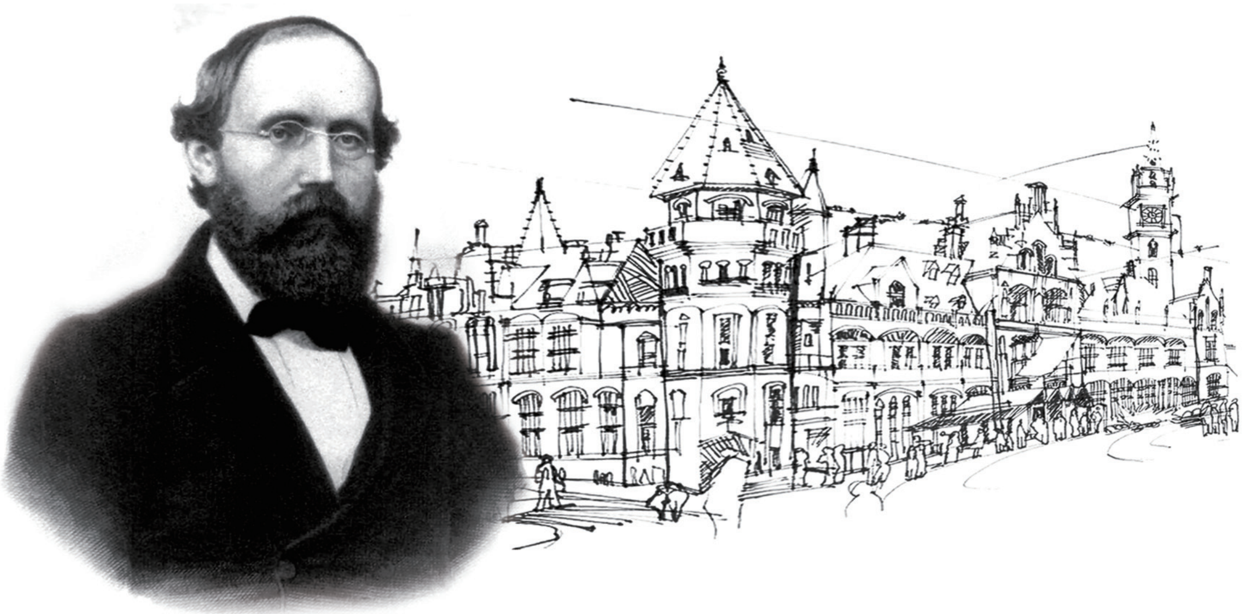
Riemann



黎曼猜想漫谈(四)

卢昌海

17 茶室邂逅

蒙哥马利（Hugh Montgomery）虽然得到了有关黎曼ζ函数非平凡零点对关联函数的猜测性结果，但这一结果究竟有何深意，对他来说还是一个谜。他觉得这个结果应该预示着什么东西，可那究竟是什么呢？他并不知道，这多少让他感到有些苦恼。

带着他的研究成果，也带着那几分苦恼，蒙哥马利于 1972 年春天飞往美国圣路易（St. Louis）参加一个解析数论会议。那趟旅行对蒙哥马利有着一举数得的意义。除会议本身外，他还到密歇根大学（University of Michigan）所在地安娜堡（Ann Arbor）买了房子，因为此前不久他已接受了一份密歇根大学的工作（蒙哥马利目前仍在密歇根大学数学系）。

至此，那趟旅行可以说已经获得了精神与物质的双重丰收。但在结束旅程前蒙哥马利还有一件事情放心不下。

我们在第三节曾经提到高斯有一个“坏毛病”，那就是常常不发表自己的工作，结果使得同时代的许多数学家在研究课题上与他“撞车”（与高斯这样的大师玩碰碰车，谁的脑袋先碰破就不必说了）。无独有偶，二十世纪的普林斯顿高等研究所也出了一位有同样“坏毛病”的数学家，那便是阿特勒 • 塞尔伯格（Atle Selberg，1917-2007）。塞尔伯格在黎曼猜想的研究中有着极为重要的地位，我们在后文中将会更多地介绍他，这里就不赘述了。让蒙哥马利放心不下的就是自己会不会与塞尔伯格“撞

Riemann

普林斯顿高等研究所 Fuld Hall（刘建亚 摄）

车”？自己的这项研究工作会不会不幸地在塞尔伯格的某一叠草稿纸上已经有了？当然，除此之外他也很想听听这位黎曼猜想研究中的顶尖高手对自己这项工作的看法，特别是对结果背后含义的理解。

于是在返回英国前他决定在普林斯顿高等研究所做短暂的停留，以便会见一下塞尔伯格。

蒙哥马利如愿见到了塞尔伯格。但塞尔伯格听完了蒙哥马利的介绍只是礼貌地表示了兴趣，却没有提出具体意见。不过他总算也没有说："干得不错，小伙子，但是 N 年之前我已经证明过这样的结果了"，还是让蒙哥马利松了一口气。

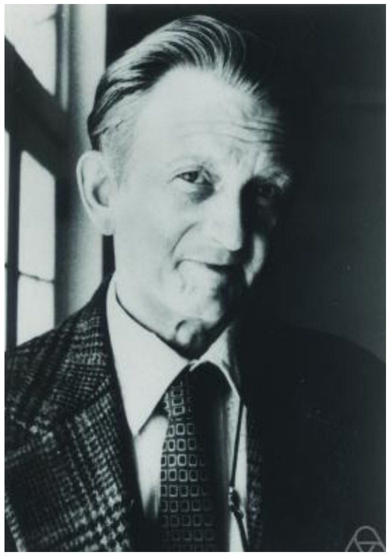
见过了塞尔伯格，蒙哥马利便和朋友周拉（Sarvadaman Chowla，1907-1995）到 Fuld Hall 去喝下午茶。喝下午茶虽是一种休闲，但在普林斯顿高等研究所的学术氛围中却是一个重要的组成部份。在这一时间里，来自世界各地、从事不同研究的学者们互相攀谈，交流看法，往往会撞击出一些意想不到的智慧火花。

