arXiv:2301.02426v2 [math.ST] 6 May 2024

已提交至伯努利

重新审视椭圆切片采样的可逆性

MAREIKE HASENPFLUG[1,a], VIACHESLAV TELEZHNIKOV[2,c]和 DANIEL RUDOLF[1,b]

[1]帕绍大学计算机科学与数学学院，*Innstraße 33, 94032* 帕绍，a*mareike.hasenplfug@uni-passau.de*，b*daniel.rudolf@uni-passau.de*[2]此前名为 *Viacheslav Natarovskii*，*Expert Analytics GmbH*，*Hubertusstraße 83, 82131* 高廷，c*viacheslav.telezhnikov@expertanalytics.de*

我们将Murray、Adams 和 MacKay（2010）提出的椭圆切片采样（一种马尔可夫链转移核）推广到无限维可分希尔伯特空间，并讨论其良定义性。我们指出一项正则性要求，给出所需可逆性的另一种证明，并证明该方法会诱导出一个正半定马尔可夫算子。证明上述结果的关键在于分析一种收缩马尔可夫链，这一分析本身也可能颇具价值。

关键词：椭圆切片采样；可逆性；收缩过程

1. 引言

在贝叶斯推断中，马尔可夫链蒙特卡罗模拟是近似采样后验分布的主要工具之一。椭圆切片采样（ESS）由 Murray、Adams 和 MacKay (2010) 提出，它是一种广受欢迎的算法转移机制，参见例如 Lie 等 (2023)、Murray 和 Graham (2016)、Nishihara、Murray 和 Adams (2014)，由此产生的马尔可夫链适用于近似采样这一目标。

椭圆切片采样的思想基于一种特定的高斯 Metropolis 随机游走，参见 Neal(1999)；这种随机游走如今有时也称为预条件 Crank-Nicolson Metropolis（参见 Cotter 等 (2013)、Rudolfand Sprungk (2018)），而收缩过程同样源自 Neal，参见 Neal (2003)。给定当前状态后，随机确定一个合适的接受区域（水平集）和一个椭圆。随后，利用上述收缩过程，在椭圆与接受区域的交集上生成马尔可夫链的下一个状态。

从计算角度看，其优势在于这个交集是一个一维紧集，其大小会以一种方式变化，使得每次迭代的跳跃距离都能根据当前位置和目标分布“自适应”。此外，这些定制的“探测区域”保证了对状态空间的充分探索。与高斯 Metropolis 随机游走相比，椭圆切片采样的优势在于不会出现拒绝，无需调整步长参数，而且可以进行更大的跳跃，因为可选的更新方式更加丰富，参见 Murray、Adams 和 MacKay (2010)。

从理论角度来看，近期 Natarovskii、Rudolf 和 Sprungk （2021a证明了在对目标分布的弱假设下，有限维空间中的几何收敛。此外，Gessner、Kanjilal 和 Hennig （2020）、Grenioux 等人 （2023）、Murray、Adams 和 MacKay （2010）、Natarovskii、Rudolf 和 Sprungk （2021a）、Ray、Pati 和 Bhattacharya （2020）的数值实验表明，ESS 具有与维度无关的性能。这促使我们探讨 ESS 在（可能是）无限维可分希尔伯特空间上的良定义性。通常，在新场景下研究马尔可夫链抽样时，几乎总是首先需要

验证正确的目标平稳分布。其重要性在于，所有收敛结果——从普通的遍历性（例如参见 Tierney （1994）到全变差距离上的收敛（参见 Meyn 和 Tweedie（2009））和 Wasserstein距离上的收敛（参见（Douc 等人，2018，第 20 节）），再到对时间平均值具有影响的谱隙（例如参见 Jones（2004）、Rudolf （2012））——都依赖于目标分布就是平稳分布。因此，我们着重验证可逆性，因为可逆性意味着平稳性，并且（相当普遍地）是后续任何收敛分析所依据的基础。我们对 ESS 的贡献包括四个方面：

1. 该算法包含一个“收缩循环”，我们为感兴趣的分布给出了该循环终止的充分条件，从而保证转移机制及相应马尔可夫链的良定义性。2. 我们阐明，椭圆上的过程实际上依赖于 $[0, 2\pi)$ 上的一个马尔可夫链；相对于经适当变换的接受区域上的均匀分布，该链是可逆的。3. 通过为（Murray、Adams 和 MacKay，2010，第 2 节 2.3 节）提供替代论证，我们证明了椭圆切片采样转移核在无限维情形下的可逆性；这尤其意味着目标测度是平稳分布。4. 我们证明，由椭圆切片采样转移核诱导的马尔可夫算子，在关于目标分布平方可积的函数空间上是半正定的。

与提出椭圆切片采样的有限维框架不同，我们考虑一个可能为无限维的情形。

对于这类情形，除了前文提到的预条件 Crank–Nicolson Metropolis 及其各种改进方法（见 Cotter 等人 (2013)、Law (2014)、Pinski 等人 (2015)、RudolfandSprungk(2018)），还有多种其他马尔可夫链蒙特卡罗方法。例如，Metropolis 调整的 Langevin 算法（见Cotter 等人(2013）、哈密顿蒙特卡罗抽样（见 Beskos 等人 (2011)、Ottobre 等人(2016)），以及融入后验分布几何结构的方法（见 Beskos 等人 (2017)、Cui、Law 和 Marzouk(2016)）。ESS 方法的一个优点或许在于，它既不需要了解梯度，也不需要调整步长参数。

目标分布定义在一个无穷维可分希尔伯特空间 $\mathcal{H}$ 上，该空间配备了相应的 Borel σ-代数 $\mathcal{B}(\mathcal{H})$。我们将看到，正如 Lie 等人 (2023) 已指出的那样，ESS 在此框架下定义良好。我们将 $\rho : \mathcal{H} \to (0, \infty)$ 视为似然函数，并将定义在 上的高斯参考测度 $\mu_0 = \mathcal{N}(0, C)$ 作为先验分布，其中 $C: \mathcal{H} \to \mathcal{H}$ 是一个非奇异协方差算子[1]。于是，所关注的概率测度，即后验/目标分布，记为 μ，其定义如下：

$$\mu(\mathrm{d}x) = \frac{1}{Z}\rho(x)\mu_0(\mathrm{d}x)$$

归一化常数为 $Z = \int_{\mathcal{H}} \rho(x)\mu_0(\mathrm{d}x)$。下面我们将考虑用于对 μ 进行近似采样的椭圆切片采样（ESS），并特别证明：若 ρ 是下半连续的，即 ρ的超水平集是开集（位于 H 内），那么收缩过程中的 while 循环会终止，且该转移机制会导出一个相对于 μ可逆的转移核。

我们简要概述本文的结构。在第 2 节开头，我们以算法形式给出一般设定和转移机制。接着，我们通过一个简单示例说明算法中所设终止准则存在的问题，并推导出与圆上的收缩过程相对应的转移核表示。在第2.3 节中，我们

[1]T这意味着 $C : \mathrm{H} \to \mathrm{H}$ 是一个有界线性、自伴、正定的迹类算子，且 $\ker C = \{0\}$.

证明上述核相对于圆的某个子集上的适当均匀分布是可逆的。最后，在第 3 节中，我们展示收缩过程的可逆性如何延续到椭圆切片采样（ESS）的转移核中，并由此得到相应马尔可夫算子的半正定性。

2. 预备知识与记号

我们以算法形式陈述椭圆切片采样转移机制的两个等价版本，并给出本文所用的记号。设 $(\Omega, \mathcal{F}, \mathbb{P})$ 为后文所用所有随机变量所依赖的概率空间。在实数轴 $\mathbb{R}$ 上，配备其标准博雷尔 σ- 代数 $\mathcal{B}(\mathbb{R})$，令 $\lambda(\cdot)$ 表示勒贝格测度。对于有界 $I \in \mathcal{B}(\mathbb{R})$，若 $\lambda(I) > 0$ 则令 U_I 为 I 上的均匀分布。算法 2.1 展示了 Murray、Adams 和 MacKay（2010）所述的椭圆切片采样转移机制。为便于下文分析，我们将其重新表述并拆分为

算法 2.1椭圆切片采样

输入： ρ 和 $x_{\text{in}} \in \mathcal{H}$，视为当前状态 ；

输出： $x_{\text{out}} \in \mathcal{H}$，视为下一状态；

1：抽取 $T \sim \mathrm{U}_{(0,\rho(x_{\text{in}}))}$，记结果为 t；2：抽取 $W \sim \mu_0 = \mathrm{N}(0, C)$，记结果为 w；3：抽取 $\Gamma \sim \mathrm{U}_{[0,2\pi)}$，记结果为 γ；4：设定 $\gamma^{\min} := \gamma - 2\pi$ 并设定 $\gamma^{\max} := \gamma$；5：当 $\rho(\cos(\gamma)x_{\text{in}} + \sin(\gamma)w) \le t$ 时，执行 6：如果 $\gamma < 0$，则执行 7：设定 $\gamma^{\min} := \gamma$；8：否则执行 9：设定 $\gamma^{\max} := \gamma$；10：结束 if；11：抽取 $\Gamma \sim \mathrm{U}_{(\gamma^{\min}, \gamma^{\max})}$，记结果为 γ；12：结束 while；13：返回 $x_{\text{out}} := \cos(\gamma)x_{\text{in}} + \sin(\gamma)w$。

tra算法 2.1 的过渡机制。为此，定义 $t \ge 0$ 关于 ρ 的（超）水平集 t as

$$\mathcal{H}(t) := \{x \in \mathcal{H} : \rho(x) > t\}$$

对于 $\alpha, \beta \in [0, 2\pi)$，令

$$I(\alpha,\beta) := \begin{cases} [0,\beta) \cup [\alpha, 2\pi), & \alpha > \beta \\ [\alpha,\beta), & \alpha < \beta \\ [0, 2\pi), & \alpha = \beta, \end{cases} \qquad J(\alpha,\beta) := \begin{cases} [0,\beta) \cup [\alpha, 2\pi), & \alpha > \beta \\ [\alpha,\beta), & \alpha < \beta \\ \emptyset, & \alpha = \beta \end{cases}$$

用来表示符合圆的几何结构的区间。注意，$I(\alpha, \beta) \cap J(\beta, \alpha) = \emptyset$且 $I(\alpha, \beta) \cup J(\beta, \alpha) = [0, 2\pi)$。此外，对于 $\alpha \neq \beta$，我们有 $I(\alpha, \beta) = J(\alpha, \beta)$。通过区分不同情形，可以直接得到一个有用的恒等式：

引理 *2.1*。设 $\alpha,\beta,\gamma\in[0,2\pi)$。若 $\gamma\neq\alpha$，则有 $\mathbf{1}_{I(\alpha,\beta)}(\gamma)=\mathbf{1}_{J(\beta,\gamma)}(\alpha)$。反之，若 $\gamma\neq\beta$，则有 $\mathbf{1}_{J(\alpha,\beta)}(\gamma)=\mathbf{1}_{I(\gamma,\alpha)}(\beta)$。

对于给定的 $x,w\in\mathrm{H}$，定义函数 $p_{x,w}$：$[0,2\pi)\to\mathrm{H}$ 如下

$$p_{x,w}(\theta):=\cos(\theta)x+\sin(\theta)w,$$

它描述了 H 中的一个椭圆，其共轭直径由 x,w 确定。我们提醒读者注意 $p_{x,w}$ 原像的定义：对于 $A\in\mathrm{B}(\mathrm{H})$，其定义如下

$$p_{x,w}^{-1}(A):=\{\theta\in[0,2\pi):p_{x,w}(\theta)\in A\}.$$

它确定 $[0,2\pi)$ 中经由 $p_{x,w}$ 映射到与 A相交的椭圆上的元素的那一部分。在前文对算法 2.1 的重新表述中，我们旨在突出椭圆切片采样方法的结构。该方法见算法2.3，并将算法 2.2 作为内置过程调用。此过程在集合 $S\in\mathcal{B}([0,2\pi))$上给出一种转移机制。

算法 2.2 收缩，调用方式为 shrink(θ_{in},S)

输入： $S\in\mathrm{B}([0,2\pi))$，$\theta_{\text{in}}\in S$ 视为当前状态；
输出： $\theta_{\text{out}}\in S$ 被视为下一状态；

1：设置 $i:=1$ 并抽取 $\Gamma_i\sim\mathrm{U}_{[0,2\pi)}$，将结果记为 γ_i；
2：设置 $\gamma_i^{\min}:=\gamma_i$ 并设置 $\gamma_i^{\max}:=\gamma_i$；
3：**当** $\gamma_i\notin S$ **时**
4：　**如果** $\gamma_i\in J(\gamma_i^{\min},\ \theta_{\text{in}})$ **那么**
5：　　设置 $\gamma_{i+1}^{\min}:=\gamma_i$ 并设置 $\gamma_{i+1}^{\max}:=\gamma_i^{\max}$；
6：　**else**
7：　　设置 $\gamma_{i+1}^{\min}:=\gamma_i^{\min}$和 $\gamma_{i+1}^{\max}:=\gamma_i$；
8：　**结束 if**
9：　抽取 $\Gamma_{i+1}\sim\mathrm{U}_I(\gamma_{i+1}^{\min},\ \gamma_{i+1}^{\max})$，将结果记为 γ_{i+1}；
10：　设置 $i:=i+1$；
11：**结束 while**
12：**返回** $\theta_{\text{out}}:=\gamma_i$。

算法 2.3重构椭圆切片采样

输入： ρ 和 $x_{\text{in}}\in\mathrm{H}$ 被视为当前状态；
输出： $x_{\text{out}}\in\mathrm{H}$ 被视为下一状态；

1：抽取 $T\sim\mathrm{U}_{(0,\rho(x_{\text{in}}))}$，将结果记为 t；
2：抽取$W\sim\mu_0=\mathrm{N}(0,\ C)$，将结果记为 w；
3：设置 $\gamma:=$收缩$(0,\ p_{x_{\text{in}},w}^{-1}(\mathrm{H}(t)))$；　　(算法 2.2)
4：**返回** $x_{\text{out}}:=\cos(\gamma)x_{\text{in}}+\sin(\gamma)w$。

比较算法 2.1 和算法 2.3 可以发现，第 1 行、第 2 行和返回行（在计算出 γ 之后）一致。给定实现 t 和 w，算法 2.1 的第 3 行到第 12 行，

