

# 随机系统的数学问题<sup>1</sup>

陈木法  
(北师大数学系)

现在在我国讲随机数学或随机学 (Stochastics) 的重要性比较容易被接受了, 因为每天的天气预报都在用“概率”. 其实, 承认“随机性”或“偶然性”是一件大事, 它本来就是哲学的几个基本要素之一. 目前, 国内、外都已形成共识, 需要尽早进行“随机性”的国民素质教育. 在各种科学领域, 甚至于在数学内部, 使用随机数学的思想和工具, 已成为一种时尚. 然而, 随机数学的研究在我国起步很晚, 普及工作更是严重不足, 因而对于随机数学的意义和价值的认识都亟待加强.

下面, 我们将介绍四个方面: 一. 概率与经济. 二. 概率与随机算法和计算复杂性. 三. 概率与几何、分析. 四. 概率与物理. 在每一部分里, 我们将解剖一些初等的有代表性的模型或问题, 说明研究的背景、价值和前景. 我们着眼于目前的一些发展方向而非对未来的展望. 回想过去的三、五十年, 每十年里都出现过一些人们预想不到的新方向、新领域. 因而对随机数学的未来的展望, 犯错误的概率会更大一些. 我们也将提及我国学者的若干重要贡献. 当然, 这必定是挂一漏万的. 一方面为题材所限, 另一方面, 对于许多学科分支, 我们都是门外汉, 既没有能力、也不应该作不负责任的评价. 关于我国学者马志明、严加安、彭实戈和黄志远等在随机分析方面的贡献, 请参考本书中的另一专题“随机分析与无穷维分析”.

我们从经济模型开始, 因为现在从上至下, 大家最关心的是经济. 经济是一个大系统, 无可非议. 经济系统含随机性, 也可容纳. 然而, 我们说, 经济在本质上是一个随机系统. 通过一个简化模型的分析, 我们将看到, 小小的随机摄动, 差之毫厘却失之千里. 遗憾的是, 人们对于经济模型随机性重要地位的认识却姗姗来迟, 以至于关于这种系统的随机数学理论还远未开发.

作为概率与物理 (特别是统计物理) 的交叉, 我们着重介绍“交互作用无穷粒子系统”. 它研究拥有大量粒子的带有交互作用的随机大系统的宏观发展行为. 中心课题是相变现象的数学理论, 这是当代数学家和物理学家所共同面临的挑战性课题. 此方向的研究与非线性偏微分方程等核心数学分支有着深刻的联系, 是无穷维数学的有代表性的一支, 是当前数学物理和概率论的最重要的主流发展方向之一. 她吸引了包括许多一流的数学家在内的庞大的研究队伍并形成了众多的新的学科分支.

第二部分介绍一个组合最优化问题, 第三部分介绍几何、分析中的若干论题. 这两部分的研究对象都是决定性的、非随机的. 然而, 使用随机数学方法, 特别是使用第四部分研究中所发展起来的数学工具, 人们却找到这些课题的令人吃惊的解答. 由此看来, 随机数学的研究对象不必是随机的.

## 一. 概率与经济

1. 投入产出法. 要了解经济情况, 先从调查入手. 首先, 把我们所关心的主要产品和劳务列为  $(x^{(1)}, x^{(2)}, \dots, x^{(d)})$ , 称为 **产综**. 我们需要调查三件事. 首先是初始产综 (如去年的产出), 记为  $x_0 = (x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(d)})$ . 其次是一年后的产出情况,  $x_1 = (x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(d)})$ . 再则是生产效率  $A_0 = (a_{ij}^{(0)})$ , 它表示每生产一个单位的  $i$  类产品需消耗  $(a_{ij}^{(0)})$  个单位的  $j$  类产品. 称  $A_0$  为

<sup>1</sup> 此文是为《当代中国数学》所写的一篇综述报告, 完成于 1997 年. 可惜该书并未出版. 部分材料后来发表于“相变的数学理论”(《21 世纪 100 个科学难题》, 李喜先主编, 吉林人民出版社 1998, 816-823) 及“谈谈概率论与其他学科的若干交叉”(数学进展 34:6, 2005, 661-672).

**效率方阵** 或 **结构方阵**. 这样, 为产出  $x_1$ , 需消耗  $\sum_{i=1}^d x_1^{(i)} a_{ij}^{(0)}$  个单位的  $j$  类产品. 现在, 我们考虑无消费的理想情形 (带消费情形才有实际意义但两者在数学上并无本质区别), 即产综  $x_0$  全部投入再生产而经一年后产出  $x_1$ . 在此条件下,  $x_0$  和  $x_1$  两者与  $A_0$  之间有如下关系:

$$x_0 = x_1 A_0.$$

依此递推, 得出

$$x_0 = x_n A_{n-1} \cdots A_0.$$

如假定  $A_n \equiv A$  (在短期内可视生产率稳定), 则上式成为  $x_0 = x_n A^n$ ,  $n \geq 0$ . 如  $A$  可逆, 则导出

$$x_n = x_0 A^{-n}, \quad n \geq 0. \quad (1)$$

换句话说, 知道初始产综和效率方阵, 我们便可预报第  $n$  年的产出  $x_n$ . 这就是著名的 **投入产出法**. 估计大家也比较熟悉. 早在 1968 年, 联合国统计司便将它列为国民经济核算的工具. 我国从 1974 年便开始编制国民经济投入产出表.

**2. 正特征向量法.** 经济如何发展好? 当然, 所谓“好”有不同的标准. 在我国, 曾长期存在一种看法, 即“快”就是“好”. 此刻, 我们不评论此标准的合理性而暂且就事论事, 如何才能快? 由 (1) 可见, 效率在短期内不变, 即  $A$  不变. 那么, 问题成为如何选初始投入  $x_0$  才能使  $x_n$  最大. 如选好投入, 即使某一产品  $x_0^{(j)}$  不够, 我们还可以进口补充之. 华罗庚先生最初研究此问题时主要是针对当时的计划经济而言的. 但他同时也指出此模型对市场经济也是适用的, 只不过产量指标  $x_n$  需换成价值指标. 但需注意在市场经济中, 随机性还要大些. 退一步说, 原模型对于宏观调控也是有参考价值的. 下面是所述问题的解答.

**华罗庚基本定理.** 设  $A$  非负不可约, 有最大特征根  $\rho = \rho(A)$ , 并以  $u$  表其左特征向量 (必定可取为正的).

(1) 如取  $x_0 = u$ , 则增长最快. 此时  $x_n = x_0 \rho^{-n}$ .

(2) 如  $x_0 \neq u$ , 则必定崩溃.

结论 (2) 甚妙. “崩溃”是指发展到某个年份,  $x_n$  必含零或负分量, 或 **崩溃时间**

$$T := \inf\{n \geq 1, \text{ 存在 } j \text{ 使 } x_n^{(j)} \leq 0\}$$

有限. 如果  $T$  总是很大, 那么我们就不要在意. 比如说, 我们只订五年计划. 请看华先生的下述例子.

例. 取  $A = \frac{1}{100} \begin{pmatrix} 25 & 14 \\ 40 & 12 \end{pmatrix}$ . 则

$$u = \left( \frac{5}{7} \sqrt{2409} + 13, 20 \right) \approx (44.34397483, 20)$$

下表给出当  $x_0$  取为  $u$  的前几位有效数字时, 对应的崩溃时间  $T$  的值.

| $x_0$             | $T$ |
|-------------------|-----|
| (44, 20)          | 3   |
| (44.344, 20)      | 8   |
| (44.34397483, 20) | 13  |

由此足以看出经济的敏感性.