蒙哥马利和周拉正在喝茶闲聊的时候，一位物理学家走了进来。

在普林斯顿高等研究所这样一个科学家阵容豪华得近乎奢侈的地方，随便哪个角落碰上的都可能是非同小可的人物。这位漫步走进茶室的物理学家也不例外。此人在二十世纪中叶曾因证明了量子电动力学的几种形式体系彼此等价，而获得了很高的声誉，也为他赢得了普林斯顿高等研究所的终生职位。而这项研究还只不过是他科学生涯中许许多多研究中的一个。他的研究涉及到核物理、凝聚态物理、天体物理，乃至天体生物学等诸多领域。这位物理学家便是弗里曼•戴森（Freeman Dyson，1923-）。在二十世纪物理殿堂的璀璨群星中戴森当然远不是最杰出的，但那个午后他和蒙哥马利的世界线在高等研究所的短暂交汇，却是科学史上一段令人难忘的佳话，对于黎曼猜想的研究来说也是一个奇峰突起的精彩篇章。

周拉是一位交际高手，一边和蒙哥马利喝茶聊天，一边仍能眼观六路、耳听八方。戴森刚一进门就被他发现了，于是他问蒙哥马利是否见过戴森，蒙哥马利说没有，周拉就说我给你引见一下。蒙哥

蒙特哥麦利
Hugh Montgomery



塞尔伯格
Atle Selberg

戴森
Freeman Dyson

马利心想自己做的东西和戴森八杆子都打不着，再说喝完茶就走人了，何必还特意打扰戴森？就说不必了。但周拉却是一个从来不把“不”字当成答案的家伙，当下二话不说就把蒙哥马利拽到了戴森跟前（谢谢周拉！）。

就这样戴森和蒙哥马利攀谈了起来。戴森问蒙哥马利最近在研究什么？蒙哥马利就把自己对黎曼 ζ 函数非平凡零点分布的研究叙述了一下。戴森礼貌地听着，他对这一领域并不熟悉。连塞尔伯格都没有发表具体的看法，蒙哥马利也并不指望这番泛泛介绍会得到比礼貌地点点头更多的回应。

但是当他介绍到自己所猜测的密度函数 $\rho(t) = 1-[\sin(\pi t)/\pi t]^2$（详见第十六节）时，戴森的眼睛猛地睁大了！

因为这个让蒙哥马利找不到北，甚至连塞尔伯格也看不出端倪来的密度函数对戴森来说却一点也不陌生，那正是随机厄密特矩阵（Random Hermitian matrices）本征值的对关联函数。物理学家们研究这类东西已经有二十年了！

而戴森本人也早在十年前就系统地研究了随机矩阵理论，是这一领域公认的先驱者之一。即使找遍整个世界，也不可能找到一个比戴森更合适的人来和蒙哥马利共喝那杯下午茶了。他们的相遇本身就是一个幸运的奇迹。

有意思的是，在与蒙哥马利的这次“茶室邂逅”的前一年（即 1972 年），戴森刚写过一篇题为“Missed Opportunity”（“错过的机会”）的文章，叙述了科学史上由于数学家与物理学家交流不够而错失发现的一些事例。

18 随机矩阵理论

身为理论物理学家的戴森如何会研究起随机矩阵理论来的呢？这当然还得从物理学说起。

我们知道在物理学上可以严格求解的问题是少之又少的。而且物理理论越发展，可以严格求解的问题就越少。举个例子来说，在牛顿引力理论中二体问题可以严格求解，但一般的三体问题就不行[注 18.1]；到了广义相对论中连一般的二体问题也解

注 18.1

这里“单体”、“二体”、“三体”指的都是点状分布或可视为点状分布的体系。

不出了，只有单体问题还可以严格求解；而到了量子场论中更是连单体问题也解不成了。

另一方面，现实物理中的体系往往既不是单体，也不是二体或三体，而是多体，少则十几、几十（比如大一点的原子、分子），多则 10^{23} 或更多（比如宏观体系）。很明显，对现实物理体系的研究离不开各种近似方法。这其中很重要的一类方法就是统计方法，由此形成了物理学的一个重要分支：统计物理。

在统计物理中，人们不再着眼于对物理体系的微观状态进行细致描述（因为这种细致描述不仅无法做到，而且对于确定体系的宏观行为来说是完全不必要的），取而代之的是"系综"的概念。所谓"系综"，指的是满足一定宏观约束条件的大量全同体系的集合，这些体系的微观状态具有一定的统计分布，我们感兴趣的体系的宏观状态就由相应物理量的系综平均值所给出。

在传统的统计物理中，组成系综的那些全同体系具有相同的哈密顿量（Hamiltonian），只有它们的微观状态才是随机的。但是随着研究的深入，物理学家们开始接触到一些连这种方法也无法处理的物理体系，其中一个典型的例子就是由大量质子中子组成的原子核。这种体系的相互作用具备了所有可以想象得到的"坏品质"（比如耦合常数很大，不是二体相互作用，不是有心相互作用等），简直是"五毒俱全"。对于这种体系，我们甚至连它的哈密顿量是什么都无法确定。这样的体系该如何处理呢？很显然还是离不开统计的方法。只不过以前在系综中只有各体系的微观状态是随机的，现在却连哈密顿量也不知道了，既然如此，那就一不做二不休，干脆把哈密顿量也一并随机化了。由于哈密顿量可以用矩阵来表示，因此这种带有随机哈密顿量的量子统计系综可以用随机矩阵理论来描述。这一点最早是由尤金·维格纳（Eugene Wigner，1902-1995）于1951 年提出的。当然，在这一领域中数学家还是要先于物理学家。随机矩阵理论在数学中最早是由威沙特（J. Wishart，1898-1956）于 1928 年提出的。

把哈密顿量随机化不等于说对哈密顿量的结构就没有任何限制了。二十世纪六十年代初，与蒙哥马利在茶室里偶遇的这位戴森对随机矩阵理论进行了深入的研究，并在 1962 年一连发表了五篇非常漂亮的论文。这些论文在随机矩阵理论的发展中具有奠基性的作用。在这些论文中戴森证明了随机矩阵理论可以按照体系在时间反演变换 T 下的性质分为三种类型：