包括 while 循环在内，对应于在算法 2.2 中调用算法 2.3 的收缩过程，输入为 $\theta_{\text{in}} = 0$ 和 $S = p_{x_{\text{in}},w}^{-1}(\mathrm{H}(t))$. 要确认这些部分也一致，请注意

$$p_{x_{\text{in}},w}^{-1}(\mathcal{H}(t)) = \{\theta \in [0, 2\pi) : \rho(\cos(\theta)x_{\text{in}} + \sin(\theta)w) > t\}$$

因此，while 循环中的终止准则保持不变。在第一次迭代中，两种算法抽取的角度始终在长度为 2π 的区间上服从均匀分布。此外，无论是在此处还是在一般情况下，构造算法 2.1 的 while 循环中的收缩区间时，都会利用函数 $p_{x_{\text{in}},w}$ 的 2π-周期性；而在算法 2.2 中，对于给定的 $\alpha, \beta \in [0, 2\pi)$，我们使用广义区间 $I(\alpha, \beta)$ 和 $J(\alpha, \beta)$。最后，为了让你确信两种算法确实执行相同的转移，说明如何在广义区间上进行均匀采样会很有帮助。具体而言，对于 $\alpha < \beta$，只需在 $[\alpha, \beta)$ 上进行均匀采样，即可得到关于 $\mathrm{U}_{I(\alpha,\beta)}$ 的一个实现值。对于 $\alpha \geq \beta$，若要在 $I(\alpha, \beta)$ 上进行均匀采样，则抽取 $V \sim \mathrm{U}_{[\alpha-2\pi,\beta)}$，得到结果 v，并在 $v \in [\alpha - 2\pi, 0)$ 时将输出设为 $v + 2\pi$，否则设为 v。在算法 2.2 中采用这一过程来实现 $\mathrm{U}_{I(\alpha,\beta)}$，并使用与算法 2.1 中区间采样相同的随机数驱动，最终会得到相同的转移和角度[2]。请注意，以算法 2.2 的形式表示椭圆切片采样，为转移机制提供了新的视角，这对于理论研究而非实际应用更有价值。例如，若按字面理解，访问 $S = p_{x_{\text{in}}}^{-1}(\mathrm{H}(t))$ 显然在计算上不可行；而在算法 2.1 中，只需一个关于 $p_{x_{\text{in}}}^{-1}(\mathrm{H}(t))$ 的成员资格预言机即可。

2.1. 收缩过程的性质与记号

算法 2.2 （以及算法 2.1）的收缩过程只有在 while 循环终止时才是良定义的。特别地，若 $\lambda(S) = 0$，则对于任意 $I(\alpha, \beta)$，满足 $\alpha, \beta \in [0, 2\pi)$，且 $V \sim \mathrm{U}_{I(\alpha,\beta)}$，我们有

$$\mathbb{P}(V \in S) = \mathcal{U}_{I(\alpha,\beta)}(S) = \frac{\lambda(S \cap I(\alpha, \beta))}{\lambda(I(\alpha, \beta))} = 0. \tag{1}$$

因此，对于输入 $S \in \mathcal{B}([0, 2\pi))$ 满足 $\lambda(S) = 0$ 的情形，算法 2.2的收缩过程几乎必然不终止，因为在第 9 行中会从某个合适的广义区间内均匀抽取，而根据（1），落入 S 的概率为零。下面通过一个玩具情景的示例，借助算法 2.3 说明良定义性问题。

例 2.2。对于 $d \in \mathbb{N}$，考虑 $\mathrm{H} = \mathbb{R}^d$，并令 $\varepsilon > 0$ 以及 $\mu = \mathrm{N}(0, I)$ 为 $\mathbb{R}^d$ 中的标准正态分布，其中 $I \in \mathbb{R}^{d\times d}$ 为单位矩阵。此外，令 $\rho\colon \mathbb{R}^d \to [\varepsilon, 1 + \varepsilon]$ 定义为

$$\rho(x) = \mathbf{1}_{[0,1]^d}(x) + \varepsilon, \qquad x \in \mathbb{R}^d.$$

注意，对于 $t > \varepsilon$，$\mathrm{H}(t) = [0, 1]^d$。我们看到，这一点是一个闭集的事实可能会（对于某些输入）导致良定义性问题。对于

$$w \in \left\{(\widetilde{w}^{(1)}, \ldots, \widetilde{w}^{(d)}) \in \mathbb{R}^d : \exists i, j \in \{1, \ldots, d\} \text{ s.t. } \widetilde{w}^{(i)} < 0,\ \widetilde{w}^{(j)} > 0\right\}$$

[2]角度与收缩后的区间在变换为 $[0, 2\pi)$ 后重合。

我们有

$$p_{0,w}([0,2\pi)) = \{\widetilde{w} \in \mathbb{R}^d \colon \widetilde{w} = sw,\ s \in [-1,1]\}, \qquad p_{0,w}([0,2\pi)) \cap \mathcal{H}(t) = \{0\} \subset \mathbb{R}^d,$$

使得 $p_{0,w}^{-1}(\mathcal{H}(t)) = \{0\} \subset [0,2\pi)$。对于随机变量 T 和 W 如算法 2.3 中所述，我们得到

$$\mathbb{P}\left(\lambda(p_{0,W}^{-1}(\mathcal{H}(T))) = 0\right) = \frac{2^d - 2}{2^d(1+\varepsilon)}.$$

因此，对于输入 $x_{\text{in}} = 0$ 和 ρ，前一种情况下，收缩过程中的 while 循环不会终止。

下面我们介绍一些数学对象，以便给出充分条件，保证上述 while 循环几乎必然终止，并确保收缩过程具有所需的可逆性。

我们先介绍一些记号。对于定义在可能不同的可测空间上的概率测度 μ，ν，相应的笛卡尔积空间上的乘积测度记为 $\mu \otimes \nu$。此外，对于（任意可测空间上）位于 ν 处的狄拉克测度，我们记为 $\delta_\nu(\cdot)$。给定两个随机变量/向量 X，Y，我们将 X 的分布记为 $\mathbb{P}_X$，将给定 Y 时 X 的条件分布记为 $\mathbb{P}_{X|Y}$。

固定 $S \in \mathrm{B}([0,2\pi))$，并令 $\theta \in S$ 满足 $\theta_{\text{in}} = \theta$。定义

$$\Lambda := \{(\gamma, \gamma^{\min}, \gamma^{\max}) \in [0,2\pi)^3 \colon \gamma \in I(\gamma^{\min}, \gamma^{\max})\},$$
$$\Lambda_\theta := \{(\gamma, \gamma^{\min}, \gamma^{\max}) \in [0,2\pi)^3 \colon \gamma, \theta \in I(\gamma^{\min}, \gamma^{\max}), \gamma \neq \theta\}.$$

将 $z_1 = (\gamma_1, \gamma_1^{\min}, \gamma_1^{\max})$ 视为算法 2.2 所得的随机向量 $Z_1 = (\Gamma_1, \Gamma_1^{\min}, \Gamma_1^{\max})$ 在 $([0,2\pi)^3, \mathcal{B}([0,2\pi)^3)$ 的一次实现，根据上述过程第 1–2 行可知，Z_1 的分布为

$$\mathbb{P}_{Z_1}(C) = \int_0^{2\pi} \delta_{(\gamma,\gamma,\gamma)}(C)\, \frac{\mathrm{d}\gamma}{2\pi}, \qquad C \in \mathcal{B}([0,2\pi)^3). \tag{2}$$

假设 Θ 是一个映射到 S 且服从分布 U_S 的随机变量，并考虑 $\theta \in S$，其中 $\theta = \theta_{\text{in}}$ 是 Θ 的一个实现。为方便起见，我们也将 Θ 称为锚点（随机变量）。给定 $\Theta = \theta$，注意到 $\Gamma_1 \in I(\Gamma_1^{\min}, \Gamma_1^{\max})$、$\theta \in S \subseteq I(\Gamma_1^{\min}, \Gamma_1^{\max})$ 以及 $\Gamma_1 \neq \theta$ 几乎必然成立，使得对于几乎所有 $\omega \in \Omega$，$Z_1(\omega) \in \Lambda_\theta$。此外，给定 $\Theta = \theta$，序列 $(z_n)_{n\in\mathbb{N}}$，其中 $z_n = (\gamma_n, \gamma_n^{\min}, \gamma_n^{\max}) \in [0,2\pi)$ 是通过迭代算法 2.2 中第 4–9 行（忽略 while 循环中的停止准则）得到的 ^{3}„ 是一个随机变量序列 $(Z_n)_{n\in\mathbb{N}}$ 的实现，并且 $Z_n = (\Gamma_n, \Gamma_n^{\min}, \Gamma_n^{\max})$。为便于说明，我们在图 1 中给出了在给定 $\Theta = \theta$ 条件下 $(Z_n)_{n\in\mathbb{N}}$ 的依赖图。

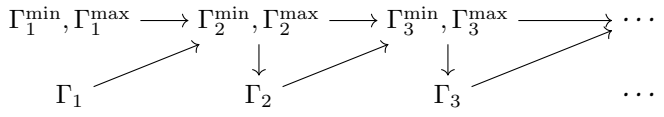

图 1。给定 $\Theta = \theta$ 时，$(Z_n)_{n\in\mathbb{N}}$ 的依赖图。

此外，我们将 $(Z_n)_{n\in\mathbb{N}}$ 称为未停止收缩序列。在给定锚点 Θ 的条件下，未停止收缩序列满足马尔可夫性质，即:

$$\mathbb{P}_{Z_{n+1}|Z_1,\ldots,Z_n,\Theta}(C) = \mathbb{P}_{Z_{n+1}|Z_n,\Theta}(C), \qquad C \in \mathcal{B}([0,2\pi)^3). \tag{3}$$

根据算法描述（参见算法 2.2），对于 $A \in \mathcal{B}([0,2\pi))$, $B \in \mathcal{B}([0,2\pi)^2)$，可以直接读出以下条件分布的性质

$$\mathbb{P}_{\Gamma_{n+1}|\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max},Z_n,\ldots,Z_1,\Theta}(A) = \mathbb{P}_{\Gamma_{n+1}|\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}}(A) = \mathcal{U}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(A), \tag{4}$$

$$\mathbb{P}_{\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}|Z_n,\Theta}(B) = \mathbf{1}_{J(\Gamma_n^{\min},\Theta)}(\Gamma_n)\delta_{(\Gamma_n,\Gamma_n^{\max})}(B) + \mathbf{1}_{I(\Theta,\Gamma_n^{\max})}(\Gamma_n)\delta_{(\Gamma_n^{\min},\Gamma_n)}(B). \tag{5}$$

由（4）可推得 $\mathbb{P}_{\Gamma_{n+1}|\Gamma_{n+1}^{\min},\ \Gamma_{n+1}^{\max},\ Z_n,\ \ldots\ldots Z_1,\ \Theta} = \mathbb{P}_{\Gamma_{n+1}|\Gamma_{n+1}^{\min},\ \Gamma_{n+1}^{\max},\ Z_n,\ \Theta}$，因此根据（5），（3）的右端可以表示为

$$\begin{aligned}
\mathbb{P}_{Z_{n+1}|Z_n,\Theta}(A\times B) &= \int_B \mathbb{P}_{\Gamma_{n+1}|\Gamma_{n+1}^{\min}=\gamma^{\min},\Gamma_{n+1}^{\max}=\gamma^{\max},Z_n,\Theta}(A)\,\mathbb{P}_{\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}|Z_n,\Theta}(\mathrm{d}\gamma^{\min}\mathrm{d}\gamma^{\max}) \\
&= \mathbf{1}_{J(\Gamma_n^{\min},\Theta)}(\Gamma_n)\int_B \mathcal{U}_{I(\gamma^{\min},\gamma^{\max})}(A)\delta_{(\Gamma_n,\Gamma_n^{\max})}(\mathrm{d}\gamma^{\min}\mathrm{d}\gamma^{\max}) \\
&\quad + \mathbf{1}_{I(\Theta,\Gamma_n^{\max})}(\Gamma_n)\int_B \mathcal{U}_{I(\gamma^{\min},\gamma^{\max})}(A)\delta_{(\Gamma_n^{\min},\Gamma_n)}(\mathrm{d}\gamma^{\min}\mathrm{d}\gamma^{\max}).
\end{aligned}$$

我们可以用转移核来改写这一表达式。给定 $\Theta=\theta$ 和当前状态 $z=(\gamma,\ \gamma^{\min},\ \gamma^{\max}) \in \Lambda_\theta$，我们在 $\Lambda_\theta \times \mathrm{B}([0,2\pi)^3)$ 上定义转移核 R_θ，如下:

$$\begin{aligned}
&R_\theta((\gamma,\gamma^{\min},\gamma^{\max}),C) := \mathbb{P}_{Z_{n+1}|Z_n=(\gamma,\gamma^{\min},\gamma^{\max}),\Theta=\theta}(C) \\
&= \mathbf{1}_{J(\gamma^{\min},\theta)}(\gamma)\int_C \mathcal{U}_{I(\alpha^{\min},\alpha^{\max})}(\mathrm{d}\alpha)\delta_\gamma(\mathrm{d}\alpha^{\min})\delta_{\gamma^{\max}}(\mathrm{d}\alpha^{\max}) \\
&\quad + \mathbf{1}_{I(\theta,\gamma^{\max})}(\gamma)\int_C \mathcal{U}_{I(\alpha^{\min},\alpha^{\max})}(\mathrm{d}\alpha)\delta_{\gamma^{\min}}(\mathrm{d}\alpha^{\min})\delta_\gamma(\mathrm{d}\alpha^{\max}), \qquad C\in\mathcal{B}([0,2\pi)^3).
\end{aligned}$$

我们注意到以下性质:

引理 2.3。对于 $\theta \in S$，任意 $z=(\gamma,\gamma^{\min},\gamma^{\max}) \in \Lambda_\theta$ 以及任意 $C\in\mathcal{B}([0,2\pi)^3)$，以锚点 $\Theta=\theta$ 为条件的未停止收缩序列的转移核R_θ 满足

$$\begin{aligned}
&R_\theta((\gamma,\gamma^{\min},\gamma^{\max}),C) \\
&= \mathbf{1}_{I(\gamma,\gamma^{\min})}(\theta)\cdot\mathcal{U}_{I(\gamma,\gamma^{\max})}\otimes\delta_{(\gamma,\gamma^{\max})}(C) + \mathbf{1}_{J(\gamma^{\max},\gamma)}(\theta)\cdot\mathcal{U}_{I(\gamma^{\min},\gamma)}\otimes\delta_{(\gamma^{\min},\gamma)}(C) \\
&= \mathbf{1}_{I(\gamma,\gamma^{\max})}(\theta)\cdot\mathcal{U}_{I(\gamma,\gamma^{\max})}\otimes\delta_{(\gamma,\gamma^{\max})}(C) + \mathbf{1}_{J(\gamma^{\min},\gamma)}(\theta)\cdot\mathcal{U}_{I(\gamma^{\min},\gamma)}\otimes\delta_{(\gamma^{\min},\gamma)}(C),
\end{aligned}$$

