arXiv:2305.02256v1 [math.ST] 3 May 2023

近似Wonham滤波器的指数收缩与鲁棒性

Samuel N. Cohen* Eliana Fausti†

2023年5月4日

摘要

我们研究如何根据含噪观测估计连续时间马尔可夫链状态的问题。我们证明，在希尔伯特射影空间中考察时，相应的最优滤波器沿路径严格收缩，并给出明确的确定性收敛速率和逐路径收敛速率。基于此，我们为最优滤波器的鲁棒性提供了另一种证明，改进了已有的误差估计，并推导出近似滤波器严格且可计算的误差界。

数学分类号：93E11, 62M05, 60J55.

关键词：非线性滤波、稳定性、模型稳健性、近似滤波器、误差界、希尔伯特射影度量、局部时。

1 引言

从不完整且含噪的观测中估计随机隐过程，是工程、信号处理、金融及许多其他应用领域中的常见问题。一般而言，信号（或状态）过程 X 随时间演化（通常假设为马尔可夫过程），无法直接测量，但需要利用观测过程 Y所提供的信息进行估计；该观测过程的动态取决于 X。计算并分析这一问题的最优解，是随机滤波理论的主要目标。

随机滤波是随机分析中的经典主题，在各种情境下都已推导出最优滤波器，例如连续时间或离散时间、有限或无限状态空间等。线性底层动力学所导出的著名卡尔曼–布西滤波器 [20, 21]，是最早在连续时间情形下得到研究的，如今已得到充分理解。另一方面，非线性滤波在理论和实践层面都面临挑战（关于非线性滤波的介绍，可参阅 Bain 和 Crisan [5] ）。

最优非线性滤波器是一个称为 非线性随机（根据上下文，也可能是 偏微分）微分方程的解，该方程称为 $Kushner–Stratonovich$ 方程。然而，在几乎所有实际应用中，都无法直接计算该解：例如， X 和 Y的模型（滤波方程明确依赖于该模型）可能存在设定错误。

*牛津大学数学研究所和艾伦·图灵研究所，cohens@maths.ox.ac.uk。

†牛津大学数学研究所，fausti@maths.ox.ac.uk。

参数，也可能完全未知。此外，即使已知真实模型，由于问题具有高维（在许多情况下甚至是无限维）且非局部的特性，数值求解滤波方程也可能难以实现。实际中，使用的往往是近似滤波器，而非最优滤波器。这就引出了两个问题：这些近似是否可靠？我们又该如何量化其相对于最优滤波器的误差？这些问题是研究非线性滤波器鲁棒性的核心。

本文将重点讨论连续时间下有限状态空间非线性滤波的情形。在这种情况下，最优滤波器是一个随机微分方程（SDE）的解，有时也称为*Wonham*滤波器 [38]。Chigansky 和 Van Handel [14] 以及 Van Handel [34]在 Wonham滤波器的鲁棒性研究方面取得了重要进展。在这些研究中，考察了真实 Wonham滤波器与模型参数设定错误的 Wonham滤波器之间的 L^1-误差。他们采用一种通过计算滤波器导数界来进行分析的方法，证明误差在无限时间范围内始终有限，并且当模型设定错误的参数趋近于真实参数时，误差会趋于零。 [14] 中的方法或可用于计算一些更一般的近似滤波器的误差界，而不仅限于由底层模型设定错误所导致的近似滤波器。然而， [14] 中的估计不够紧，无法给出有用的定量界（参见 [14, 注 2.8] 和 [34, 注 3.3.8]），因此这些结果仍主要具有定性意义。

鲁棒性问题始终与非线性滤波器的稳定性 问题相关。与鲁棒性相比，稳定性仅关注滤波方程初始条件设定错误所导致的误差。如果真实滤波器与“初始化错误”的滤波器之间的误差随着时间推移趋于零，那么该滤波器便会“忘却”初始误差，并称为渐近稳定。这对近似滤波有重要影响：直观地说，在离散时间设定下，如果非线性滤波器稳定，那么使用近似滤波器本质上等同于使用最优滤波器，只是在每个时间步引入了近似误差。如果所有近似误差都有界，稳定性便能确保这些误差也会随着时间推移而被“忘却”，从而使总误差保持有界，并得到稳健性型估计。

在离散时间的有限状态空间非线性滤波中，鲁棒性估计确实就是这样推导出来的（参见 Budhiraja 和 Kushner [11], Le Gland 和 Mevel [26] 以及 Le Gland 和 Oudjane [27]）。与连续时间情形的主要区别在于，文献中针对 Wonham滤波器的稳定性估计还不够强，无法直接应用这类方法。事实上，要在连续时间中实现这一论证，需要稳定性误差呈指数压缩（或具有类似性质）。本文的首要目标是为 Wonham滤波器建立这样的压缩结果。第二个目标是利用我们的稳定性估计，为近似滤波器提供可计算误差界。

1.1 已知结果讨论

自20世纪90年代以来，滤波稳定性一直是一个活跃的研究领域。文献中的一篇重要论文是 [29], ，Ocone和Pardoux在其中确立了卡尔曼滤波器的稳定性与信号-观测线性控制系统的可检测性/可稳定性之间的关系。然而，他们关于非线性情形下稳定性的论证依赖于Kunita 的一项结果 [24], ，后来发现该结果存在错误（详见Baxendale、Chigansky和Liptser [7, 第 2] 节，其中给出了详细解释和一个反例；另请参阅Budhiraja [10], 了解

对其在非线性滤波稳定性背景下相关性的分析）。Van Handel弥补了Kunita证明中的漏洞，针对不同情形确立了非线性滤波器稳定性的必要条件（包括离散时间和连续时间下的遍历信号 [36], 、具有紧状态空间的非遍历信号 [35], ，以及具有波兰状态空间的非遍历信号 [37]）。近来，Kim、Mehta和Meyn [23] 以及Kim和Mehta [22], 证明，Wonham滤波器的稳定性等价于某个对偶控制问题的可稳定性，这是将 [29]推广至非线性情形的结果。感兴趣的读者可参阅Chigansky [12]，其中对非线性滤波稳定性结果（离散时间、有限状态空间）作了全面综述；也可参阅 [16, 第 3]部分，其中汇集了大量综述论文。与本文研究情形尤为相关的是，Chigansky、Liptser和Van Handel [13]对 [35–37]的稳定性结果作了通俗易懂的介绍。

尽管上述结果在尽可能一般的条件下（且只需尽可能弱的假设）保证了滤波器的稳定性，但其定性性质使其不适合用于理解一般的近似误差。另一方面，如果愿意对信号过程施加相对较强的遍历性假设，文献中已有明确的衰减率，至少对 Wonham滤波器这一特定情形如此。Delyon 和 Zeitouni [17] 开创了对 Wonham滤波器最大Lyapunov指数的研究，并证明在模型参数满足某些条件时，该指数为负。Atar 和 Zeitouni [3, 4], 扩展了这一方法：在对信号采用相当强的混合假设后，他们计算出了稳定性误差的显式指数衰减率。运用 [3], Baxendale、Chigansky 和 Lipster 的方法，他们通过证明当所有状态彼此连通时衰减率几乎必然为负，略微放宽了遍历性假设 X [7, 定理 4.1]（不过，这样一来便无法得到显式的衰减率）。最后，通过研究平滑过程 （例如 Liptser 和 Shiryayev [28, 定理 9.5]中所述），他们针对混合信号，给出了以其遍历分布 [7, 定理 4.2], 为依据的显式指数衰减率，以及稳定性误差的非渐近 指数上界 [7,定理 4.3], ，其衰减率与 [3, 4]相同。

据我们所知， [7,定理 4.3]中的界是文献中唯一一个针对连续时间Wonham滤波器稳定性误差的非渐近界。指数衰减项的前因子与Wonham随机微分方程的维数以及真实初始分布和“错误”初始分布的Radon–Nikodym导数成正比；从定量角度来看，这个前因子过大，使得该界难以发挥实际作用。Van Handel对其进行了显著改进（尽管结果仍远未达到压缩），而关于前因子的最佳估计可通过结合 [14, 命题 3.5] 和 [34, 推论 2.3.2]得到。这一稳定性结果是 [14]中Wonham滤波器鲁棒性分析的核心。另一方面，我们之前提到的离散时间非线性滤波器鲁棒性结果 [11, 26, 27]，建立在Atar和Zeitouni关于稳定性的研究之上（ [3, 4]中同时分析了离散时间和连续时间情形）。