- 如果体系不具有时间反演不变性，则演化算符为幺正矩阵（Unitary Matrices）。
- 如果体系具有时间反演不变性，且 $T^2=I$，则演化算符为正交矩阵（Orthogonal Matrices）。
- 如果体系具有时间反演不变性，且 $T^2=-I$，则演化算符为辛矩阵（Symplectic Matrices）。

这里戴森用演化算符 U 取代了哈密顿量 H，这两者之间由 $U=\exp(-iHt)$ 相联系。用演化算符的好处是它的参数空间是紧致的。

除了按照对称性对演化算符的结构进行分类外，还有一个需要解决的问题就是哈密顿量的分布函数。戴森引进的是高斯型分布，这是数学物理中比较常见的一种分布。在这种分布下具有上述三种对称性的系综分别被称为：Gaussian Unitary Ensemble (GUE)，Gaussian Orthogonal Ensemble (GOE) 和 Gaussian Symplectic Ensemble (GSE)。

戴森在得知了蒙哥马利的密度函数时猛然想起的"随机厄密矩阵"所描述的正是这三种系综中的一种——Gaussian Unitary Ensemble——的哈密顿量（幺正演化算符对应的哈密顿量是厄密的），它的几率测度定义为高斯型分布：

$$P(H)dH = C\exp\left(-\frac{tr(H^2)}{2\sigma^2}\right)dH,$$

其中 C 为归一化常数，H 为体系的哈密顿量，σ 为标准差，通常取为 $2^{-1/2}$。

对于一个量子体系，能级分布是在理论与观测上都极其重要的性质。这也是随机矩阵理论中物理学家们最感兴趣的东西之一。物理学家所说的能级用数学术语来说就是哈密顿量的本征值。那么随机厄密矩阵的本征值是怎样分布的呢？分析表明，一个 N 阶随机厄密矩阵的本征值分布密度为：

$$P(\lambda_1,\cdots,\lambda_N)=C\exp\left(-\sum_i \lambda_i^2\right)\prod_{j>k}(\lambda_j-\lambda_k)^2,$$

其中 $\lambda_1, \ldots, \lambda_N$ 为本征值，C 为归一化常数。

通过对这一分布密度的积分，我们可以计算出随机厄密矩阵本征值的各种关联函数。但是这些关联函数的表观复杂程度与本征值的平均间距有很大关系，因此我们要先对本征值做一点处理，以便简化结果。这一处理所依据的是维格纳曾经证明过的一个结果，那就是当矩阵阶数 $N\to\infty$ 时，n 阶随机厄密矩阵的本征值趋向于区间 $[-2(2n)^{1/2},\ 2(2n)^{1/2}]$ 上的半圆状分布，即

$$P(\lambda)d\lambda=\sqrt{8n-\lambda^2}\,\frac{d\lambda}{4\pi},$$

其中 $P(\lambda)d\lambda$ 为区间 $(\lambda,\ \lambda+d\lambda)$ 上的本征值个数。这一规律被称为维格纳半圆律（Wigner Semicircle Law）。利用这一规律，我们可以对本征值做一个标度变换，引进

$$\mu=\frac{\lambda\sqrt{8n-\lambda^2}}{4\pi},$$

可以证明（请读者自己证明），这一变换就象我们在第十六节中对黎曼 ζ 函数零点虚部所做的处理那样，将本征值的间距归一化为：$\Delta\mu\sim1$。在这种间距归一化的本征值下，关联函数的形式变得相对简单，其中对关联函数的计算结果为：

$$P_2(\mu_1,\mu_2)=1-\left(\frac{\sin(\pi|\mu_2-\mu_1|)}{\pi|\mu_2-\mu_1|}\right)^2.$$

看到这里，大家想必也和戴森一样看出来了，随机厄密矩阵本征值的对关联函数正是我们在 第十六节 中介绍过的，蒙哥马利所猜测的黎曼 ζ 函数非平凡零点的对关联函数！当然那时候蒙哥马利用的不是象“对关联函数”这样摩登的术语，事实上“对关联函数”这一术语蒙哥马利在和戴森交谈前连听都没听说过，他自己用的是象“我正在研究零点间距”这样土得掉渣的“白话文”。

有的读者可能会提出这样一个问题，那就是哈密顿量的分布为什么要选择成高斯型分布？对于这个问题，实用主义的回答是：高斯型分布是数学上比较容易处理的（不要小看这样的理由，当问题复杂到一定程度时这种理由有时是最具压倒性的）；稍为深刻一点的回答则是：高斯型分布在固定的 |H|2 系综平均值及标准差下具有最大的熵，换句话说它描述的是在一定约束下具有最大随机性的体系；但是最深刻的回答却是：我们其实并不需要特意选择高斯型分布！ 随机矩阵理论的一个非常引人注目的特点便是：在矩阵阶数 $N\to\infty$ 的极限下它的本征值分布具有普适性（即不依赖于哈密顿量的特定分布）。正是这种普适性使得随机矩阵理论在从复杂量子体系的能级分布到无序介质中的波动现象，从神经网络系统到量子混沌，从 $N_c\to\infty$ 的 QCD 到二维量子引力的极为广阔的领域中都得到了应用。

但是把随机矩阵理论的所有这些不同尺度、不同维度的应用加在一起，也比不上它与黎曼 ζ 函数非平凡零点分布之间的关联来得神奇。蒙哥马利曾经为不知道自己的结果预示着什么而苦恼，现在他知道了那样的结果也出现在由随机矩阵理论所描述的一系列物理现象中。

但这与其说是解惑，不如说是一种更大的困惑。象黎曼 ζ 函数非平凡零点分布这样最纯粹的数学性质，怎么会与象复杂量子体系、无序介质那样最现实的物理现象扯上关系的呢？这种神奇的关联本身又预示着什么呢？

