数学通报， 2002 第 8 期， 0–3

第一特征值问题

陈木法

(北京师范大学数学系，随机数学中心，北京 100875)
(Fax: (010)62209447; E-mail: mfchen@bnu.edu.cn)
(主页： http://www.bnu.edu.cn/~chenmf)
(June 9, 2002)

摘要 本文就矩阵的特殊情形，以尽可能初等的语言，介绍将在 2002 年国际数学家大会上演讲的部分内容：所研究的问题、难度、价值、一个主要结果、部分证明和进一步论题.

1 问题.

考虑如下无限矩阵

$$Q=(q_{ij})=\begin{pmatrix} -b_0 & b_0 & 0 & 0 & \dots \\ a_1 & -(a_1+b_1) & b_1 & 0 & \dots \\ 0 & a_2 & -(a_2+b_2) & b_2 & \dots \\ \vdots & \ddots & \ddots & \ddots & \end{pmatrix}$$

其中 a_k, $b_k>0$. 这里的行、列标记为 $0,1,2,\cdots$. 因为仅有处于对角线附近的三条线的元素非零，故称为三对角阵. 注意矩阵的每一行和为零，因此它与元素恒为 1 的常值列向量 $\mathbf{1}$ 的乘积为元素恒为零的列向量 $\mathbf{0}$: $Q\mathbf{1}=\mathbf{0}=0\cdot\mathbf{1}$. 即矩阵 Q 有平凡特征值 $\lambda_0=0$. 其次，若考虑它的前 n+1 阶子矩阵 Q_n, 则 $-Q_n$ 有 n+1 个特征值： $0=\lambda_0<\lambda_1\leqslant\cdots\leqslant\lambda_n$. 我们所关心的是 λ_1, 即第一个非平凡特征值. 当然，同样的问题也适用于有限维或无穷维空间上的微分算子.

2 难点.

为使大家对于此问题的难度有点具体的感受，让我们看看一些简单例子. 先从平凡的两点情形开始. 此时 $Q=\begin{pmatrix} -b_0 & b_0 \\ a_1 & -a_1\end{pmatrix}$, $-Q$ 的第一特征值是 $\lambda_1=a_1+b_0$. 结果很好， λ_1 随 a_1 或 b_0 的增加而增加. 再看三点情形. 此时有 4 个参数 b_0,b_1 和 a_1,a_2,

$$\lambda_1=2^{-1}\big[a_1+a_2+b_0+b_1-\sqrt{(a_1-a_2+b_0-b_1)^2+4a_1b_1}\big].$$

∗ 基金项目：国家自然科学基金 (批准号： 10121101); 973 项目；教育部博士点专项研究基金资助项目

陈木法 (1946–), 男，福建人，北京师范大学数学系教授，从事随机过程及相关领域研究

但诸参数对于 λ_1 的贡献已经相当含混．最后看四点情形，此时有 6 个参数：$b_0, b_1, b_2, a_1, a_2, a_3$. 第一特征值是

$$\lambda_1 = \frac{D}{3} - \frac{C}{3\cdot 2^{1/3}} + \frac{2^{1/3}\,\left(3\,B - D^2\right)}{3\,C},$$

其中 D, B, C 三个量的表达式并不复杂：

$$\begin{aligned}
&D = a_1 + a_2 + a_3 + b_0 + b_1 + b_2,\\
&B = a_3\,b_0 + a_2\,(a_3 + b_0) + a_3\,b_1 + b_0\,b_1 + b_0\,b_2 + b_1\,b_2 + a_1\,(a_2 + a_3 + b_2)\,,\\
&C = \left(A + \sqrt{4(3\,B - D^2)^3 + A^2}\right)^{1/3},
\end{aligned}$$