**3. 随机模型.** 在实际中,  $A$  并非决定性的, 总存在各种偶然性因素. 例如农业生产依赖于气候. 即使忘掉客观因素, 对于大矩阵  $A$ , 特征值  $\rho$  也不可能精确算出. 误差总是存在的. 试问这些因素影响如何? 首先的猜测可能是无所谓. 回到上述例子, 在每一年, 以随机矩阵  $A_n = (a_{ij}^{(n)})$  代替上述的  $A = (a_{ij})$ , 其中  $a_{ij}^{(n)}$  以概率  $2/3$  与决定性情形相同, 即仍然取值  $a_{ij}$ .

但  $a_{ij}^{(n)}$  以  $1/6$  的概率分别取值  $(1 \pm 0.01)a_{ij}$ . 即以  $1/3$  小概率作  $1\%$  的小摄动. 我们还是从  $x_0 = (44.344, 20)$  出发, 其第  $n$  年崩溃的概率如下表.

| $n$ | 1 | 2    | 3    |
|-----|---|------|------|
| 概率  | 0 | 0.09 | 0.65 |

这表明在三年内崩溃的概率达  $0.74$ , 与前面的  $T = 8$  相比, 相距甚远. 事实上, 可以证明, 如固定  $A$  而不及时调整, 则随机模型常以概率  $1$  走向崩溃<sup>[6]</sup>. 以上讨论多少已经表明, 依据所具备的客观条件, 经济有它自身的合理的发展速度. 太快了不好, 太慢了也不行. 其实点破了也就很容易明白. 我们可以在一夜之间办起无数的工厂, 但没有足够的电力和交通, 这些厂子能运转起来吗? 我们可创办很多很多的大学, 但没有教师, 没有足够财力的支持, 资料室是空的, 实验仪器一样也没有, 能算得上大学吗? 再则, 以上讨论也表明, 在经济学的研究中, 如不考虑随机因素, 就会造成很大的失真. 这从一定角度上也解释了为什么人们普遍觉得投入产出法不好用的一种原因. 从数学上讲, 我们对此课题感兴趣的主要原因在于: 这里所要用到的随机矩阵理论, 乃是当今概率论十分活跃的发展方向.

应当说, 目前经济数学的研究与实践依然有相当大的距离. 经济毕竟是一个极其庞大的复杂系统, 是一个有待开垦的大领域. 近年来有“金融数学”新方向, 很热门. 我国学者彭实戈在此方向的贡献已有相当的国际地位.

在上述经济模型中, 因为客观上存在着随机性, 我们被迫要使用随机数学. 这是被动的. 但新近出现了反面, 即人们主动使用随机数学, 这里客观上并不存在随机性. 这就是下一部分的主题.

## 二. 概率与随机算法和计算复杂性

现在, 让我们转到计算机科学. 在计算机科学中, 把计算时间与计算规模的多项式同阶的算法称为多项式算法. 这种问题也称为多项式问题, 即 **P问题**. 非多项式阶的问题统称为 **NP问题**. 还是从一具体问题谈起.

**货郎担问题**. 给定全国  $144$  个城市, 找出一条经过所有这些城市而每城市只经过一次的最短闭路.

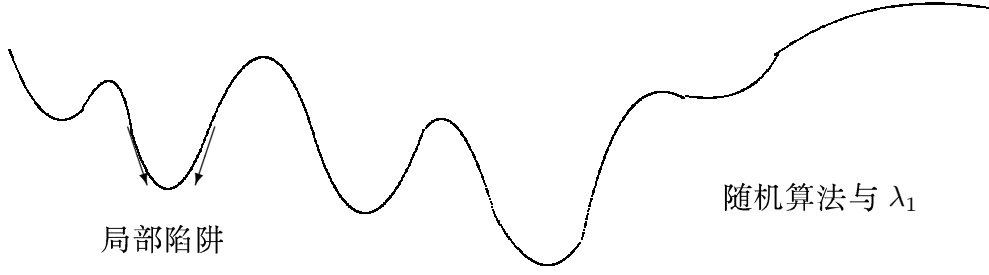
总共  $144$  个城市, 这个数字并不大. 但它们组合起来的闭路则有  $143!$  条. 要知道这个数有多大, 我们让每秒能够计算一亿条路 (注意这远快于每秒一亿次) 的计算机来算, 那么所需要的计算时间是

$$\frac{143!}{10^9 \times 365 \times 24 \times 60 \times 60} \approx 10^{231} (\text{年}).$$

这个组合优化问题就是一个典型的 NP 问题.

如何处理 NP 问题? 这当然是一个严峻的挑战! 似乎无路可循. 正是在这个人们以为“山穷水尽疑无路”的地方, S. Smale 的工作把我们带来了“柳暗花明又一村”. 他的主要想法是: 如果我们允许算法以小概率犯错误, 则我们可对 NP 问题找出多项式算法<sup>[20],[26]</sup>. 他首次将随机性引入计算复杂性理论, 并以这项成就荣获计算机科学的最高奖 (图灵奖).

让我们再回到组合最优化问题. 目的是要寻求最小值. 麻烦在于这里有众多极小值. 已有的算法常常会掉入局部陷阱 (每步总往取值较小的方向走). 近些年发展出一套随机算法 (如 **模拟退火算法**). 这种算法的要点是使用马氏链 (详见下段. 所以也称为 **马氏链 Monte Carlo**), 从一个极小值的附近出发, 以一定的概率向该极小值处靠拢, 但总允许以某个小概率往远处走走. 这就避免了落入局部极小值的陷阱而不能自拔. 有趣的是, 这种算法非常有效、实用. 如上述的货郎担问题, 使用这种算法可以很快找出一条  $30421$  公里的闭路, 这与目前所知的最好结果  $30380$  公里相差无几, 而真正的全局最优解依然是人们目前所力不能及的. 详见 [17]. 换言之, 若把寻求最优解放松为较优解, 则随机算法可把 NP 难度化为 P 难度. 在我国, 由钱敏平所领导的研究集体, 在随机算法方面已取得一大批出色成果<sup>[28],[29]</sup>.



下面介绍马氏链的一点简单知识. 所谓马氏链, 是取值于可数集  $E = \{i, j, k, \dots\}$  上的 (连续时间) 随机过程. 从  $i$  出发、经时间  $t$  转移到  $j$  的概率记为  $p_{ij}(t)$ . 这样, 我们有  $p_{ij}(t) \geq 0$  及  $\sum_{j \in E} p_{ij}(t) = 1$ . 其零点的导数记为  $q_{ij} = p'_{ij}(t)|_{t=0}$ . 这个矩阵  $Q = (q_{ij})$  常有如下简单性质: 非对角线元素非负 ( $q_{ij} \geq 0, i \neq j$ ) 而行和为零 ( $\sum_j q_{ij} = 0$ ). 由后一性质立知  $Q$  作用于常值向量  $\mathbf{1}$  为 0, 即  $Q\mathbf{1} = 0 = 0 \cdot \mathbf{1}$ , 即矩阵  $Q$  有平凡特征值  $\lambda_0 = 0$ . 自然要问,  $-Q$  的下一个特征值  $\lambda_1$  是什么? 当然, 可能是复的. 但我们现在只考虑实的情形. 对于无穷矩阵,  $-Q$  的谱可能连续, 未必有  $\lambda_1 > \lambda_0 (=0)$ . 后者是真正有用的、我们所感兴趣的情况. 这是因为, 常有

$$p_{ij}(t) \rightarrow \pi_j \quad \text{当 } t \rightarrow \infty. \quad (*)$$

若进一步问, 收敛速度有多快, 那么, 答案就是  $\lambda_1$ . 即  $|p_{ij}(t) - \pi_j| \sim e^{-\lambda_1 t}, t \rightarrow \infty$ . 对于随机算法, 通常把  $(\pi_i)$  取为“反比”于在  $i$  处的函数值, 即函数值越小, 算法停留在该处的可能性越大. 这样, 算法的有效性主要就取决于所构造的马氏链  $(p_{ij}(t))$  逼近此分布的速度. 换言之,  $\lambda_1$  越大, 相应的随机算法就越好. 可见第一 (非平凡) 特征值  $\lambda_1$  在这里起着根本的作用. 有关此论题的更详细的论述可参考 [21]. 顺便提及,  $(*)$  中的极限常唯一且满足方程

$$\pi_j = \sum_i \pi_i p_{ij}(t), \quad t \geq 0, j \in E.$$

由此出发, 也能证明前一部分所讲到的华先生的基本定理. 这相隔很远的两件事, 却能够联系起来, 实在是数学的奥妙特征 (详见 [6]).

