

随机矩阵在量子输运中的应用

王东

中国科学院大学

wangdong@wangd-math.xyz

2023 年 7 月 14 日

随机矩阵，顾名思义就是各项有随机性的一个矩阵。我们知道，矩阵在线性代数以及它的各种应用中有核心地位，而随机性是世界中无处不在的现象。所以，随机矩阵作为二者的结合，能在数学理论以及它的应用中占重要地位，是不足为奇的。

在目前，随机矩阵研究的主战场是统计，以及进一步衍生的机器学习等新兴领域。统计理论中，只要数据不是一维的（或者说，Excel 表格中不止有一个单元有数据），矩阵就无可避免地被引入。而随机性的考量显然是统计理论中不言而喻的。从这个角度看，随机矩阵和我们的生活大有关系。

但是，随机矩阵理论的提出和早期发展，却不是统计学家领的头，而是理论物理学家。这个历史发展过程并非巧合，而是科学发展的宏观历史的缩影：虽然科学对我们的日常生活息息相关，可是科学的发端并不是源自对日上生活的观察和积累，而是“仰望星空”。对天文的观测和解释，最终导致了哥白尼革命。因为日常生活中的科学，或者数学，往往更难。

随着一个世纪前量子力学的建立，物理学家了解到，原子，以及原子核的物理性质，由它们的哈密顿算子决定。这个哈密顿算子，用通俗的语言讲，就是一个矩阵——没有随机性的矩阵——由原子（核）唯一确定。量子力学的基本原理要求这个矩阵是一个厄米矩阵。

学过《大学物理》的同学都知道怎样写出氢原子（本质上就是一个电子在有心力场）的哈密顿算子，并且精确求出它的能级（也就是哈密顿算子的特征根）。深入学习过《量子力学》的同学也知道怎样分析氦原子（本质上是两个电子）的哈密顿算子，但是这时精确求解已经不可能，只能使用近似方法。物理学家在二战后的年代，最关心的是铀原子核——230 多个核子，其中 92 个质子，其余是中子——的哈密顿算子。显然，就算我们了解清楚质子、中子间所有的相互作用，并且能精确写出这个原子核的哈密顿算子，对其精确求解几乎不可能，甚至做近似求解，做数值分析，也有极大的困难。考虑到那个时期的超级计算机算力不到现在千元老人手机的算力的零头，这个困难就更大了。

物理学家 Wigner 和 Dyson 等有了一个相当有洞见（或者相当偷懒？）的想法：既然这个哈密顿算子复杂到无法分析，为什么不用一个完全随机的矩阵去建模这个哈密顿算子？这样的话，也许最后得到的结果虽不中亦不远呢。

因为基本原理要求哈密顿算子是厄米算子，所以，体现最大随机性的随机矩阵应该是各项为独立同分布的（复）正态随机变量，如下所示：

$$G = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{pmatrix}, \quad g_{i,i} = N(0, 1), \quad g_{i,j} = \overline{g_{j,i}} = \frac{1}{2}(N(0, 1) + iN(0, 1)).$$

很多物理模型同时具有时间反演对称性，也就对哈密顿算子又有了新的约束条件。这种情况下，具有最大随机性的矩阵就成为类似的实对称随机矩阵，如下所示：

$$G = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{pmatrix}, \quad g_{i,i} = N(0, 1), \quad g_{i,j} = g_{j,i} = N(0, 1).$$

这两种随机矩阵，被命名为高斯酉系综（Gaussian Unitary Ensemble, GUE）和高斯正交系综（Gaussian Orthogonal Ensemble, GOE）。最后，有些物理模型既有时间反演对称性，又有半整数自旋，需要用高斯辛系综（Gaussian Symplectic Ensemble, GSE）去建模。这三种矩阵模型，对应了物理系统的关于时间反演的三个对称类型。