[3, 4] 的根本贡献是引入希尔伯特射影距离（见下文的 [3, 式 9], 或（3））作为概率测度空间上的度量，用于对非线性滤波器进行稳定性估计。使用希尔伯特度量的一个关键优势是，正线性算子在该距离下会发生收缩：这是 Birkhoff 的一项结果（见 [8] 或 [9, 第十六章]）。Seneta 关于正线性算子乘积的研究（见 [32], 或 [31, 第 3]章），涵盖了离散时间马尔可夫链遍历性的分析，对于理解在恰当的背景下使用希尔伯特距离这一工具的强大之处尤为启发深刻。鉴于离散时间马尔可夫链的生成元是随机矩阵，Birkhoff 和 Seneta 的研究使稳定性结果

对于离散时间非线性滤波而言，直观上很简单。

Atar 和 Zeitouni 给出了离散时间和连续时间两种情形下滤波器稳定性误差衰减的渐近速率。在这些思想以及 Seneta 的工作的基础上，Le Gland 和 Mevel [25,26], 随后又与 Oudjane [27] 分别证明了离散时间情形下非渐近、非对数的稳定性界，其条件是信号过程满足强混合假设。在 [27], 中，他们还能够处理非线性滤波器的鲁棒性问题（离散时间情形），并特别研究了相互作用粒子近似滤波过程的全局误差。本文的结果大致沿着相同思路展开，不过研究的是连续时间情形。此外，我们的方法与 [3,4,25–27]中的方法有本质区别；两者唯一的共同点是在稳定性分析中使用希尔伯特度量。

1.2 主要贡献与论文结构

我们的第一个主要贡献是针对连续时间、有限状态空间非线性滤波器的稳定性误差，给出了希尔伯特距离下的指数收缩估计（见定理 2.2）。事实上，我们的结论更强，因为我们可以证明希尔伯特距离的双曲正切具有压缩性，而这可以直接推出前述结果。据我们所知，这两项结果都是新的，并且显著改进了现有文献中关于误差的定量估计。我们还提出了一种在希尔伯特距离下研究连续时间滤波器稳定性误差的替代方法，该方法不依赖于 Atar 和 Zeitouni 的论证。相反，受 Amari [2], 的启发，我们将引入一个从概率单纯形到 $\mathbb{R}^n$的坐标变换，并研究 Wonham滤波器在新坐标系中的演化。正如我们将看到的，尽管采用了不同的方法，我们的论证与 [7, 定理 4.3],的证明存在一些相似之处。

我们的第二项主要贡献是针对 Wonham滤波器的稳健性型估计（见定理 2.3）。与 [14], 相比，我们针对一般的近似滤波器，并以希尔伯特距离为度量给出了误差界。由于希尔伯特距离比 L_1-范数更强，而 [14], 使用的是该范数，因此我们给出的误差界更紧（尽管仍非最优，正如我们将在第3.4节和第 4.2.1 节中讨论的那样）。我们也认为，我们的证明方法本身就很有意义，因为它远比 [14]中的论证简单：它仅依赖标准的随机分析工具，而 [14] 中的作者则需要借助 Malliavin 微积分来处理预见性随机积分。

最后，定理 2.2 中的结果表明，在研究近似滤波器的误差时，希尔伯特距离的双曲正切（而非希尔伯特距离本身）或许是最优度量。特别地，当近似滤波器的选取使其随机项与 Wonham随机微分方程的随机项完全匹配时，我们便能得到更紧的逐路径误差界，并在定理 2.4 中证明了这一点。

本文组织如下：在下一节中，我们将建立滤波方程并定义希尔伯特范数，然后陈述我们的假设和主要结果。第 3 节专门用于证明定理 2.2，并对我们的稳定性结果进行一些讨论。第 4 节分为两个部分：在第一部分中，我们以某种方式重新获得 Chigansky 和 Van Handel 关于模型参数设定错误时鲁棒性的研究结果（参见定理 4.1，并与 [14, 定理 1.1]进行比较）；在第二部分中，我们证明定理 2.3 中给出的一般近似滤波器的误差界，以及定理 2.4 中的逐路径界，同时还展示并讨论一些数值实验。

2 滤波设定与主要结果

令 $(\Omega, \mathcal{F}, \mathbb{P})$ 为一个满足通常条件的概率空间，并配备滤过 $\{\mathcal{F}_t, t \geq 0\}$。考虑一个 $\{\mathcal{F}_t\}$-适应的连续时间、时间齐次马尔可夫链 $X = (X_t)_{t\geq 0}$，其具有有限状态空间

$$\mathbb{S} = \{a_0, \ldots, a_n\}$$

$Q = (q_{ij}) \in \mathbb{R}^{(n+1)\times(n+1)}$，以及相应的转移强度矩阵 (Q)。我们分别用 $\mathcal{M}(\mathbb{S})$ 和 $\mathcal{P}(\mathbb{S})$ 表示 $\mathbb{S}$ 上的非负测度和概率测度。设 X 的初始分布为 $\mu^i = \mathbb{P}(X_0 = a_i)$。

回顾一下，Q-矩阵定义为转移率矩阵，其每一行的元素之和为 0，且非对角元素非负，即 $\sum_j q_{ij} = 0$ 对所有 i，以及 $q_{ij} \geq 0$ 对所有 $i,\ j \leq n+1,\ i \neq j$，并且

$$M_t^{\varphi} = \varphi(X_t) - \varphi(X_0) - \int_0^t Q\varphi(X_s)\,\mathrm{d}s, \quad t \geq 0$$

对于所有有界函数 $\varphi : \mathbb{S} \to \mathbb{R}$，它都是一个 $\{\mathcal{F}_t\}$-适应的右连续鞅。

设 $h = (h_i)_{i=1}^d : \mathbb{S} \to \mathbb{R}^d$ 是一个有界函数，且 $\sigma \neq 0$。假设 W 是一个标准的、$\{\mathcal{F}_t\}$-适应的 d 维布朗运动，与 X 独立，并令 $Y = (Y_t)_{t\geq 0}$ 为满足以下随机微分方程的过程

$$Y_t = Y_0 + \int_0^t h(X_s)\,\mathrm{d}s + \sigma W_t. \tag{1}$$

令 $\{\mathcal{Y}_t\}_{t\geq 0}$ 为由观测过程 Y 生成的（完备化）自然滤过。这描述了在时间区间 $[0,t]$ 内观测 Y 所能获得的信息。

按照惯例，我们将状态空间 $\mathbb{S}$ 与标准基 $\{e_0, \ldots, e_n\}$ 对应起来，基底属于 $\mathbb{R}^{n+1}$。记 $\pi_t = \mathbf{E}[X_t | \mathcal{Y}_t]$ 为给定 $\mathcal{Y}_t$ 时 X 的条件期望。换言之，稍微滥用一下记号，$\pi_t^i = \mathbb{P}(X_t = a_i | \mathcal{Y}_t)$。

过程 π_t 满足 Wonham 形式的 ***Kushner–Stratonovich*** 方程（例如参见 [5, 式 3.53]）：

$$\mathrm{d}\pi_t = Q^\top \pi_t\,\mathrm{d}t + \frac{1}{\sigma^2}\sum_{k=1}^d \left(H^k - \pi_t^\top h_k\, \mathbb{I}_{n+1}\right)\pi_t \left(\mathrm{d}Y_t^k - \pi_t^\top h_k\,\mathrm{d}t\right), \quad \pi_0 = \mu, \tag{2}$$

其中，对于 $k = 1, \ldots, d$，$H^k = \mathrm{diag}(h_k(a_i))$ 是一个 $(n+1)\times(n+1)$ 维对角矩阵，$\mathbb{I}_{n+1}$ 是单位矩阵。注意，(2) 以 $\mu = \mathrm{law}(X_0) = \mathbf{E}[X_0]$ 为初值。

概率 π_t（$t \geq 0$）是 $(n+1)$ 维（列）向量，因此 (2) 是一个 $(n+1)$ 维非线性随机微分方程。事实上，由于分量 π_t^i 对所有 $t \geq 0$的和必须为 1，随机微分方程 (2) 描述了 n维概率单纯形 $\mathcal{S}^n$上的流，其中

$$\mathcal{S}^n = \left\{ x \in \mathbb{R}^{n+1} : \sum_i x_i = 1,\ x_i \geq 0 \right\}.$$