这里, 我们用随机数学研究 NP 问题, 应当属于随机数学的范畴. 然而, 我们的研究对象毫无疑问是非随机的. 这或多或少跳出了随机数学的传统定义“从数量上研究随机现象的规律性的学科”, 因为后者把研究对象限于“随机现象”.

我们已经多次提到马氏链, 因而, 借此机会介绍我国在马氏链方面的若干成果. 当然, 这些工作未必于随机算法有关. 人们常把马氏链在概率论中的地位比喻为数论在数学中的地位, 易懂但不易做. 也许除日本而外, 每一个概率论强的国家其马氏链都有很高的水准. 我国在马氏链的研究方面自王梓坤教授六十年代的工作开始, 已经有了长期的积累和雄厚的基础并在国际上处于前列. 例如王梓坤和杨向群、侯振挺和郭青峰分别在 Springer 出版了两本研究专著 [22], [16], 杨向群在 John-Wiley 出版了一本研究专著 [23], 陈木法在新加坡世界科学出版社出版了一本研究专著 [5]. 对于这些专著, 几种主要的数学评论杂志都给予了很高的评价. 例如, 美国的《Math. Reviews》和德国的《Math. Abstracts》在评论陈木法的专著 [5] 时写道: 他们创建或领导了我国的一个“学派” (School). T. M. Liggett (1994) 把此书列为粒子系统的五本基本著作之一; C. Maes 和 S. B. Shlosman (1993) 则把此书列为来自美、俄、中的三本标准参考书之一 (standard references). 为总结过去三十年来在马氏链方面的研究成就, Springer 出版了 W. J. Anderson (1991) [1] 的总结性论著. 书中专辟一章介绍侯振挺的“Q 过程的唯一性准则”, 还收入了侯振挺、陈木法关于可逆 Q 过程的一条定理和杨向群关于单流出 Q 过程的一条构造定理.

在已往马氏链的研究中, 常归结为非负整数集. 这失去了许多几何性质, 特别是到了高维情形更是如此, 以至于许多结果无法使用. 例如最著名的唯一性准则, 此时就无法直接验证.

基于此, 陈木法提出了几个非常有效的唯一性充分条件. 这些工作后来被 S. P. Meyn 和 R. L. Tweedie (1993), K. Hamza 和 F. C. Klebaner (1995), G. Kersting 和 F. C. Klebaner (1995) 等继续和发展. 也是 D. A. Dawson, S. Feng 和 X. G. Zheng 研究平均场模型的七篇文章 (1992-1996)、宋京生研究无界情形马氏决策 (1987) 以及下面将要谈到的反应扩散过程等一系列研究的基石. 此项成果以及陈木法关于马氏链耦合的基本定理, 均被收入上述 Anderson 的总结性论著.

我国在马氏链方面的另一项重要成果是戴永隆于 1991 年最终解决了著名的 Davison 猜想 (1968)<sup>[12]</sup>. 所有这一切, 充分反映出我国学者在马氏链研究方面的国际地位.

### 三. 概率与分析、几何

下面介绍概率论与基础数学及其它学科的一些交叉研究成果. 当然, 这种交叉已不是什么新闻了. 概率方法在其它数学分支的应用已经相当多了. 这里只能讲讲几个重大成果.

1. **Atiyah-Singer 指标定理**. 大家知道, 这是近几十年来数学的重大成就之一. 1984 年, J. M. Bismut 给出了一种概率证明. 他的证明用到不久前发展起来的 Malliavin Calculus. 因为用概率方法能证出这个大定理, 此事在概率学界是很出名的. 基于他的概率想法, 两年后, E. Getzler (1986) 给出了一个简单直接的分析证明. 题目就是 “A short proof of the Atiyah-Singer index theorems”, 共 7 页. 有了这么简短的证明, 其内涵也就很清楚了. 因此, Bismut 的证明是一件很有价值的工作, 至少在概率界是一件值得纪念的事.

2. **亚椭圆性的 Hörmander 定理**. 上面所提到的 Malliavin Calculus 旨在建立一种无穷维概率空间上的微分和 Sobolev 光滑性理论. P. Malliavin (1976) 所做的第一件事就是以此工具重新证明线性 PDE 中的一条大定理, 即 Hörmander 定理: 在适当条件下, 二阶退化 PDE 有光滑解. 此方向由许多人继续和发展, 例如 Kusuoka-Stroock (1985) 等.

3. **完全非线性方程的 Krylov-Safonov 估计**. 1980 年所得到的这个估计, 对非线性方程产生很大影响. 例如见 [11]. 虽然这一结果目前已有分析证明, 但它首先是用概率方法发现的. N. V. Krylov 本人是一位著名的概率学家, 他对扩散过程理论有根本性贡献.

4. **第一特征值估计**. 上面已讲到马氏链的  $\lambda_1$ , 它就是第一 (非零) 特征值. 同样的问题也适用于欧氏空间或流形上的微分算子. 事实上, 三个方面都有了很好的进展. 让我们还是从一个小例子、从最简单的情形开始.

考虑二阶方阵 (两点情形):

$$Q = \begin{pmatrix} -b & b \\ a & -a \end{pmatrix}, \quad a, b > 0.$$

平凡特征值  $\lambda_0 = 0$ . 第一特征值  $\lambda_1 = a + b$ , 它是两个参数  $a$ 、 $b$  之和. 这时  $\lambda_1$  有很好的直觉, 不管那一个参数增大,  $\lambda_1$  都增大. 让我们往前再走一步, 看看三点情形:

$$Q = \begin{pmatrix} -b_0 & b_0 & 0 \\ a_1 & -(a_1 + b_1) & b_1 \\ 0 & a_2 & -a_2 \end{pmatrix}$$

共四个正参数  $b_0, b_1, a_1$  和  $a_2$ . 此时

$$\lambda_1 = \frac{1}{2} \left[ a_1 + a_2 + b_0 + b_1 - \sqrt{(a_1 - a_2 + b_0 - b_1)^2 + 4a_1b_1} \right],$$

此时  $\lambda_1$  关于参数 (例如说  $a_1$ ) 的单调性就没有了. 我们看不出来为使  $\lambda_1$  大, 究竟  $a_1$  是大一点好还是小一点好.

再看无穷阶的例子. 我们关心的是雅可比矩阵 (三对角阵):  $q_{i,i+1} = i + 1$  ( $i \geq 0$ ),  $q_{i,i-1} = 2i$  ( $i \geq 1$ ),  $q_{ii} = -q_{i,i+1} - q_{i,i-1}$ , 其它的  $q_{ij} = 0$ . 此时  $\lambda_1 = 1$ . 若保持  $q_{i,i+1}$  不动, 把  $q_{i,i-1}$  改为  $2i + 1.1$ , 即作小扰动, 那么  $\lambda_1$  变成 2. 相应的特征函数也由一阶多项式变成二阶多项式. 可见

$\lambda_1$  关于参数的变化极为敏感. 难怪特征值问题纷繁复杂, 其研究经久不衰. 容易看出, 对于任意的  $Q$ , 要找出  $\lambda_1$  的精确值是没有指望的, 只能做些估计.