19 Montgomery-Odlyzko 定律

蒙哥马利关于黎曼 ζ 函数非平凡零点分布的论文于 1973 年发表在美国数学学会的 Proc. Sym. Pure Math. 上。但最初几年里它并不曾吸引多少眼球，因为无论这种存在于零点分布与随机矩阵理论间的关联有多奇妙，在当时它还只是一个纯粹的猜测，既没有严格的数学证明，也没有直接的数值证据。我们曾在第 十三、十四 节中介绍过零点计算的简史。在蒙哥马利的论文发表之初，人们对零点的计算还只进行到几百万个，而且——如我们在第十五节中所说——那些计算大都只是验证了“前 N 个零点”位于 critical line 上，却不曾涉及零点的

Riemann

欧德里考
Andrew Michael Odlyzko

具体数值。既然没有具体数值，自然就无法用来检验蒙哥马利的对关联假设。更何况——如我们在第十六节中所说——为了检验后者，我们需要研究虚部很大的零点，这显然也是当时的计算所远远不及的。因此当时就连蒙哥马利自己也觉得对他的猜测进行数值验证将是极为遥远的将来的事情。

但是蒙哥马利和我们在第十四节中提到的输掉葡萄酒的查基尔（Don Zagier，1951-）一样大大低估了计算机领域的发展速度。在他的论文发表五年之后的一天，他又来到了普林斯顿。不过这次不是为了觐见塞尔伯格，而是来做一个有关黎曼 ζ 函数零点分布的演讲。在那次演讲的听众中有一位来自 32 英里外的贝尔实验室（Bell Labs）Murray Hill 研究中心的年轻人，他被蒙哥马利讲述的零点分布与随机矩阵理论间的关联深深地吸引住了。而他所在的实验室恰好拥有当时著名的 Cray 巨型计算机。这位年轻人便是我们在第十六节中提到的欧德里考（Andrew Michael Odlyzko，1949-）。

普林斯顿真是蒙哥马利的福地，五年前与戴森在这里的相遇，使他了解到了零点分布与随机矩阵理论间的神秘关联，从而为他的研究注入了一种奇异的魅力。五年后又是在这里，这种魅力打动了欧德里考，从而有了我们在第十六节中介绍的他对黎曼 ζ 函数零点的大规模计算分析。这些计算为蒙哥马利所猜测的零点分布与随机矩阵理论间的关联提供了大量的数值证据[注 19.1]。这种关联，即经过适当的归一化后黎曼 ζ 函数非平凡零点的间距分布与 Gaussian Unitary Ensemble（参阅第十八节）的本征值间距分布相同，也因此渐渐被人们称为蒙哥马利 - 欧德里考定律（Montgomery-Odlyzko Law）[注 19.2]。蒙哥马利 - 欧德里考定律虽然是用 Gaussian Unitary Ensemble 来表述的，但我们在第十八节 中曾经提到，随机矩阵理论的本征值分布在矩阵阶数 N → ∞时具有普适性。因此蒙哥马利 - 欧德里考 定律所给出的关联并不限于 Gaussian Unitary Ensemble。不仅如此，这种本征值分布的普适性还有一层含义，那就是它不仅在各种系综下都相同，而且对系综中任何一个典型的系统——即任何一个典型的随机矩阵——都相同。换句话说，我们不仅不需要指定系综的分布函数，甚至连系综本身都不需要，只要随便取出一个随机矩阵就可以了[注 19.3]。因此蒙哥马利 - 欧德里考 定律实际上意味着**黎曼 ζ 函数非平凡零点的分布可以用任何一个典型随机厄密矩阵的本征值分布来描述**。

蒙哥马利当初的研究——如我们在第十六节中介绍的——只涉及零点分布的对关联函数。在他之后，人们对零点分布的高阶关联函数也作了研

注 19.1

这种数值证据之一便是我们在第十六节中给出的关于 Montgomery 零点对关联函数的拟合曲线。

注 19.2

这“定律”二字通常在物理学中用得比在数学中多，它很贴切地表达了这一命题虽有大量的数值证据，却缺乏数学意义上的严格证明这一特点。

注 19.3

当然，别忘了 $N\to\infty$，以及矩阵为幺正（对演化算符而言）或厄密（对哈密顿量而言）这些条件。

究。1996 年，Z. Rudnick 与 P. Sarnak 及 E. B. Bogomolny 与 J. P. Keating 分别“证明”了零点分布的高阶关联函数也与相应的随机矩阵的本征值关联函数相同。美中不足的是，我们不得不对这种“证明”加上引号，因为它们和蒙哥马利的研究一样，并不是真正严格的证明，它们或是引进了额外的限制条件（如 Z. Rudnick 与 P. Sarnak 的研究），或是运用了本身尚未得到证明的黎曼猜想及强孪生素数猜想（如 E. B. Bogomolny 与 J. P. Keating 的研究）。

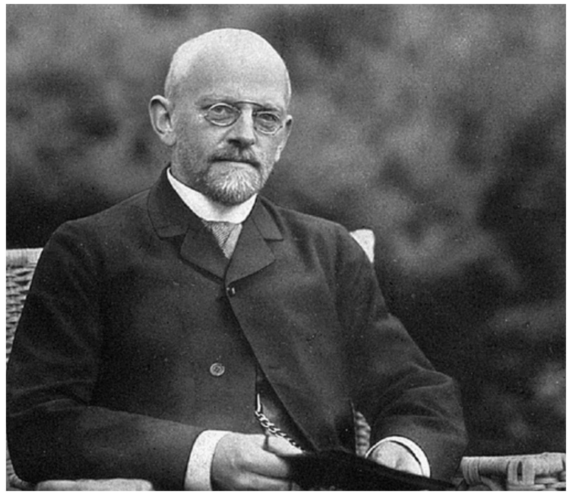
但即便如此，所有这些理论及计算的结果还是非常清楚地显示出黎曼 ζ 函数非平凡零点的分布与随机矩阵的本征值分布 —— 从而与由随机矩阵理论所描述的一系列复杂物理体系的性质 —— 间的确存在着令人瞩目的关联。蒙哥马利 - 欧德里考定律在“经验”意义上的成立几乎已是一个毋庸置疑的事实。