这反映了算法 2.2 的分类讨论。此外，我们有 $R_\theta((\gamma,\gamma^{\min},\gamma^{\max}),\Lambda_\theta)=1$。

证明。第一个等式由引理 2.1 得出，第二个等式则是考虑到 $z\in\Lambda_\theta$，特别是，$\theta\in I(\gamma^{\min},\ \gamma^{\max})$而得。还需证明 $R_\theta((\gamma,\ \gamma^{\min},\ \gamma^{\max}),\ \Lambda_\theta)=1$。注意到

$$\mathcal{U}_{I(\gamma,\gamma^{\max})}\otimes\delta_{(\gamma,\gamma^{\max})}(\Lambda_\theta) = \mathbf{1}_{I(\gamma,\gamma^{\max})}(\theta)\cdot\mathcal{U}_{I(\gamma,\gamma^{\max})}\left(I(\gamma,\gamma^{\max})\setminus\{\theta\}\right) = \mathbf{1}_{I(\gamma,\gamma^{\max})}(\theta)$$

并且根据相同的论证 $\mathrm{U}_{I(\gamma^{\min},\ \gamma)} \otimes \delta_{(\gamma^{\min},\gamma)}(\Lambda_\theta) = \mathbf{1}_{I(\gamma^{\min},\ \gamma)}(\theta)$。由于 $\gamma \in I(\gamma^{\min},\ \gamma^{\max})$，我们有

$$I(\gamma^{\min},\gamma^{\max}) = J(\gamma^{\min},\gamma) \cup I(\gamma,\gamma^{\max}) \quad \text{and} \quad J(\gamma^{\min},\gamma) \cap I(\gamma,\gamma^{\max}) = \emptyset,$$

使得

$$R_\theta(z,\Lambda_\theta) = \mathbf{1}_{I(\gamma,\gamma^{\max})}(\theta) + \mathbf{1}_{J(\gamma^{\min},\gamma)\cap I(\gamma^{\min},\gamma)}(\theta) = \mathbf{1}_{I(\gamma,\gamma^{\max})}(\theta) + \mathbf{1}_{J(\gamma^{\min},\gamma)}(\theta) = 1. \qquad \square$$

根据前一个引理，很容易得到转移核关于随机变量的一个有用表示。

引理 2.4. 对于任意 $\theta \in S$、任意 $z \in \Lambda_\theta$、任意 $C \in \mathcal{B}([0,2\pi)^3)$ 以及任意 $n \geq 2$，未停止收缩序列的转移核 R_θ 可表示为前几次迭代和锚点 $\Theta = \theta$ 的条件期望，如下所示：

$$R_\theta(z,C) = \mathbb{E}[\mathbf{1}_{I(\Gamma_n^{\min},\Gamma_n^{\max})}(\theta)\, \mathcal{U}_{I(\Gamma_n^{\min},\Gamma_n^{\max})} \otimes \delta_{(\Gamma_n^{\min},\Gamma_n^{\max})}(C) \mid Z_{n-1} = z, \Theta = \theta]. \tag{6}$$

我们补充一个关于 Θ 在给定 Z_1、……、Z_n 时的分布的性质，该性质的证明见附录 A.1

引理 2.5. 对于任意 $n \in \mathbb{N}$ 和 $A \in \mathcal{B}([0,2\pi))$，未停止收缩序列中锚点 Θ 在给定 $Z_1 = (\Gamma_1, \Gamma_1^{\min}, \Gamma_1^{\max})$、……、$Z_n = (\Gamma_n, \Gamma_n^{\min}, \Gamma_n^{\max})$ 时的条件分布为

$$\mathbb{P}_{\Theta|Z_1,\ldots,Z_n}(A) = \mathbb{P}_{\Theta|\Gamma_n^{\min},\Gamma_n^{\max}}(A) = \frac{\mathbb{P}_\Theta(A \cap I(\Gamma_n^{\min},\Gamma_n^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min},\Gamma_n^{\max}))}. \tag{7}$$

2.2. 收缩过程的停止

现在，我们的目标是将算法2.2中 while 循环里的停止准则纳入考虑。为此，我们引入 σ-代数 $\mathcal{F}_n := \sigma(Z_1, \ldots\ldots, Z_n)$ 以及 $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ 的自然滤过，其中 $(Z_n)_{n\in\mathbb{N}}$。我们将 while 循环的（随机）终止时间 τ_S 定义为首个满足 $n \in \mathbb{N}$ 的时刻，其中 Γ_n 属于 S，即：

$$\tau_S := \inf\left\{n \in \mathbb{N} : Z_n \in S \times [0,2\pi)^2\right\}, \tag{8}$$

按惯例，$\inf\emptyset = \infty$。注意，τ_S 是相对于自然滤过的停止时，因为

$$\{\tau_S = n\} = \bigcap_{k=1}^{n-1} \left\{Z_k \notin S \times [0,2\pi)^2\right\} \cap \left\{Z_n \in S \times [0,2\pi)^2\right\} \in \mathcal{F}_n,$$

对任意 $n \in \mathbb{N}$均成立。此外，当给定输入2.2 和 S 时，如果在事件 θ 条件下 while 循环几乎必然终止，则算法 $\Theta = \theta$ 的转移机制可以用转移核表示，即 $\mathbb{P}(\tau_S < \infty \mid \Theta = \theta) = 1$。为说明这一性质的充分条件，我们需要以下非正式定义。

定义 2.6. 集合 $S \in \mathrm{B}([0,2\pi))$ 若满足以下条件，则称其为圆周上的开集：对所有 $\theta \in S$，都存在一个 $\varepsilon > 0$，使得 $I(\theta - \varepsilon \text{ 模 } 2\pi,\ \theta + \varepsilon \text{ 模 } 2\pi) \subseteq S$。

换言之，如果 $S \in \mathrm{B}([0,2\pi))$ 是某个开集在 $[0,2\pi)$ 与一维球面之间的自然双射下的像，则称其为圆周上的开集。

引理 2.7. 假设 $S \in \mathcal{B}([0,2\pi))$ 是圆周上的非空开集。那么，对于任意 $\theta \in S$，我们有 $\mathbb{P}(\tau_S < \infty \mid \Theta = \theta) = 1$，即算法 2.2 中 *while* 循环在终止前的迭代次数所对应的停止时几乎必然有限。

证明。 根据 S 是圆周上的开集这一事实，以及 $\theta \in S$ 存在一个 $\varepsilon > 0$ 使得 $I_\theta := I(\theta_{\varepsilon,-}, \theta_{\varepsilon,+}) \subseteq S$，其中

$$\theta_{\varepsilon,-} = \theta - \varepsilon \mod 2\pi, \qquad \theta_{\varepsilon,+} = \theta + \varepsilon \mod 2\pi,$$

其中 $\theta \in I_\theta$。此外，注意到 $\lambda(I_\theta) = 2\varepsilon$。令 $\widetilde{S} := S \times [0,2\pi)^2$，并注意，对于任意 $\gamma^{\min}, \gamma^{\max} \in [0,2\pi)$，若 $\theta \in I(\gamma^{\min}, \gamma^{\max})$，则有

$$\begin{aligned}
\mathcal{U}_{I(\gamma^{\min},\gamma^{\max})}(S) &= \frac{\lambda(S \cap I(\gamma^{\min},\gamma^{\max}))}{\lambda(I(\gamma^{\min},\gamma^{\max}))} \\
&\geq \begin{cases} 1 & \gamma^{\min}, \gamma^{\max} \in I_\theta \\ \dfrac{\varepsilon}{\lambda(I(\gamma^{\min},\gamma^{\max}))} & \gamma^{\min} \in I_\theta, \gamma^{\max} \notin I_\theta \quad \text{or} \quad \gamma^{\min} \notin I_\theta, \gamma^{\max} \in I_\theta \\ \dfrac{2\varepsilon}{\lambda(I(\gamma^{\min},\gamma^{\max}))} & \gamma^{\min}, \gamma^{\max} \notin I_\theta \end{cases} \\
&\geq \frac{\varepsilon}{2\pi}.
\end{aligned}$$

利用这个估计，我们对任意 $z = (\gamma, \gamma^{\min}, \gamma^{\max}) \in \Lambda_\theta$ 得到

$$R_\theta(z, \widetilde{S}) = \mathbf{1}_{I(\gamma,\gamma^{\max})}(\theta)\mathcal{U}_{I(\gamma,\gamma^{\max})}(S) + \mathbf{1}_{J(\gamma^{\min},\gamma)}(\theta)\mathcal{U}_{I(\gamma^{\min},\gamma)}(S) \geq \frac{\varepsilon}{2\pi}.$$

回忆一下，Z_1 其中 $\mathbb{P}_{Z_1}$ 由（2）给出，满足 $Z_1 \in \Lambda_\theta$ 几乎必然成立。现在迭代应用前面的估计可得

$$\begin{aligned}
\mathbb{P}(Z_1, \ldots, Z_n \notin \widetilde{S}^n \mid \Theta = \theta) &= \overbrace{\int_{\widetilde{S}^c} \cdots \int_{\widetilde{S}^c}}^{n-1} R_\theta(z_{n-1}, \widetilde{S}^c) R_\theta(z_{n-2}, \mathrm{d}z_{n-1}) \cdots R_\theta(z_1, \mathrm{d}z_2) \mathbb{P}_{Z_1}(\mathrm{d}z_1) \\
&\leq \left(1 - \frac{\varepsilon}{2\pi}\right) \mathbb{P}(Z_1, \ldots, Z_{n-1} \notin \widetilde{S}^{n-1} \mid \Theta = \theta) \leq \cdots \leq \left(1 - \frac{\varepsilon}{2\pi}\right)^{n-1} \mathbb{P}_{Z_1}(\widetilde{S}^c) \leq \left(1 - \frac{\varepsilon}{2\pi}\right)^{n-1},
\end{aligned}$$

使得

$$\mathbb{P}(\tau_S = \infty \mid \Theta = \theta) \leq \lim_{n\to\infty} \mathbb{P}(\tau_S > n \mid \Theta = \theta) \leq \lim_{n\to\infty} \mathbb{P}(Z_1, \ldots, Z_n \notin \widetilde{S}^n \mid \Theta = \theta) \leq 0$$

证明至此完成。 □

推论 2.8. 假设似然函数/未归一化密度 $\rho\colon \mathcal{H} \to (0, \infty)$ 在 S 上可耕种的。

希尔伯特空间 $\mathcal{H}$ 是下半连续的，也就是说，所有水平集 $\mathcal{H}(t)$ 都是开集。那么

$$\mathbb{P}(\tau_{P_{x,w}^{-1}(\mathcal{H}(t))} < \infty \mid \Theta = 0) = 1, \qquad \forall x, w \in \mathcal{H} \quad \textit{and} \quad t \in (0, \rho(x)),$$

也就是说，在椭圆切片采样（*ESS*）中调用的算法 2.2的 *while* 循环直至终止所经历的迭代次数对应的停止时几乎必然有限。

证明。 由p_x的连续性w，以及 2π-周期性和$\mathcal{H}(t)$ 是开集这一事实，还可得$p_x^{-1}(w(\mathcal{H}(t))) \subseteq [0,2\pi)$是$0 \in p_x^{-1}(w(\mathcal{H}(t)))$所在圆周上的开集。因此，由引理 2.7可得结论。 □

前一个推论告诉我们，只要 ρ 是下半连续的，那么以2.2为输入调用算法 $S = p_x^{-1}(w(\mathcal{H}(t)))$ 和 $\theta_{\text{in}} = 0$ 几乎必然终止，因此算法 2.3 也会终止且定义良好。

注 2.9。 通常，未能终止的问题似乎不会对实际应用产生很大影响，因为大多数感兴趣的密度函数都具有开水平集。例如，每个连续密度都是下半连续的。即使它们存在一些离群点，使算法无法终止，在实践中，算法也会不断收缩；由于计算机以机器精度运算，收缩后的区间最终将无法与当前状态区分，因此算法最终会接受该区间，并将当前状态作为下一个样本返回。从这个意义上说，算法不会陷入无限循环。不过在这种情况下，算法的数学描述与实际实现并不一致。

2.3. 停止收缩过程的性质

现在，我们引入马尔可夫链 $(Z_n)_{n\in\mathbb{N}}$ 的停止随机变量 Z_{τ_S}。为正式定义事件 $\tau_S = \infty$ 上的该变量，取任意随机变量 Z_∞，并假设其关于 $\mathrm{F}_\infty := \sigma(Z_k,\ k \in \mathbb{N})$ 可测。我们定义

$$Z_{\tau_S}(\omega) := Z_\infty(\omega)\mathbf{1}_{\{\tau_S=\infty\}}(\omega) + \sum_{k=1}^{\infty} Z_k(\omega)\mathbf{1}_{\{\tau_S=k\}}(\omega), \qquad \omega \in \Omega.$$

注意，Z_{τ_S} 确实关于由 τ_S 诱导的 σ-代数可测

$$\mathcal{F}_{\tau_S} := \{A \in \mathcal{F} : A \cap \{\tau_S = k\} \in \mathcal{F}_k,\ k \in \mathbb{N}\},$$

因为对于任意 $A \in \mathrm{F}$ 和 $k \in \mathbb{N}$，我们有

$$\{Z_{\tau_S} \in A\} \cap \{\tau_S = k\} = \{Z_k \in A\} \cap \{\tau_S = k\} \in \mathcal{F}_k.$$

因此，$Z_{\tau_S} = (\Gamma_{\tau_S},\ \Gamma_{\tau_S}^{\min},\ \Gamma_{\tau_S}^{\max})$ 是一个取值于 $[0,2\pi)^3$ 的随机变量，其分量 Γ_{τ_S}，$\Gamma_{\tau_S}^{\min}$，$\Gamma_{\tau_S}^{\max}$ 是概率空间 $(\Omega,\ \mathrm{F}_{\tau_S},\ \mathbb{P})$ 上取值于 $[0,2\pi)$ 的随机变量。

