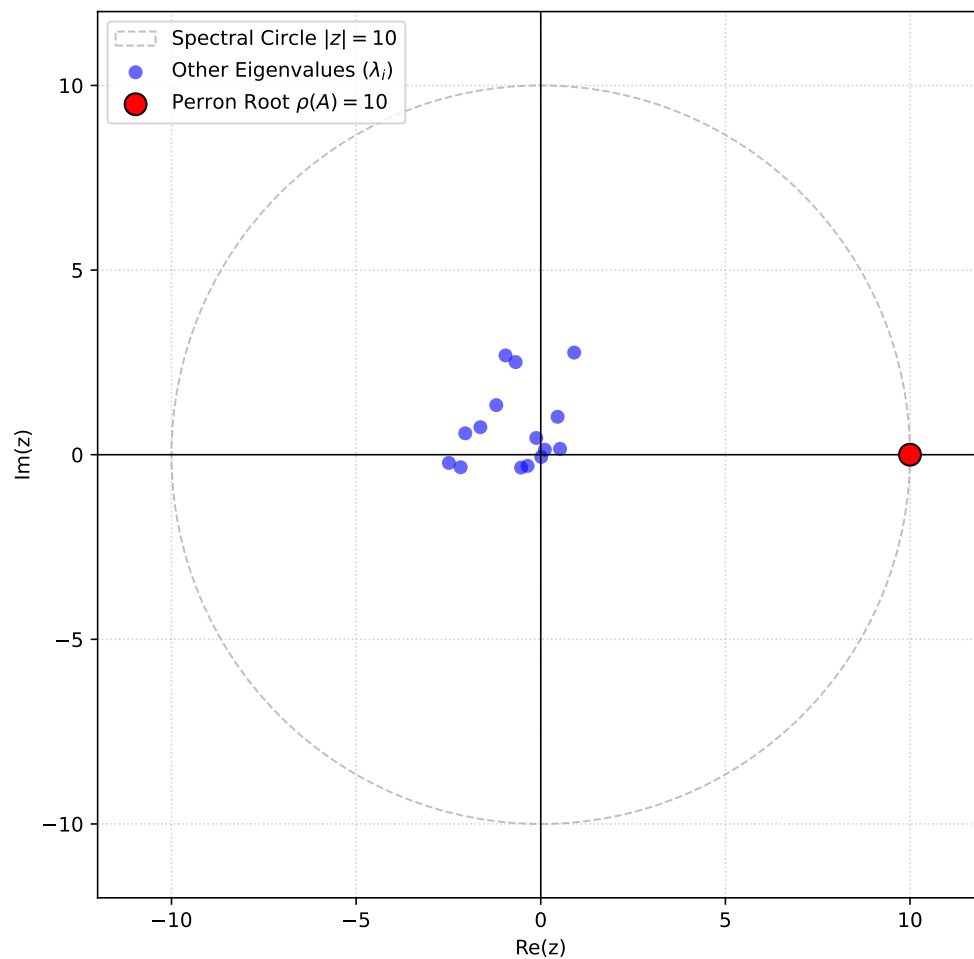


图 2: 正矩阵的谱

这种现象并非偶然。事实上，我们有如下重要定理：

定理 2.1 (Perron 定理). 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 是正矩阵，则

1. $\rho(\mathbf{A}) > 0$.
2. $\exists \mathbf{x} > 0$ 使得 $\mathbf{A}\mathbf{x} = \mathbf{x}\rho(\mathbf{A})$; $\exists \mathbf{y} > 0$ 使得 $\mathbf{y}^T \mathbf{A} = \rho(\mathbf{A})\mathbf{y}^T$.
3. $\forall \lambda \in \text{spec}(\mathbf{A}) \setminus \{\rho(\mathbf{A})\}$, 有 $|\lambda| < \rho(\mathbf{A})$.
4. $m_{\text{alge}}(\rho(\mathbf{A})) = 1$.
5. $\lim_{k \rightarrow \infty} \left(\frac{\mathbf{A}}{\rho(\mathbf{A})} \right)^k = \frac{\mathbf{x}\mathbf{y}^T}{\mathbf{y}^T \mathbf{x}}$.