但

$$\begin{aligned}
A = &-2\,a_1^3 - 2\,a_2^3 - 2\,a_3^3 + 3\,a_3^2\,b_0 + 3\,a_3\,b_0^2 - 2\,b_0^3 + 3\,a_3^2\,b_1 - 12\,a_3\,b_0\,b_1 + 3\,b_0^2\,b_1\\
&+ 3\,a_3\,b_1^2 + 3\,b_0\,b_1^2 - 2\,b_1^3 - 6\,a_3^2\,b_2 + 6\,a_3\,b_0\,b_2 + 3\,b_0^2\,b_2 + 6\,a_3\,b_1\,b_2 - 12\,b_0\,b_1\,b_2\\
&+ 3\,b_1^2\,b_2 - 6\,a_3\,b_2^2 + 3\,b_0\,b_2^2 + 3\,b_1\,b_2^2 - 2\,b_2^3 + 3\,a_1^2\,(a_2 + a_3 - 2\,b_0 - 2\,b_1 + b_2)\\
&+ 3\,a_2^2\,[a_3 + b_0 - 2\,(b_1 + b_2)]\\
&+ 3\,a_2\left[a_3^2 + b_0^2 - 2\,b_1^2 - b_1\,b_2 - 2\,b_2^2 - a_3(4\,b_0 - 2\,b_1 + b_2) + 2\,b_0(b_1 + b_2)\right]\\
&+ 3\,a_1\left[a_2^2 + a_3^2 - 2\,b_0^2 - b_0\,b_1 - 2\,b_1^2 - a_2(4\,a_3 - 2\,b_0 + b_1 - 2\,b_2)\right.\\
&\left.+ 2\,b_0\,b_2 + 2\,b_1\,b_2 + b_2^2 + 2\,a_3(b_0 + b_1 + b_2)\right].
\end{aligned}$$

这样, 诸参数对于 λ_1 的贡献就完全糊涂了. 当然, 对于 6 阶或 6 阶以上的情形, 根据伽罗瓦理论 (因 $\lambda_0 = 0$ 而多一阶), 根本不可能写出显式解. 因此，不可能指望把 λ_1 准确地算出来.

既然如此, 我们退而求其次, 即尝试估计 λ_1. 现在考虑无限矩阵. 以 $D(g)$ 表示 λ_1 的特征向量 g 的主阶 (如 g 为多项式). 下表的三个例子给出了 λ_1 和 $D(g)$ 的摄动.

$b_i\ (i \geqslant 0)$	$a_i\ (i \geqslant 1)$	λ_1	$D(g)$
$i + c\,(c > 0)$	$2i$	1	1
$i + 1$	$2i + 3$	2	2
$i + 1$	$2i + \left(4 + \sqrt{2}\right)$	3	3

表中的第一行是著名的线性模型，$\lambda_1 = 1$ 而与常数 $c > 0$ 无关, 相应的特征向量 g 是一次多项式函数．其次，保持 $b_i = i + 1$ 不变．那么，当 a_i 从 $2i$ 变到 $2i + 3$ 再变到 $2i + 4 + \sqrt{2}$ 时， λ_1 依次从 1 跳到 2 再跳到 3. 更奇妙的是特征向量依次从一次跳到二次再跳到三次多项式．至于 a_i 取值介于 $2i$, $2i + 3$ 和 $2i + 4 + \sqrt{2}$ 之间时，情况更糟，我们根本不知道 λ_1 为何值，因为此时特征向量 g 并非多项式而难于计算．这样，在一般情况下，要估计 λ_1 也是极端艰难的.

这些例子给我们形成一个印象，关于此论题似乎无事可做？！然而，世界上总有一些怪人，“明知山有虎，偏向虎山行”.

3 意义.

在给出上述问题的解答之前，我们先说明第一特征值研究的价值．众所周知，谱理论在数学各分支和物理学中都占有极重要的地位．而第一特征值正是谱的主阶，从而是一个传统的研究对象，已有百年以上的积累和无数文献．从经典的投入产出法，到目前相当时尚的随机算法的有效性和相变现象的研究，都要用到第一特征值估计．这些背景正是我们研究的原始动机．事实上，由于广泛的需求，特征值计算早已成为计算数学专门的一支，有一批成熟的算法．与此不同的是：我们这里所关心的是解析估计而非近似计算．