我们用 $\mathring{\mathcal{S}}^n$ 表示单纯形的内部，即 $x \in \mathring{\mathcal{S}}^n$，若 $x \in \mathcal{S}^n$，且对所有 i 都有 $x^i > 0$。

我们通过介绍所选度量来结束本节的设定。给定两个在 $\mathbb{R}^{n+1}$ 中表示为非负向量的非负测度 $\mu, \nu \in \mathcal{M}^+(\mathbb{S})$，希尔伯特射影距离 $\mathcal{H}$ 定义为

$$\mathcal{H}(\mu,\nu) = \begin{cases} \log\left(\dfrac{\max_{j:\nu^j>0} \frac{\mu^j}{\nu^j}}{\min_{i:\nu^i>0} \frac{\mu^i}{\nu^i}}\right), & \mu \sim \nu, \\ \infty, & \mu \nsim \nu. \end{cases} \tag{3}$$

希尔伯特距离是 $\mathcal{M}^+(\mathbb{S})$ 上的伪度量：它非负、对称，满足三角不等式，并且 $\mathcal{H}(\mu,\nu)=0$ 当且仅当 $\nu=c\mu$ 对某个常数 $c\in\mathbb{R}^+$成立。它是概率单纯形 $\mathcal{S}^n$上的度量。关于进一步讨论，请参阅 [9, 第十六章] 和 [31, 第 3] 章。以下这一性质将证明很有用，它将希尔伯特度量与 $\mathbb{R}^{n+1}$ 上的欧几里得度量联系起来。

引理 2.1。 对于任意$(\mu,\nu)\in\mathcal{S}^n\times\mathcal{S}^n$，均有 *JSImport*

(i) 对于所有在欧几里得度量下收敛的序列 $(\mu_m,\nu_m)\to(\mu,\nu)$，$\liminf_{m\to\infty}\mathcal{H}(\mu_m,\nu_m)\geq\mathcal{H}(\mu,\nu)$；*(ii)* 存在一个序列 $(\mu_m,\nu_m)\in\mathcal{S}^n\times\mathcal{S}^n$，使得 $\lim_{m\to\infty}\mathcal{H}(\mu_m,\nu_m)=\mathcal{H}(\mu,\nu)$，且 $(\mu_m,\nu_m)\to(\mu,\nu)$ 在欧几里得度量下；*(iii)* 若 $\mu,\nu\in\mathcal{S}^n$，则对于所有在欧几里得度量下收敛的序列 $(\mu_m,\nu_m)\to(\mu,\nu)$，$\lim_{m\to\infty}\mathcal{H}(\mu_m,\nu_m)=\mathcal{H}(\mu,\nu)$。

证明。 如果 $(\mu,\nu)\in\mathcal{S}^n\times\mathcal{S}^n$，则根据希尔伯特度量的定义以及除法和对数运算的连续性，结论立即成立，从而证明了*(iii)*。现在考虑一对在欧几里得度量下收敛的数列 $\{\mu_m\}$, $\{\nu_m\}\in\mathring{\mathcal{S}}^n$，其相应的极限分别为 $\mu,\nu\in\mathcal{S}^n$。首先假设 $\mu\nsim\nu$，则容易验证，或者 $\max_i\{\mu_m^i/\nu_m^i\}\to\infty$，或者 $\min_i\{\mu_m^i/\nu_m^i\}\to 0$，因此 $\mathcal{H}(\mu_m,\nu_m)\to\infty=\mathcal{H}(\mu,\nu)$。

假设情况相反，即 $\mu\sim\nu$，令 $J=\{j:\ \mu^j=\nu^j=0\}$. 若

$$\limsup_{m\to\infty}\max_{j\in J}\{\mu_m^j/\nu_m^j\}\leq\max_{i:\nu^i>0}\{\mu^i/\nu^i\}\quad\text{and}\quad\liminf_{m\to\infty}\min_{j\in J}\{\mu_m^j/\nu_m^j\}\geq\min_{i:\nu^i>0}\{\mu^i/\nu^i\},$$

然后直接计算可得 $\mathcal{H}(\mu_m,\ \nu_m)\to\mathcal{H}(\mu,\ \nu)$。由于我们总能选择满足上述两个不等式的 $\{\mu_m\}$ and$\{\nu_m\}$，这就证明了 *(ii)*。

如果给定的序列 $(\mu_m,\ \nu_m)$ 不满足上述不等式，则取任意子序列，仍以 m 为索引，使得 $\mathcal{H}(\mu_m,\ \nu_m)$ 在 $[0,\infty]$ 中收敛，并且上述不等式中至少有一个对该子序列中的每一项均不成立；特别地，假设对于某个 $\epsilon>0$，对所有 m,

$$\max_{j\in J}\{\mu_m^j/\nu_m^j\}>\max_{i:\nu^i>0}\{\mu^i/\nu^i\}+\epsilon.$$

Then

$$\lim_{m\to\infty}\mathcal{H}(\mu_m,\nu_m)\geq\log\left(\frac{\max_{i:\nu^i>0}\{\mu^i/\nu^i\}+\epsilon}{\min_{i:\nu^i>0}\{\mu^i/\nu^i\}}\right)>\mathcal{H}(\mu,\nu).$$

对于任意满足 $\min_{j\in J}\{\mu_m^j/\nu_m^j\}<\min_{i:\nu^i>0}\{\mu^i/\nu^i\}-\epsilon$ 的子序列，类似的论证同样成立。因此，我们得出 $\liminf_{m\to\infty}\mathcal{H}(\mu_m,\nu_m)\geq\mathcal{H}(\mu,\nu)$，这就是 *(i)*。 □

2.1 主要结果

全文中，我们对第 2 节所述的非线性滤波系统作出以下假设。

(A1). X 是一个具有 $n+1$ 个状态的时间齐次连续时间马尔可夫链。

(A2). $h=(h_i)_{i=1}^d$ *is bounded for all* i.

(A3). $\sigma = 1$ 和 $d = 1$ 分别出现在(1) 和(2)中。

最后一个假设仅用于简化记号——我们的所有结果都很容易推广到 Y 为多维且 $\sigma \in \mathbb{R}^{d\times d}$ 可逆的情形。同样，只要前者始终有正的下界，后者对所有 t 都保持有界，我们也可以很容易地允许 σ 和 h 依赖于时间；此外，还可以允许 X 的马尔可夫链动力学具有时间非齐次性。

记号。 鉴于我们将观测 Y 视为一维的，传感器函数h：$\mathbb{S} \to \mathbb{R}$ 可以看作一个向量 $h \in \mathbb{R}^{n+1}$，其分量为 $h^i = h(a_i)$，其中 $i = 0, \ldots\ldots, n$。从现在起，我们将采用这一记号。我们还用 $H = \mathrm{diag}(h)$ 表示其元素为 $(H)_{ii} = h^i$ 的对角矩阵。一般而言，我们始终用上标表示向量（或向量值过程）的分量。我们用 $\mathbf{N}$ 表示自然数集 $\{0, \ldots, n\}$。~ ~有时，对于定义在 $[0, \infty)$ 上的两个勒贝格–斯蒂尔切斯测度 A_t 和 A_t，我们会写作 $\mathrm{d}A_t \le \mathrm{d}A_t$，其含义是：对所有 $0 \le s < t < \infty$，均有 $\int_s^t \mathrm{d}A_r \le \int_s^t \mathrm{d}\,\tilde{A}_r$。

作为参考，在上述假设和记号下，我们在此重写Wonham滤波器的方程(2)

$$\mathrm{d}\pi_t = Q^\top \pi_t \,\mathrm{d}t + \left(H - \pi_t^\top h\, \mathbb{I}_{n+1}\right) \pi_t \left(\mathrm{d}Y_t - \pi_t^\top h \,\mathrm{d}t\right), \quad \pi_0 = \mu. \tag{4}$$

我们首先考虑 π_t 与 $\tilde{\pi}_t$之间误差的长期行为，其中 $\tilde{\pi}_t$是采用“错误”初始数据 $\tilde{\pi}_0 = \nu \neq \mu$ 、但与 π具有相同动力学的滤波器。 $\tilde{\pi}_t$ 的演化方程为

$$\mathrm{d}\tilde{\pi}_t = Q^\top \tilde{\pi}_t \,\mathrm{d}t + \left(H - \tilde{\pi}_t^\top h\, \mathbb{I}_{n+1}\right) \tilde{\pi}_t \left(\mathrm{d}Y_t - \tilde{\pi}_t^\top h \,\mathrm{d}t\right), \quad \tilde{\pi}_0 = \nu. \tag{5}$$