对这三种看似简单的随机矩阵的研究，确实揭示了复杂物理模型哈密顿算子的一些特点。我们知道，氢原子的能级满足一些非常简单的关系。但是复杂一些的哈密顿算子的能级，则很明显杂乱无章。也许最容易脱口而出的猜测，就是这些能级的局部分布大约等同于泊松点过程。很不幸，这是错的。实验数据表明，能级之间似乎有某种“排斥力”，使它们不像泊松过程中的粒子一样经常靠得很近。（见图1）

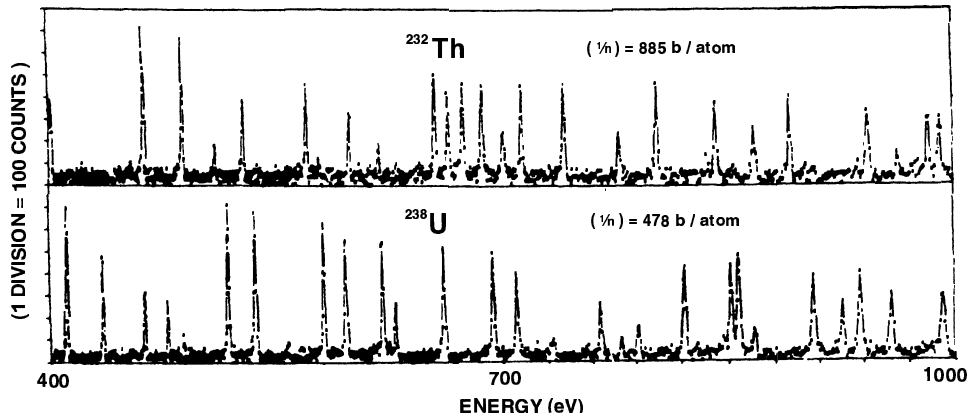


Figure 1.1. Slow neutron resonance cross-sections on thorium 232 and uranium 238 nuclei. Reprinted with permission from The American Physical Society, Rahn et al., Neutron resonance spectroscopy, X, *Phys. Rev. C* 6, 1854–1869 (1972).

图 1: 取自 Mehta 的专著 *Random Matrices* (第三版) 第一章第一节。

物理学家对 GUE 和 GOE 的特征根分布进行了理论计算。因为这些高斯系综具有旋转不变性，所以熟悉经典李群的物理学家很容易地剥离出特征根的分布函数，并且发现确实它们之

间体现出一个排斥力，而且这个排斥力的大小，在 GUE 中恰好为 GOE 中的 2 倍。（在 GSE 中，这个力的大小是 GOE 中的 4 倍。）并且，特征根的经验密度分布的极限（半圆率）与局部极限分布也都计算出来。其中局部极限分布和我们熟悉的正态分布，指数分布等常用极限极限都不同（见图2），有一个复杂的解析公式。最后，和实验数据对比，竟然是对的！