首先我们来解释一下定理的内容：第一条说明正矩阵的谱半径是正的，亦即正矩阵一定有非零特征值；第二条说明正矩阵的最大特征值对应的左右特征向量都可以取为正向量；第三条说明正矩阵的最大特征值在模意义下是唯一的；第四条说明正矩阵的最大特征值代数重数为 1，也就是说它只对应于一个一维的特征子空间；第五条说明正矩阵经过适当归一化后，乘幂会收敛到一个秩 1 矩阵，这个秩 1 矩阵由最大特征值对应的左右特征向量张成。

这个定理实在是强得令人吃惊：唯一性和占优性居然仅仅依赖于所有元素大于零这一个条件。事实上，矩阵的正性隐含了非常强的耦合关系，它会使得矩阵的谱结构变得非常简单。正矩阵是一种严格正向的作用。从哲学的角度来看，当一种相互作用是纯粹的、普遍的、建设性的时，它将从数学的混沌中，必然地、唯一地孕育出一种最高效、最稳定、最和谐的秩序状态。

接下来，我们就来证明 Perron 定理。

证明. (1) 由于 \mathbf{A} 是正矩阵，设 $\mathbf{e} = [1, 1, \dots, 1]^T \in \mathbb{R}^n$ ，则有

$$\mathbf{A}\mathbf{e} > 0.$$

于是，存在一个数 $\alpha > 0$ ，使得

$$\mathbf{A}\mathbf{e} \geq \alpha\mathbf{e}.$$

由 Gelfand 定理，我们有

$$\rho(\mathbf{A}) = \lim_{k \rightarrow \infty} \|\mathbf{A}^k\|^{1/k} \geq \lim_{k \rightarrow \infty} \|\alpha^k \mathbf{I}\|^{1/k} \geq \alpha > 0.$$

(2) 定理的第二条比较难证。我们需要用到 Brouwer 不动点定理。我们不妨定义

$$\Omega = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq 0\}, \quad \Omega_1 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq 0, \sum_{i=1}^n x_i = 1\}.$$

显然， Ω_1 是将 Ω 中的向量在 1-范数意义下归一化后的集合，它是一个闭凸集。我们知道， $\mathbf{A}(\Omega) \subset \Omega$ ——这是因为 \mathbf{A} 的作用是完全正向的。现在我们希望定义一个映射 \mathbf{f} ，使得 $\mathbf{f}(\Omega_1) \subset \Omega_1$ 。为此，我们定义

$$\mathbf{f}(\mathbf{x}) = \frac{\mathbf{A}\mathbf{x}}{\|\mathbf{A}\mathbf{x}\|_1}, \quad \mathbf{x} \in \Omega_1.$$

显然， \mathbf{f} 是一个从 Ω_1 到 Ω_1 的映射。由于 Ω_1 是闭凸集，且 \mathbf{f} 是连续映射，根据 Brouwer 不动点定理， \mathbf{f} 至少有一个不动点 $\mathbf{x}_\star \in \Omega_1$ ，使得

$$\mathbf{f}(\mathbf{x}_\star) = \mathbf{x}_\star \implies \mathbf{A}\mathbf{x}_\star = \mathbf{x}_\star \|\mathbf{A}\mathbf{x}_\star\|_1 := \mathbf{x}_\star \rho.$$

下面我们只需说明 $\rho = \rho(\mathbf{A})$ 即可。这部分证明比较繁琐，读者可参考 Horn 和 Johnson 的 *Matrix Analysis* 第 8.2 节的定理 8.2.2。

(3)(4) 可参考 Horn 和 Johnson 的 *Matrix Analysis* 第 8.2 节的定理 8.2.7。

(5) 本质上是乘幂法的收敛性。由 (3), $\rho(\mathbf{A})$ 是主导的特征值, 其他特征值的模严格小于谱半径。因此, 乘幂法收敛到主导特征子空间。由于 (4) 保证了主导特征子空间是一维的, 因此 $\mathbf{A}^k/\rho(\mathbf{A})$ 能够收敛到一个秩 1 矩阵。那么这个秩 1 矩阵是什么呢? 我们设 \mathbf{y}, \mathbf{x} 是对应于 $\rho(\mathbf{A})$ 的左右特征向量, 满足归一化条件 $\mathbf{y}^T \mathbf{x} = 1$ 。由乘幂法原理我们有

$$\lim_{k \rightarrow \infty} \left(\frac{\mathbf{A}^k}{\rho(\mathbf{A})^k} \right) = \text{最大特征值对应的特征子空间}$$