下面我们想着重讲一下几何方面的工作, 因为这方面的工作最完整. 有三本很好的书 [2], [3] 和 [25]. 这三本书都花了相当多的篇幅讨论第一特征值问题. 其中第一本还收集了大约 2000 篇参考文献.

特征值估计是几何学的传统课题. 我们不禁想起 M. Kac 以“你能听出鼓的形状吗?” 为题的著名论文. 想法是使用几何特征表出特征值. 此处我们仅考虑紧连通黎曼流形. 我们要用到三个几何量: 流形的维数  $d$ , 直径  $D$  和 Ricci 曲率下界  $K$ . 一维情形  $K = 0$ , 高维球面  $K > 0$ , 双曲空间  $K < 0$ . 我们希望使用这三个几何量估计流形上拉氏算子第一特征值的下界. 我们不谈上界估计, 它相对容易一些. 单就下界估计而言, 我们也只能谈谈一些著名估计. 第一个结果是差不多 40 年前由 A. Lichnerowicz(1958) 所得到的.

$$\lambda_1 \geq \frac{d}{d-1}K, \quad K > 0, \quad (2)$$

用到两个几何量  $d$  和  $K$ . 这是一个优美结果, 因为对于高维空间的单位球面, 它是精确的. 经过 28 年之后, 此估计才由 P. Bérard, G. Besson 和 S. Gallot(1985) 部分地改进为

$$\lambda_1 \geq \frac{dK}{d-1} \left\{ \frac{\int_0^{\pi/2} \cos^{d-1} u \, du}{\int_0^{D/2} \cos^{d-1} u \, du} \right\}^{2/d}, \quad K \geq d-1. \quad (3)$$

当  $K = 0$  时, 两者均失效. 八十年代开始时的一项重要进展属于 P. Li 和 S. T. Yau(1980). 他们得出

$$\lambda_1 \geq \frac{\pi^2}{2D^2}, \quad K \geq 0.$$

前者使用等周不等式和 Cheeger 不等式 (1970), 后者使用梯度估计方法. 两种方法后来都有一系列的应用. 后一估计由钟家庆 — 杨洪苍 (1984) 改进为

$$\lambda_1 \geq \frac{\pi^2}{D^2}, \quad K \geq 0, \quad (4)$$

在零曲率时达到最优.

关于负曲率情形, 有 Li-Yau(1980) 估计

$$\lambda_1 \geq \frac{\exp(-1 - \sqrt{1 - 4D^2K(d-1)})}{D^2(d-1)}, \quad K \leq 0.$$

由这个表达式也可感觉出来负曲率情形要麻烦得多, 这是几何上常见的事. 随后的进一步发展有蔡开泉 (1991):

$$\lambda_1 \geq \frac{\pi^2}{D^2} + K, \quad K \leq 0; \quad (5)$$

杨洪苍 (1989) 和贾方 (1991),

$$\lambda_1 \geq \begin{cases} \frac{\pi^2}{D^2} e^{-\alpha/2}, & d \geq 5; \\ \frac{\pi^2}{2D^2} e^{-\alpha'/2}, & 2 \leq d \leq 4. \end{cases} \quad (6)$$

此处  $\alpha = D\sqrt{-K(d-1)}$ ,  $\alpha' = D\sqrt{-K((d-1) \vee 2)}$ . 杨文 (1989) 把这一估计称为丘成桐猜测.

通过以上的简要介绍已经可以看出, 几何上关于第一特征值的研究已相当完整. 面对这样研究得很好的领域, 人们还能再做些什么? 最初考虑这个课题的动机是为了研究具有交互作用的无穷粒子系统的相变行为 (后面将进一步介绍). 希望从几何学家那里寻找工具和方法. 后来发现用研究粒子系统所发展起来的概率方法也能够用于研究几何特征值. 陈木法和王凤雨 [10] 使用概率方法, 发现了一个一般公式. 我们准备介绍这期间的许多中间步骤, 而只谈谈这个公式. 先介绍两个简单记号

$\mathcal{F}$  = 在  $[0, D]$  上连续、在内部为正的函数  $f$  的全体.

$$C(r) = \cosh^{d-1} \left( \frac{r}{2} \sqrt{\frac{K}{d-1}} \right), \quad r \in [0, D],$$

后者是双曲余弦的  $d-1$  次幂. 有了这两个记号, 便可写出如下一般公式.

$$\lambda_1 \geq \sup_{f \in \mathcal{F}} \inf_{r \in (0, D)} \frac{4f(r)}{\int_0^r C(s)^{-1} ds \int_s^D C(u)f(u) du}.$$

留心这里关于函数  $f \in \mathcal{F}$  取上确界, 因而对每一试验函数  $f$ , 只要不太糟糕, 都可以得到一个不坏的下界. 例如取平凡的  $f \equiv 1$ , 所得出的估计在几何上也是非平凡的. 从这个意义讲, 此公式乃是一种变分公式. 问题的关键在于, 这种用于下界估计的公式, 先前从未有过. 对比上界估计早有的经典的极大极小公式, 可见来之不易. 为理解此公式的效力, 只需将它依次用于如下初等函数:

$$f(r) = \sin(\beta r), \sin(\alpha r), \sin(\beta r), \cosh^{d-1}(\alpha r) \sin(\beta r),$$

此处  $\beta = \frac{\pi}{2D}$ ,  $\alpha = \frac{1}{2} \sqrt{\frac{|K|}{d-1}}$ , 便可得出如下四个估计:

$$\lambda_1 \geq \frac{\pi^2}{D^2} + \max \left\{ \frac{\pi}{4d}, 1 - \frac{2}{\pi} \right\} K, \quad K \geq 0. \quad (7)$$

$$\lambda_1 \geq \frac{dK}{d-1} \{1 - \cos^d[\alpha D]\}^{-1}, \quad K \geq 0. \quad (8)$$

$$\lambda_1 \geq \frac{\pi^2}{D^2} + \left( \frac{\pi}{2} - 1 \right) K, \quad K \leq 0. \quad (9)$$

$$\lambda_1 \geq \frac{\pi^2}{D^2} \sqrt{1 - \frac{2D^2 K}{\pi^4}} \cosh^{1-d}[\alpha D], \quad K \leq 0. \quad (10)$$

与前面所列举的各种最优估计比较, 显然有 (7) 优于 (4), (9) 优于 (5), (10) 优于 (6). 因  $K > 0$  时,  $\alpha D \leq \pi/2$ , 故 (8) 优于 (2). 还可以证明 (8) 优于 (3). 这样, 前面所述的各种最优估计都得到了改进. 须知为取得这些估计, 几何学家曾付出了近四十年的艰辛努力.

应当指出, 所用的证明方法跟以前完全不同. 到目前为止, 我们还不知上述公式的非概率证明. 此外, 新方法具有普遍意义, 不仅适用于紧流形, 也适用于非紧流形; 不仅适用于连续空间, 也适用于离散空间; 不仅适用于有限维空间, 也适用于无限维空间. 关于陈木法和王凤雨等对几何、分析和马氏链的谱隙估计、对数 Sobolev 常数估计等的诸多研究成果, 请参考综合报告 [8]. 更近些, 王凤雨关于这两个不等式是否成立的判定取得重要进展.

这项工作再一次显示出人们常说的“数学上的美感”和“数学的整体性”. 两个记号加一个公式, 非常简洁, 没有拖泥带水的东西. “一种概率方法”和“黎曼几何的第一特征值估计”, 两者相距那么远, 却出奇地走到了一起, 这之间的微妙联系, 并不是一下子就能够领悟得到的. 一方面, 这需要长期的积累. 另一方面, 也说明科学研究存在着偶然性. 那就是数学的发展是很难完全预料的. 我们这里所讲的研究方向, 多半就没有超过卅年. 我们已讲了三个题目: 概率与经济, 概率与计算机科学, 概率与几何、分析. 下面讲讲

## 四. 概率与物理

这个题目很大, 此处侧重讲几点.