20 希尔伯特 – 波利亚猜想

那么在黎曼 ζ 函数非平凡零点这样的纯数学客体与由随机矩阵理论所描述的纯物理现象之间为什么会出现象蒙哥马利 - 欧德里考定律那样的关联呢？这却是一个我们至今也未能完全理解的谜团。但有意思的是，虽然在距离蒙哥马利的论文发表已有三十余年的今天我们仍未能彻底理解蒙哥马利 - 欧德里考定律的本质，可是远在蒙哥马利的论文发表六十余年前的二十世纪一二十年代，数学界就流传着一个与蒙哥马利 - 欧德里考定律极有渊源的猜想 —— 希尔伯特 - 波利亚猜想：

> **希尔伯特 – 波利亚猜想**：黎曼 ζ 函数的非平凡零点与某个厄密算符的本征值相对应。

更确切地讲，希尔伯特 - 波利亚猜想指的是：如果把黎曼 ζ 函数的非平凡零点写成 $\rho = 1/2 + it$，则所有这些 t 与某个厄密算符的本征值一一对应（自第十一节引进 t 以来，当我们提到黎曼 ζ 函数的非平凡零点时往往指的是 t，这一点读者应该很容易从上下文中判断出来）。我们知道，厄密算符的本征值都是实数。因此如果所有的 t 都与某个厄密



希尔伯特
David Hilbert

算符的本征值相对应，则它们必定全都是实数，从而所有非平凡零点 $\rho = 1/2 + it$ 的实部都等于 1/2，这正是黎曼猜想的内容。因此如果希尔伯特 - 波利亚猜想成立，则黎曼猜想也必定成立。

我们在上节中提到，蒙哥马利 - 欧德里考定律表明黎曼 ζ 函数非平凡零点的分布可以用任何一个典型随机厄密矩阵的本征值分布来描述。这种描述虽然奇妙，终究只是统计意义上的描述。但是如果希尔伯特 - 波利亚猜想成立，则黎曼 ζ 函数的非平凡零点干脆直接与某个厄密矩阵的本征值一一对应了。这是严格意义上的对应，有了这种对应，统计意义上的对应自然就不在话下。因此希尔伯特 - 波利亚猜想虽然比蒙哥马利 - 欧德里考定律早了六十余年，却是一个比蒙哥马利 - 欧德里考定律更强的命题！

从二十世纪初开始流传的希尔伯特 - 波利亚猜想在无形之中与半个多世纪后才出现蒙哥马利 - 欧德里考定律做了跨越时间的遥远呼应。

但是这一呼应委实是太过遥远了，蒙哥马利的论文尚且因为缺乏证据而遭冷场，希尔伯特 - 波利亚猜想就更无人问津了。这种冷落是如此地彻底，以至于当蒙哥马利的论文及后续研究重新燃起人们对希尔伯特 - 波利亚猜想的兴趣，从而开始追溯它的起源时，大家惊讶地发现不仅希尔伯特（David Hilbert，1862-1943） 和 乔 治 • 波 利 亚（George

Riemann

波利亚
George Pólya

Pólya，1887-1985）不曾在人们找寻得到的任何发表物或手稿中留下过一丝一毫有关 希尔伯特 - 波利亚猜想的内容。而且在蒙哥马利之前所有其他人的文字中竟也找不到任何与这一猜想相关的叙述。一个隐约流传了大半个世纪的数学猜想竟似没有落下半点文字记录，却一直流传了下来，真是一个奇迹！

1981 年 12 月 8 日，欧德里考给波利亚发去了一封信，询问希尔伯特 - 波利亚猜想的来龙去脉。当时波利亚已是九十四岁高龄，卧病在床，基本不再执笔回复任何信件，但欧德里考的信却很及时地得到了他的亲笔回复。毕竟，对一位数学家来说，自己的名字能够与伟大的希尔伯特出现在同一个猜想中是一种无上的荣耀。波利亚在回信中写道：

> 很感谢你 12 月 8 日的来信。我只能叙述一下自己的经历。
>
> 1914 年初之前的两年里我在哥廷根。我打算向兰道学习解析数论。有一天他问我："你学过一些物理，你知道任何物理上的原因使黎曼猜想必须成立吗？" 我回答说，如果 ξ - 函数的非平凡零点与某个物理问题存在这样一种关联，使得黎曼猜想等价于该物理问题中所有本征值都是实数这一事实，那么黎曼猜想就必须成立。

s 波利亚提到的 ξ- 函数应该是指我们在第五节的 [注 5.1] 中提到的黎曼本人所定义的 ξ 函数。黎曼猜想等价于那个 ξ 函数的零点为实数。

三年后波利亚离开了人世，他的这封信便成了迄今所知有关希尔伯特 - 波利亚猜想的唯一文字记录。至于早已过世的希尔伯特在什么场合下提出过类似的想法，则也许将成为数学史上一个永远的谜团了。

21 黎曼体系何处觅？

如上所述，假如希尔伯特 - 波利亚猜想成立，则黎曼 ζ 函数的非平凡零点将与某个厄密算符的本征值一一对应。我们知道厄密算符可以用来表示量子力学体系的哈密顿量，而厄密算符的本征值则对应于该量子力学体系的能级。因此如果希尔伯特 - 波利亚猜想成立，则黎曼 ζ 函数的非平凡零点有可能对应于某个量子力学体系的能级，非平凡零点的全体则对应于该量子力学体系的能谱。我们把这一特殊的量子力学体系称为黎曼体系，把这一体系的哈密顿量称为黎曼算符。

严格讲，量子力学中所有的可观测量都是由厄密算符表示的，哈密顿量只是其中之一。不仅如此，由厄密算符的本征值所描述的物理量甚至并不限于量子力学中的物理量。从波利亚给欧德里考的信中也可以看到，波利亚当年并没有对与黎曼 ζ 函数非平凡零点相对应的"物理问题"做具体的猜测。因此从希尔伯特 - 波利亚猜想到黎曼 ζ 体系是后人所做的进一步猜测。之所以做这种猜测，除了哈密顿量对物理体系所具有的重要性外，或许是因为随机

注 21.1

Bohigas 猜想的原始表述是只针对 Gaussian Orthogonal Ensemble 的。

Riemann

矩阵理论最初是在研究原子核能级时被引入物理学中的。另一方面，量子体系的能级是自然界中含义最为深刻的离散现象之一，这或许也是人们把注意力集中到这一方向上的原因之一。