因此, 这个极限矩阵可以表示为 $\mathbf{x} \mathbf{z}^T$ 的形式, 其中 \mathbf{z} 是某个单位向量。为了确定 \mathbf{z} , 我们考虑

$$\mathbf{y}^T \mathbf{x} \mathbf{z}^T.$$

由于 $\mathbf{y}^T \mathbf{A} = \rho \mathbf{y}^T$, 我们有

$$\mathbf{y}^T \mathbf{x} \mathbf{z}^T = \lim_{k \rightarrow \infty} \mathbf{y}^T \left(\frac{\mathbf{A}^k}{\rho^k} \right) = \lim_{k \rightarrow \infty} \frac{\mathbf{y}^T \mathbf{A}^k}{\rho^k} = \mathbf{y}^T.$$

但另一方面

$$\mathbf{y}^T \mathbf{x} \mathbf{z}^T = (\mathbf{y}^T \mathbf{x}) \mathbf{z}^T = \mathbf{z}^T.$$

结合以上两个式子, 我们得到 $\mathbf{z}^T = \mathbf{y}^T$ 。因此, 极限矩阵为

$$\lim_{k \rightarrow \infty} \left(\frac{\mathbf{A}}{\rho(\mathbf{A})} \right)^k = \mathbf{x} \mathbf{y}^T.$$

如果左右特征向量没有归一化, 那么极限矩阵是

$$\lim_{k \rightarrow \infty} \left(\frac{\mathbf{A}}{\rho(\mathbf{A})} \right)^k = \frac{\mathbf{x} \mathbf{y}^T}{\mathbf{y}^T \mathbf{x}}.$$

事实上, 这是一个斜投影矩阵, 它把任何向量都投影到主导特征子空间上。从几何角度, 这也是非常合理的: 谱投影算子必须唯一, 因为它要同时满足值域是主导特征子空间, 而零空间又要是其他特征子空间的直和。在这一斜投影算子下, 右特征向量 \mathbf{x} 描述了系统的状态, 而左特征向量 \mathbf{y} 则描述了系统的观测方式, 二者的配对是天然符合数学美感的。□

我们称 $\rho(\mathbf{A})$ 为 Perron 根, 称对应的左右特征向量为左右 Perron 向量。

例 2.2. 设

$$\mathbf{A} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

则 \mathbf{A} 是双随机矩阵。由 Gershgorin 圆盘定理可知, $\rho(\mathbf{A}) \leq 1$ 。另一方面, 设 $\mathbf{e} = [1, 1, 1]^T$, 则有

$$\mathbf{A} \mathbf{e} = \mathbf{e}, \quad \mathbf{e}^T \mathbf{A} = \mathbf{e}^T.$$

由 Perron 定理可知, $\rho(\mathbf{A}) = 1$, 且对应的左右 Perron 向量均为 \mathbf{e} 。因此

$$\lim_{k \rightarrow \infty} \mathbf{A}^k = \frac{\mathbf{e} \mathbf{e}^T}{\mathbf{e}^T \mathbf{e}} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

2.2 非负矩阵

正矩阵的作用都是严格正向的，而非负矩阵则不然，零意味着连接的断裂，意味着系统存在着冷漠、隔绝或壁垒。这时，秩序就可能瓦解为多个孤立的子系统，各自为政，互不干涉——Perron 定理的许多结论都不再成立。

具体地，第一条谱半径不再严格大于零，例如零矩阵的谱半径就是零。第二条左右 Perron 向量也不一定是正向量，只能保证非负。例如，二阶 Jordan 块

$$\mathbf{J} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

的左特征向量为 $[0, 1]^T$ ，右特征向量为 $[1, 0]^T$ ，均非负但不正。第三条占优性也不再成立，例如 Jordan 块

$$\mathbf{J} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

的两个特征值均为 1。第四条代数重数也不再是 1，例如上面的 Jordan 块的代数重数为 2。第五条乘幂法也不一定收敛，因为模最大的特征值不唯一。

所以，对非负矩阵而言，Perron 定理唯一能保住的结论是：谱半径对应的特征值至少有一个非负的左右特征向量。这实在无法令我们满意。那么，我们需要给非负矩阵添加什么条件，才能让 Perron 定理的其余结论成立呢？

前面我们提到，非负矩阵的缺陷在于，系统中存在着孤立的子系统。从数学的角度来看，这也就是说矩阵可以通过某种合同排列，变成分块上三角的形式：

$$\mathbf{P}^T \mathbf{A} \mathbf{P} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \cdots & \mathbf{A}_{1,m} \\ 0 & \mathbf{A}_{2,2} & \cdots & \mathbf{A}_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}_{m,m} \end{bmatrix}$$