下面，仅以概率论中的遍历性观点来说明第一特征值的重要性，这是以概率论为工具的各种应用的一个基点．还是从两点的平凡情形谈起．此时 $Q = \begin{pmatrix} -b & b \\ a & -a \end{pmatrix}$, $a,\, b > 0$. 我们有半群 $P_t = (p_{ij}(t)) = e^{tQ} := \sum_{n=0}^{\infty} t^n Q^n/n!$. 即

$$P_t = \frac{1}{a+b}\begin{pmatrix} a + be^{-\lambda_1 t} & b\big(1-e^{-\lambda_1 t}\big) \\ a\big(1-e^{-\lambda_1 t}\big) & b + ae^{-\lambda_1 t} \end{pmatrix}.$$

这可由线性方程 $\frac{\mathrm{d}}{\mathrm{d}t}P_t = QP_t$ 及初值条件 $P_0 = I$(单位矩阵) 解出．关于 P_t 显然有如下性质：

遍历: 当 $t \to \infty$ 时， $p_{ij}(t) \to \pi_j$. 其中 $\pi_0 = \dfrac{a}{a+b}$, $\pi_1 = \dfrac{b}{a+b}$. 即当 $t \to \infty$ 时，矩阵 P_t 的每一个元素趋于一个正的极限．事实上，我们还有

指数遍历: $|p_{ij}(t) - \pi_j| \sim e^{-\lambda_1 t}$ (即以指数式速度收敛) 以及

一致遍历或强遍历: $\sup_i |p_{ij}(t) - \pi_j| \sim e^{-\lambda_1 t}$ (即收敛关于指标 i 一致).

由此可见，第一特征值对于刻画遍历性起着重要作用．事实上，可以证明 λ_1 即是指数式遍历的速度．我们指出，对于无限矩阵，λ_1 可能为零．在一般情况下，强遍历的速度小于指数式遍历的速度．又见本文的末节．

4 结果.

此处，我们仅写出无限情形的解答，有限矩阵情形更易些．定义 $\mu_0 = 1$, $\mu_i = b_0 \cdots b_{i-1}/(a_1 \cdots a_i)\,(i \geqslant 1)$. 此时需假定半群唯一：

$$\sum_{k=0}^{\infty} \frac{1}{b_k \mu_k} \sum_{i=0}^{k} \mu_i = \infty \qquad \text{以及} \qquad \mu = \sum_{i=0}^{\infty} \mu_i < \infty. \tag{4.1}$$

记 $\pi_i = \mu_i/\mu$, $L^2(\pi) = \big\{f : \pi(f^2) < \infty\big\}$, 这里及随后常用缩写 $\pi(f) = \sum_{i=0}^{\infty} \pi_i f_i$, 还有 $\pi(f^2) = \sum_{i=0}^{\infty} \pi_i f_i^2, \pi(w) = \sum_{i=0}^{\infty} \pi_i w_i$ 等等. 命 $D(f) = \sum_{i=0}^{\infty} \pi_i b_i (f_{i+1} - f_i)^2$. 作为有限矩阵特征值概念的推广，无限情形的第一特征值由下述古典变分公式

$$\lambda_1 = \inf\big\{D(f) : \pi(f) = 0, \pi\big(f^2\big) = 1\big\} \tag{4.2}$$

给出．为陈述我们的主要结果，还需引进几个记号．命

$$\begin{aligned}
\mathscr{W}'' &= \{w : w_0 = 0,\ w_i \text{ 为 } i \text{ 的严格增函数}\},\\
\mathscr{W}' &= \{w : w_0 = 0, \text{存在 } k : 1 \leqslant k \leqslant \infty \text{ 使得 } w_i = w_{\min\{i,k\}}\\
&\qquad\quad \text{且 } w \text{ 在 } [0,k] \text{ 上严格增}\},\\
I_i(w) &= \frac{1}{\mu_i b_i (w_{i+1} - w_i)} \sum_{j=i+1}^{\infty} \mu_j w_j,
\end{aligned}$$

这里本质上只有两个记号 $\mathscr{W}''$ 和 $I(w)$, $\mathscr{W}'$ 只是将 $\mathscr{W}''$ 中的函数 (数列) 从后面拉平. 记 $\bar{w}_i = w_i - \pi(w)$, $i \geq 0$. 那么，我们有如下结果：

定理 在 (4.1) 的条件下，我们有

(1) 对偶变分公式: $\inf\limits_{w\in\mathscr{W}'}\sup\limits_{i\geqslant 1} I_i(\bar{w})^{-1} = \lambda_1 = \sup\limits_{w\in\mathscr{W}''}\inf\limits_{i\geqslant 0} I_i(\bar{w})^{-1}$.