1. **量子概率**. 简单地讲, 它是概率论与量子物理相结合的产物. 但因为量子物理的本质是非交换的, 它与传统的概率论有着根本的区别, 以至于概率论的 Kolmogorov 公理系统已不适用. 而且很多数学工具都有很大不同, 如这里的基本工具之一是  $C^*$  代数. 还有从量子物理出发, 由 R. L. Hudson 和 K. R. Parthasarathy(1983) 所引进的量子随机积分. 这一专题已由后者出版了专著. 量子概率目前已有一支不小的队伍, 主要在欧洲. 由 L. Accardi 等人所主编的文集已出版了九卷之多. 关于这方向, 我国学者卢云刚 (他目前在意大利工作), 已经很有成就.

2. **渗流理论(Percolation)**. 按照通常辨别好数学问题的标准“易于表述而又难于解决”(俗话所述“好说不好做”)来看, 这里存在大量的优美数学问题. 正因为模型简单, 才有可能成为研究其它问题的基本工具. 从这个意义上讲, 这是一个很好的研究方向. 现在, 它是概率界、数学物理学界很时髦的方向. 领袖是 Cornell 大学的 H. Kesten, 他出版了专著 [18]. 关于这方向, 我国学者张雨 (他目前在美国工作), 已经很有成就.

还是从简单模型谈起.

a) **边模型**. 考虑  $d(\geq 2)$  维整点  $\mathbb{Z}^d$ . 两紧邻点构成一条边. 假定每条边独立地以概率  $p(1 \geq p \geq 0)$ “开”, 当然也就独立地以概率  $1-p$ “闭”. 把开的边顺序连接起来的折线称为一条路, 其长度定义为该路所包含边的条数. 我们关心的是是否有一条“走向无穷的路”. 容易想象, 这代表地下石油在地层中的渗流. 当然, 当  $p=1$  时, 所有的边都开着, 走向无穷的路很多; 当  $p=0$  时, 所有的边都闭着, 没有走向无穷的路. 问题是: 对于其它  $p$ , 是否以概率 1 存在走向无穷的路? 可见, 这里的假设和问题都很容易说清楚. 事实上, 这是一个非常初等的概率问题. 显然,  $p$  越大, 越容易开, 故存在一个最小值, 叫临界值, 记为  $p_c$ . 那么, 我们容易猜出来  $p_c = \frac{1}{2}$ , 很好记. 当  $p > \frac{1}{2}$  时, 总有一条走向无穷的路; 当  $p < \frac{1}{2}$  时, 则不行, 存在走向无穷的路的概率为 0. 不过, 我们强调此刻讲的是二维 ( $d=2$ ) 情形.

b) **点模型**. 假如我们把上述模型稍许改一下, 把“边”改为“点”. 即每点独立地以概率  $p$ “开”(而以概率  $1-p$ “闭”). 如紧邻两点都开, 则称此边开. 于是可如上定义开路并提出同样的问题. 同样地, 我们可定义临界点  $p_c$ . 对于二维情形, 这个点模型的  $p_c$  是多少? 我们可大致猜出  $p_c > \frac{1}{2}$  (与边模型比较), 但这个  $p_c$  究竟等于多少则是一个至今未能解决的难题. 顺便提一句, 如能找出精确解, 也会产生无数的应用.

进一步, 如果把  $d=2$  换成  $d \geq 3$ , 那么无论是边模型或点模型, 其临界值都是未知的, 目前仅知道一个大概的范围. 这样, 我们已经跟大家介绍了现在很热门的研究方向中的一个“小”问题.

3. **交互作用无穷粒子系统**. 简称为粒子系统. 这是我们要着重介绍的一个概率与物理的交叉学科. 先复习前面已讲过的两点情形:  $E = \{-1, +1\}$ . 设  $a, b > 0$ ,

$$Q = \begin{pmatrix} -b & b \\ a & -a \end{pmatrix}, \quad \lambda_1 = a + b.$$

由  $Q$  导出转移概率矩阵

$$P(t) = (p_{ij}(t)) = e^{tQ} = \frac{1}{a+b} \begin{pmatrix} a + be^{-\lambda_1 t} & b[1 - e^{-\lambda_1 t}] \\ a[1 - e^{-\lambda_1 t}] & b + ae^{-\lambda_1 t} \end{pmatrix}.$$

命  $t \rightarrow \infty$ , 得出

$$P(t) \rightarrow \begin{pmatrix} \pi \\ \pi \end{pmatrix} = \begin{pmatrix} \pi_{-1} & \pi_{+1} \\ \pi_{-1} & \pi_{+1} \end{pmatrix},$$

其中  $\pi_{-1} = \frac{a}{a+b}$ ,  $\pi_{+1} = \frac{b}{a+b}$ . 我们曾说过这个  $\pi = (\pi_{-1}, \pi_{+1})$  满足方程

$$\pi_j = \sum_i \pi_i p_{ij}(t), \quad t \geq 0, j \in E.$$



右方表示从  $\pi$  出发, 过程在时刻  $t$  的概率分布. 此式表明, 分布  $\pi$  与  $t$  无关, 故称为 **平稳分布**, 在物理上称为 **平衡态**. 由这些讨论, 得出如下结论:

- a) 平衡态  $\pi$  存在且唯一.
- b) 收敛于平衡态的速度为  $e^{-\lambda_1 t}$ .

往前走一步, 考虑有限乘积空间. 把  $\{-1, +1\}$  换成  $E = \{-1, +1\}^S$  ( $S$  有限). 那么上述两条结论依然成立.

再往前走一步, 考虑无限 (可数) 乘积空间  $E = \{-1, +1\}^{\mathbb{Z}^d}$ . 简单理解如下: 考虑平面上的整点, 每一点上放一磁单子, 它有正、负极之分. 整个平面代表一个铁磁体. 我们所关心的基本问题是何时铁磁体可被磁化, 即从“无序”变为“有序”. 这一磁化过程叫做 **相变**. 相变现象处处可见. 水在零度时固化、在  $100^\circ\text{C}$  时汽化, 都是相变. 相变研究的价值是熟知的, 超导即是相变. 现在, 我们用  $x = (x_u : u \in \mathbb{Z}^d)$  来表示  $E$  中的点, 把  $u \in \mathbb{Z}^d$  称为位置或坐标.  $x$  在  $u$  处的值只取  $\pm 1$  两个值. 给定  $x \in E$  和  $u \in \mathbb{Z}^d$ , 记  ${}_u x$  为  $E$  中的一点, 它与  $x$  的差别只在  $u$  处变号. 即  $({}_u x)_v = x_v$ , 如  $v \neq u$ ; 而  $({}_u x)_u = -x_u$ . 上面我们曾用矩阵  $Q = (q_{ij})$  表示系统的发展速率, 即从  $i \rightarrow j$  ( $i \neq j$ ) 的速率为  $q_{ij}$ . 然后由它决定半群  $P(t) = e^{tQ}$ . 在目前的情况下, 转移速率  $c(u, x)$  取为

$$x \rightarrow {}_u x : \quad \text{转移速率 } c(u, x) = \exp \left[ -\beta \sum_{v: |u-v|=1} x_u x_v \right],$$

此处  $\beta > 0$  称为 **反温度**. 我们不考虑其它方式的转移. 这种取法不是随便的, 而有其物理意义. 代替原先的矩阵  $Q$  (因此时空间  $E$  不可数), 我们使用算子

$$\Omega f(x) = \sum_{u \in \mathbb{Z}^d} c(u, x) [f({}_u x) - f(x)].$$

由此出发, 可导出半群  $P(t)$ , 形式上写成  $P(t) = e^{t\Omega}$ . 进而有平衡态  $\pi$ :

$$\pi = \pi P(t), \quad t \geq 0.$$

当  $d = 1$  时, 平衡态依然唯一. 但当  $d \geq 2$  时, 出现了十分有趣的现象: 存在临界值  $\beta_c^{(d)}$ , 使得  
如  $\beta < \beta_c^{(d)} (< \infty)$ , 则平衡态唯一;  
如  $\beta > \beta_c^{(d)} (> 0)$ , 则平衡态非唯一.