那么这个黎曼体系——如果存在的话——会是一个什么样的量子力学体系呢？

有关这个问题最重要的线索显然来自蒙哥马利-欧德里考定律。由于这一定律表明黎曼 ζ 函数的非平凡零点分布与随机厄密矩阵的本征值分布相同，因此我们不难猜测，黎曼算符是一个随机厄密矩阵。那么由随机厄密矩阵所描述的量子力学体系具有什么特点呢？这个问题自二十世纪七十年代末以来有许多人研究过。1984 年，O. Bohigas 提出了一个猜想，即由随机矩阵理论描述的量子体系在经典近似下对应于经典混沌体系[注 21.1]。这一猜想已经有了许多数值计算的支持，但至今仍未得到严格的证明。不过从物理角度上讲，与经典混沌体系相对应的量子体系的波函数会在一定程度上秉承经典轨迹的混沌性，从而使得哈密顿量的矩阵元呈现随机性，这正是随机矩阵的特点。

由此看来黎曼体系很可能是一个与经典混沌体系相对应的量子体系。那么这个与黎曼体系相对应的经典混沌体系又会具有什么样的特征呢？这个问题人们也做过一些研究。由于我们所知有关黎曼体系最明确的信息是黎曼 ζ 函数的非平凡零点，即黎曼体系的能谱。因此寻找黎曼体系的努力显然要从能谱入手。描述量子体系能谱的一个很有用的工具是所谓的能级密度函数：

$$\rho(E)=\sum_n \delta(E-E_n)\ .$$

早在二十世纪六十年末和七十年代初，M. C. Gutzwiller 就对这一能级密度函数的经典极限做了研究，得到了一个我们现在称为 Gutzwiller 求迹公式（Gutzwiller Trace Formula）的结果。在对应的经典体系具有混沌性的情形下，Gutzwiller 求迹公式为：

$$\rho(E)=\bar{\rho}(E)+2\sum_p\sum_k A_{p,k}\cos\left(\frac{2\pi k S_p}{h}+\alpha_p\right)\ ,$$

其中 h 为普朗克常数，$\bar{\rho}(E)$ 是一个平均密度。我们感兴趣的是第二项，它包含一个对经典极限下所有闭合轨道 p 及正整数 k（沿闭合轨道的绕转数）的双重求和。求和式中的 S_p 是闭合轨道 p 的作用量，α_p 是一个被称为 Maslov phase 的相位。而 $A_{p,k}$ 与闭合轨道的性质有关，可以表示为：

$$A_{p,k}=\frac{T_p}{h\sqrt{\det\left(M_p^k-1\right)}}$$

其中 T_p 是闭合轨道 p 的周期，M_p 则是描述闭合轨道 p 稳定性的单值矩阵（monodromy matrix）。

另一方面，我们也可以计算黎曼 ζ 函数非平凡零点的密度函数：

$$\rho(t)=\sum_n \delta(t-t_n)\ .$$

1985 年，M. V. Berry 给出了这一计算的结果：

$$\rho(t)=\bar{\rho}(t)-2\sum_p\sum_k \frac{\ln(p)}{2\pi}\exp\left(-\frac{k\ln(p)}{2}\right)\cos\left[kt\ln(p)\right]$$

要注意的是，这里的 p 是素数而非一般的自然数！将这个结果与前面有关量子体系能级密度的计算相比较，我们发现为使两者一致，必须：

$$\alpha_p=\pi,\quad T_p=\ln(p),\quad S_p=\frac{ht}{2\pi}T_p\ ,$$

$$A_{p,k}=\frac{T_p}{2\pi\exp\left(kT_p/2\right)}$$

这其中最简洁而漂亮的关系式就是 $T_p=\ln(p)$，它表明与黎曼体系相对应的经典体系具有周期等于素数对数 $\ln(p)$ 的闭合轨道！ 这无疑是这一体系最奇异的特征之一。

研究黎曼体系的努力仍在继续着，在一些数学物理学家心目中，它甚至已经成为了一种证明黎曼猜想的新的努力方向，即物理证明。会不会有一天人们在宇宙的某个角落里发现一个奇特的物理体系，它的经典基本周期恰好是 ln2, ln3, ln5, ... ？或者它的量子能谱恰好包含 14.1347251, 21.0220396, 25.0108575, ... ？我们不知道。也许并不存在这样的体系，但如果存在的话，它无疑将是大自然最美丽的奇迹之一。只要想到象素数和黎曼 ζ 函数非平凡零点这样纯粹的数学元素竟有可能出现在物理的天空里，变成优美的轨道和绚丽的光谱线，我们就

玻尔
Harald August Bohr

兰道
Edmund G. H. Landau

不能不惊叹于数学与物理的神奇，惊叹于大自然的无穷造化。而这一切，正是科学的伟大魅力所在。

22 玻尔－兰道定理

在前面的 21 节里，我们介绍了黎曼 ζ 函数的定义及其零点，介绍了它们与素数分布之间的关联，也介绍了黎曼 ζ 函数非平凡零点的计算。沿着零点计算这一方向，我们介绍了人们对零点分布的统计研究，以及由此而发现的零点分布与物理之间出人意料的关联。这无疑是整个旅程中最令人惊叹的风景，事实上也正是这一段风景使我萌生了写作这一系列的念头，从而使得整个旅程成为可能。

看过了这些风景，现在让我们重新回到纯数学的领地中来。从纯数学的角度讲，对一个数学猜想最直接的研究莫过于是寻求它的证明（或否证）。可惜的是，黎曼猜想直到今天也还没有一个得到数学界公认的证明（或否证）。因此我们所能介绍的只是数学家们试图逼近黎曼猜想——或者说逼近 critical line——的过程。