现在，对于给定的 $S \in \mathcal{B}([0,2\pi))$ 以及任意的 $\theta_{\text{in}} \in S$，完成整个构造后，我们可以给出转移核 Q_S，它定义在 $S \times \mathcal{B}(S)$上，也称为收缩核，对应于算法 2.2 的转移机制。对于 $\theta_{\text{in}} \in S$ 和 $F \in \mathcal{B}(S)$，其定义如下

$$Q_S(\theta_{\text{in}}, F) = \mathbb{P}(\Gamma_{\tau_S} \in F, \tau_S < \infty \mid \Theta = \theta_{\text{in}}). \tag{9}$$

我们表述关于转移核 Q_S 的主要结果，该结果实质上用于验证椭圆切片采样的可逆性。

定理 2.10。 设 $S \in \mathcal{B}([0,2\pi))$ 是圆周上的非空开集。那么，收缩核Q_S 关于定义在 S 上的均匀分布 $\mathcal{U}_S$ 可逆。

证明。 根据 $(Z_n)_{n\in\mathbb{N}}$ 在给定 $\Theta=\theta$ 以及引理 2.4 时的马尔可夫性质，我们有

$$
\begin{aligned}
&\mathbb{P}(\Gamma_{\tau_S} \in F, \tau_S < \infty \mid \Theta=\theta) = \sum_{k=1}^{\infty} \mathbb{P}[\Gamma_{\tau_S} \in F, \tau_S = k \mid \Theta=\theta] \\
&= \sum_{k=1}^{\infty} \mathbb{P}[\Gamma_k \in F\cap S, \Gamma_1 \in S^c, \ldots, \Gamma_{k-1} \in S^c \mid \Theta=\theta] = \sum_{k=1}^{\infty} \mathbb{E}\left[\mathbf{1}_F(\Gamma_k) \prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i) \mid \Theta=\theta\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_F(\Gamma_k) \prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i) \mid Z_1, \ldots, Z_{k-1}, \Theta\right] \mid \Theta=\theta\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i) \mathbb{E}\left[\mathbf{1}_F(\Gamma_k) \mid Z_1, \ldots, Z_{k-1}, \Theta\right] \mid \Theta=\theta\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i) \mathbb{E}\left[\mathbf{1}_{F\times[0,2\pi)^2}(Z_k) \mid Z_{k-1}, \Theta\right] \mid \Theta=\theta\right] \\
&\underset{(6)}{=} \sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i) \mathbb{E}\left[\mathbf{1}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(\theta)\, \mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(F) \mid Z_{k-1}, \Theta\right] \mid \Theta=\theta\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i) \mathbb{E}\left[\mathbf{1}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(\theta)\, \mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(F) \mid Z_1, \ldots, Z_{k-1}, \Theta\right] \mid \Theta=\theta\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i) \mathbf{1}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(\theta)\, \mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(F) \mid Z_1, \ldots, Z_{k-1}, \Theta\right] \mid \Theta=\theta\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i) \mathbf{1}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(\theta)\, \mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(F) \mid \Theta=\theta\right].
\end{aligned}
$$

对于 F， $G \in \mathrm{B}(S)$，根据条件分布的定义可得

$$
\int_G Q_S(\theta, F)\, \mathcal{U}_S(\mathrm{d}\theta) = \sum_{k=1}^{\infty} \mathbb{E}\left[\mathbf{1}_G(\Theta) \prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i) \mathbf{1}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(\Theta)\, \mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(F)\right], \tag{10}
$$

因为 $\Theta \sim \mathrm{U}_S$。注意，利用 $F \in \mathrm{B}(S)$，我们得到

$$
\mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(F) = \frac{\lambda(F\cap I(\Gamma_k^{\min},\Gamma_k^{\max}))}{\lambda(I(\Gamma_k^{\min},\Gamma_k^{\max}))} = \frac{\mathcal{U}_S(F\cap I(\Gamma_k^{\min},\Gamma_k^{\max}))}{\mathcal{U}_S(I(\Gamma_k^{\min},\Gamma_k^{\max}))}. \tag{11}
$$

利用这一点和引理 2.5 我们修改求和中的期望，得到

$$
\mathbb{E}\left[\mathbf{1}_G(\Theta) \prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i) \mathbf{1}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(\Theta)\, \mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(F)\right]
$$

$$
\begin{aligned}
&= \mathbb{E}\left[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i)\mathbf{1}_{G\cap I(\Gamma_k^{\min},\Gamma_k^{\max})}(\Theta)\,\frac{\mathcal{U}_S(F\cap I(\Gamma_k^{\min},\Gamma_k^{\max}))}{\mathcal{U}_S(I(\Gamma_k^{\min},\Gamma_k^{\max}))}\right]\\
&= \mathbb{E}\Big[\mathbb{E}\Big[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i)\mathbf{1}_{G\cap I(\Gamma_k^{\min},\Gamma_k^{\max})}(\Theta)\,\frac{\mathcal{U}_S(F\cap I(\Gamma_k^{\min},\Gamma_k^{\max}))}{\mathcal{U}_S(I(\Gamma_k^{\min},\Gamma_k^{\max}))} \mid Z_1,\dots,Z_k\Big]\Big]\\
&= \mathbb{E}\Big[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i)\,\frac{\mathcal{U}_S(F\cap I(\Gamma_k^{\min},\Gamma_k^{\max}))}{\mathcal{U}_S(I(\Gamma_k^{\min},\Gamma_k^{\max}))}\mathbb{E}\big[\mathbf{1}_{G\cap I(\Gamma_k^{\min},\Gamma_k^{\max})}(\Theta) \mid Z_1,\dots,Z_k\big]\Big]\\
&\underset{(7)}{=} \mathbb{E}\Big[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i)\,\frac{\mathcal{U}_S(F\cap I(\Gamma_k^{\min},\Gamma_k^{\max}))}{\mathcal{U}_S(I(\Gamma_k^{\min},\Gamma_k^{\max}))}\,\frac{\mathcal{U}_S(G\cap I(\Gamma_k^{\min},\Gamma_k^{\max})}{\mathcal{U}_S(I(\Gamma_k^{\min},\Gamma_k^{\max})}\Big], \qquad (12)
\end{aligned}
$$

其中在最后一个等式中，我们还用到了 $\mathbb{P}_\Theta = \mathrm{U}_S$。现在我们可以交换 F 和 G的角色，从而反向推理可得

$$\int_G Q_S(\theta,F)\,\mathcal{U}_S(\mathrm{d}\theta) = \int_F Q_S(\theta,G)\,\mathcal{U}_S(\mathrm{d}\theta),$$

这表明了所声称的可逆性。 □

下面我们证明与转移核 Q_S相对应的马尔可夫算子的半正定性。为得出这一结论，我们先引入所需的对象。定义希尔伯特空间

$$L_S^2 := \{f : [0,2\pi) \to \mathbb{R} : \int_S f(\theta)^2\,\mathcal{U}_S(\mathrm{d}\theta) < \infty\}$$

其内积为 $f,g_S := \int_S f(\theta)g(\theta)\,\mathrm{U}_S(\mathrm{d}\theta)$，对所有 $f,g \in L_S^2$均成立。转移核Q_S 诱导出一个线性算子 Q_S：$L_S^2 \to L_S^2$，其定义为

$$f \mapsto \left[\alpha \mapsto \int_S f(\theta)\,Q_S(\alpha,\mathrm{d}\theta)\right], \quad f \in L_S^2. \qquad (13)$$

引理 2.11。 假设 $S \in \mathcal{B}([0,2\pi))$ 是圆周上的开集且非空。则，由收缩核导出的算子 Q_S定义于(13)，是半正定的，即对所有 $f \in L_S^2$均有 $\langle \mathrm{Q}_S f, f\rangle_S \geq 0$。

证明。 根据（10）、（11）和（12），对于任意 F，$G \in \mathcal{B}(S)$，有

$$\langle \mathrm{Q}_S\mathbf{1}_F, \mathbf{1}_G\rangle_S = \sum_{k=1}^{\infty} \mathbb{E}\Big[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i)\,\mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(F)\,\mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(G)\Big].$$

对于固定的 $F \in \mathrm{B}(S)$，借助积分理论的标准工具（利用线性性和单调收敛），我们将前述表示从 $g = \mathbf{1}_G$ 推广到非负的 $g \in L_S^2$，使得

$$\langle \mathrm{Q}_S\mathbf{1}_F, g\rangle_S = \sum_{k=1}^{\infty} \mathbb{E}\Big[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i)\,\mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(F)\,\mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(g)\Big],$$

其中 $\mathrm{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(g) := \int_{I(\Gamma_k^{\min},\Gamma_k^{\max})} g(\theta)\mathrm{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(\mathrm{d}\theta)$。对于固定的非负 $g \in L_S^2$，同样借助积分理论的标准工具，我们将后一个恒等式从 $f = \mathbf{1}_F$ 推广到非负的 $f \in L_S^2$，使得

$$\langle \mathrm{Q}_S f, g\rangle_S = \sum_{k=1}^{\infty} \mathbb{E}\Big[\prod_{i=1}^{k-1} \mathbf{1}_{S^c}(\Gamma_i)\, \mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(f)\, \mathcal{U}_{I(\Gamma_k^{\min},\Gamma_k^{\max})}(g)\Big].$$

最后，利用内积的双线性性，以及 f，g 可以表示为其正部与负部之差这一点，我们还可以得到 $f, g \in L_S^2$ 的后一种表达式。从这个最后的表示式可直接推出，对所有 $f \in L_S^2$，均有 $\mathrm{Q}_S f,\ f_S \geq 0$。□

本节最后陈述转移核 Q_S 的推前不变性。关于推前转移核的一般性质，请参阅Rudolf and Sprungk （2022）。

引理 2.12. 设 $S \in \mathcal{B}([0,2\pi))$ 是圆周上的非空开集。对于 $\theta \in S$，定义函数 $g_\theta\colon [0,2\pi) \to [0,2\pi)$ 为 $g_\theta(\alpha) = (\theta - \alpha)$ 模 2π。那么，收缩核 Q_S 满足

$$Q_{g_\theta^{-1}(S)}(g_\theta^{-1}(\alpha), g_\theta^{-1}(B)) = Q_S(\alpha, B), \qquad \alpha \in S,\ B \in \mathcal{B}(S).$$

前一个引理的证明移至附录，见第 A.2节。

注 2.13。 转移核 Q_S 不仅在椭圆切片采样中有意义。收缩过程也可用于其他马尔可夫链蒙特卡罗方法，以替代从可参数化为 $[0,2\pi)$ 的集合的子集中进行均匀采样。事实上，球面上的测地线切片采样（见Habeck 等人 (2023)）以及 Gibbs 极坐标切片采样（见Schär、Habeck 和 Rudolf (2023) ）都采用收缩，从 $d-1$ 维欧几里得单位球面的若干大圆上生成点。当然，在此类情形下，可以利用本节中已确立的收缩核 Q_S 的所有性质。

3. 椭圆切片采样的可逆性与半正定性

借助算法 2.2 中收缩过程的转移机制关于转移核 Q_S 的表示，我们可以给出椭圆切片采样的转移核，记作 H，其对应于算法 2.3 的转移机制。对于所有 $x_{\mathrm{in}} \in \mathcal{H}$ 以及 $A \in \mathcal{B}(\mathcal{H})$，其定义如下：

$$H(x_{\mathrm{in}}, A) = \frac{1}{\rho(x_{\mathrm{in}})} \int_0^{\rho(x_{\mathrm{in}})} \int_{\mathcal{H}} Q_{S(x_{\mathrm{in}},w,t)}(0, p_{x_{\mathrm{in}},w}^{-1}(\mathcal{H}(t) \cap A))\, \mu_0(\mathrm{d}w)\mathrm{d}t, \tag{14}$$

其中，对于任意 $w \in \mathrm{H}$ 和 $t \in (0,\infty)$，$S(x_{\mathrm{in}}, w, t) := p_{x_{\mathrm{in}},w}^{-1}(\mathrm{H}(t))$。我们还考虑由 H 诱导的线性算子。为此，定义希尔伯特空间

$$L_\mu^2 := \{f : \mathcal{H} \to \mathbb{R} \colon \int_{\mathcal{H}} f(x)^2 \mu(\mathrm{d}x) < \infty\}$$

其内积为 $\langle f,g\rangle_\mu := \int_{\mathcal{H}} f(x)g(x)\mu(\mathrm{d}x)$，其中 $f,g \in L^2_\mu$。然后，可将 H 扩展为线性算子 $\mathrm{H}: L^2_\mu \to L^2_\mu$，定义为

$$f \mapsto \left[x \mapsto \int_{\mathcal{H}} f(y)H(x,\mathrm{d}y)\right], \quad f \in L^2_\mu. \tag{15}$$

这里我们验证，关于 $S(x_{\text{in}}, w, t)$ 上均匀分布的收缩过程的可逆性和半正定性，可推广为 H 关于 μ 的可逆性以及 H 的半正定性。

定理 *3.1.* 设 H 为希尔伯特空间，$C: H \to H$ 为非奇异协方差算子，定义 $\mu_0 = N(0, C)$。此外，设 $\rho: H \to (0,\infty)$ 为下半连续函数，且 $Z := \int H\ \rho(x)\mu_0(dx) < \infty$。那么，式 *(13)* 中给出的椭圆切片采样器的转移核 H 关于……可逆。

$$\mu(\mathrm{d}x) = \frac{1}{Z}\rho(x)\mu_0(\mathrm{d}x).$$

此外，由 H: $L^2_\mu \to L^2_\mu$ 诱导的 H（见(15)）是半正定的。

在证明主定理之前，我们先给出一些辅助工具。

引理 3.2. 设 X 和 Y 是取值于希尔伯特空间 $\mathcal{H}$ 的独立随机变量，且各自服从 $\mu_0 = \mathcal{N}(0, C)$，其中 $C: \mathcal{H} \to \mathcal{H}$ 是非奇异协方差算子。对于任意 $\theta \in [0, 2\pi)$，令 $T_{(\theta)}: \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ 定义如下：

$$T^{(\theta)}(x,y) := (x\cos\theta + y\sin\theta, x\sin\theta - y\cos\theta).$$

Then

$$\mathbb{E}(f(\theta, X, Y)) = \mathbb{E}(f(\theta, T^{(\theta)}(X,Y))), \quad \theta \in [0, 2\pi), \tag{16}$$

对于任意 f: $[0,2\pi) \times \mathcal{H}^2 \to \mathbb{R}$，只要其中一个期望存在，即两个相同中心正态分布的乘积测度在 $T_{(\theta)}$ 下保持不变。