我们的关键结果是关于滤波器稳定性的以下逐路径估计。

定理 2.2 ($\mathcal{H}(\pi_t, \tilde{\pi}_t)$的收缩率）。令 π_t 为(4) 的解， $\tilde{\pi}_t$ 为(5)的解。假设对所有 $i \neq j$都有 $q_{ij} > 0$ 。则对所有 $t < \infty$,

$$\tanh\left(\frac{\mathcal{H}(\pi_t, \tilde{\pi}_t)}{4}\right) \le \tanh\left(\frac{\mathcal{H}(\mu, \nu)}{4}\right) e^{-\lambda t},$$

其中 $\lambda = 2$的最小值$_{i\neq j} \sqrt{q_{ij} q_{ji}}$。特别地,

$$\mathcal{H}(\pi_t, \tilde{\pi}_t) \le \mathcal{H}(\mu, \nu) e^{-\lambda t}.$$

不出所料，我们的收缩率与 [4], 中的渐近速率以及 [7,定理 4.3], 中的非渐近速率相同，并且存在同样的问题：只有当 Q 的所有非对角元素都为（严格）正值时，该速率才是（严格）正的。这是关于 X的一个非常强的混合假设；然而，要计算显式的收缩率，这一假设似乎是必要的。事实上，在离散时间情形的 [27] 中也采用了类似的假设（见 [27,定义 3.2]）。

基于这些稳定性估计，我们可以继续探讨下一个难题：理解近似滤波器的误差。考虑如下形式的一般近似滤波器:

$$\mathrm{d}\tilde{\pi}_t = \tilde{f}_t \,\mathrm{d}t + \tilde{g}_t \,\mathrm{d}Y_t, \quad \tilde{\pi}_0 = \nu, \tag{6}$$

其中， $\tilde{f}_t, \tilde{g}_t$ 是取值于 $\mathbb{R}^{n+1}$ 的 $\{\mathcal{Y}_t\}$-可预测过程，并且对于所有 t，都有 $\tilde{\pi}_t \in \mathring{\mathcal{S}}^n$ ；关于 f_t、 $\tilde{g}_t$ 和 $\tilde{\pi}_t$ 的必要假设，请参见~第 4.2 节。

定理 2.3（期望希尔伯特误差的界）。令 π_t 为（4）的解，并令 $\tilde{\pi}_t$ 为（6）的解。假设对于所有 $i \neq j$，μ、$\nu \in \mathcal{S}^n$ 和 $q_{ij} > 0$。假设（6）中具有充分的可积性（见假设 *(A4)*），对于所有 $t < \infty$，我们有

$$\begin{aligned}\mathbf{E}\left[\mathcal{H}(\pi_t, \tilde{\pi}_t)\right] \leq \mathcal{H}(\mu,\nu)e^{-\lambda t} &+ \int_0^t e^{-\lambda(t-s)}\mathbf{E}\left[\max_{i,k}\left\{\mathcal{E}_s^{1,i} - \mathcal{E}_s^{1,k} - \frac{1}{2}\big(\mathcal{E}_s^{2,i} - \mathcal{E}_s^{2,k}\big)\right\}\right] \mathrm{d}s \\ &+ \max_j |h^j| \int_0^t e^{-\lambda(t-s)}\mathbf{E}\left[\max_{i,k}\left\{\mathcal{E}_s^{3,i} - \mathcal{E}_s^{3,k}\right\}\right] \mathrm{d}s \\ &+ \frac{1}{4}\sum_{(i,k)}\sum_{(j,l)\neq(i,k)} \mathbf{E}\left[\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}L_s^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot))\right],\end{aligned}$$

其中 $\lambda = 2\min_{i\neq k}\sqrt{q_{ik}q_{ki}}$ 是定理 *2.2* 中的确定性收缩率。对于 $j \in \mathbf{N}$，误差项为

$$\mathcal{E}_t^{1,j} = \left(\sum_{m=0}^n q_{mj}\frac{\tilde{\pi}_t^m}{\tilde{\pi}_t^j}\right) - \frac{\tilde{f}_t^j}{\tilde{\pi}_t^j}, \quad \mathcal{E}_t^{2,j} = (h^j)^2 - \frac{(\tilde{g}_t^j)^2}{(\tilde{\pi}_t^j)^2}, \quad \mathcal{E}_t^{3,j} = h^j - \frac{\tilde{g}_t^j}{\tilde{\pi}_t^j}, \tag{7}$$

过程 $(\Delta_{ik}(t))_{t\geq 0}$ 对于 $(i,k) \in N \times N$ 定义为 $\Delta_{ik}(t) = \log\frac{\pi_t^i}{\pi_t^k} - \log\frac{\tilde{\pi}_t^i}{\tilde{\pi}_t^k}$。集合 $\mathcal{I}_t = \{(i,k) : \Delta_{ik}(t) = \mathcal{H}(\pi_t, \tilde{\pi}_t)\}$ 是这些过程对所有 $t < \infty$ 的最大值点，而 $L_t^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot))$ 表示差分过程 $(\Delta_{ik} - \Delta_{jl})$ 在 *0* 处的局部时，对所有 (i,k)、$(j,l) \in N \times N$ 均成立。

假设比较 (6) 和 (4) 时随机项中没有误差，则可以得到更强的结果。

定理 2.4（希尔伯特误差的逐路径衰减率）。在与定理 *2.3* 相同的假设下，假设在（7）中定义的误差项 $\mathcal{E}^{3,i}_t$ 对所有 $i \in$ **自然数集** 以及所有$t \geq 0$均为零，并且 $\tilde{\pi}_t$ 是可观测的。令 $u_t \in (0, 1)$ 为具有随机系数的常微分方程的唯一解，其定义如下：

$$\frac{\mathrm{d}u_t}{\mathrm{d}t} = -\tilde{\lambda}^\star(t,u_t)u_t + \max_{i,k}\left\{\mathcal{E}_s^{1,i} - \mathcal{E}_s^{1,k}\right\}(1-u_t^2), \qquad u_0 = \tanh\left(\frac{\mathcal{H}(\mu,\nu)}{4}\right), \tag{8}$$

其中

$$\tilde{\lambda}^\star(t,u_t) = \min_{i\neq k}\left\{\left(q_{ik}\frac{\tilde{\pi}_t^i}{\tilde{\pi}_t^k} + \sum_{\substack{j\neq i,k,\\ j\notin\tilde{\mathcal{J}}_k^i(t,u_t)}} q_{jk}\frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^k}\right)\frac{1+u_t}{1-u_t} + \left(q_{ki}\frac{\tilde{\pi}_t^k}{\tilde{\pi}_t^i} + \sum_{\substack{j\neq i,k,\\ j\in\tilde{\mathcal{J}}_k^i(t,u_t)}} q_{ji}\frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^i}\right)\frac{1-u_t}{1+u_t}\right\},$$

且 $\tilde{\mathcal{J}}_k^i(t,u_t) := \left\{j \in \text{自然数集} : \frac{q_{jk}}{\tilde{\pi}_t^k} \geq \frac{q_{ji}}{\tilde{\pi}_t^i}\left(\frac{1-u_t}{1+u_t}\right)^2\right\}$。则对于所有 $t < \infty$，

$$\tanh\left(\frac{\mathcal{H}(\pi_t,\tilde{\pi}_t)}{4}\right) \leq u_t.$$

~ ~特别地， $\lambda^\star(t,\ u_t) \geq \lambda_t^\star$，其中

$$\tilde{\lambda}_t^\star := 2\min_{i\neq k}\left\{\min_{S\subseteq\mathbf{N}}\sqrt{q_{ik}q_{ki}+\sum_{j\in S,\,j\neq i,k}q_{ik}q_{ji}\frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^k}+\sum_{l\notin S,\,l\neq i,k}q_{ki}q_{lk}\frac{\tilde{\pi}_t^l}{\tilde{\pi}_t^i}+\sum_{\substack{j\in S,\\ j\neq i,k}}\sum_{\substack{l\notin S,\\ l\neq i,k}}q_{ji}q_{lk}\frac{\tilde{\pi}_t^j\tilde{\pi}_t^l}{\tilde{\pi}_t^i\tilde{\pi}_t^k}}\right\},$$

$$\geq 2\min_{i\neq k}\left\{\sqrt{q_{ik}q_{ki}+\sum_{j\neq i,k}\min\left\{\frac{q_{ji}q_{ik}}{\tilde{\pi}_t^k},\frac{q_{jk}q_{ki}}{\tilde{\pi}_t^i}\right\}\tilde{\pi}_t^j}\right\}=:\tilde{\lambda}_t, \tag{9}$$