(2) 显式估计: $\mu\delta^{-1} \geqslant \lambda_1 \geqslant (4\delta)^{-1}$, 其中 $\delta = \sup\limits_{i\geqslant 1}\sum\limits_{j\leqslant i-1}(\mu_j b_j)^{-1}\sum\limits_{j\geqslant i}\mu_j$.

(3) 逼近程序: 可构造出显式序列 η_n' 和 η_n'' 使得 ${\eta_n'}^{-1} \geqslant \lambda_1 \geqslant {\eta_n''}^{-1} \geqslant (4\delta)^{-1}$.

容易看出，第一条中 λ_1 的左、右两端分别用于上、下界估计：对于每一个严格增的正数列 $w_i\,(w_0 = 0)$, 代入 (1) 式的右端，便可得出 λ_1 的一个下界估计 $\inf_{i\geqslant 0} I_i(\bar{w})^{-1}$. 这就是 "变分" 一词得含义. "对偶" 一词意指：若交换 "sup" 和 "inf", 则上、下界的表达式互换，只是 $\mathscr{W}'$ 和 $\mathscr{W}''$ 略有差别. 留意此公式与古典公式 (4.2) 完全不同，因而先前从未出现过. "显式" 一词意指表达式只依赖于系数 a_k 和 b_k. 将此定理与第 2 节中所讨论的例子作一对比，人们会不由自主地惊叹所得解答的简洁和彻底.

下界变分公式来自 [1], 最早是使用概率方法证明的. 其中的一些重要想法来自 [2]. 分析证明发表于 [3]. 定理的其余部分来自 [4] 和 [5].

5 证明.

此处，我们只证明 $\lambda_1 \geqslant \sup\limits_{f\in\mathscr{W}''}\inf\limits_{i\geqslant 0} I_i(\bar{w})^{-1}$ 以说明一种主要想法. 这个证明虽然初等，却展示了 Cauchy–Schwarz 不等式的巧妙应用.

(a) 先证对于给定的 $w \in \mathscr{W}''$, 对于一切 $i \geqslant 1$ ，有 $I_i(\bar{w}) > 0$. 等价地，对于一切 $i \geqslant 0$ ，有 $\sum_{j=i+1}^{\infty}\mu_j\bar{w}_j > 0$. 反证之，设有 i_0 使得 $\sum_{j=i_0+1}^{\infty}\mu_j\bar{w}_j \leqslant 0$. 则由 $\bar{w}_j$ 的严格单调性知 $w_{i_0} < 0$, 而且

$$0 = \sum_{j=0}^{\infty}\mu_j\bar{w}_j = \sum_{j=0}^{i_0}\mu_j\bar{w}_j + \sum_{j=i_0+1}^{\infty}\mu_j\bar{w}_j \leqslant \sum_{j=0}^{i_0}\mu_j\bar{w}_j \leqslant \bar{w}_{i_0}\sum_{j=0}^{i_0}\mu_j < 0$$

导致矛盾.

(b) 对于每 $i \geqslant 0$, 定义一条边 $e_i := \langle i, i+1\rangle$. 其次，对于每一对 $i, j(i < j)$, 定义 γ_{ij} 为由诸边 $e_i, e_{i+1}, \cdots, e_{j-1}$ 所构成的一条路. 给定 $w \in \mathscr{W}''$, 定义边上的正的权函数 $(w(e))$: $w(e_i) = w_{i+1} - w_i$, 再定义路的长度 $|\gamma_{ij}|_w = \sum_{e\in\gamma_{ij}} w(e)$. 命

$$J(w)(e) = \frac{1}{a(e)w(e)}\sum_{\{i,j\}\gamma_{ij}\ni e}|\gamma_{ij}|_w\pi_i\pi_j,$$

其中 $a(e_i) = \pi_i b_i$. 同时，也简记 $f(e_i) = f_{i+1} - f_i$.