证明。 由于 $X, Y \sim \mu_0 = \mathrm{N}(0, C)$ 相互独立，因此随机向量 $\begin{pmatrix} X \\ Y \end{pmatrix}$ 在 H $\times$H 上服从 $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix}\right)$. 注意到 $T_{(\theta)}(x,y)^t = \begin{pmatrix} \cos\theta I & \sin\theta I \\ \sin\theta I & -\cos\theta I \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$，其中 I: H $\to$H 表示恒等算子。因此，根据高斯测度的线性变换定理，参见例如（DaPrato and Zabczyk，2002，命题 1.2.3），可得向量 $T_{(\theta)}(X,Y)^t$ 服从

$$N\left(\begin{pmatrix} \cos\theta I & \sin\theta I \\ \sin\theta I & -\cos\theta I \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \cos\theta I & \sin\theta I \\ \sin\theta I & -\cos\theta I \end{pmatrix}\begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix}\begin{pmatrix} \cos\theta I & \sin\theta I \\ \sin\theta I & -\cos\theta I \end{pmatrix}\right) = N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix}\right).$$

因此，(X, Y) 和 $T_{(\theta)}(X, Y)$ 的分布相同，因此（16）成立。 □

以下关于内积的有用表示，使我们能够将收缩过程的结果应用于椭圆切片采样器。

引理 3.3。对于函数 $f\colon \mathcal{H}\to\mathbb{R}$，定义

$$f_{x,y,t}:[0,2\pi)\to\mathbb{R},\quad \theta\mapsto \mathbf{1}_{S(x,y,t)}(\theta)(f\circ p_{x,y})(\theta),\qquad x,y\in\mathcal{H},t\in(0,\infty).$$

在定理 *3.1* 的假设下，对于所有 $f\in L^2_\mu$，几乎所有 $(x,y,t)\in\{(x,y,t)\in H^2\times(0,\infty):\ x\in H(t)\}$（相对于 $\mu_0\otimes\mu_0\otimes\lambda$）均有 $f_{x,y,t}\in L^2_S(x,y,t)$。此外，对于所有 $f,g\in L^2_\mu$，有

$$\langle \mathrm{H}f,g\rangle_\mu=\frac{1}{Z}\int_0^\infty\int_{\mathcal{H}}\int_{\mathcal{H}(t)}\langle \mathrm{Q}_{S(x,y,t)}f_{x,y,t},g_{x,y,t}\rangle_{S(x,y,t)}\ \mu_0(\mathrm{d}x)\mu_0(\mathrm{d}y)\mathrm{d}t. \tag{17}$$

证明。我们先作一些在证明中反复用到的预备考虑。对于任意 $\theta\in[0,2\pi)$，令 $g_\theta(\alpha):=(\theta-\alpha)\bmod 2\pi$，其中 $\alpha\in[0,2\pi)$，并注意到，利用三角函数的和角公式，我们有

$$p_{T^{(\theta)}(x,y)}(\alpha)=p_{x,y}(g_\theta(\alpha)),\qquad \forall\alpha\in[0,2\pi).$$

利用前一个等式，对于所有 x，$y\in\mathrm{H}$ 和 $t\in(0,\infty)$，有

$$\alpha\in S(T^{(\theta)}(x,y),t)\iff g_\theta(\alpha)\in S(x,y,t).$$

因此，$S(T(\theta)(x,y),t)=g_\theta^{-1}(S(x,y,t))$。特别地，我们有 $\lambda(S(T(\theta)(x,y),t))=\lambda(S(x,y,t))$，以及 $\mathbf{1}_{S(T(\theta)(x,y),t)}(\theta)=\mathbf{1}_{S(x,y,t)}(0)=\mathbf{1}_{\mathcal{H}(t)}(x)$。

对于 $f\in L^2_\mu$ 和 X，Y 为相互独立的 μ_0分布随机变量，通过纳入“活跃项”，我们得到

$$\int_{\mathcal{H}}f(x)^2\mu(\mathrm{d}x)=\frac{1}{Z}\int_0^{2\pi}\int_0^\infty\mathbb{E}[h_t^{(1)}(\theta,X,Y)]\mathrm{d}t\mathrm{d}\theta,$$

其中 $\mathrm{h}_t^{(1)}(\theta,x,y)=\mathbf{1}_{\mathcal{H}(t)}(x)\mathbf{1}_{S(x,y,t)}(\theta)\ f(x)^2\lambda(S(x,y,t))^{-1}$，这里还用到了 $\mathbf{1}_{(0,\rho(x))}(t)=\mathbf{1}_{\mathcal{H}(t)}(x)$ 以及 $\lambda(S(x,y,t))>0$。后者成立是因为，由于 ρ 的下半连续性，$S(x,y,t)$ 是圆周上的开集且非空（至少对于上式中出现的那些 t 而言）。于是，引理 3.2 和前面的考虑得到

$$\begin{aligned}
&\int_{\mathcal{H}}f(x)^2\mu(\mathrm{d}x)=\frac{1}{Z}\int_0^{2\pi}\int_0^\infty\mathbb{E}[h_t^{(1)}(\theta,T^{(\theta)}(X,Y))]\mathrm{d}t\mathrm{d}\theta\\
&=\int_0^{2\pi}\int_0^\infty\int_{\mathcal{H}}\int_{\mathcal{H}}\frac{\mathbf{1}_{\mathcal{H}(t)}\left(p_{x,y}(\theta)\right)\mathbf{1}_{S(T^{(\theta)}(x,y),t)}(\theta)f\left(p_{x,y}(\theta)\right)^2}{\lambda(S(T^{(\theta)}(x,y),t))Z}\ \mu_0(\mathrm{d}x)\,\mu_0(\mathrm{d}y)\,\mathrm{d}t\,\mathrm{d}\theta\\
&=\int_0^{2\pi}\int_0^\infty\int_{\mathcal{H}}\int_{\mathcal{H}}\frac{\mathbf{1}_{\mathcal{H}(t)}(x)f_{x,y,t}(\theta)^2}{\lambda(S(x,y,t))Z}\ \mu_0(\mathrm{d}x)\,\mu_0(\mathrm{d}y)\,\mathrm{d}t\,\mathrm{d}\theta\\
&=\frac{1}{Z}\int_0^\infty\int_{\mathcal{H}}\int_{\mathcal{H}}\mathbf{1}_{\mathcal{H}(t)}(x)\int_{S(x,y,t)}f_{x,y,t}(\theta)^2\,\mathcal{U}_{S(x,y,t)}(\mathrm{d}\theta)\mu_0(\mathrm{d}x)\,\mu_0(\mathrm{d}y)\,\mathrm{d}t.
\end{aligned}$$

该等式意味着，对于几乎所有的 $(x,y,t)\in\{(x,y,t)\in\mathrm{H}^2\times(0,\infty):\ x\in\mathrm{H}(t)\}$，$\mu_0\otimes\mu_0\otimes\lambda(x,y,t)$ $x,y,t(x,y,t)$。为证明 (17)，令 $g\in L^2_\mu$。与上文类似，根据 H 的定义，有

$$\mathrm{Q}_{S(x,y,t)}(f\circ p_{x,y})(0)=\int_{S(x,y,t)}f\left(p_{x,y}(\alpha)\right)\ Q_{S(x,y,t)}(0,\mathrm{d}\alpha),$$

我们得到

$$\begin{aligned}\langle \mathrm{H}f, g\rangle_\mu &= \frac{1}{Z}\int_{\mathcal{H}} \mathrm{H}f(x)g(x)\rho(x)\ \mu_0(\mathrm{d}x)\\ &= \frac{1}{Z}\int_0^\infty \int_{\mathcal{H}}\int_{\mathcal{H}} g(x)\mathbf{1}_{\mathcal{H}(t)}(x)\mathrm{Q}_{S(x,y,t)}(f\circ p_{x,y})(0)\ \mu_0(\mathrm{d}x)\mu_0(\mathrm{d}y)\mathrm{d}t\\ &= \frac{1}{Z}\int_0^\infty \int_0^{2\pi} \mathbb{E}\left[h_t^{(2)}(\theta, X, Y)\right]\ \mathrm{d}\theta\,\mathrm{d}t,\end{aligned}$$

其中 X，Y 是相互独立且服从 μ_0 分布的随机变量，并且

$$h_t^{(2)}(\theta,x,y) = \mathbf{1}_{S(x,y,t)}(\theta)g(x)\mathbf{1}_{\mathcal{H}(t)}(x)\mathrm{Q}_{S(x,y,t)}(f\circ p_{x,y})(0)\,\lambda(S(x,y,t))^{-1}.$$

于是，根据引理3.2可得

$$\begin{aligned}\langle \mathrm{H}f, g\rangle_\mu &= \frac{1}{Z}\int_0^\infty \int_0^{2\pi} \mathbb{E}\left[h_t^{(2)}(\theta, T^{(\theta)}(X,Y))\right]\mathrm{d}\theta\mathrm{d}t\\ &= \frac{1}{Z}\int_0^\infty \int_{\mathcal{H}}\int_{\mathcal{H}}\int_0^{2\pi} h_t^{(2)}(\theta, T^{(\theta)}(x,y))\ \mathrm{d}\theta\,\mu_0(\mathrm{d}x)\,\mu_0(\mathrm{d}y)\,\mathrm{d}t.\end{aligned}$$

考虑到证明开头的讨论，引理 2.12 蕴含

$$\begin{aligned}&\mathrm{Q}_{S(T^{(\theta)}(x,y),t)}(f\circ p_{T^{(\theta)}(x,y)})(0) = \int_0^{2\pi} \mathbf{1}_{S(T^{(\theta)}(x,y),t)}(\alpha) f(p_{T^{(\theta)}(x,y)}(\alpha)) Q_{S(T^{(\theta)}(x,y),t)}(0,\mathrm{d}\alpha)\\ &= \int_0^{2\pi} \mathbf{1}_{S(x,y,t)}\left(g_\theta(\alpha)\right) f\left(p_{x,y}(g_\theta(\alpha))\right) Q_{g_\theta^{-1}(S(x,y,t))}(g_\theta^{-1}(\theta),\mathrm{d}\alpha)\\ &= \int_0^{2\pi} \mathbf{1}_{S(x,y,t)}\left(\alpha\right) f\left(p_{x,y}(\alpha)\right) Q_{S(x,y,t)}(\theta,\mathrm{d}\alpha)\\ &= \int_{S(x,y,t)} f_{x,y,t}(\alpha) Q_{S(x,y,t)}(\theta,\mathrm{d}\alpha) = \mathrm{Q}_{S(x,y,t)} f_{x,y,t}(\theta),\end{aligned}$$

并进一步得到

$$\begin{aligned}h_t^{(2)}(\theta, T^{(\theta)}(x,y)) &= \frac{\mathbf{1}_{S(T^{(\theta)}(x,y),t)}(\theta)\cdot g\left(p_{x,y}(\theta)\right)\cdot \mathbf{1}_{\mathcal{H}(t)}\left(p_{x,y}(\theta)\right)\cdot \mathrm{Q}_{S(x,y,t)} f_{x,y,t}(\theta)}{\lambda(S(T^{(\theta)}(x,y),t))}\\ &= \mathbf{1}_{\mathcal{H}(t)}(x)\cdot g_{x,y,t}(\theta)\cdot \mathrm{Q}_{S(x,y,t)} f_{x,y,t}(\theta)\cdot \lambda(S(x,y,t))^{-1}.\end{aligned}$$

因此

$$\begin{aligned}\int_0^{2\pi} h_t^{(2)}(\theta, T^{(\theta)}(x,y))\ \mathrm{d}\theta &= \mathbf{1}_{\mathcal{H}(t)}(x)\int_{S(x,y,t)} g_{x,y,t}(\theta)\ \mathrm{Q}_{S(x,y,t)} f_{x,y,t}(\theta)\ \mathcal{U}_{S(x,y,t)}(\mathrm{d}\theta)\\ &= \mathbf{1}_{\mathcal{H}(t)}(x)\langle \mathrm{Q}_{S(x,y,t)} f_{x,y,t}, g_{x,y,t}\rangle_{S(x,y,t)}.\end{aligned}$$

综上，我们得到

$$\langle \mathrm{H}f, g\rangle_\mu = \frac{1}{Z}\int_0^\infty \int_{\mathcal{H}}\int_{\mathcal{H}(t)} \langle \mathrm{Q}_{S(x,y,t)} f_{x,y,t}, g_{x,y,t}\rangle_{S(x,y,t)}\ \mu_0(\mathrm{d}x)\,\mu_0(\mathrm{d}y)\,\mathrm{d}t.$$

□

结合前面的结果，我们证明了主定理。

定理 3.1 的证明。设 $A, B \in \mathrm{B}(\mathrm{H})$。由于 ρ 是下半连续的， $S(x,y,t)$ 在圆周上是开集，且对所有 $(x, y, t) \in \{(x,y,t) \in \mathrm{H}^2 \times (0,\infty) : x \in \mathrm{H}(t)\}$ 都非空。由定理 2.10 可得

$$
\begin{aligned}
&\langle \mathrm{Q}_{S(x,y,t)} \mathbf{1}_{p_{x,y}^{-1}(A \cap \mathcal{H}(t))}, \mathbf{1}_{p_{x,y}^{-1}(B \cap \mathcal{H}(t))} \rangle_{S(x,y,t)} \\
&= \int_{p_{x,y}^{-1}(B \cap \mathcal{H}(t))} Q_{S(x,y,t)}(\theta, p_{x,y}^{-1}(A \cap \mathcal{H}(t)))\, \mathcal{U}_{S(x,y,t)}(\mathrm{d}\theta) \\
&= \int_{p_{x,y}^{-1}(A \cap \mathcal{H}(t))} Q_{S(x,y,t)}(\theta, p_{x,y}^{-1}(B \cap \mathcal{H}(t)))\, \mathcal{U}_{S(x,y,t)}(\mathrm{d}\theta) \\
&= \langle \mathrm{Q}_{S(x,y,t)} \mathbf{1}_{p_{x,y}^{-1}(B \cap \mathcal{H}(t))}, \mathbf{1}_{p_{x,y}^{-1}(A \cap \mathcal{H}(t))} \rangle_{S(x,y,t)}.
\end{aligned}
$$