这意味着，对于所有 $t<\infty$，我们都有以下两个界

$$\tanh\left(\frac{\mathcal{H}(\pi_t,\tilde{\pi}_t)}{4}\right)\leq\tanh\left(\frac{\mathcal{H}(\mu,\nu)}{4}\right)e^{-\int_0^t\tilde{\lambda}_s^\star\,\mathrm{d}s}+\frac{1}{4}\int_0^t e^{-\int_s^t\tilde{\lambda}_r^\star\,\mathrm{d}r}\max_{i,k}\left\{\mathcal{E}_s^{1,i}-\mathcal{E}_s^{1,k}\right\}\mathrm{d}s, \tag{10}$$

and

$$\mathcal{H}(\pi_t,\tilde{\pi}_t)\leq\mathcal{H}(\mu,\nu)e^{-\int_0^t\tilde{\lambda}_s^\star\,\mathrm{d}s}+\int_0^t e^{-\int_s^t\tilde{\lambda}_r^\star\,\mathrm{d}r}\max_{i,k}\left\{\mathcal{E}_s^{1,i}-\mathcal{E}_s^{1,k}\right\}\mathrm{d}s. \tag{11}$$

注：定理 2.4 提出了一种构造误差相对较小的近似滤波器的方法。若已知 h，并且我们可以选择 $\tilde{g}_t = H\tilde{\pi}_t$，使得误差项 $\mathcal{E}^{2,\,i}$ 和 $\mathcal{E}^{3,\,i}$ 对所有 $i\in$**自然数集**均消失，则随机项导致的误差以及局部时间项也会消失。从数值角度来看，这等价于消除 $\sqrt{\mathrm{d}t}$ 阶的无穷小误差；在寻找 Wonham随机微分方程的近似解时，这是很自然的做法。

注：与定理 2.2 的证明类似，通过使用引理 2.1，只要在构造常微分方程（8）的解时足够谨慎，就可以解除定理 2.4 中的°假设，即 μ、$\nu\in\mathcal{S}^n$。

下面的推论说明，在计算衰减率时，π_t 与 $\tilde{\pi}_t$ 之间可以互换。

推论 2.4.1。假设 *Wonham*滤波器 π_t 是可观测的。定理 *2.4* 等价地成立~ ~，只需在 (8) 中用 $\lambda^\star(t,u_t)$ 替换 $\lambda^\star(t,u_t)$，并在 (10) 和(11)中用 $\lambda_t^\star$替换 $\lambda_t^\star$，其中

$$\lambda^\star(t,u_t)=\min_{i\neq k}\left\{\left(q_{ik}\frac{\pi_t^i}{\pi_t^k}+\sum_{\substack{j\neq i,k,\\ j\in\mathcal{J}_k^i(t,u_t)}}q_{jk}\frac{\pi_t^j}{\pi_t^k}\right)\frac{1-u_t}{1+u_t}+\left(q_{ki}\frac{\pi_t^k}{\pi_t^i}+\sum_{\substack{j\neq i,k,\\ j\in\mathcal{J}_k^i(t,u_t)}}q_{ji}\frac{\pi_t^j}{\pi_t^i}\right)\frac{1+u_t}{1-u_t}\right\},$$

以及 $\mathcal{J}_k^i(t,u_t)$：$=\left\{j\in$**自然数集**：$\frac{q_{jk}}{\pi_t^k}\leq\frac{q_{ji}}{\pi_t^i}\left(\frac{1+u_t}{1-u_t}\right)^2\right\}$，并且 $\lambda_t^\star$等价地定义为将 $\tilde{\pi}_t$替换为 π_t后的(9)。

3 希尔伯特射影度量中的收缩率

本节旨在证明定理 2.2，并给出一些与非线性滤波器相对于初始条件的稳定性相关的其他结果。

我们首先介绍一族坐标变换，将°概率单纯形 $\mathcal{S}^n$ 的内部映射到 $\mathbb{R}^n$，把离散概率分布映射为其自然参数。我们推导出Wonham滤波器在这些新参数化下的演化方程，

接着考察以 μ =law(X_0)为初始分布的Wonham滤波器的自然参数，与以 $\nu \neq \mu$为初始分布的「错误」Wonham滤波器的自然参数之间的差异。通过将各参数化下该差异的 ℓ_∞ 范数最大值与希尔伯特射影度量联系起来，我们得以明确计算出Wonham滤波器在希尔伯特度量下的指数收缩率。在证明定理 2.2之前，我们将经常额外假设 $\pi_0 = \mu$ 和 $\tilde{\pi}_0 = \nu$ 属于单纯形的内部。

备注。 在本节全文中，$(\pi_t)_{t\geq 0}$ 表示初始于$\pi_0 = \mu$的Wonham滤波器，而$(\tilde{\pi}_t)_{t\geq 0}$ 表示初始于 $\tilde{\pi}_0 = \nu$的Wonham滤波器。

3.1 坐标变换

我们将在此处考虑的坐标变换，将概率分布映射为统计学中所谓的自然（或规范）参数。自然参数通常是指数族分布所采用的参数化方式，其概率密度通常可写成如下的一般形式

$$p(x,\theta) = \exp\{\theta \cdot c(x) + k(x) - \psi(\theta)\}, \tag{12}$$

其中， $\theta \in \mathbb{R}^n$ 是一个 n维自然参数向量， $c(x)$是该分布的充分统计量 向量，且其 n 个分量线性无关； $k(x)$是 x 的函数，而 $\psi(\theta)$是对数配分函数。通过将测度 $\mathrm{d}x$ 变换为 $\mathrm{d}\upsilon(x) = \exp\{k(x)\}\mathrm{d}x$，我们可以忽略 $k(x)$，只要将 $p(x,\theta)$理解为相对于测度 $\mathrm{d}\upsilon(x)$的密度即可。为简单起见，我们假设 $k(x) = 0$ 。

我们选择在自然参数的坐标系 θ 中研究滤波方程，是受 Amari 的信息几何理论 [1,2]启发。如果 μ， $\nu \in \mathcal{S}^n$，那么滤波过程 π_t 位于 $\mathcal{S}^n$，即概率单纯形的内部。用信息几何的术语来说， $\mathcal{S}^n$ 是一个 n维统计流形，其整体坐标图为 θ （及其对偶仿射坐标——期望参数 η）。 $\mathcal{S}^n$ 上的黎曼度量是费希尔信息，它在无穷小意义下与 KL散度一致。本文中，我们不会使用 Amari 为 $\mathcal{S}^n$ 发展出的微分几何结构，因为我们选择用希尔伯特范数来衡量概率向量之间的距离，而该范数不允许定义光滑几何。不过，使用由 θ 参数化给出的整体坐标系仍然很方便。

考虑一个概率向量 $p \in \mathring{\mathcal{S}}^n \subset \mathbb{R}^{n+1}$。注意， $\mathring{\mathcal{S}}^n$ 是一个 n 维指数族，我们可以通过固定 $k \in$**自然数集**并选择 $c^i(x) = \delta_{a_i}(x)$（对于 $a_i \in \mathbb{S}$），将离散分布 $p \in \mathcal{S}^n$ 写成形式 (12)。选择 $k \in$ **自然数集**，我们定义微分同胚 $\theta_k \colon \mathcal{S}^n \to \mathbb{R}^n$ ，它按如下方式映射 $p \mapsto \theta_k$ ：

$$\theta_k^i = \log \frac{p^i}{p^k}, \quad \forall i \in \mathbf{N}. \tag{13}$$

我们指出， $\theta_k^k = 0$ 可以作为向量 θ_k 的一个分量忽略（也可以"跳过"），因此确实有 $\theta_k \in \mathbb{R}^{k-1} \times \{0\} \times \mathbb{R}^{n-k} \cong \mathbb{R}^n$。