其中每个对角块 $\mathbf{A}_{i,i}$ 都是方阵。这些块对应的空间都是不变子空间，导致块与块之间老死不相往来。我们把这样的矩阵称为 **可约矩阵**。相反地，如果一个非负矩阵无法通过合同排列变成分块上三角形式，我们就称它为 **不可约矩阵**。从直觉上来说，由于可约矩阵的各个对角块之间没有相互作用，因此它们的最大特征值可能相等，这就直接导致了 Perron 定理的失败。从图论的角度来看，可约矩阵意味着从某个点出发，可能永远无法到达系统中的另一个点。另一方面，不可约矩阵在图论中对应着强连通图，意味着系统中的任何一个点都可以通过某种路径到达系统中的任何另一个点，这种强耦合关系正是 Perron 定理成立的基础。

顺道一提，不可约矩阵 \mathbf{A} 的高次幂可能是正矩阵，即存在某个正整数 k 使得 \mathbf{A}^k 是正矩阵。这是因为不可约矩阵对应的图是强连通图，而强连通图经过有限步长的路径之后，任意两个节点之间

都存在路径相连。这样就能够把所有零元填满，从而得到正矩阵。进一步地，下面两个矩阵函数也会是正矩阵：

$$e^{\mathbf{A}} = \mathbf{I} + \frac{\mathbf{A}}{1!} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \cdots,$$

$$(\mathbf{I} + \mathbf{A})^{n-1} = \mathbf{I} + \binom{n-1}{1}\mathbf{A} + \binom{n-1}{2}\mathbf{A}^2 + \cdots + \binom{n-1}{n-1}\mathbf{A}^{n-1}.$$

现在，我们来陈述不可约非负矩阵的 Perron–Frobenius 定理：

定理 2.3 (不可约非负矩阵的 Perron–Frobenius 定理). 设 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 是不可约非负矩阵，则

1. $\rho(\mathbf{A}) > 0$.
2. $\exists \mathbf{x} > 0$ 使得 $\mathbf{A}\mathbf{x} = \mathbf{x}\rho(\mathbf{A})$; $\exists \mathbf{y} > 0$ 使得 $\mathbf{y}^T \mathbf{A} = \rho(\mathbf{A})\mathbf{y}^T$.
3. $m_{\text{alge}}(\rho(\mathbf{A})) = 1$.

我们仍称 $\rho(\mathbf{A})$ 为 Perron 根，称对应的左右特征向量为左右 Perron 向量。

与正矩阵的 Perron 定理相比，不可约非负矩阵的 Perron–Frobenius 定理少了两条结论：占优性和乘幂法收敛性。这个定理的证明比较复杂，读者可参考 Horn 和 Johnson 的 *Matrix Analysis* 第 8.4 节的定理 8.4.4。

2.3 Perron 根

下面我们就来研究一下 Perron 根的性质。

第一条性质自然是 Gelfand 公式：

定理 2.4 (Gelfand 公式). 设 $\mathbf{A} \in \mathbb{C}^{n \times n}$ ，则有

$$\lim_{k \rightarrow \infty} \|\mathbf{A}^k\|^{1/k} = \rho(\mathbf{A}).$$

证明我们在第三节课中已经讲过，这里不再赘述。

第二条性质是单调性：

命题 2.5. 设 $\mathbf{A} \geq \mathbf{B} \geq 0$ ，则有

$$\rho(\mathbf{A}) \geq \rho(\mathbf{B}).$$

若 $\mathbf{A} > \mathbf{B} \geq 0$ ，则有

$$\rho(\mathbf{A}) > \rho(\mathbf{B}).$$

证明. (1) 由 Gelfand 公式可知

$$\begin{aligned}\rho(\mathbf{A}) &= \lim_{k \rightarrow \infty} \|\mathbf{A}^k\|^{1/k} \\ &\geq \lim_{k \rightarrow \infty} \|\mathbf{B}^k\|^{1/k} \\ &= \rho(\mathbf{B}).\end{aligned}$$