由 $(\mathbf{1}_A)_{x,\ y,\ t} = \mathbf{1}_{p_x^{-1,\ y}(A \cap \mathcal{H}(t))}$，结合（17）可得

$$
\begin{aligned}
&\int_B H(x,A)\mu(\mathrm{d}x) = \langle \mathrm{H}\mathbf{1}_A, \mathbf{1}_B \rangle_\mu \\
&= \frac{1}{Z} \int_0^\infty \int_{\mathcal{H}} \int_{\mathcal{H}(t)} \langle \mathrm{Q}_{S(x,y,t)} \mathbf{1}_{p_{x,y}^{-1}(A \cap \mathcal{H}(t))}, \mathbf{1}_{p_{x,y}^{-1}(B \cap \mathcal{H}(t))} \rangle_{S(x,y,t)}\ \mu_0(\mathrm{d}x)\, \mu_0(\mathrm{d}y)\, \mathrm{d}t \\
&= \frac{1}{Z} \int_0^\infty \int_{\mathcal{H}} \int_{\mathcal{H}(t)} \langle \mathrm{Q}_{S(x,y,t)} \mathbf{1}_{p_{x,y}^{-1}(B \cap \mathcal{H}(t))}, \mathbf{1}_{p_{x,y}^{-1}(A \cap \mathcal{H}(t))} \rangle_{S(x,y,t)}\ \mu_0(\mathrm{d}x)\, \mu_0(\mathrm{d}y)\, \mathrm{d}t \\
&= \langle \mathrm{H}\mathbf{1}_B, \mathbf{1}_A \rangle_\mu = \int_A H(x,B)\mu(\mathrm{d}x),
\end{aligned}
$$

这验证了相对于 μ 所要求的可逆性。现在我们转向半正定性。结合引理3.3和引理 2.11，可得对所有 $f \in L^2_\mu$ 均有

$$
\langle \mathrm{H}f, f \rangle_\mu = \frac{1}{Z} \int_0^\infty \int_{\mathcal{H}} \int_{\mathcal{H}(t)} \langle \mathrm{Q}_{S(x,y,t)} f_{x,y,t}, f_{x,y,t} \rangle_{S(x,y,t)}\ \mu_0(\mathrm{d}x)\mu_0(\mathrm{d}y)\mathrm{d}t \geq 0. \qquad \square
$$

4. 总结与展望

我们来总结一下主要发现。我们证明了椭圆切片采样（ESS）的可逆性和半正定性，其中相应马尔可夫链的底层状态空间可以是无限维希尔伯特空间。在此过程中，我们指出，似然函数满足一个（弱）定性正则条件（ρ 假设其下半连续），该条件可确保其中的 while 循环终止，从而得到定义良好的转移核。此外，借助（14），我们构造了椭圆切片采样转移核的一种表示。我们的方法从半正定性的角度揭示了混合切片抽样结构（参见 Łatuszyński 和 Rudolf(2014)，见引理2.11；同时也揭示了收缩过程相对于角度空间子集上的均匀分布的可逆性，见定理 2.10） $[0, 2\pi$。

此前提出的表示方法和工具或许能为分析椭圆切片采样的谱间隙及其与维度无关的行为铺平道路。严格为正的谱间隙是马尔可夫链的一项理想性质，既关系到其混合性质，也关系到从理论上评估马尔可夫链蒙特卡罗对期望的近似所产生的均方误差，参见 μ，对于

详情参见，例如 Rudolf (2012）。针对简单切片抽样提出的耦合构造 Natarovskii、Rudolf 和 Sprungk (2021b)或许是验证在 $\mathcal{H}$ 上存在此类正谱间隙的一种有前景的方法。此外，进一步探索梅特罗波利斯–黑斯廷斯与切片抽样方法在结构上的相似性似乎也很有益。尤其值得关注的是依赖于似然函数代理评估的近似（椭圆）切片抽样。在这方面，Habeck、Rudolf 和 Sprungk(2020)、Sprungk (2020)提出的稳定性研究或可用于为椭圆切片采样推导扰动理论结果，正如 Medina-Aguayo、Rudolf 和 Schweizer (2020)、RudolfandSchweizer (2018)针对近似梅特罗波利斯–黑斯廷斯所做的那样。最终，对椭圆切片采样的理论研究或许有助于验证其他依赖收缩过程的切片抽样方案是否也具有可逆性和半正定性。

附录A：技术性证明

A.1. 引理2.5的证明

证明。通过对 $n \in \mathbb{N}$ 进行归纳，我们证明

$$\mathbb{E}[\mathbf{1}_A(\Theta) \mid Z_1, \dots, Z_n] = \frac{\mathbb{P}_\Theta(A \cap I(\Gamma_n^{\min}, \Gamma_n^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min}, \Gamma_n^{\max}))}, \tag{18}$$

由此可立即得到结论。我们从基础情形开始，即考虑 $n = 1$。注意，$Z_1 = (\Gamma_1, \Gamma_1^{\min}, \Gamma_1^{\max})$ 与 Θ 独立，并且 $I(\Gamma_1^{\min}, \Gamma_1^{\max}) = I(\Gamma_1, \Gamma_1) = [0, 2\pi)$。利用这些性质可得

$$\mathbb{E}(\mathbf{1}_A(\Theta) \mid Z_1) = \mathbb{P}_\Theta(A) = \frac{\mathbb{P}_\Theta(A \cap I(\Gamma_1, \Gamma_1))}{\mathbb{P}_\Theta(I(\Gamma_1, \Gamma_1))} = \frac{\mathbb{P}_\Theta(A \cap I(\Gamma_1^{\min}, \Gamma_1^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_1^{\min}, \Gamma_1^{\max}))},$$

这验证了 18 对于 $n = 1$ 成立。

假设 18 对于 n 成立，我们将证明它对于 $n+1$ 也成立。注意，由于 $Z_n \in \Lambda_\Theta$ 几乎必然成立，我们有 $\Theta \in I(\Gamma_n^{\min}, \Gamma_n^{\max})$，$\Gamma_n \in I(\Gamma_n^{\min}, \Gamma_n^{\max})$ 并且有以下两个蕴含关系

$$\Theta \in I(\Gamma_n, \Gamma_n^{\max}) \implies \Gamma_n \in J(\Gamma_n^{\min}, \Theta) \implies \Gamma_{n+1}^{\min} = \Gamma_n,\ \Gamma_{n+1}^{\max} = \Gamma_n^{\max}, \tag{19}$$

$$\Theta \in J(\Gamma_n^{\min}, \Gamma_n) \implies \Gamma_n \in I(\Theta, \Gamma_n^{\max}) \implies \Gamma_{n+1}^{\min} = \Gamma_n^{\min},\ \Gamma_{n+1}^{\max} = \Gamma_n. \tag{20}$$

此外，根据归纳假设，事实是 $\Gamma_n \in I(\Gamma_n^{\min}, \Gamma_n^{\max})$ 几乎必然成立，并利用测度分解，见（Kallenberg，2002，第 6 章，定理 6.4 和式（6）），我们得到

$$\mathbb{E}[\mathbf{1}_{I(\Gamma_n, \Gamma_n^{\max})}(\Theta) \mid Z_1, \dots, Z_n] = \frac{\mathbb{P}_\Theta(I(\Gamma_n, \Gamma_n^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min}, \Gamma_n^{\max}))}, \tag{21}$$

$$\mathbb{E}[\mathbf{1}_{J(\Gamma_n^{\min}, \Gamma_n)}(\Theta) \mid Z_1, \dots, Z_n] = \frac{\mathbb{P}_\Theta(J(\Gamma_n^{\min}, \Gamma_n))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min}, \Gamma_n^{\max}))} = \mathbf{1}_{\{\Gamma_n^{\min} \neq \Gamma_n\}} \frac{\mathbb{P}_\Theta(I(\Gamma_n^{\min}, \Gamma_n))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min}, \Gamma_n^{\max}))}. \tag{22}$$

对于任意 $A \in \mathrm{B}([0, 2\pi))$，$C_i \in \mathrm{B}([0, 2\pi)^3)$ 且 $i = 1, \dots\dots, n+1$，我们验证

$$\mathbb{E}\left[\mathbf{1}_A(\Theta) \prod_{i=1}^{n+1} \mathbf{1}_{C_i}(Z_i)\right] = \mathbb{E}\left[\prod_{i=1}^{n+1} \mathbf{1}_{C_i}(Z_i) \frac{\mathbb{P}_\Theta(A \cap I(\Gamma_{n+1}^{\min}, \Gamma_{n+1}^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_{n+1}^{\min}, \Gamma_{n+1}^{\max}))}\right]. \tag{23}$$

因此，根据条件分布/期望的定义，以及上述形式的笛卡尔积集合生成了 σ-代数 $\mathcal{B}([0,2\pi)^{3(n+1)})$，我们得到 (18)。为证明 (23)，我们注意到

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n+1}\mathbf{1}_{C_i}(Z_i)\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n+1}\mathbf{1}_{C_i}(Z_i)\mid Z_1,\dots,Z_n,\Theta\Big]\Big]\\
&= \mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathbb{P}(Z_{n+1}\in C_{n+1}\mid Z_1,\dots,Z_n,\Theta)\Big]\\
&= \mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathbb{P}(Z_{n+1}\in C_{n+1}\mid Z_n,\Theta)\Big] = \mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,R_\Theta(Z_n,C_{n+1})\Big].
\end{aligned}$$

根据引理 2.3 由前面的计算可得

$$\begin{aligned}
\mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n+1}\mathbf{1}_{C_i}(Z_i)\Big] &= \mathbb{E}\Big[\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathbf{1}_{A\cap I(\Gamma_n,\Gamma_n^{\max})}(\Theta)\,\mathcal{U}_{I(\Gamma_n,\Gamma_n^{\max})}\otimes\delta_{(\Gamma_n,\Gamma_n^{\max})}(C_{n+1})\Big]\\
&\quad+\mathbb{E}\Big[\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathbf{1}_{A\cap J(\Gamma_n^{\min},\Gamma_n)}(\Theta)\,\mathcal{U}_{I(\Gamma_n^{\min},\Gamma_n)}\otimes\delta_{(\Gamma_n^{\min},\Gamma_n)}(C_{n+1})\Big]\\
&= \mathbb{E}\Big[\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathcal{U}_{I(\Gamma_n,\Gamma_n^{\max})}\otimes\delta_{(\Gamma_n,\Gamma_n^{\max})}(C_{n+1})\,\mathbb{E}[\mathbf{1}_{A\cap I(\Gamma_n,\Gamma_n^{\max})}(\Theta)\mid Z_1,\dots,Z_n]\Big]\\
&\quad+\mathbb{E}\Big[\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathcal{U}_{I(\Gamma_n^{\min},\Gamma_n)}\otimes\delta_{(\Gamma_n^{\min},\Gamma_n)}(C_{n+1})\,\mathbb{E}[\mathbf{1}_{A\cap J(\Gamma_n^{\min},\Gamma_n)}(\Theta)\mid Z_1,\dots,Z_n]\Big].
\end{aligned}$$

为简化记号，定义

$$\begin{aligned}
H_{\max}(Z_1,\dots,Z_n) &:= \prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathcal{U}_{I(\Gamma_n,\Gamma_n^{\max})}\otimes\delta_{(\Gamma_n,\Gamma_n^{\max})}(C_{n+1}),\\
H_{\min}(Z_1,\dots,Z_n) &:= \prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathcal{U}_{I(\Gamma_n^{\min},\Gamma_n)}\otimes\delta_{(\Gamma_n^{\min},\Gamma_n)}(C_{n+1}).
\end{aligned}$$

注意到 $\mathbb{P}_\Theta(A\cap J(\Gamma_n^{\min},\Gamma_n)) = \mathbf{1}_{\{\Gamma_n^{\min}\neq\Gamma_n\}}\,\mathbb{P}_\Theta(A\cap I(\Gamma_n^{\min},\Gamma_n))$。因此，利用归纳假设，以及 $\Gamma_n\in I(\Gamma_n^{\min},\Gamma_n^{\max})$ 几乎必然蕴含 $J(\Gamma_n^{\min},\Gamma_n)\subset I(\Gamma_n^{\min},\Gamma_n^{\max})$ 且 $I(\Gamma_n,\Gamma_n^{\max})\subset I(\Gamma_n^{\min},\Gamma_n^{\max})$ 几乎必然这一事实，我们有

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n+1}\mathbf{1}_{C_i}(Z_i)\Big]\\
&= \mathbb{E}\Big[H_{\max}(Z_1,\dots,Z_n)\,\frac{\mathbb{P}_\Theta(A\cap I(\Gamma_n,\Gamma_n^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min},\Gamma_n^{\max}))}\Big]\\
&\quad+\mathbb{E}\Big[H_{\min}(Z_1,\dots,Z_n)\mathbf{1}_{\{\Gamma_n^{\min}\neq\Gamma_n\}}\,\frac{\mathbb{P}_\Theta(A\cap I(\Gamma_n^{\min},\Gamma_n))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min},\Gamma_n^{\max}))}\Big]
\end{aligned}$$

$$= \mathbb{E}\Big[H_{\max}(Z_1,\dots,Z_n)\,\frac{\mathbb{P}_\Theta(I(\Gamma_n,\Gamma_n^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min},\Gamma_n^{\max}))}\,\frac{\mathbb{P}_\Theta(A\cap I(\Gamma_n,\Gamma_n^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_n,\Gamma_n^{\max}))}\Big]$$

$$+\mathbb{E}\Big[H_{\min}(Z_1,\dots,Z_n)\mathbf{1}_{\{\Gamma_n^{\min}\neq\Gamma_n\}}\,\frac{\mathbb{P}_\Theta(I(\Gamma_n^{\min},\Gamma_n))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min},\Gamma_n^{\max}))}\,\frac{\mathbb{P}_\Theta(A\cap I(\Gamma_n^{\min},\Gamma_n))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min},\Gamma_n))}\Big].$$