逆映射 θ_k^{-1} 由以下公式给出。

$$p^i = \frac{\exp \theta_k^i}{1 + \sum_{j\neq k} \exp \theta_k^j}, \quad \forall i \in \mathbf{N}. \tag{14}$$

现在，我们希望将坐标变换（13）应用于 $(\pi_t)_{t\geq 0}$ 和 $(\tilde{\pi}_t)_{t\geq 0}$，并推导出参数 $\theta_k(\pi_t)$ 和 $\theta_k(\tilde{\pi}_t)$ 的演化方程。对于所有 $k \in$**自然数集**，为简化符号，定义

$$\theta_k(t) := \theta_k(\pi_t), \quad \tilde{\theta}_k(t) := \theta_k(\tilde{\pi}_t),$$

因此，逐分量地，我们有

$$\theta_k^i(t) := \log \frac{\pi_t^i}{\pi_t^k}, \quad \tilde{\theta}_k^i(t) := \log \frac{\tilde{\pi}_t^i}{\tilde{\pi}_t^k}, \quad \forall (i,k) \in \mathbf{N} \times \mathbf{N}.$$

下面的引理保证这些过程对所有$t < \infty$几乎必然都有定义。证明参见 [14]。

引理 3.1 ([14]中的引理 2.1**)**。记$\pi_{s,t}(\mu)$ 为在时间 $t \geq 0$ 时、以 $\pi_s = \mu$为初值从时间 $s \leq t$ 开始的方程(4) 的解。那么

$$\mathbb{P}\big(\pi_{s,t}(\mu) \in \mathring{\mathcal{S}}^n \text{ for all } \mu \in \mathring{\mathcal{S}}^n \text{ and all } 0 \leq s \leq t < \infty\big) = 1.$$

°**推论 3.1.1**。假设 $\mu, \nu \in \mathcal{S}^n$。我们有，几乎必然地，

$$\mathcal{H}(\pi_t, \tilde{\pi}_t),\ |\theta_k^i(t)|,\ |\tilde{\theta}_k^i(t)| < \infty, \quad \forall\, (i,k) \in \mathbf{N} \times \mathbf{N},$$

对所有时间 $0 \leq t < \infty$均成立。

引理 3.1 的证明还直接给出了以下替代结果。

引理 3.2。 记 $\pi_t(\mu)$ 为在时间 $0 \leq t$ 以$\pi_0 = \mu$为初值的方程 (4) 在时间 $t \geq 0$ 的解。假设对所有 $i \neq j$， $q_{ij} > 0$ 。那么

$$\mathbb{P}\big(\pi_t(\mu) \in \mathring{\mathcal{S}}^n \text{ for all } \mu \in \mathcal{S}^n \text{ and all } 0 < t < \infty\big) = 1.$$

现在我们来研究自然参数 $\theta_k^i(t)$ 和 $\tilde{\theta}_k^i(t)$ 的动力学。对于所有指标对 $(i,\ k) \in$ **自然数集** $\times$ **自然数集**，定义差分过程

$$\Delta_{ik}(t) := (\theta_k^i(t) - \tilde{\theta}_k^i(t))_{t\geq 0}, \tag{15}$$

其中，对所有 $i \in$**自然数集**， $\Delta_{ii} = 0$ 。

我们从以下命题开始。

°**命题 3.3.** 假设 $\mu, \nu \in \mathcal{S}^n$。对于所有 $0 \leq t < \infty$ 以及所有指标对 $(i,k) \in$**自然数集** $\times$**自然数集**，过程 $\Delta_{ik}(t)$ 关于时间是 C^1，其动力学方程为

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta_{ik}(t) = -\sum_{\substack{j=0\\ j\neq k}}^{n} q_{jk}\big(e^{\theta_k^j} - e^{\tilde{\theta}_k^j}\big) + \sum_{\substack{j=0\\ j\neq i}}^{n} q_{ji}\big(e^{\theta_i^j} - e^{\tilde{\theta}_i^j}\big),$$

$$\Delta_{ik}(0) = \log \frac{\mu^i}{\mu^k} - \log \frac{\nu^i}{\nu^k}. \tag{16}$$

证明。对于 $i = k$，过程 Δ_{kk} 恒等于 0，因此结论显然成立。假设$i \neq k$。考虑 $\theta_k^i(t) = \log(\pi_t^i/\pi_t^k)$，其中 $i \neq k$。应用伊藤公式可得，对于任意选取的 $k \in$ **自然数集**以及 $i \neq k$，有

$$\begin{aligned}
\mathrm{d}\log\frac{\pi^i}{\pi^k}(t) &= -\sum_{\substack{j=0\\ j\neq k}}^{n} q_{jk}\frac{\pi_t^j}{\pi_t^k}\,\mathrm{d}t + \sum_{\substack{j=0\\ j\neq i}}^{n} q_{ji}\frac{\pi_t^j}{\pi_t^i}\,\mathrm{d}t + (q_{ii}-q_{kk})\,\mathrm{d}t + (h^i-h^k)\,\mathrm{d}B_t \\
&\quad + \frac{1}{2}\Big((h^k)^2 - (h^i)^2 + 2(h^i-h^k)\pi_t^\top h\Big)\,\mathrm{d}t, \\
\log\frac{\pi^i}{\pi^k}(0) &= \log\frac{\mu^i}{\mu^k},
\end{aligned} \tag{17}$$

其中，为便于阅读，我们引入了创新过程 $B_t = Y_t - \int_0^t \pi_s^\top h\mathrm{d}s$，它是一个 $\{\mathcal{Y}_t\}$-适应的布朗运动（参见例如 [5,命题 2.30]）。类似地，

$$\begin{aligned}
\mathrm{d}\log\frac{\tilde\pi^i}{\tilde\pi^k}(t) &= -\sum_{\substack{j=0\\ j\neq k}}^{n} q_{jk}\frac{\tilde\pi_t^j}{\tilde\pi_t^k}\,\mathrm{d}t + \sum_{\substack{j=0\\ j\neq i}}^{n} q_{ji}\frac{\tilde\pi_t^j}{\tilde\pi_t^i}\,\mathrm{d}t + (q_{ii}-q_{kk})\,\mathrm{d}t + (h^i-h^k)\,\mathrm{d}B_t \\
&\quad + \frac{1}{2}\Big((h^k)^2 - (h^i)^2 + 2(h^i-h^k)\tilde\pi_t^\top h\Big)\,\mathrm{d}t + (h^i-h^k)\big(\pi_t^\top h - \tilde\pi_t^\top h\big)\,\mathrm{d}t, \\
\log\frac{\tilde\pi^i}{\tilde\pi^k}(0) &= \log\frac{\nu^i}{\nu^k}.
\end{aligned}$$

将两个方程相减，我们可以看到，其差值具有绝对连续的动力学。

$$\begin{aligned}
\mathrm{d}\left(\log\frac{\pi^i}{\pi^k}(t) - \log\frac{\tilde\pi^i}{\tilde\pi^k}(t)\right) &= -\sum_{\substack{j=0\\ j\neq k}}^{n} q_{jk}\left(\frac{\pi_t^j}{\pi_t^k} - \frac{\tilde\pi_t^j}{\tilde\pi_t^k}\right)\mathrm{d}t + \sum_{\substack{j=0\\ j\neq i}}^{n} q_{ji}\left(\frac{\pi_t^j}{\pi_t^i} - \frac{\tilde\pi_t^j}{\tilde\pi_t^i}\right)\mathrm{d}t, \\
\log\frac{\pi^i}{\pi^k}(0) - \log\frac{\tilde\pi^i}{\tilde\pi^k}(0) &= \log\frac{\mu^i}{\mu^k} - \log\frac{\nu^i}{\nu^k}.
\end{aligned} \tag{18}$$

注意到上述方程的右端关于时间连续（因为 π_t 和 $\tilde\pi_t$ 都连续），因此对于每个 $t \geq 0$，Δ_{ik} 的导数存在且连续，于是得到（16）。

□

3.2 希尔伯特误差

比较式 (15) 与式 (3)，我们现在观察到，希尔伯特范数可以通过最大过程来表示。

$$\Delta_\infty(t) := \max_{k\in\mathbf{N}} \left\|\theta_k(t) - \tilde\theta_k(t)\right\|_{\ell^\infty} = \max_{(i,k)\in\mathbf{N}\times\mathbf{N}} \Delta_{ik}(t).$$