(c) 作为 Cauchy–Schwarz 不等式的一种好的用法，我们有

$$(f_i - f_j)^2 = \Big(\sum_{e\in\gamma_{ij}} f(e)\Big)^2 = \Big(\sum_{e\in\gamma_{ij}}\frac{f(e)}{\sqrt{w(e)}}\cdot\sqrt{w(e)}\Big)^2 \leqslant \Big(\sum_{e\in\gamma_{ij}}\frac{f(e)^2}{w(e)}\Big)|\gamma_{ij}|_w.$$

这样，对于每一满足 $\pi(f)=0$ 和 $\pi(f^2)=1$ 的 f, 我们有

$$
\begin{aligned}
1 &= \frac{1}{2}\sum_{i,j}\pi_i\pi_j(f_i-f_j)^2 = \sum_{\{i,j\}}\pi_i\pi_j\bigg(\sum_{e\in\gamma_{ij}} f(e)\bigg)^2 \leqslant \sum_{\{i,j\}}\pi_i\pi_j\bigg(\sum_{e\in\gamma_{ij}}\frac{f(e)^2}{w(e)}\bigg)|\gamma_{ij}|_w \\
&= \sum_e a(e)f(e)^2\frac{1}{a(e)w(e)}\sum_{\{i,j\}:\,\gamma_{ij}\ni e}|\gamma_{ij}|_w\pi_i\pi_j \leqslant D(f)\sup_e J(w)(e),
\end{aligned}
$$

其中 $\{i,j\}$ 表示 i 和 j 的无序对. 将和式展开易证第一个等式，交换求和次序得出最后一个等号. 注意 $|\gamma_{k\ell}|_w=(w_{k+1}-w_k)+\cdots+(w_\ell-w_{\ell-1})=w_\ell-w_k$, $\ell>k$. 如果路 $\gamma_{k\ell}\,(k<\ell)$ 包含边 $e_i=\langle i,i+1\rangle$, 则 k 只能取值 $0,1,\cdots,i$ 而 ℓ 只能取值 $i+1,i+2,\cdots$. 这样，只要 $\pi(w)\geqslant 0$, 就有

$$
\begin{aligned}
&\sum_{\{k,\ell\}:\,\gamma_{k\ell}\ni e_i}|\gamma_{k\ell}|_w\pi_k\pi_\ell \\
&\qquad= \sum_{k=0}^{i}\sum_{\ell=i+1}^{\infty}\pi_k\pi_\ell(w_\ell-w_k) = \sum_{k=0}^{i}\pi_k\sum_{\ell=i+1}^{\infty}\pi_\ell w_\ell - \sum_{k=0}^{i}\pi_k w_k\sum_{\ell=i+1}^{\infty}\pi_\ell \\
&\qquad= \sum_{\ell=i+1}^{\infty}\pi_\ell w_\ell - \bigg(\sum_{k=i+1}^{\infty}\pi_k\bigg)\sum_{\ell=i+1}^{\infty}\pi_\ell w_\ell - \sum_{k=0}^{i}\pi_k w_k\sum_{\ell=i+1}^{\infty}\pi_\ell \\
&\qquad= \sum_{\ell=i+1}^{\infty}\pi_\ell w_\ell - \bigg(\sum_{k=i+1}^{\infty}\pi_k\bigg)\bigg(\sum_{\ell=0}^{\infty}\pi_\ell w_\ell\bigg) \leq \sum_{\ell=i+1}^{\infty}\pi_\ell w_\ell, \qquad i\geqslant 0.
\end{aligned}
$$