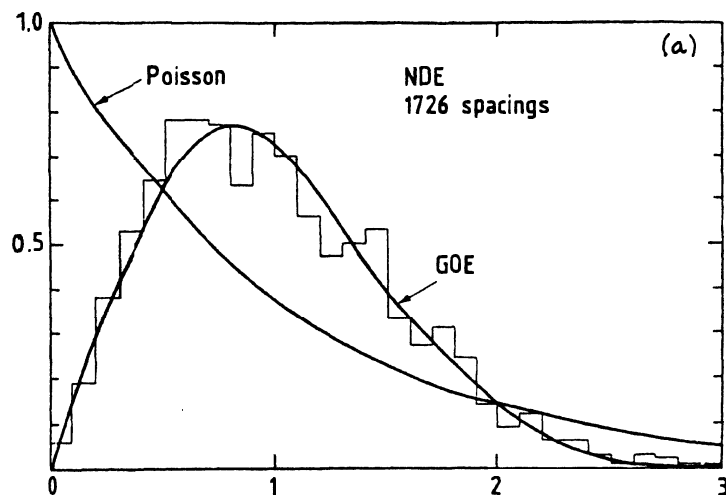


Figure 1.4. Level spacing histogram for a large set of nuclear levels, often referred to as nuclear data ensemble. The data considered consists of 1407 resonance levels belonging to 30 sequences of 27 different nuclei: (i) slow neutron resonances of Cd(110, 112, 114), Sm(152, 154), Gd(154, 156, 158, 160), Dy(160, 162, 164), Er(166, 168, 170), Yb(172, 174, 176), W(182, 184, 186), Th(232) and U(238); (1146 levels); (ii) proton resonances of Ca(44) ($J = 1/2+$), Ca(44) ($J = 1/2-$), and Ti(48) ($J = 1/2+$); (157 levels); and (iii) (n, γ)-reaction data on Hf(177) ($J = 3$), Hf(177) ($J = 4$), Hf(179) ($J = 4$), and Hf(179) ($J = 5$); (104 levels). The data chosen in each sequence is believed to be complete (no missing levels) and pure (the same angular momentum and parity). For each of the 30 sequences the average quantities (e.g. the mean spacing, spacing/mean spacing, number variance μ_2 , etc., see Chapter 16) are computed separately and their aggregate is taken weighted according to the size of each sequence. The solid curves correspond to the Poisson distribution, i.e. no correlations at all, and that for the eigenvalues of a real symmetric random matrix taken from the Gaussian orthogonal ensemble (GOE). Reprinted with permission from Kluwer Academic Publishers, Bohigas O., Haq R.U. and Pandey A., Fluctuation properties of nuclear energy levels and widths, comparison of theory with experiment, in: *Nuclear Data for Science and Technology*, Bökhoff K.H. (Ed.), 809–814 (1983).

图 2: 取自 Mehta 的专著 *Random Matrices* (第三版) 第一章第四节。

以上是对随机矩阵理论发端的正统描述。实际情况如何呢？用先驱者 Dyson 的话说¹：

All of our struggles were in vain. 82 levels were too few to give a statistically significant test of the model. As a contribution to the understanding of nuclear physics, random matrix theory was a dismal failure. By 1970 we had decided that random matrix theory was a beautiful piece of pure mathematics having nothing to do with physics. Random matrix theory went temporarily to sleep.

因为重原子核的结构虽然复杂，它包括的质子和中子加在一起顶多也就二百多个。用纸和

¹Dyson 的引文，以及上面随机矩阵研究开端的历史回顾，都可以参见 Dyson 给 *Oxford Handbook of Random Matrix Theory* 写的序言。

笔做分析自然难比登天，可是我们有计算机，还有摩尔定律。所以，当代的核物理研究，尤其是核武器研发，应该用不到随机矩阵的。

当然，Dyson 只是说 sleep，并且是在物理学领域暂时 sleep。随机矩阵理论的以后在统计领域开始发挥重要作用，甚至影响到了数论等纯数学研究。因为与本文主旨不直接相关，就不赘述了。

至于随机矩阵作为 a beautiful piece of pure mathematics 的研究，Dyson 自己就做了相当多的贡献。这里举其中一例。回忆我们已经提到过的作为随机矩阵的 GUE 和 GOE。它们的各项都是相互独立的正态分布随机变量。一个概率理论中熟知的事实是：如果我们把一系列独立同分布的正态分布随机变量加起来，我们就得到一个简单的随机过程，并且经过适当地比例变换，这个随机过程收敛到布朗运动。那么，如果我们不是把一维的随机变量，而是把高维的随机矩阵相加呢？比如我们考虑 GOE。将一系列独立同分布的 GOE 随机矩阵加在一起，经过比例变换，我们得到

$$G(t) = \begin{pmatrix} g_{11}(t) & g_{12}(t) & \cdots & g_{1n}(t) \\ g_{21}(t) & g_{22}(t) & \cdots & g_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1}(t) & g_{n2}(t) & \cdots & g_{nn}(t) \end{pmatrix}, \quad g_{i,i} = B_{i,i}(t), \quad g_{i,j}(t) = g_{j,i}(t) = B_{i,j}(t).$$