这给出了相变的一种数学刻画. 回顾前面所说, 有限维 (即有限乘积) 空间上平衡态唯一, 因而无相变. 这说明我们不得不要研究无穷维. 上述例子是统计物理的典型模型, 叫做 Ising 模型 (1925), 是每一本统计物理教材都要讲到的 (当然, 使用的是物理语言). 关于二维情形, 临界值已由 L. Onsager (1944) 定出:

$$\beta_c^{(2)} = \frac{1}{2} \log(1 + \sqrt{2}) \approx 0.44.$$

值得一提的是李政道、杨振宁 (1952) 曾有杰出贡献. 当  $d \geq 3$  时,  $\beta_c^{(d)}$  却是未知的, 这是数学界和物理学界普遍了解的经典难题. 当然, 数学上的兴趣并不完全在于确定这些临界值, 还在于它们是检验新理论、新方法的重要标尺.

物理学家研究统计物理的常用方法是所谓“热力学极限”. 这跟上述讲法不同. 后者所用的是现代的“交互作用粒子系统”(简称为“粒子系统”)的数学语言. 1965 年, R. L. Dobrushin 使用概率论工具, 严格证明了高维 Ising 模型相变的存在性. 从此诞生了“随机场”及其姐妹学科“交互作用无穷粒子系统”. 前者是静态的, 后者是动态的 (即含有时间  $t$ ). 三十多年来, 经过一大批著名数学家的努力, 形成了拥有粒子系统、随机场、渗流理论、超过程、流体动力学极限、随机介质 (或随机环境) 中的随机过程和高分子模型等众多学科分支的研究领域和一支庞大的研究队伍. 并已出版了一批总结性论著. 例如 T. M. Liggett<sup>[19]</sup>, R. Durrett<sup>[13]</sup>, 严士健<sup>[24]</sup> 和陈木法的 [4] 和 [5]. 关于随机场, 则有 H. O. Georgii<sup>[14]</sup>. 无疑地, 这是现代数学与物

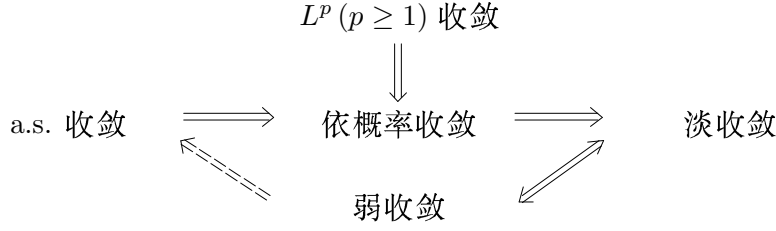
理重新汇合的一个发端领域，也是当前数学物理和概率论的最重要的主流发展方向之一。这只要随便翻翻“J. of Statistical Physics”和“Communications in Mathematical Physics”两种杂志便可看出，更不用说概率论杂志了。如下事实也是一个佐证：(1) 最近几届数学家大会几乎每次都有这方面的一小时报告(如 Dobrushin, Fröhlich, Sinai 和 Varadhan 等)；其中 Dobrushin 的报告题目是“作为概率论一个分支的统计力学”。上届概率论方面的 45 分钟报告有 3/5 属本领域。(2) 英国“牛顿数学科学院”成立不久(1993)，即举办了为期半年的“概率与相变”的专题活动。(3) 1996 年 7 月由美国数学会在普林斯顿主办的暑期概率学校，6 个报告人中有 3 人属本领域。这两学科的牢靠的根基，决定了它们自身必定有远大前景。顺便说一句，前面所讲的模拟退火算法，有两项基础：一是马氏链，二是来自统计物理的 Gibbs 分布。粒子系统的研究现状也正如 Dobrushin 所说：“一方面，这里有若干数学理论的大陆，有一些数学成果的小岛屿；另一方面，这里有着未解决问题的汪洋大海。”粒子系统的迷人之处和广阔前景主要就在于它有太多的未解决的重要问题。

我们已经看到，大多粒子系统模型都有两个区域：平衡态唯一区域和非唯一区域。这两区域的面孔可不大一样。这里用得上托尔斯泰的名言：“幸福的家庭家家相似，不幸的家庭各个不同”。为证明相变的存在性，就需要针对具体模型，使用不同的手法。经过近二、三十年的努力，已形成若干重要理论，其中较近的一种方法就是第一特征值估计。在唯一性区域， $\lambda_1 > 0$  而在非唯一区域  $\lambda_1 = 0$ 。这样，通过估计  $\lambda_1$  从正退化为零，可刻画相变。这就是研究  $\lambda_1$  估计的最初的动机。

我国关于粒子系统的研究主要集中于无穷维反应扩散过程，它来源于当今统计物理中非常活跃的非平衡系统而与以 Ising 模型为代表的平衡态统计物理根本不同。在我国，这一研究方向最早是由严士健提出的(1980)。区别于国外工作的主要特征是状态空间非紧而且过程一般不可逆，粒子系统的研究中最基本的 Liggett 构造定理不适用。陈木法首次给出了上述非平衡统计物理的典型模型的构造定理，并命名为“反应扩散过程”。这些工作也正是陈研究耦合方法(见下条)的发端。在 W. D. Ding、R. Durrett 和 T. M. Liggett (1990)，C. Neuhauser (1990) 和 T. S. Mountford (1992) 的论文中，引用了陈在这方面的五篇论文，并把上述过程的构造称为“陈氏构造”。十多年来，曾有几位学者力图改进这项成果但均未获本质进步。国外曾有人在专著中自己证之，却完全错掉。可见其难度。陈关于反应扩散过程的遍历性成果也处于领先地位。所用的工具依然是耦合方法。在这些工作中，最核心的是陈所找到的“有限矩条件”和“半群的 Lipschitz 连续条件”，是反应扩散过程及其平均场模型的已有研究都要用到的。国内参加此项研究的还有丁万鼎、唐守正、郑小谷等一大批学者。刚刚提到的丁万鼎等三人的论文，证明了在可逆的条件下，反应扩散过程总遍历。这个重要成果的物理意义在于：反应扩散过程本质上不可逆(即非平衡)。关于此方向十二年来的研究成果，可参考综述报告 [9]。还应指出，郭懋正与 Varadhan 和 Papanicolou 的文章 [15]，早已是 Hydrodynamical limit 方面每文必引用的名作，陈大岳、冯建峰和钱敏平关于亚稳态研究取得优美成果 [27]。我国在渗流理论和流体力学的数学理论等方面，特别是由王梓坤和吴荣所领导的研究集体关于超过程、马氏过程与位势理论的研究，都已作出了一批有份量的工作，在此不能一一列举。

粒子系统属于无穷维的数学。因为是无穷维，许多有限维的数学工具用不上，变得极为艰难。有些人觉得这里缺乏系统的理论。其实，无穷维数学极为丰富多采，将来即使形成象有限维数学那么“完整”的理论，恐怕这些理论的名目也会极其繁多。另一方面，为研究无穷维，也促使人们去寻找新方法、新工具。作为一个例证，下面介绍

**4. 概率距离与耦合方法。**回忆在初等概率论(或实变函数论)里，我们曾讲过各种收敛性。对于定义在概率空间  $(\Omega, \mathcal{F}, \mathbb{P})$  上的随机变量序列  $\xi_n$ ，我们有如下各种收敛性：