通过观察可轻易看出

$$\begin{aligned}
\Delta_\infty(t) &= \max_{(i,k)\in\mathbf{N}\times\mathbf{N}} \left(\log\frac{\pi^i}{\pi^k}(t) - \log\frac{\tilde\pi^i}{\tilde\pi^k}(t)\right) = \max_{(i,k)\in\mathbf{N}\times\mathbf{N}} \left(\log\frac{\pi^i}{\tilde\pi^i}(t) - \log\frac{\pi^k}{\tilde\pi^k}(t)\right) \\
&= \max_{i\in\mathbf{N}} \log\frac{\pi^i}{\tilde\pi^i}(t) - \min_{k\in\mathbf{N}} \log\frac{\pi^k}{\tilde\pi^k}(t) = \mathcal{H}(\pi_t, \tilde\pi_t),
\end{aligned} \tag{19}$$

其中最后一个等式由对数的单调性得出。

我们希望研究随机过程 $\Delta_\infty(t)$ 随时间的演化。这里我们借鉴了 [7], 的一些论证，因为事实证明，命题 3.3 中的差分过程 Δ_{ik} 的动力学与 [7,第 5.2 节的式 5.6 和式 5.7]所考虑的平滑过程方程有些相似。

接下来我们需要以下引理。

引理 3.4 (Dupuis 和 Ellis 中的定理 A.6.3 [18])。 设 $g:\ [0,1]\to\mathbb{R}$ 是一个绝对连续函数。那么对每个实数 $r\in\mathbb{R}$，集合 $\{t:\ g(t)=r, \dot g(t)\neq 0\}$ 的勒贝格测度为0。

在下文中，当我们说适应随机过程 $Z(t,\omega)$ 是绝对连续的或具有绝对连续路径（几乎处处）时，我们不仅指它可以表示为 $\mathrm{d}Z(t,\omega) = g(t,\omega)\,\mathrm{d}t$，其中 $g\in L^1([0,t])$ 对所有 $t>0$（几乎处处）成立，还指弱导数 $g(t,\omega)$ 联合可测且适应于底层滤过。下一个引理证实，过程 $\Delta_\infty(t,\omega)$ 确实满足这一条件。

引理 3.5。 假设 $\mu,\nu\in\mathcal{S}^n$。随机过程 $(t,\omega)\mapsto\Delta_\infty(t,\omega)$ 具有绝对连续路径（特别是，它是可预测的）。

证明。 固定任意一个 $k\in$ **自然数集**。首先考虑过程 $\Delta^\star_{i,k}(t)=\Delta_{0k}\vee\Delta_{1k}\vee\cdots\vee\Delta_{ik}$，其中 $i\in$ **自然数集**。我们通过归纳法证明 $\Delta^\star_{n,k}(t)=\max_{i\in\mathbf{N}}\Delta_{ik}$ 的绝对连续性。显然，$\Delta^\star_{0,k}(t)=\Delta_{0k}(t)$ 是绝对连续的，因为根据定义，它要么恒为 0（若 $k=0$），要么根据命题 3.3 是绝对连续的（若 $k\neq 0$）。考虑情形 $i=1$，其中 $\Delta^\star_{1,k}(t)=\Delta_{0k}(t)\vee\Delta_{1k}(t)$。回忆 $a\vee b=\frac{1}{2}(a+b+|a-b|)$。于是

$$\Delta^\star_{1,k}(t)=\frac{1}{2}\big(\Delta_{0k}(t)+\Delta_{1k}(t)+|\Delta_{0k}(t)-\Delta_{1k}(t)|\big).$$

根据命题 3.3，我们有 $\Delta_{0k}(t)$ 和 $\Delta_{1k}(t)$ 关于时间是 C^1 绝对连续的，并且关于 ω 是 $\mathcal{F}_t$-可测的。根据弱可微函数的链式法则，如果 $F(t)$ 是绝对连续的，其弱导数为 $f(t)$，那么

$$\mathrm{d}|F(t)|=\operatorname{sign}(F(t))f(t)\,\mathrm{d}t. \tag{20}$$

因此，我们有 $|\Delta_{0k}(t)-\Delta_{1k}(t)|$ 关于时间是绝对连续的（对每个 ω），并且从（20）的形式显然可知，其弱导数关于 (t,ω) 是联合可测的，并且是 $\mathcal{F}_t$-适应的。因此，$\Delta^\star_{1,k}(t)$ 也具有相同性质。

现在 n 注意到 $\Delta^\star_{i,k}(t)=\Delta^\star_{i-1,k}(t)\vee\Delta_{ik}(t)$ 对所有 $2\le i\le n$ 都成立，如前所述，我们可以写成

$$\Delta^\star_{i,k}(t)=\frac{1}{2}\big(\Delta^\star_{i-1,k}(t)+\Delta_{ik}(t)+|\Delta^\star_{i-1,k}(t)-\Delta_{ik}(t)|\big),$$

根据归纳法可知，$\Delta^\star_{n,k}(t)=\max_{i\in\mathbf{N}}\Delta_{ik}$ 具有绝对连续路径。

由于上述论证与我们对 k 的选择无关，因此对于所有 $k\in$N，$\Delta^\star_{n,k}(t)$ 都是绝对连续的。现在，我们只需对所有 $k\in$ N 取 $\Delta^\star_{n,k}(t)$ 的最大值，并证明它同样是绝对连续的。考虑当 $k\in$N 时的过程 $\Delta^\star_k(t)=\Delta^\star_{n,0}\vee\Delta^\star_{n,1}\vee\cdots\vee\Delta^\star_{n,k}$。完全按照上述方法，通过利用过程 $\Delta^\star_{n,k}$ 的绝对连续性进行归纳推理，我们最终得到过程 $\Delta^\star_n(t)=\max^{k\in\mathbf{N}}\Delta^\star_{n,k}$ 在 (t,ω) 上是可测的，并且关于时间是绝对连续的。注意到 $\Delta^\star_n(t)=\Delta_\infty(t)$，证毕。

□

°**引理3.6**。 假设 $\mu, \nu \in \mathcal{S}^n$。存在一个 $\{\mathcal{Y}_t\}$-可预测的指标选择 $(t,\omega) \mapsto (i^\star(t,\omega)\ k^\star(t,\omega))$ 使得

$$\Delta_\infty(t,\omega) = \Delta_{i^\star(t,\omega)k^\star(t,\omega)}(t,\omega) \quad \textit{for all } t,\omega.$$

此外， $\Delta_\infty(t,\omega)$ 的动力学由下式给出

$$\begin{aligned} \mathrm{d}\Delta_\infty(t) &= \sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{N}} \mathbf{1}_{\{(i^\star,k^\star)(t)=(i,k)\}} \frac{\mathrm{d}}{\mathrm{d}t}\Delta_{ik}(t)\,\mathrm{d}t, \\ \Delta_\infty(0) &= \log\frac{\mu^{i^\star(0)}}{\mu^{k^\star(0)}} - \log\frac{\nu^{i^\star(0)}}{\nu^{k^\star(0)}}. \end{aligned} \tag{21}$$

证明。考虑可测空间（$M, \mathcal{M}$），其中 $M = ([0,\infty) \times \Omega)$，且 $\mathcal{M}$ 是 $\{\mathcal{Y}_t\}$-可预测 σ-代数。令 $U = \mathbf{N} \times \mathbf{N}$ 赋予离散拓扑。考虑函数 $f\colon M \times U \to \mathbb{R}$，使得 $f((t,\omega),(i,k)) = \Delta_{ik}(t,\omega)$。注意，对所有 $(i,k) \in U$， $z(\cdot,(i,k)) = \Delta_{ik}(\cdot)$ 都是 $\mathcal{M}$-可测的，这由命题 3.3 得出。此外， $z((t,\omega)\ \cdot) = \Delta_\cdot(t,\omega)$ 作为 $U \to \mathbb{R}$ 的函数是连续的（因为它定义在离散空间 $U = \mathbf{N} \times \mathbf{N}$ 上）。根据引理 3.5，函数 $\Delta_\infty\colon M \to \mathbb{R}$ 是 $\mathcal{M}$-可测的。由于 $\Delta_\infty = \max_{(i,k)\in\mathbf{N}\times\mathbf{N}} \Delta_{ik}$，我们必然有 Δ_∞ 的像包含在 f 的像中。换言之，我们有

$$\Delta_\infty(t,\omega) \in f((t,\omega),U) \quad \forall (t,\omega) \in M.$$

于是，根据 Filippov 隐函数引理（参见例如 [15, 定理 A.10.2]），存在一个$\mathcal{M}$-可测（即 $\{\mathcal{Y}_t\}$-可预测）映射 $u\colon M \to U$，它将 $(t,\ \omega) \mapsto (i^\star(t,\ \omega)\ k^\star(t,\ \omega))$ 映射为，使得

$$\Delta_\infty(t,\omega) = f\big((t,\omega), u(t,\omega)\big) = f\big((t,\omega),(i^\star(t,\omega),k^\star(t,\omega))\big) = \Delta_{i^\star(t,\omega)k^\star(t,\omega)}(t,\omega).$$