这里 $B_{i,j}(t)$ 代表相互独立的布朗运动。这样定义的 $G(t)$ 的特征根 $x_1(t), x_2(t), \dots, x_n(t)$ 所满足的随机过程就被称为 Dyson 布朗运动。GUE 和 GSE 显然也可以定义类似的对应 Dyson 布朗运动。

Dyson 布朗运动可以通过 SDE 来表示：

$$dx_i(t) = dB_i + \frac{\beta}{2} \sum_{j \neq i} \frac{dt}{x_i(t) - x_j(t)}.$$

其中参数 $\beta = 1, 2, 4$ 时，这个 SDE 就表示 GOE, GUE 和 GSE 定义的 $G(t)$ 的特征根的随机过程。对这个 SDE 的研究推进了概率中很多方向的发展，尤其是又促进了随机矩阵本身的理论研究——当然，可能不包括 Dyson 所期望的物理学研究。

注意：用 $\beta = 1, 2, 4$ 标志三种对称类型是随机矩阵研究的标准做法，下文中还会出现。

以上关于随机矩阵和物理学的关系都是引子。现在我们言归正传，考虑量子输运 (quantum transport)²。量子输运主要考虑的问题是电子在导体中的传输，尤其是导体的尺度到了纳米量级时。这时，把电流看做经典物理量（也就是连续的，确定的量）已经不能很好地解释物理现象，而必须把传输的电子，以及传输导体从量子力学的角度建模。这时候出现了技术困难：导体虽然很小，但也包含很多原子，其形状很难规整，更有不可避免的杂质在里面，所以精确地写出它的哈密顿算子几乎不可能。并且因为导体各个不同，针对某一个导体对具体的哈密顿算子做数值计算也没什么意义。这时，用随机矩阵去建模各不相同的纳米导体，就是自然地思路。这个领域比起核物理来，明显更是随机矩阵的用武之地。

²本文关于量子输运的介绍主要基于 Beenakker 在 Review of Modern Physics 上的综述 Random matrix theory of quantum transport。

纳米导体的量子性质体现在什么地方呢？我们以一维的纳米导线为例，考虑它的电导（也就是电阻的倒数）。经典的导体有一个确定的电导，并且如果我们改变导体的一些环境参数，其电导或者不变，或者连续有规律变化。但是如果我们考虑一个一维的纳米导线——不需要固定它的长度，也不需要假设具体的粗细形状，只要它属于“导体”这个范畴，我们就可以观测到这样一个经典物理不能解释的现象：如果我们连续改变某个环境参数，比如把导体置于外加磁场中，则其电导会不规则变化，围绕着平均电导一会变大一会变小，并且其变化规律不平滑，其图像很类似布朗运动的路径（见图3）。

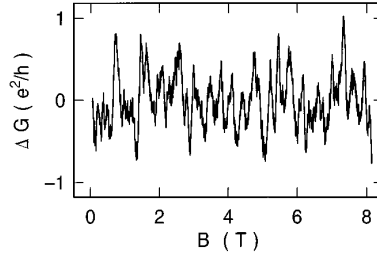


FIG. 4. Fluctuations as a function of perpendicular magnetic field of the conductance of a 310 nm long and 25 nm wide Au wire at 10 mK. The trace appears random but is completely reproducible from one measurement to the next. The root mean square of the fluctuations is $0.3 e^2/h$, which is not far from the theoretical result $\sqrt{1/15} e^2/h$ [Eq. (51) with $\beta=2$ due to the magnetic field and a reduced conductance quantum of e^2/h due to the strong spin-orbit scattering in Au]. After Washburn and Webb (1986).