它们之间的关系如上图所示. 这里, 淡收敛也称为依分布收敛, 如已知  $\xi_n$  的分布淡收敛于某随机变量  $\xi$  的分布, 则此时淡收敛也就等价于弱收敛了. 后者等价于, 对于每一有界连续函数  $f$ ,

$$\int f dP_n \rightarrow \int f dP, \quad n \rightarrow \infty,$$

此处  $P_n$  和  $P$  分布是  $\xi_n$  和  $\xi$  的分布. 这些收敛性之间, 有一重要差别: 淡收敛和弱收敛不依赖于参考标架  $(\Omega, \mathcal{F}, \mathbb{P})$ , 而其它收敛性则严格依赖于此标架 (所谓概率空间, 乃是一参考坐标系). 如固定参考标架, 那么 a.s. 收敛性严格强于弱收敛. 反之, 若已知分布  $P_n$  弱收敛于  $P$ , 那么有一著名定理断言: 我们总可以构造出一个概率空间  $(\Omega', \mathcal{F}', \mathbb{P}')$  及其上的随机变量  $\xi'_n$  和  $\xi'$ , 使得  $\xi'_n$  有分布  $P_n$ ,  $\xi$  有分布  $P$ , 而且在  $(\Omega', \mathcal{F}', \mathbb{P}')$  上,  $\xi'_n$  a.s. 收敛于  $\xi'$ . 这是图中虚箭头所表示的蕴涵关系. 这样, 从内在 (即不依赖于参考标架) 的观点来看, 后面这四种收敛性都是等价的. 换句话说, 如果我们要寻找一种不同的、不依赖于坐标标架的距离, 那么自然会想到去寻找  $L^p$  收敛的类似物. 以下限于  $p=1$  好了. 回忆  $L^1$  距离的定义

$$\|\xi_1 - \xi_2\|_{L^1} = \mathbb{E} |\xi_1 - \xi_2| = \int |x_1 - x_2| \tilde{P}(dx_1, dx_2)$$

此处  $\tilde{P}$  是  $\xi_1$  和  $\xi_2$  的联合分布. 若把  $\xi_k$  的分布记为  $P_k$  ( $k=1,2$ ), 那么  $\tilde{P}$  就称为  $P_1$  和  $P_2$  的耦合 (coupling). 对于给定的  $P_1$  和  $P_2$ , 存在很多的耦合  $\tilde{P}$ , 换言之, 上式右方还没有脱离参考标架. 为避免这一点, 改定义

$$W(P_1, P_2) = \inf_{\tilde{P}} \int |x_1 - x_2| \tilde{P}(dx_1, dx_2), \quad (11)$$

此处  $\tilde{P}$  跑遍  $P_1$  和  $P_2$  的一切耦合, 我们便得出  $P_1$  和  $P_2$  的一种距离  $W(P_1, P_2)$ , 称为 Wasserstein(1969)距离. 其实这一距离可以追溯到五十年代甚至上一世纪, 因而有几种不同的名称. 我们不去管它. 当然, 我们可以把上面的长度  $|x_1 - x_2|$  改为任一距离函数, 一切照样. 要看看这一距离的优点, 取  $\xi_1 \in \mathbb{R}^d$ , 其分布为  $P_1$ . 再取  $\xi_2 = \xi_1 + a$ ,  $a \in \mathbb{R}^d$ , 即  $\xi_2$  是  $\xi_1$  的平移. 那么易证  $W(P_1, P_2) = |a|$ , 即分布的  $W$  距离就等于平移的长度. 可见这种距离包含着重要的几何内涵. 当然, 精确地算出两个分布  $P_1$  和  $P_2$  的距离远不是这么容易的. 然而, 由定义 (11) 式知, 任何耦合  $\tilde{P}$  自动给出  $W(P_1, P_2)$  的上界估计. 对于许多应用来说, 上界估计就够了 (特别对于收敛性  $W(P_n, P) \rightarrow 0$ ). 为此, 适当选择耦合  $\tilde{P}$ , 则常常是切实可行的. 由此也可以看出耦合方法与概率距离之间的紧密联系.

耦合是什么? 我们可以给出如下一般定义. 给定一个可测空间  $(E, \mathcal{E})$  上的概率测度  $P_1$  和  $P_2$ , 所谓  $P_1$  和  $P_2$  的一个耦合  $\tilde{P}$ , 乃是乘积空间  $(E^2, \mathcal{E}^2)$  上的一个概率测度, 满足如下的边缘性:

$$\begin{aligned} \tilde{P}(A \times E) &= P_1(A), \\ \tilde{P}(E \times A) &= P_2(A), \quad A \in \mathcal{E}. \end{aligned}$$

对于给定的  $(E, \mathcal{E})$  上的转移概率函数  $p(t, x, A)$ , 当  $t$  和  $x$  固定时,  $p(t, x, \cdot)$  是一概率测度. 这样, 对于固定的  $t, x$  和  $y$ , 我们可定义  $p(t, x, \cdot)$  和  $p(t, y, \cdot)$  的耦合, 进而定义它们的 Wasserstein

距离. 可惜, 这样的定义没有多少用处, 因为  $p(t, x, \cdot)$  几乎都是没有显式表示的. 我们的做法是, 考虑过程的无穷小算子的耦合 (即马氏耦合). 还是以矩阵情形, 说明一下大体的想法. 前面已讲过, 对于  $Q = (q_{ij})$ ,  $q_{ij}$  表示过程从状态  $i$  跳到  $j (\neq i)$  的速度. 我们可以简单地表成

$$i \rightarrow j (\neq i) \quad \text{速率} \quad q_{ij}.$$

现在, 所谓耦合, 就是要构造乘积空间  $E^2$  上的矩阵. 仍用矩阵写法就很麻烦了, 但用上述表示则十分容易.

**例 1 (古典耦合).** 当  $i_1 \neq i_2$  时, 取

$$\begin{aligned} (i_1, i_2) &\rightarrow (j_1, i_2) & \text{速率} & q_{i_1 j_1} \quad (j_1 \neq i_1) \\ &\rightarrow (i_1, j_2) & \text{速率} & q_{i_2 j_2} \quad (j_2 \neq i_2) \end{aligned}$$

否则, 取

$$(i, i) \rightarrow (j, j) \quad \text{速率} \quad q_{ij} \quad (j \neq i).$$

此例的第一行表示, 保持耦合过程的第 2 个分量不动, 第 1 分量以速率  $q_{i_1 j_1}$  从  $i_1$  跳到  $j_1 \neq i_1$ . 第二行的意义类似. 这些是指两分量的出发点不同 ( $i_1 \neq i_2$ ) 的情形. 如出发点相同, 则让两分量一起跳好了, 这正是上述的末行. 对于后面的例子, 这第三行是一样的因而略去不写. 我们把这个耦合命名为“一见钟情”耦合, 因为两分量相遇之前, 各走各的路 (独立), 而从相遇那一刻开始, 就有了共同的生活道路.

**例 2 (齐步走耦合).** 设  $E = \{0, 1, 2, \dots\}$ .

$$\begin{aligned} (i_1, i_2) &\rightarrow (i_1 + k, i_2 + k) & \text{速率} & q_{i_1, i_1+k} \wedge q_{i_2, i_2+k} \\ &\rightarrow (i_1 + k, i_2) & \text{速率} & (q_{i_1, i_1+k} - q_{i_2, i_2+k})^+ \\ &\rightarrow (i_1, i_2 + k) & \text{速率} & (q_{i_2, i_2+k} - q_{i_1, i_1+k})^+, \end{aligned}$$

此处约定  $q_{ii} = 0$ ,  $q_{ij} = 0$  若  $j < 0$ .

此耦合的关键是第一行, 它的意义是: 两分量尽可能保持与出发时相同的距离. 如同部队操练时, 一旦下达“齐步走”的命令, 战士们在行进中总是尽可能保持他们之间的距离不变. 有了第一行之后, 后两行可自动写出来, 因而下例中我们将略去后两行不写.