(2) 设 $\mathbf{A} \geq \alpha \mathbf{B} > \mathbf{B} \geq 0$. 由 (1) 可知

$$\rho(\mathbf{A}) \geq \alpha \rho(\mathbf{B}) > \rho(\mathbf{B}).$$

□

很多时候我们计算特征值是不够方便的, 那么我们有没有一套行之有效的方法来估计 Perron 根呢? 首先由 Gershgorin 圆盘定理可知, Perron 根一定落在所有 Gershgorin 圆盘的联合中。因此, 它的上界可以轻松地估计出来:

$$\rho(\mathbf{A}) \leq \max_{1 \leq i \leq n} \sum_{j=1}^n a_{i,j}.$$

那么下界呢? 我们自然会猜测

$$\rho(\mathbf{A}) \geq \min_{1 \leq i \leq n} \sum_{j=1}^n a_{i,j}.$$

这个猜测对不对呢? 我们先来考虑一种特殊情况: 所有行和都相等的矩阵。此时,

$$\mathbf{A} \mathbf{e} = \mathbf{e} \rho(\mathbf{A}).$$

因此, $\rho(\mathbf{A})$ 就是行和。现在, 我们来考虑一般情况。我们希望尽可能地把一般情况规约到行和相等的特殊情况。假设最大行和是 R , 最小行和是 $r = r_{j_0}$, 则我们定义一个对角矩阵

$$\mathbf{D} = \begin{bmatrix} \frac{r}{r_1} & & & \\ & \frac{r}{r_2} & & \\ & & \ddots & \\ & & & \frac{r}{r_n} \end{bmatrix}$$

则有

$$\mathbf{D} \mathbf{A} \leq \mathbf{A}.$$

由单调性可知

$$\rho(\mathbf{D} \mathbf{A}) \leq \rho(\mathbf{A}).$$

另一方面, $\mathbf{D} \mathbf{A}$ 的每一行的和都是 r , 因此

$$\rho(\mathbf{D} \mathbf{A}) = r.$$

综上所述，我们得到了 Perron 根的行和界：

$$\min_{1 \leq i \leq n} \sum_{j=1}^n a_{i,j} \leq \rho(\mathbf{A}) \leq \max_{1 \leq i \leq n} \sum_{j=1}^n a_{i,j}.$$

我们还可以对这个不等式进行改进。我们刚刚是直接把一行的所有元素加起来，那么如果我们想要加一点权重呢？具体地，设 $\mathbf{x} > 0$ ，那么加权的行和就是 \mathbf{Ax} 。为了能够使用我们上面的行和估计结果，我们考虑将 \mathbf{Ax} 变换为

$$\mathbf{Ax} = \mathbf{A} \begin{bmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{bmatrix} \mathbf{e}.$$

但是，这样我们就改变了矩阵 \mathbf{A} 的谱结构。为了补偿这个变化，我们考虑相似变换

$$\left(\begin{bmatrix} x_1^{-1} & & & \\ & x_2^{-1} & & \\ & & \ddots & \\ & & & x_n^{-1} \end{bmatrix} \mathbf{A} \begin{bmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{bmatrix} \right) \mathbf{e}.$$

于是，由行和估计可知

$$\min_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^n a_{i,j} x_j \leq \rho(\mathbf{A}) \leq \max_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^n a_{i,j} x_j.$$

如果 \mathbf{x} 是右 Perron 向量，那么上面的不等式就变成了等式，即有

$$\max_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^n a_{i,j} x_j = \rho(\mathbf{A}) = \min_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^n a_{i,j} x_j.$$

这就是 Collatz–Wielandt 公式。

2.4 随机矩阵

随机矩阵的英文是 *Stochastic Matrices*，不是 *Random Matrices*！

在随机过程中，我们经常会遇到一种特殊的非负矩阵，称为随机矩阵。随机矩阵有三种类型：行随机矩阵、列随机矩阵和双随机矩阵。行随机矩阵是指每一行的元素和为 1 的非负矩阵；列随机矩阵是指每一列的元素和为 1 的非负矩阵；双随机矩阵是指每一行和每一列的元素和都为 1 的非负矩阵。作为高代课，我们不会深入探讨随机矩阵的应用背景，只介绍它的一些基本性质。

命题 2.6 (Birkhoff 定理). 双随机矩阵是有限个排列阵的凸组合。

证明参考 Horn 和 Johnson 的 *Matrix Analysis* 第 8.7 节的定理 8.7.2。

命题 2.7. 若 $Q \in \mathbb{R}^{n \times n}$ 是所有非对角元素都非负且行和为零的矩阵, 则 e^{tQ} 是行随机矩阵, $\forall t \geq 0$.

证明. 由 Q 行和为零, 我们有 $Qe = 0$. 因此,

$$e^{tQ}e = \left(I + \frac{tQ}{1!} + \frac{(tQ)^2}{2!} + \cdots \right) e = e, \quad \forall t \geq 0.$$

另一方面, 由于 Q 的所有非对角元素都非负, e^{tQ} 的所有元素也都是非负的。因此, e^{tQ} 是行随机矩阵。 \square