综合以上两式，只要 $\pi(w)\geq 0$, 就有

$$
\begin{aligned}
1 &\leqslant D(f)\sup_{i\geqslant 0}J(w)(e_i) \leqslant D(f)\sup_{i\geqslant 0}\frac{1}{a(e_i)w(e_i)}\sum_{j=i+1}^{\infty}\pi_j w_j \\
&= D(f)\sup_{i\geqslant 0}\frac{1}{\pi_i b_i(w_{i+1}-w_i)}\sum_{j=i+1}^{\infty}\pi_j w_j = D(f)\sup_{i\geqslant 0}\frac{1}{\mu_i b_i(w_{i+1}-w_i)}\sum_{j=i+1}^{\infty}\mu_j w_j.
\end{aligned}
$$

因为 $\pi(\bar{w})=0$, 以 $\bar{w}$ 代替 w, 此式更成立. 由此及 (a) 得

$$D(f)\geqslant \inf_{i\geqslant 0}I_i(\bar{w})^{-1}.$$

然后由条件 $\pi(f)=0$, $\pi(f^2)=1$ 及 (4.2) 式得出

$$\lambda_1\geqslant \inf_{i\geqslant 0}I_i(\bar{w})^{-1}.$$

但 $w\in\mathscr{W}''$ 任意，故有

$$\lambda_1\geqslant \sup_{w\in\mathscr{W}''}\inf_{i\geqslant 0}I_i(\bar{w})^{-1}.$$

即所欲证者.

6 进一步论题.

本文只讨论矩阵情形, 对于欧氏空间上的椭园算子的第一特征值估计, 见 [2]; 对于黎曼流形上的拉氏算子的第一特征值估计, 见 [6]. 在第 3 节中, 我们讨论了三种古典的遍历性. 三者之间有如下关系: 强遍历强于指数遍历, 指数遍历强于普通遍历. 简单例子如次.

例. 设当 i 充分大时 $a_i = b_i = i^\gamma$ $(\gamma > 0)$. 则

(1) 半群 P_t 遍历当且仅当 $\gamma > 1$.
(2) 半群 P_t 指数遍历当且仅当 $\gamma \geqslant 2$.
(3) 半群 P_t 强遍历当且仅当 $\gamma > 2$.

事实上, 还有更多的新型遍历性. 我们已完成 9 种遍历性之间的蕴涵关系图. 对于无限情形, 如上例所示, 某些遍历性可能不成立, 因而需要寻求各种遍历性成立的判别准则, 特别是如第 4 节定理的第 2 条所示的显式判别准则. 我们将在 ICM2002 的演讲中报告这些结果 [7]. 更多的文章可在作者的主页 (http://www.bnu.edu.cn/~chenmf) 和我们研究群体的主页 (http://math.bnu.edu.cn/probab/) 中找到.

参考文献

[1] Chen, M. F., Estimation of spectral gap for Markov chains, Acta Math. Sin. New Series 12:4(1996), 337-360.
[2] Chen, M. F. and Wang, F. Y., Estimation of spectral gap for elliptic operators, Trans. Amer. Math. Soc. 349:3(1997), 1239-1267.
[3] Chen, M. F., Analytic proof of dual variational formula for the first eigenvalue in dimension one, Sci. in China (A) 1999, 29:4, 327–336 (中文版); 42:8 (English Edition), 805–815.
[4] Chen, M. F., Explicit bounds of the first eigenvalue, Sci. China (A) 2000, 39:9 (中文版), 769–776; 43:10 (English Edition), 1051–1059
[5] Chen, M. F., Variational formulas and approximation theorems for the first eigenvalue, Sci. China (A) 2001, 31:1 (中文版), 28–36; 44:4 (English Edition), 409–418
[6] Chen, M. F. and Wang, F. Y., General formula for lower bound of the first eigenvalue, Sci. Sin. 1997, 27:1, 34–42 (中文版); 1997, 40:4, 384–394 (English Edition)
[7] Chen, M. F., Ergodic convergence rates of Markov processes — eigenvalues, inequalities and ergodic theory, Proceedings of ICM 2002 (Ⅲ), Higher Edu. Press 2002, 41–52