从以上两个构造中, 人们多少可以感受到现代概率论 (随机过程) 的味道, 一种很强的概率直觉. 从这个意义上, 概率论更接近于几何. 对于三对角阵, 本质上由两列元素决定:  $q_{i, i+1} = b_i > 0$ ,  $q_{i, i-1} = a_i > 0$ . 对角线等于这两列和的反号. 对这种特殊情形, 我们还有另一种耦合.

**例 3 (反射耦合).** 如  $i_2 = i_1 + 1$ , 取

$$\begin{aligned} (i_1, i_2) &\rightarrow (i_1 - 1, i_2 + 1), & \text{速率} & a_{i_1} \wedge b_{i_2} \\ &\rightarrow (i_1 + 1, i_2), & \text{速率} & b_{i_1} \\ &\rightarrow (i_1, i_2 - 1), & \text{速率} & a_{i_2} \end{aligned}$$

如  $i_2 \geq i_1 + 2$ , 取

$$\begin{aligned} (i_1, i_2) &\rightarrow (i_1 - 1, i_2 + 1), & \text{速率} & a_{i_1} \wedge b_{i_2} \\ &\rightarrow (i_1 + 1, i_2 - 1), & \text{速率} & b_{i_1} \wedge a_{i_2}. \end{aligned}$$

由对称性, 可写出  $i_1 > i_2$  时的速率.

此构造的最后一行是自然的,因为它要求两个分量往尽可能靠近的地方走.然而,第一、四两行却很不正常,它要求两分量往相离远些的地方走.这是我们很长时间不能接受的.出乎意料的是,在许多场合,这种耦合却是最好的<sup>[7]</sup>.对于其中的奥妙,此处不可能详述.只是说明一句:所有这些想法,正是最终找到第一特征值下界估计变分公式的主要武器.

虽然概率距离可追溯到百年以上,耦合方法可追溯到1937年.然而,它们的兴起和发展毫无疑问应归功于随机场和粒子系统研究.事实上,它们已是后两学科研究的最重要的数学工具.概率距离和耦合方法是二十多年来概率论的最重要的发展之一,其重要性已逐渐被大家所认识,并且已经写进好几本研究生的基本教材.陈木法发展了耦合理论,主要有三方面:1)最早考察问题的根本点:即相应于给定的边缘无穷小算子,耦合无穷小算子的构造.得到连续时间随机过程的耦合的第一条基本定理,即跳过程耦合的基本定理.这些成果也已被收入Anderson的总结性论著<sup>[1]</sup>.2)把概率度量与耦合方法两方面的研究结合起来.首次找到了耦合的分类和优化的一种途径.发现了一些意想不到的最优马氏耦合.这结束了只限于变差距离的耦合的研究状况.3)更进一步,解决了距离关于耦合的分类和优化问题.以上耦合研究的“三部曲”,更新了耦合理论,也是研究上述椭圆算子和流形上的谱隙估计等问题的基石.事实上,上述最优估计全部来源于最优耦合.耦合方法是贯穿陈和他的研究集体的一系列研究工作的主要数学工具.详见综述报告[8]及[5].新近,张绍义和张余辉关于最优耦合的存在性和构造取得了重要进展.

### 参考文献

- [1]. Anderson, W. J. (1991), *Continuous-Time Markov Chains*, Springer Series in Statistics.
- [2]. Bérard, P. H. (1986), *Spectral Geometry: Direct and Inverse Problem*, LNM. vol. 1207, Springer-Verlag.
- [3]. I. Chavel (1984), *Eigenvalues in Riemannian Geometry*, Academic Press.
- [4]. 陈木法 (1986), 跳过程与粒子系统, 北京师大出版社, 现代数学丛书.
- [5]. Chen, M. F. (1992), *From Markov Chains to Non-Equilibrium Particle Systems*, World Scientific. 国内重印版: 1994, 世界图书出版公司.
- [6]. 陈木法 (1992), 经济最优化的随机模型 (I), (II), 应用概率统计杂志 8:3, 289-294; 8:4, 374-377.
- [7]. Chen, M. F. (1994), *Optimal Markovian couplings and applications*, Acta Math. Sin. New Ser. 10:3, 260-275.
- [8]. Chen, M. F. (1997), *Coupling, spectral gap and related topics* (I)–(III), Chin. Sci. Bulletin, (I): 42:14, 1472–1477 (中文版), 42:16, 1321–1327 (英文版); (II): 42:15, 1585–1591 (中文版), 42:17, 1409–1416 (英文版); (III): 42:16, 1696–1703 (中文版), 42:18, 1497–1505 (英文版).
- [9]. Chen, M. F. (1997), *Reaction-diffusion processes*, Chin. Sci. Bulletin.
- [10]. Chen, M. F. and Wang, F. Y. (1997), *General formula for lower bound of the first eigenvalue*, Sci. Sin. 27:1, 34–42 (中文版); 40:4, 384–394 (英文版).
- [11]. 陈亚浙、吴兰成 (1991), 二阶椭圆型方程与椭圆型方程组, 科学出版社.
- [12]. 戴永隆 (1993), 马尔可夫振动问题, 广东科技出版社, 广州.
- [13]. Durrett, R. (1988), *Lecture Notes on Particle Systems and Percolation*, Wadsworth and Brooks / Cole, Pacific Grove, Calif.
- [14]. Georgii, H. O. (1979), *Canonical Gibbs States*, vol. 761, Springer-Verlag, LNM..
- [15]. Guo, M. Z., Papanicolaou, G. C. and Varadhan, S. R. S. (1988), *Non-linear diffusion limit for a system with nearest neighbor interactions*, Comm. Math. Phys. **118**, 31–59.
- [16]. Hou, Z. T. and Guo, Q. F. (1978), *Time-Homogeneous Markov Processes with Countable State Space* (In Chinese), Beijing Sci. Press. English translation (1988): Beijing Sci. Press and Springer-Verlag.
- [17]. 康立山等著 (1994), 非数值并行算法 (第一册), 科学出版社.
- [18]. Kesten, H. (1982), *Percolation Theory for Mathematicians*, Birkhäuser.
- [19]. Liggett, T. M. (1985), *Interacting Particle Systems*, Springer-Verlag.
- [20]. Smale, S., 参见《数学译林》(1986), 5:1 和 5:2.

- [21]. Sinclair, A. (1993), *Algorithms for Random Generation and Counting: A Markov Chain Approach*, Birkhäuser.
- [22]. 王梓坤和杨向群 (1992), *The Birth and Death Processes and Markov Chains*, Springer-Verlag, Science Press.
- [23]. 杨向群 (1990), *The Construction Theory of Denumerable Markov Processes*, John Wiley & Sons.
- [24]. 严士键 (1989), 无穷粒子马尔可夫过程引论, 北京师范大学出版社.
- [25]. 丘成桐、孙理察 (1988), 微分几何, 科学出版社.
- [26]. Blum, L., Cucker, F., Shub, M. and Smale, S. (1998), *Complexity and Real Computation*, preprint.
- [27]. 陈大岳、冯建峰和钱敏平 (1997), 三维随机 *Ising* 模型的亚稳态 (I), (II), 中国科学 27:6, 504–513; 27:7, 618–623.
- [28]. Fang, H. T., Gong, G. L., and Qian, M. P. (1997), *Annealing of iterated stochastic schemes*, Siam J. Optim. & Control.
- [29]. Fang, H. T., Gong, G. L., and Qian, M. P. (1997), *An improved annealing method and its large time behavior*, Stoch. Proc. & their Appl.
- [30]. 侯振挺等 (1994), 马尔可夫过程的 *Q* 矩阵问题, 湖南科技出版社.