在前面各节中，我们曾经介绍过两个具有普遍意义的零点分布结果。一个是第五节中提到的：黎曼 ζ 函数的所有非平凡零点都位于复平面上 $0 \le \mathrm{Re}(s) \le 1$ 的区域内。这是欧拉乘积公式 的一个简单推论。另一个则是第七节中提到的：黎曼 ζ 函数的所有非平凡零点都位于复平面上 $0 < \mathrm{Re}(s) < 1$ 的区域（即 critical strip）内。这是在证明素数定理的过程中由阿达马（Jacques Hadamard，1865-1963）与 de la Vallée-Poussin 所证明的，比前面的结果略进一步，时间是 1896 年。这两个结果与黎曼猜想虽然还相距很远，但它们是普遍而严格的结果，适用于所有的零点，在这一点上远远胜过有关

注 22.1

在 1914 年之前也有过一些值得一提的结果，比较著名的一个是 Ernst Lindelöf (1870-1946) 于 1908 年提出的有关虚部 t 趋于无穷时 $|\zeta(\sigma+it)|$ 渐进行为的猜想——Lindelöf 猜想。1918 年 Backlund 证明了 Lindelöf 猜想等价于这样一个命题：黎曼 ζ 函数在复平面上 $\{1/2<\sigma\le\mathrm{Re}(s)\le1, T\le t\le T+1\}$ 的非平凡零点数目为 $N(\sigma,T)=o(\ln T)$。读者们可以对比第五节中黎曼三个命题中的第一个来思考一下这一猜想的含义。不过 Lindelöf 猜想虽然远比黎曼猜想弱，其证明却困难得出乎意料，直到今天也还只是一个猜想 (1998 年 N. V. Kuznetsov 曾提出过一个长达 89 页的证明，但后来被发现是错误的)，因此我们只在这里简略地提一下。

零点的数值计算。

在阿达马与 de la Vallée- Poussin 之后的第十八个年头，即 1914 年，数学家们在零点分布的研究上又取得了两个重大进展[注 22.1]。取得这两个重大进展的数学家正是我们在旅程伊始提到的哈代（Godfrey Harold Hardy，1877-1947），玻尔（Harald August Bohr，1887-1951）和兰道（Edmund Georg Hermann Landau，1877-1938）。在本节中我们先来介绍玻尔与兰道的工作，即玻尔 - 兰道定理。

但是在介绍玻尔 - 兰道定理之前，让我们先对零点分布的基本对称性做一个简单分析。我们在 [注 8.1] 中曾经提到黎曼 ζ 函数在上半复平面与下半复平面的非平凡零点是一一对应的。具体地讲，这种一一对应是通过以 s=1/2（即实轴与 critical line 的交汇点）为原点的反演对称性实现的。这种对应性来源于零点与黎曼 ζ 函数非平凡零点重合的函数 $\xi(s)$ 所满足的关系式 $\xi(s)=\xi(1-s)$（参阅第五节）。除了这一反演对称性外，黎曼 ζ 函数非平凡零点的分布还满足一个对称性，那就是关于实轴的反射对称性。这是由于 $\xi(s)$ 除满足 $\xi(s)=\xi(1-s)$ 外，还满足一个关系式：$\overline{\xi(s)}=\xi(\bar{s})$（请读者自行证明）。由这两个对称性可知黎曼 ζ 函数非平凡零点的分布相对于 critical line 也具有反射对称性。这些对称性的存在表明：要研究零点的分布，只需研究 critical strip 的四分之一，即 $\{\mathrm{Re}(s)\geq 1/2, \mathrm{Im}(s)\geq 0\}$ 的区域就行了。我们以前介绍过的零点计算就是针对这一区域的，下面要介绍的玻尔 - 兰道定理的表述也是如此。

玻尔与兰道所证明的是这样一个定理：

玻尔 - 兰道定理：如果 $|\zeta(s)|^2$ 在直线 $\mathrm{Re}(s)=\sigma$ 上的平均值对 $\sigma>1/2$ 有界，且对 $\sigma\geq\sigma_0>1/2$ 一致有界，则对于任何 $\delta>0$，位于 $\mathrm{Re}(s)\geq 1/2+\delta$ 的非平凡零点所占的比例为无穷小。

注 22.2

玻尔与兰道实际证明的结果比这更具体，他们证明了对于任何 $\delta>0$，位于 $\{\mathrm{Re}(s)\geqslant 1/2+\delta, 0\leqslant t\leqslant T\}$ 的非平凡零点的数目不超过 KT(从而所占的比例为无穷小——请读者思考这是为什么？)。

这里我们所用的表述与玻尔与兰道所用的略有差异。他们的表述是针对 $(1-2^{1-s})\,\zeta(s)$ 的平均值而给出的。在进一步讨论之前，我们先来解释或定义一下定理中所涉及的一些术语。首先解释一下什么叫做"$|\zeta(s)|^2$ 在直线 $\mathrm{Re}(s)=\sigma$ 上的平均值"。这个平均值是由

$$\lim_{T\to\infty}\frac{1}{T-1}\int_1^T\left|\zeta(\sigma+it)\right|^2 dt$$

来定义的。这个定义与函数平均值的普遍定义——即函数在区间上的积分除以区间的长度——是完全一致的。只不过由于 $\mathrm{Re}(s)=\sigma$ 的长度无限，因此在定义中涉及一个极限。此外由于我们真正关心的是 t 很大的区域，因此积分下限的选择并不重要，为了避免 $\zeta(s)$ 在 $s=1$ 处的极点对定理的表述造成不必要的麻烦，我们选了一个非零的积分下限。

其次，什么叫做 $|\zeta(s)|^2$ 在直线 $\mathrm{Re}(s)=\sigma$ 上的平均值"对 $\sigma>1/2$ 有界，且对 $\sigma\geq\sigma_0>1/2$ 一致有界"？"对 $\sigma>1/2$ 有界"很简单，就是说对任何 $\sigma>1/2$，存在常数 T_0 及 C 使得：

$$\frac{1}{T-1}\int_1^T\left|\zeta(\sigma+it)\right|^2 dt<C$$

对所有 $T>T_0$ 成立。而"对 $\sigma\geq\sigma_0>1/2$ 一致有界"是说对任何 $\sigma_0>1/2$，存在与 σ 无关的常数 T_0 及 C 使得上式对所有 $\sigma\geq\sigma_0$ 及 $T>T_0$ 都成立。