图3: 取自 Beenakker 在 Review of Modern Physics 上的综述 Random matrix theory of quantum transport。

这些变化的电导有个平均值，对应的是经典意义的电导。围绕均值，可以测量电导的变化的方差，这也是纳米导线的一个重要性质，但是不能用经典物理解释。测量的结果发现，不同的纳米导线平均电导各不相同，但是电导的方差则不依赖于纳米导线的具体尺寸、形状和材质。取了一个模型无关的单位之后，这个方差总是等于

$$\frac{2}{15} \frac{1}{\beta},$$

这里 $\beta = 1, 2, 4$ ，对应纳米导线对时间反演的三种对称性类型。能不能解释这个 $2/15$ ，是理论是否成功的一个硬性标准。随机矩阵模型能不能通过这个检验呢？

首先，我们先建立一个最简单的模型。假设这个纳米导体没有任何确定的内部结构，于是我们用最具一般性的随机矩阵去模拟它。注意，量子输运模型并不是考虑导体本身的性质，而是考虑通过它的电子所收到的影响。所以，我们不是模拟导体的哈密顿算子，而是把它看做对于电子的散射矩阵——一个酉矩阵，而不是厄米矩阵。所以，如果要考虑一个最一般的随机矩阵，我们应该考虑一个服从 Haar 测度的随机酉矩阵（如果有时间反演对称性，就考虑服从 Haar 测度的随机正交矩阵或者随机辛矩阵）。

散射矩阵，包括了（向前）传输（transmission，记作 t ）和（向后）反射（reflection，记作 r ）两部分信息。显然我们关心的是电子的传输。所以，我们只需要随机酉/正交/辛矩阵（记为

S) 的一部分:

$$S = \begin{pmatrix} r & t' \\ t & r' \end{pmatrix}, \quad \mathcal{T} = tt^\dagger.$$

也就是说, 散射矩阵的一部分 t , 定义了我们需要的传输矩阵 \mathcal{T} 。令 \mathcal{T} 的特征值 (也就是 t 的奇异值的平方) 为 T_1, T_2, \dots, T_n 。则电导的均值/方差, 就是 $T_1 + T_2 + \dots + T_n$ 这个随机变量的均值/方差。而这个随机变量是随机矩阵理论可以刻画的。

结果差强人意。当 $n \rightarrow \infty$ (也就是纳米导体的尺度比原子要大很多) 时, 用随机酉/正交/辛矩阵来建模得出来的电导方差为³

$$\frac{1}{8} \frac{1}{\beta}.$$

虽然 $1/8$ 和 $2/15$ 的差别不是很大, 只有不到 7%。但是物理学是精确科学, 7% 的误差是不可忍受的。为什么有这个误差呢? 因为, 我们用最一般的随机矩阵去建模导体, 却忽略了一个重要信息: 我们要考虑的纳米导线是一维的, 而这个维度信息, 建模时并未考虑进去。其实, 我们已经得到的结果是“量子点”, 也就是各个方向没有区别的纳米尺寸导体的模型。

由点很自然地就可以连成线。相应地, 我们可以把一维纳米导线看做很多个量子点连接在一起, 然后就可以从量子点的输运模型出发, 得到一维纳米导线的的散射矩阵以及传输矩阵。具体的物理学推导过程在这里不详述, 我们只需要指出, 一维导体的传输矩阵的特征根的分布可以用一个随机过程表示。如果用 t 表示导线的长度, 而 $T_1(t), T_2(t), \dots, T_n(t)$ 表示长度为 t 时导线的传输矩阵特征值, 则它们满足一个 SDE, 形式上和 Dyson 布朗运动所满足的 SDE 类似。这个相似性不难直观想象: Dyson 布朗运动是很多相互独立的高斯酉/正交/辛系综相加构造的, 而一维导线的散射矩阵, 似乎应该是很多相互独立的量子点散射矩阵相乘构造的。(这个类比其实是错误的, 因为依 Haar 测度的随机酉/正交/辛矩阵的乘积结果仍然服从同样的 Haar 测度。这里只是打个比方。)