利用（21）以及（22）可得

$$\mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n+1}\mathbf{1}_{C_i}(Z_i)\Big]$$

$$= \mathbb{E}\Big[H_{\max}(Z_1,\dots,Z_n)\,\mathbb{E}[\mathbf{1}_{I(\Gamma_n,\Gamma_n^{\max})}(\Theta)\mid Z_1,\dots,Z_n]\cdot\frac{\mathbb{P}_\Theta(A\cap I(\Gamma_n,\Gamma_n^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_n,\Gamma_n^{\max}))}\Big]$$

$$+\mathbb{E}\Big[H_{\min}(Z_1,\dots,Z_n)\,\mathbb{E}[\mathbf{1}_{J(\Gamma_n^{\min},\Gamma_n)}(\Theta)\mid Z_1,\dots,Z_n]\cdot\frac{\mathbb{P}_\Theta(A\cap I(\Gamma_n^{\min},\Gamma_n))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min},\Gamma_n))}\Big]$$

$$= \mathbb{E}\Big[\mathbb{E}\Big[H_{\max}(Z_1,\dots,Z_n)\,\mathbf{1}_{I(\Gamma_n,\Gamma_n^{\max})}(\Theta)\cdot\frac{\mathbb{P}_\Theta(A\cap I(\Gamma_n,\Gamma_n^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_n,\Gamma_n^{\max}))}\mid Z_1,\dots,Z_n\Big]\Big]$$

$$+\mathbb{E}\Big[\mathbb{E}\Big[H_{\min}(Z_1,\dots,Z_n)\,\mathbf{1}_{J(\Gamma_n^{\min},\Gamma_n)}(\Theta)\cdot\frac{\mathbb{P}_\Theta(A\cap I(\Gamma_n^{\min},\Gamma_n))}{\mathbb{P}_\Theta(I(\Gamma_n^{\min},\Gamma_n))}\mid Z_1,\dots,Z_n\Big]\Big].$$

记

$$\mathcal{T}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(A) := \frac{\mathbb{P}_\Theta(A\cap I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}))}$$

并利用（19）以及（20）可得

$$\mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n+1}\mathbf{1}_{C_i}(Z_i)\Big]$$

$$= \mathbb{E}\Big[\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathcal{U}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}\otimes\delta_{(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(C_{n+1})\,\mathbf{1}_{I(\Gamma_n,\Gamma_n^{\max})}(\Theta)\mathcal{T}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(A)\Big]$$

$$+\mathbb{E}\Big[\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathcal{U}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}\otimes\delta_{(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(C_{n+1})\,\mathbf{1}_{J(\Gamma_n^{\min},\Gamma_n)}(\Theta)\mathcal{T}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(A)\Big].$$

由 $\Theta,\Gamma_n\in I(\Gamma_n^{\min},\Gamma_n^{\max})$ 几乎必然成立这一事实，可得 $\mathbf{1}_{I(\Gamma_n,\Gamma_n^{\max})}(\Theta)+\mathbf{1}_{J(\Gamma_n^{\min},\Gamma_n)}(\Theta)=1$ 几乎必然成立，使得

$$\mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n+1}\mathbf{1}_{C_i}(Z_i)\Big] = \mathbb{E}\Big[\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathcal{U}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}\otimes\delta_{(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(C_{n+1})\,\mathcal{T}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(A)\Big].$$

根据（4），我们有

$$\mathbb{E}[\mathbf{1}_{C_{n+1}}(Z_{n+1})\mid\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}] = \mathcal{U}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}\otimes\delta_{(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(C_{n+1})$$

$$= \mathbb{E}[\mathbf{1}_{C_{n+1}}(Z_{n+1})\mid Z_1,\dots,Z_n,\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}],$$

使得

$$\begin{aligned}\mathbb{E}\Big[\mathbf{1}_A(\Theta)\prod_{i=1}^{n+1}\mathbf{1}_{C_i}(Z_i)\Big] &= \mathbb{E}\Big[\prod_{i=1}^{n}\mathbf{1}_{C_i}(Z_i)\,\mathbb{E}[\mathbf{1}_{C_{n+1}}(Z_{n+1}) \mid Z_1,\dots,Z_n,\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}]\,\mathcal{T}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(A)\Big]\\ &= \mathbb{E}\Big[\mathbb{E}\Big[\prod_{i=1}^{n+1}\mathbf{1}_{C_i}(Z_i)\,\mathcal{T}_{I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max})}(A) \mid Z_1,\dots,Z_n,\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}\Big]\Big]\\ &= \mathbb{E}\Big[\prod_{i=1}^{n+1}\mathbf{1}_{C_i}(Z_i)\,\frac{\mathbb{P}_\Theta(A\cap I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}))}\Big].\end{aligned}$$

最后注意到 $\frac{\mathbb{P}_\Theta(A\cap I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}))}{\mathbb{P}_\Theta(I(\Gamma_{n+1}^{\min},\Gamma_{n+1}^{\max}))}$ 相对于 $\sigma(Z_1,\dots,Z_{n+1})$ 可测，因此我们已证明（18），整个陈述得证。 □

A.2. 引理 2.12 的证明2.12

我们先说明记号，并证明一些用于引理 2.12 的实际证明中的辅助结果。对于 α，$\beta\in[0,2\pi)$ 令

$$I^\circ(\alpha,\beta) := \begin{cases}(\alpha,\beta), & \alpha<\beta\\ [0,\beta)\cup(\alpha,2\pi), & \alpha>\beta\\ [0,2\pi), & \alpha=\beta.\end{cases}$$

我们给出两个引理，其中第一个可通过分类讨论得到。

引理 A.1. 设 $\theta\in[0,2\pi)$ 并定义 $g_\theta\colon\ [0,2\pi)\to[0,2\pi)$ 为 $g_\theta(\alpha)=(\theta-\alpha)$ 模 2π。那么，对于任意 $\alpha,\beta\in[0,2\pi)$，有

$$g_\theta(I^\circ(\alpha,\beta)) = I^\circ(g_\theta(\beta),g_\theta(\alpha)).$$

引理 A.2. 设 α，$\theta\in(0,\ 2\pi]$ 且

$$G_\alpha := \{(\gamma,\gamma^{\min},\gamma^{\max})\in[0,2\pi)^3 \colon \alpha,\gamma\in I^\circ(\gamma^{\min},\gamma^{\max}),\alpha\neq\gamma,\alpha\neq\gamma^{\min},\alpha\neq\gamma^{\max}\}.$$

定义 $\widetilde{g}_\theta\colon\ [0,2\pi)^3\to[0,2\pi)^3$ 通过 $\widetilde{g}_\theta(\gamma_1,\gamma_2,\gamma_3)=(g_\theta(\gamma_1),g_\theta(\gamma_3),g_\theta(\gamma_2))$，其中 g_θ 是引理 A.1 中的映射。那么，对于 R_α（在引理 2.3 之前引入），任意 $z\in G_\alpha$ 均有 $R_\alpha(z,G_\alpha)=1$ 且

$$R_{g_\theta^{-1}(\alpha)}(\widetilde{g}_\theta(z),\widetilde{g}_\theta(C)) = R_\alpha(z,C), \tag{24}$$

对于任意 $C\in\mathcal{B}([0,2\pi)^3)$。

证明。 注意，对所有 $\alpha,\beta\in[0,2\pi)$，均有 $\mathrm{U}_{I(\alpha,\beta)}=\mathrm{U}_{I^\circ(\alpha,\beta)}$，并且对所有 $z=(\gamma,\gamma^{\min},\gamma^{\max})$，$\in G_\alpha$ 成立

$$\mathbf{1}_{I(\gamma,\gamma^{\max})}(\alpha)=\mathbf{1}_{I^\circ(\gamma,\gamma^{\max})}(\alpha),\qquad \mathbf{1}_{J(\gamma^{\min},\gamma)}(\alpha)=\mathbf{1}_{[0,2\pi)\setminus\{\gamma^{\min}\}}(\gamma)\,\mathbf{1}_{I^\circ(\gamma^{\min},\gamma)}(\alpha).$$

因此，由引理 2.3 可得，对所有 $z=(\gamma,\gamma^{\min},\gamma^{\max})\in G_\alpha$ 以及任意 $C\in\mathcal{B}([0,2\pi)^3)$ 满足

$$\begin{aligned}R_\alpha((\gamma,\gamma^{\min},\gamma^{\max}),C)&=\mathbf{1}_{I^\circ(\gamma,\gamma^{\max})}(\alpha)\cdot\mathcal{U}_{I^\circ(\gamma,\gamma^{\max})}\otimes\delta_{(\gamma,\gamma^{\max})}(C)\\&\quad+\mathbf{1}_{[0,2\pi)\setminus\{\gamma^{\min}\}}(\gamma)\,\mathbf{1}_{I^\circ(\gamma^{\min},\gamma)}(\alpha)\cdot\mathcal{U}_{I^\circ(\gamma^{\min},\gamma)}\otimes\delta_{(\gamma^{\min},\gamma)}(C).\end{aligned}\tag{25}$$

现在固定 $z=(\gamma,\gamma^{\min},\gamma^{\max})\in G_\alpha$ 以及 $C\in\mathcal{B}([0,2\pi)^3)$。由于 $z\in G_\alpha$ 蕴含 $\widetilde{g}_\theta(z)\in G_{g_\theta^{-1}(\alpha)}$ 且 $g_\theta=g_\theta^{-1}$，$\widetilde{g}_\theta=\widetilde{g}_\theta^{-1}$，由式（25）可得

$$\begin{aligned}R_{g_\theta^{-1}(\alpha)}(\widetilde{g}_\theta(z),\widetilde{g}_\theta(C))&=\mathbf{1}_{g_\theta(I^\circ(g_\theta(\gamma),g_\theta(\gamma^{\min})))}(\alpha)\cdot\mathcal{U}_{g_\theta(I^\circ(g_\theta(\gamma),g_\theta(\gamma^{\min})))}\otimes\delta_{(\gamma^{\min},\gamma)}(C)\\&\quad+\mathbf{1}_{[0,2\pi)\setminus(g_\theta(\gamma^{\max}))}(g_\theta(\gamma))\mathbf{1}_{g_\theta(I^\circ(g_\theta(\gamma^{\max}),g_\theta(\gamma)))}(\alpha)\cdot\mathcal{U}_{g_\theta(I^\circ(g_\theta(\gamma^{\max}),g_\theta(\gamma)))}\otimes\delta_{(\gamma,\gamma^{\max})}(C).\end{aligned}$$

引理 A.1 再次给出 $g_\theta=g_\theta^{-1}$ 可得

$$\begin{aligned}R_{g_\theta^{-1}(\alpha)}(\widetilde{g}_\theta(z),\widetilde{g}_\theta(C))&=\mathbf{1}_{I^\circ(\gamma^{\min},\gamma)}(\alpha)\cdot\mathcal{U}_{I^\circ(\gamma^{\min},\gamma)}\otimes\delta_{(\gamma^{\min},\gamma)}(C)\\&\quad+\mathbf{1}_{[0,2\pi)\setminus\{\gamma^{\max}\}}(\gamma)\,\mathbf{1}_{I^\circ(\gamma,\gamma^{\max})}(\alpha)\cdot\mathcal{U}_{I^\circ(\gamma,\gamma^{\max})}\otimes\delta_{(\gamma,\gamma^{\max})}(C).\end{aligned}$$

注意到 $z\in G_\alpha$ 意味着 $\gamma=\gamma^{\max}$ 或 $\gamma=\gamma^{\min}$ 只有在 $\gamma=\gamma^{\min}=\gamma^{\max}$ 时才可能。因此

$$\begin{aligned}R_{g_\theta^{-1}(\alpha)}(\widetilde{g}_\theta(z),\widetilde{g}_\theta(C))&=\mathbf{1}_{[0,2\pi)\setminus\{\gamma^{\min}\}}(\gamma)\,\mathbf{1}_{I^\circ(\gamma^{\min},\gamma)}(\alpha)\cdot\mathcal{U}_{I^\circ(\gamma^{\min},\gamma)}\otimes\delta_{(\gamma^{\min},\gamma)}(C)\\&\quad+\mathbf{1}_{I^\circ(\gamma,\gamma^{\max})}(\alpha)\cdot\mathcal{U}_{I^\circ(\gamma,\gamma^{\max})}\otimes\delta_{(\gamma,\gamma^{\max})}(C)=R_\alpha(z,C),\end{aligned}$$

其中，最后一个等式由式（25）得出。

接下来讨论 $R_\alpha(z,G_\alpha)=1$。同样，由式（25）以及 G_α 的定义可得

$$\begin{aligned}&R_\alpha((\gamma,\gamma^{\min},\gamma^{\max}),G_\alpha)\\&=\mathbf{1}_{I^\circ(\gamma,\gamma^{\max})}(\alpha)\cdot\mathcal{U}_{I^\circ(\gamma,\gamma^{\max})}\left(I^\circ(\gamma,\gamma^{\max})\setminus\{\alpha\}\right)\\&\quad+\mathbf{1}_{[0,2\pi)\setminus\{\gamma^{\min}\}}(\gamma)\,\mathbf{1}_{I^\circ(\gamma^{\min},\gamma)}(\alpha)\cdot\mathcal{U}_{I^\circ(\gamma^{\min},\gamma)}\left(I^\circ(\gamma^{\min},\gamma)\setminus\{\alpha\}\right)\\&=\mathbf{1}_{I^\circ(\gamma,\gamma^{\max})}(\alpha)+\mathbf{1}_{J(\gamma^{\min},\gamma)}(\alpha)=1.\end{aligned}$$

□

回顾核与概率测度的张量积概念。对于 $\alpha\in[0,2\pi)$ 令 $(Z_k)_{k\in\mathbb{N}}$ 为转移核为 R_α、定义在 Λ_α 上且初始分布为 $\mathbb{P}_{Z_1}$（见（2））的收缩马尔可夫链。那么，对于 $k\in\mathbb{N}$ 乘积测度 $\mathbb{P}_{Z_1}\otimes R_\alpha^{\otimes(k-1)}$ 定义在乘积空间 $([0,2\pi)^{3k}$ 上，$\mathcal{B}([0,2\pi)^{3k}))$ 是满足下列条件的唯一确定的概率测度

$$\begin{aligned}\mathbb{P}_{Z_1}\otimes R_\alpha^{\otimes(k-1)}(A_1\times\cdots\times A_k)&=\int_{A_1}\int_{A_2}\cdots\int_{A_k}R_\alpha(z_{k-1},\mathrm{d}z_k)\ldots R_\alpha(z_1,\mathrm{d}z_2)\mathbb{P}_{Z_1}(\mathrm{d}z_1)\\&=\mathbb{P}(Z_1\in A_1,\ldots,Z_k\in A_k\mid\Theta=\alpha)\end{aligned}$$

对于任意 $A_1,\ldots,A_k\in\mathrm{B}([0,2\pi)^3)$。对于 $k\geq2$，我们利用以下事实：

$$\mathbb{P}_{Z_1}\otimes R_\alpha^{\otimes(k-1)}(A_1\times\cdots\times A_k)=\int_{A_1\times\cdots\times A_{k-1}}R_\alpha(z_{k-1},A_k)\,\mathbb{P}_{Z_1}\otimes R_\alpha^{\otimes(k-2)}(\mathrm{d}z_1\ldots\mathrm{d}z_{k-1}).$$