最后，"位于 $\mathrm{Re}(s)\geq 1/2+\delta$ 的非平凡零点所占的比例为无穷小"指的是位于 $\{\mathrm{Re}(s)\geq 1/2+\delta, 0\leq t\leq T\}$ 的非平凡零点的数目与位于 $\{\mathrm{Re}(s)\geq 1/2, 0\leq t\leq T\}$ 的非平凡零点的数目之比在 $T\to\infty$ 时趋于零[注 22.2]。

现在我们对玻尔 - 兰道定理的字面含义已经有了一些了解。它实质上是在 $|\zeta(s)|^2$ 的平均值与 $\zeta(s)$ 的零点分布之间建立了一种联系。这种存在于复变函数的模与零点之间的关联并不鲜见，1899 年 J. L. Jensen 提出的 Jensen 公式就是一例，它把一个亚纯函数（Meromorphic Function）在一个圆域内的零点和极点与函数的模在圆域边界上的性质联系在了一起。这一公式也正是玻尔与兰道在证明他们的定理时用到的主要公式。

Riemann

很明显，我们感兴趣的是玻尔 - 兰道定理中有关非平凡零点分布的叙述，即“对于任何 $\delta>0$，位于 $\mathrm{Re}(s)\geq 1/2+\delta$ 的非平凡零点所占的比例为无穷小”。但是这一叙述是否成立还有赖于玻尔 - 兰道定理的前提，即“$|\zeta(s)|^2$ 在直线 $\mathrm{Re}(s)=\sigma$ 上的平均值对 $\sigma>1/2$ 有界，且对 $\sigma\geq\sigma_0>1/2$ 一致有界”的成立与否。

幸运的是，这一前提可以证明是成立的。为了看到这一点，我们来分析一个比较简单的情形，即 $\sigma\geq\sigma_0>1$ 的情形。用我们在上文提到的关系式 $\overline{\xi(s)}=\xi(\bar{s})$，及 $\sigma>1$ 时 $\zeta(\sigma+it)$ 的级数展开式 $\sum_n n^{-\sigma-it}$ 可得：

$$|\zeta(\sigma+it)|^2=\zeta(\sigma+it)\zeta(\sigma-it)=\sum_n\sum_m n^{-\sigma-it}m^{-\sigma+it}\ .$$

另一方面，由于 $\sigma\geq\sigma_0>1$ 时 $\zeta(s)$ 在 $s=1$ 处的极点对计算没有影响，因此我们可以将 $|\zeta(\sigma+it)|^2$ 的平均值定义中的积分下限取为 $-T$（相应的将 $1/(T-1)$ 改为 $1/(2T)$）以利于计算积分（这里再次用到了 $\overline{\xi(s)}=\xi(\bar{s})$）。将上面有关 $|\zeta(\sigma+it)|^2$ 双重求和表达式代入平均值的定义，并先交换积分与求和的顺序，再交换求和与极限 $T\to\infty$ 的顺序（请读者自行证明这样做的合理性），可以发现只有 $m=n$ 的项才对结果有贡献，而它们的贡献一致收敛于 $\sum_n n^{-2\sigma}=\zeta(2\sigma)$（请读者自行证明）。这表明对所有 $\sigma\geq\sigma_0>1$ 玻尔 - 兰道定理中的前提都是成立的。

显然这样的简单证明不适用于 $\sigma\leq 1$ 的情形（因为 $\zeta(\sigma+it)$ 的级数展开式不再适用），但我们可以注意到证明结果中的 $\zeta(2\sigma)$ 对所有 $\sigma>1/2$ 都有意义。因此读者们也许会猜测这一结果的适用范围可以由 $\sigma\geq\sigma_0>1$ 拓展到 $\sigma\geq\sigma_0>1/2$。事实也正是如此。可以证明，对于任何 $\sigma_0>1/2$ 及 $\varepsilon>0$，存在与 σ 无关的常数 T_0 使得：

$$\left|\frac{1}{T-1}\int_1^T|\zeta(\sigma+it)|^2dt-\zeta(2\sigma)\right|<\varepsilon$$

对所有 $\sigma\geq\sigma_0$ 及 $T>T_0$ 都成立。这一结果显然表明（请读者自行证明）玻尔 - 兰道定理中的前提成立。这一点在玻尔 - 兰道定理之前就已经被证明，并出现在 1909 年出版的兰道的名著《素数分布理论手册》中。

既然前提成立，那么玻尔 - 兰道定理的结论也就成立了。这样我们就得到了继阿达马与 de la Vallée-Poussin 之后又一个有关零点分布的结果：对于任何 $\delta>0$，位于 $\mathrm{Re}(s)\geq 1/2+\delta$ 的非平凡零点所占的比例为无穷小。或者换句话说，在包含 critical line 的无论多小的带状区域中都包含了几乎所有的非平凡零点。

看到这里，有些读者也许会问：既然包含 critical line 的无论多小的带状区域都包含了几乎所有的非平凡零点，那么通过将这个带状区域无限逼近 critical line，我们是不是就可以把那些零点“逼”到 critical line 上，从而证明几乎所有的非平凡零点都落在 critical line 上呢？ 很遗憾，我们不能。事实上单单从玻尔 - 兰道定理所给出的描述中我们不仅无法证明几乎所有的非平凡零点都落在 critical line 上，甚至无法证明哪怕有一个零点落在 critical line 上！ 零点的分布完全有可能满足玻尔 - 兰道定理所给出的描述，却没有一个真正落在 critical line 上（请读者想一想这是为什么）。这是数学中与无穷有关的无数微妙细节中的一个。

但尽管如此，玻尔 - 兰道定理对非平凡零点分布的描述比十八年前阿达马与 de la Vallée-Poussin 所证明的结果还是要强得多。它虽然没能直接证明 critical line 上有任何零点（阿达马与 de la Vallée-Poussin 的结果也同样不能证明这一点），但它非常清楚地显示出了 critical line 在非平凡零点分布中的独特地位，即 critical line 起码是黎曼 ζ 函数非平凡零点的汇聚中心。这是一个沉稳而扎实的进展，数学家们正在一步步地逼近着 critical line。

未完待续