下面我们只考虑最简单的时间反演对称性破缺, 也就是 $\beta = 2$ 的情况。这时, 经过变量替换 $\lambda_k(t) = (1 - T_k(t))/T_k(t)$, $\lambda_k(t)$ 满足的 SDE 为

$$d\lambda_k(t) = \left(2\lambda_k(t) + 1 + \frac{1}{n} \sum_{l \neq k} \frac{2\lambda_k(t)\lambda_l(t) + \lambda_k(t) + \lambda_l(t)}{\lambda_k(t) - \lambda_l(t)} \right) dt - \sqrt{\frac{2\lambda_k(t)(\lambda_k(t) + 1)}{n}} B_k.$$

这个 SDE 称为 Dorokhov-Mello-Pereyra-Kumar (DMPK) 方程。

这个 DMPK 方程是否是一维导线的合理模型呢? 我们希望它能在电导方差与实验数据吻合。因为我们的 $\lambda_k(t)$ 是 $T_k(t)$ 的函数, 我们需要的是 $T_1(t) + T_2(t) + \dots + T_n(t)$ 的方差, 也就是

$$C(t) = \frac{1}{\lambda_1(t) + 1} + \frac{1}{\lambda_2(t) + 1} + \dots + \frac{1}{\lambda_n(t) + 1}$$

的方差。最后希望的结果是在先取 $n \rightarrow \infty$ (也就是导线直径远大于原子尺度) 然后再取 $t \rightarrow \infty$ (也就是导线的长度远大于直径, 成为名副其实的“线”) 之后, $C(t)$ 的方差趋近于 $1/15$ (因为我们取 $\beta = 2$, 抵消掉了分子 2)。

³我国南科大姜铁峰教授在这方面做了严格的计算, 见 Physics Letters A, 373(25):2117。

上述的 DMPK 方程很类似于 Dyson 布朗运动。所以，对 Dyson 布朗运动的深入研究，肯定会有助于证明上面的结果⁴，也就是说，从真正的数学严格意义上，验证随机矩阵在一维纳米导线的量子输运问题上成功应用。不过数学的发展往往是太慢了，跟不上物理学家的急切心情。于是物理学家抄了一个近道：当 $n, t \rightarrow \infty$ 而且 $t \ll n$ 时，如果做变量替换 $\lambda_k(t) = \sinh^2(\sqrt{x_k(t)})$ ，则（注意：这是我们这篇小文章最后一个公式了）

$$C(t) = \frac{1}{\cosh^2 \sqrt{x_1(t)}} + \frac{1}{\cosh^2 \sqrt{x_2(t)}} + \cdots + \frac{1}{\cosh^2 \sqrt{x_n(t)}}$$

而且 $x_1(t), x_2(t), \dots, x_n(t)$ 近似满足一个联合分布

$$P(x_1(t), \dots, x_n(t)) = \prod_{1 \leq i < j \leq n} (x_i(t) - x_j(t))(f(x_i(t)) - f(x_j(t))) \prod_{i=1}^n W(x_i(t)),$$

这里

$$f(x) = \sinh^2(\sqrt{x}) = \frac{1}{4} (e^{2\sqrt{x}} - 2 + e^{-2\sqrt{x}}), \quad W(x) = e^{-\frac{n}{t}x} \left(\frac{e^{2\sqrt{x}} - e^{-2\sqrt{x}}}{4x} \right)^{\frac{1}{2}}.$$

虽然这个联合分布看上去还是很复杂，但是从随机矩阵的角度看，属于现有技术可以操作的领域了。现在该是大结局时间了：应用这个近似联合分布之后，物理学家已经用物理中的数学（也就是不严格的数学）论证了 $C(t)$ 在 $n \rightarrow \infty$ 且 $t \rightarrow \infty$ 时趋近于 $1/15$ 。至于严格的数学证明，让我们拭目以待吧⁵。

⁴我国中科院李向东研究员在这方面做了相关研究，见 *Journal of Statistical Physics* (2020) 181:1277。

⁵最新进展，可以参考笔者拙文 <https://arxiv.org/abs/2307.03720>。