由此，我们再补充一个辅助观察。

引理 A.3. 令 $\alpha, \theta \in (0, 2\pi]$ 并令 g_θ, $\widetilde{g}_\theta$ 与引理 A.2 中的定义相同。那么，对于任意 $k \in \mathbb{N}$ 和 $C_1, \ldots, C_k \in \mathcal{B}([0,2\pi)^3)$ 我们有

$$\mathbb{P}_{Z_1} \otimes R_{g_\theta^{-1}(\alpha)}^{\otimes(k-1)} \left(\widetilde{g}_\theta(C_1) \times \ldots \times \widetilde{g}_\theta(C_k)\right) = \mathbb{P}_{Z_1} \otimes R_\alpha^{\otimes(k-1)} \left(C_1 \times \ldots \times C_k\right). \tag{26}$$

证明。 对于 G_α （如引理 A.2 中所定义），对于任意 $k \in \mathbb{N}$ 我们通过归纳法证明

$$\mathbb{P}_{Z_1} \otimes R_\alpha^{\otimes(k-1)} \left(G_\alpha \times \ldots \times G_\alpha\right) = 1. \tag{27}$$

对于 $k = 1$，我们有

$$\mathbb{P}_{Z_1}(G_\alpha) = \int_{[0,2\pi)} \mathbf{1}_{G_\alpha}(\gamma, \gamma, \gamma) \, \frac{\mathrm{d}\gamma}{2\pi} = \mathcal{U}_{[0,2\pi)}([0,2\pi) \setminus \{\alpha\}) = 1.$$

假设（27）对 k成立。然后，利用引理 A.2 中的 R_α（z_k, G_α）$= 1$，可得

$$\begin{aligned}\mathbb{P}_{Z_1} \otimes R_\alpha^{\otimes k} \left(G_\alpha \times \ldots \times G_\alpha\right) &= \int_{G_\alpha \times \cdots \times G_\alpha} R(z_k, G_\alpha) \, \mathbb{P}_{Z_1} \otimes R^{\otimes(k-1)} (\mathrm{d}z_1 \ldots \mathrm{d}z_k) \\ &= \mathbb{P}_{Z_1} \otimes R_\alpha^{\otimes(k-1)} \left(G_\alpha \times \ldots \times G_\alpha\right) = 1.\end{aligned}$$

现在我们再次通过归纳法证明（26）。事实上，通过变量替换 $\widetilde{\gamma} = g_\theta(\gamma)$，对于所有$C_1 \in$ B（$[0,2\pi)$ ^{3}），我们有

$$\mathbb{P}_{Z_1}\left(\widetilde{g}_\theta(C_1)\right) = \int_0^{2\pi} \mathbf{1}_{C_1}\left(g_\theta(\gamma), g_\theta(\gamma), g_\theta(\gamma)\right) \frac{\mathrm{d}\gamma}{2\pi} = \int_0^{2\pi} \mathbf{1}_{C_1}\left(\widetilde{\gamma}, \widetilde{\gamma}, \widetilde{\gamma}\right) \frac{\mathrm{d}\widetilde{\gamma}}{2\pi} = \mathbb{P}_{Z_1}\left(C_1\right),$$

这给出了 $k = 1$的结论。假设（26）对 $k \in \mathbb{N}$成立。那么

$$\begin{aligned}&\mathbb{P}_{Z_1} \otimes R_{g_\theta^{-1}(\alpha)}^{\otimes k} \left(\widetilde{g}_\theta(C_1) \times \ldots \times \widetilde{g}_\theta(C_{k+1})\right) \\ &= \int_{\widetilde{g}_\theta(C_1) \times \ldots \times \widetilde{g}_\theta(C_k)} R_{g_\theta^{-1}(\alpha)}\left(z_k, \widetilde{g}_\theta(C_{k+1})\right) \mathbb{P}_{Z_1} \otimes R_{g_\theta^{-1}(\alpha)}^{\otimes(k-1)} (\mathrm{d}z_1 \ldots \mathrm{d}z_k) \\ &= \int_{C_1 \times \ldots \times C_k} R_{g_\theta^{-1}(\alpha)}\left(\widetilde{g}_\theta(z_k), \widetilde{g}_\theta(C_{k+1})\right) \mathbb{P}_{Z_1} \otimes R_\alpha^{\otimes(k-1)} (\mathrm{d}z_1 \ldots \mathrm{d}z_k) \\ &\underset{(27)}{=} \int_{(C_1 \times \ldots \times C_k) \cap G_\alpha^k} R_{g_\theta^{-1}(\alpha)}\left(\widetilde{g}_\theta(z_k), \widetilde{g}_\theta(C_{k+1})\right) \mathbb{P}_{Z_1} \otimes R_\alpha^{\otimes(k-1)} (\mathrm{d}z_1 \ldots \mathrm{d}z_k) \\ &\underset{(24)}{=} \int_{(C_1 \times \ldots \times C_k) \cap G_\alpha^k} R_\alpha\left(z_k, C_{k+1}\right) \mathbb{P}_{Z_1} \otimes R_\alpha^{\otimes(k-1)} (\mathrm{d}z_1 \ldots \mathrm{d}z_k) \underset{(27)}{=} \mathbb{P}_{Z_1} \otimes R_\alpha^{\otimes k} \left(C_1 \times \ldots \times C_{k+1}\right),\end{aligned}$$

证明至此完成。 □

现在我们转而证明引理2.12。

证明。 将（26）代入 $Q_{g_\theta^{-1}(S)}$ 的定义，可得对所有 $B \in \mathrm{B}(S)$,

$Q_{g_\theta^{-1}(S)}$ $(g_\theta^{-1}(\alpha),$ $g_\theta^{-1}(B))$

$$= \sum_{k=1}^{\infty} \mathbb{P}\left[\Gamma_k \in g_\theta^{-1}(B) \cap g_\theta^{-1}(S), \Gamma_1 \in g_\theta^{-1}(S)^c, \ldots, \Gamma_{k-1} \in g_\theta^{-1}(S)^c \mid \Theta = g_\theta^{-1}(\alpha)\right]$$

$$= \sum_{k=1}^{\infty} \mathbb{P}_{Z_1} \otimes R_{g_\theta^{-1}(\alpha)}^{\otimes(k-1)} \left(\widetilde{g}_\theta(S^c \times [0,2\pi)^2)^{k-1} \times \widetilde{g}_\theta(S \cap B \times [0,2\pi)^2)\right)$$

$$\underset{(26)}{=} \sum_{k=1}^{\infty} \mathbb{P}_{Z_1} \otimes R_\alpha^{\otimes(k-1)} \left((S^c \times [0,2\pi)^2) \times \ldots \times (S^c \times [0,2\pi)^2) \times (S \cap B \times [0,2\pi)^2)\right) = Q_S(\alpha, B),$$

这就验证了该结论。 □

致谢

作者感谢 Michael Habeck、Philip Schär 和 Björn Sprungk 对稿件初稿提出意见，并就这一主题进行了富有成效的讨论。Mareike Hasenpflug 和 Daniel Rudolf 衷心感谢德国研究联合会通过项目 432680300 –SFB 1456 子项目 B02 提供的资助。

参考文献

BESKOS, A., PINSKI, F. J., SANZ-SERNA, J. M. 和 STUART, A. M. (2011)。希尔伯特空间上的混合蒙特卡罗。《随机过程及其应用》121 2201–2230。https://doi.org/10.1016/j.spa.2011.06.003 BESKOS, A., GIROLAMI, M., LAN, S., FARRELL, P. E. 和 STUART, A. M. (2017)。无限维逆问题的几何马尔可夫链蒙特卡罗。《计算物理学杂志》335 327–351。https://doi.org/10.1016/j.jcp.2016.12.041 COTTER, S. L., ROBERTS, G. O., STUART, A. M. 和 WHITE, D. (2013)。函数的马尔可夫链蒙特卡罗方法：改进旧算法以提升速度。《统计科学》424–446。https://doi.org/10.1214/13-STS421 CUI, T., LAW, K. J. H. 和 MARZOUK, Y. M. (2016)。维数无关的似然信息马尔可夫链蒙特卡罗。《计算物理学杂志》304 109–137。https://doi.org/j.jcp.2015.10.008 DA PRATO, G. 和 ZABCZYK, J. (2002)。希尔伯特空间中的二阶偏微分方程 293。剑桥大学出版社，剑桥。https://doi.org/10.1017/CBO9780511543210 DOUC, R., MOULINES, E., PRIOURET, P. 和 SOULIER, P. (2018)。马尔可夫链。施普林格，Cham。https: // doi.org/10.1007/978-3-319-97704-1 GESSNER, A., KANJILAL, O. 和 HENNIG, P. (2020)。线性区域约束下的高斯积分。载于《第23届国际人工智能与统计会议论文集》2764–2774。GRENIOUX, L., O. DURMUS, A., MOULINES, E. 和 GABRIÉ, M. (2023)。使用近似传输映射进行采样。载于《第40届国际机器学习会议论文集》，机器学习研究论文集 202 11698–11733。HABECK, M., RUDOLF, D. 和 SPRUNGK, B. (2020)。双重难处理分布的稳定性。《电子概率通讯》25 1–13。https://doi.org/10.1214/20-ECP341 HABECK, M., HASENPFLUG, M., KODGIRWAR, S. 和 RUDOLF, D. (2023)。球面上的测地线切片采样。arXiv:2301.08056。预印本。https://doi.org/10.48550/arXiv.2301.08056 JONES, G. (2004)。关于马尔可夫链中心极限定理。《概率调查》1 299–320。https://doi.org/10.1214/ 154957804100000051 KALLENBERG, O. (2002)。现代概率论基础，第二版。概率及其应用。施普林格出版社，纽约。https://doi.org/10.1007/978-3-030-61871-1 ŁATUSZY ´NSKI, K. 和 RUDOLF, D. (2024)。通过谱隙研究混合切片采样的收敛性。《应用概率杂志》。已接收。https://doi.org/10.48550/arXiv.1409.2709 LAW, K. J. H. (2014)。加速函数空间马尔可夫链蒙特卡罗的提议。《计算与应用数学杂志》262 127–138。https://doi.org/10.1016/j.cam.2013.07.026

LIE, H. C., RUDOLF, D., SPRUNGK, B. 和 SULLIVAN, T. J. (2023)。球面上的维数无关马尔可夫链蒙特卡罗。《斯堪的纳维亚统计学杂志》50 1818–1858。https://doi.org/10.1111/sjos.12653
MEDINA-AGUAYO, F., RUDOLF, D. 和 SCHWEIZER, N. (2020)。通过受限近似获得蒙特卡罗方法在梅特罗波利斯算法中的扰动界。《随机过程及其应用》130 2200–2227。https://doi.org/10.1016/j.spa.2019.06.015 MEYN, S. 和 TWEEDIE, R. L. (2009)。马尔可夫链与随机稳定性。剑桥大学出版社, 剑桥。https://doi.org/10.1007/978-1-4471-3267-7 MURRAY, I., ADAMS, R. P. 和 MACKAY, D. J. C. (2010)。椭圆切片采样。载于《第13届国际人工智能与统计会议论文集》541–548。MURRAY, I. 和 GRAHAM, M. (2016)。伪边际切片抽样。载于《第19届国际人工智能与统计会议论文集》911–919。
NATAROVSKII, V., RUDOLF, D. 和 SPRUNGK, B. (2021a)。椭圆切片采样的几何收敛。载于《第38届国际机器学习会议论文集》，机器学习研究论文集 7969–7978。NATAROVSKII, V., RUDOLF, D. 和 SPRUNGK, B. (2021b)。简单切片抽样的定量谱隙估计与Wasserstein收缩。《应用概率年刊》31 806–825。https://doi.org/10.1214/20-AAP1605 NEAL, R. M. (1999)。使用高斯过程先验进行回归与分类。载于《贝叶斯统计学 6：第6届瓦伦西亚国际会议论文集》475-501。
https://doi.org/10.1093/oso/9780198504856.003. 0021 NEAL, R. M. (2003)。切片抽样。《统计年刊》31 705–767。https://doi.org/10.1214/aos/1056562461 NISHIHARA, R., MURRAY, I. 和 ADAMS, R. P. (2014)。采用广义椭圆切片采样的并行马尔可夫链蒙特卡罗。《机器学习研究杂志》15 2087–2112。https://doi.org/10.48550/arXiv.1210.7477 OTTOBRE, M., PILLAI, N. S., PINSKI, F. J. 和 STUART, A. M. (2016)。具有二阶朗之万扩散极限的函数空间HMC算法。《伯努利》22 60–106。
https://doi.org/10.3150/14-BEJ621 PINSKI, F. J., SIMPSON, G., STUART, A. M. 和 WEBER, H. (2015)。无限维概率测度的Kullback-Leibler近似算法。《SIAM科学计算杂志》37 a2733–a2757。https: //doi.org/10.1137/14098171X RAY, P., PATI, D. 和 BHATTACHARYA, A. (2020)。利用约束高斯过程先验进行高效的贝叶斯形状约束函数估计。《统计计算》30 839–853。
https://doi.org/10.1007/s11222-020-09922-0 RUDOLF, D. (2012)。马尔可夫链蒙特卡罗的显式误差界。《数学论文集》485 1–93。https: //doi.org/10.4064/dm485-0-1 RUDOLF, D. 和 SCHWEIZER, N. (2018)。通过Wasserstein距离研究马尔可夫链扰动理论。《伯努利》24 2610–2639。
https://doi.org/10.3150/17-BEJ938 RUDOLF, D. 和 SPRUNGK, B. (2018)。关于预条件Crank-Nicolson Metropolis算法的一种推广。《计算数学基础》18 309–343。
https://doi.org/10.1007/s10208-016-9340-x RUDOLF, D. 和 SPRUNGK, B. (2022)。适用于集中后验的稳健类随机游走Metropolis-Hastings算法。arXiv:2202.12127。预印本。
https://doi.org/10.48550/arXiv.2202.12127 SCHÄR, P., HABECK, M. 和 RUDOLF, D. (2023)。吉布斯极坐标切片抽样。载于《第40届国际机器学习会议论文集》30204–30223。SPRUNGK, B. (2020)。贝叶斯逆问题的局部Lipschitz稳定性。《逆问题》36 31页。https://doi.org/10.1088/1361-6420/ab6f43 TIERNEY, L. (1994)。用于探索后验分布的马尔可夫链。《统计年刊》22 1701–1728。https: //doi.org/10.1214/aos/1176325750