为了证明引理的第二部分，回想由引理 3.5 可知， $\Delta_\infty(t,\ \omega)$ 是绝对连续的。于是，对某个密度 $g(t)$，有 $\mathrm{d}\Delta_\infty(t) = g(t)\,\mathrm{d}t$，且对每个 $t \geq 0$， $\int_0^t |g(s)|\mathrm{d}s < \infty$ 几乎处处成立。由于 $\sum_{i\in\text{自然数集}} \sum_{k\in\text{自然数集}} \mathbf{1}_{\{(i^\star,k^\star)(t)=(i,\ k)\}} = 1$，我们可以写成

$$\Delta_\infty(t) = \Delta_\infty(0) + \int_0^t \sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{N}} \mathbf{1}_{\{(i^\star,k^\star)(t)=(i,k)\}} g(s)\,\mathrm{d}s.$$

因此，如果我们能够证明，对于任意 $(i,k) \in \mathbf{N} \times \mathbf{N}$ 以及任意 $t > 0$，我们都有

$$\int_0^t \mathbf{1}_{\{(i^\star,k^\star)(t)=(i,k)\}} \Big| g(s) - \frac{\mathrm{d}}{\mathrm{d}t}\Delta_{ik}(s)\Big|\,\mathrm{d}s = 0 \quad \text{a.s.}, \tag{22}$$

我们完成了。将上述等式左侧改写后，得到

$$\begin{aligned} 0 &\leq \int_0^t \mathbf{1}_{\{(i^\star,k^\star)(t)=(i,k)\}} \Big| g(s) - \frac{\mathrm{d}}{\mathrm{d}t}\Delta_{ik}(s)\Big|\,\mathrm{d}s \\ &\leq \int_0^t \mathbf{1}_{\{\Delta_\infty(s)-\Delta_{ik}(s)=0\}} \Big| g(s) - \frac{\mathrm{d}}{\mathrm{d}t}\Delta_{ik}(s)\Big|\,\mathrm{d}s \\ &= \int_0^t \mathbf{1}_{\left\{\Delta_\infty(s)-\Delta_{ik}(s)=0,\, g(s)-\frac{\mathrm{d}}{\mathrm{d}t}\Delta_{ik}(s)\neq 0\right\}} \Big| g(s) - \frac{\mathrm{d}}{\mathrm{d}t}\Delta_{ik}(s)\Big|\,\mathrm{d}s, \end{aligned}$$

并且由于集合 $\big\{s\colon\ \Delta_\infty(s) - \Delta_{ik}(s) = 0,\quad g(s) - \overline{\mathrm{d}\ \mathrm{d}t}\Delta_{ik}(s) \neq 0\big\}$ 由引理 3.4 可知测度为 0，因此我们得出 (22) 成立，证明完成。 □

最后，(19) 使我们得以陈述以下引理，它们将在后文中派上用场。

引理3.7。 假设 $\mu, \nu \in \mathcal{S}^n$。对于所有 $t < \infty$，指标 $i^\star(t,\omega)$ 和 $k^\star(t,\omega)$ 分别使 $\frac{\pi_t^j}{\tilde{\pi}_t^j}$ 在 $j \in$ **自然数集**上取最大值和最小值。此外，我们有 $\frac{\pi_t^{i^\star}}{\tilde{\pi}_t^{i^\star}} =: M_t \geq 1$以及 $\frac{\pi_t^{k^\star}}{\tilde{\pi}_t^{k^\star}} =: \frac{1}{m_t} \leq 1$ 对所有 $t < \infty$均成立。

证明。 固定 $(t, \omega) \in [0,\infty) \times \Omega$。回顾引理3.6，根据 Δ_{ik}的定义和（19）可知

$$\log \frac{\pi_t^{i^\star(t,\omega)}}{\tilde{\pi}_t^{i^\star(t,\omega)}}(\omega) - \log \frac{\pi_t^{k^\star(t,\omega)}}{\tilde{\pi}_t^{k^\star(t,\omega)}}(\omega) = \Delta_{i^\star(t,\omega)k^\star(t,\omega)}(t,\omega)$$
$$= \Delta_\infty(t,\omega) = \log \max_{i\in\mathbf{N}} \frac{\pi_t^i}{\tilde{\pi}_t^i}(\omega) - \log \min_{k\in\mathbf{N}} \frac{\pi_t^k}{\tilde{\pi}_t^k}(\omega),$$

因此，引理的第一部分得证。对于第二部分，反设存在 $\omega \in \Omega$，使得 $M_t(\omega) < 1$。那么，对于所有 $j \in$ **自然数集**

$$\frac{\pi_t^j}{\tilde{\pi}_t^j}(\omega) \leq M_t(\omega) < 1 \Longrightarrow \pi_t^j < \tilde{\pi}_t^j,$$

这意味着 $\sum_j \pi_t^j < 1$，这与 π_t 是概率分布这一事实相矛盾。 $1/m_t(\omega)$ 的论证类似。

□

引理 3.8。 假设 $\mu, \nu \in \mathring{\mathcal{S}}^n$。对于所有 $i, k \in$ **自然数集** $\times$ **自然数集**，定义 $T_{ik}(t) := \frac{\pi_t^i}{\pi_t^k} - \frac{\tilde{\pi}_t^i}{\tilde{\pi}_t^k}$。对于所有$t < \infty$，我们有 $T_{ji^\star}(t) \leq 0$ 以及 $T_{jk^\star}(t) \geq 0$，其中 $i^\star = i^\star(t,\omega)$ 和 $k^\star = k^\star(t,\omega)$ 分别是引理 *3.7* 中的最大化/最小化指标。

证明。</style}显然， $T_{i^\star i^\star}(t) = T_{k^\star k^\star}(t) = 0$。现在考虑 $j \neq k^\star$ 的 $T_{jk^\star}(t)$。注意到

$$T_{jk^\star}(t) = \frac{\pi_t^j}{\pi_t^{k^\star}} - \frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^{k^\star}} = \left(\frac{\pi_t^j}{\tilde{\pi}_t^j} - \frac{\pi_t^{k^\star}}{\tilde{\pi}_t^{k^\star}}\right)\frac{\tilde{\pi}_t^j}{\pi_t^{k^\star}}, \quad \forall j \in \mathbf{N}, j \neq k^\star.$$

根据引理 3.7， $k^\star$ 在**自然数集**上使 $\frac{\pi_t^j}{\tilde{\pi}_t^j}$ 取得最小值，因此对所有 $j \neq k^\star$，都有 $\frac{\pi_t^j}{\tilde{\pi}_t^j} - \frac{\pi_t^{k^\star}}{\tilde{\pi}_t^{k^\star}} \geq 0$ 。此外， $\tilde{\pi}_t^j/\pi_t^{k^\star} > 0$ 也成立，因为 π_t 和 $\tilde{\pi}_t$ 在 $t < \infty$处的分量均为正。由此可得，对所有 $t < \infty$, $T_{jk^\star}(t) \geq 0$ 。

对于 $T_{ji^\star}(t)$ 的情形，同样可以通过注意到以下事实来论证：

$$T_{ji^\star}(t) = \frac{\pi_t^j}{\pi_t^{i^\star}} - \frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^{i^\star}} = -\left(\frac{\pi_t^{i^\star}}{\tilde{\pi}_t^{i^\star}} - \frac{\pi_t^j}{\tilde{\pi}_t^j}\right)\frac{\tilde{\pi}_t^j}{\pi_t^{i^\star}}, \quad \forall j \in \mathbf{N}, j \neq i^\star,$$

并利用 $i^\star$ 使 $\frac{\pi_t^j}{\tilde{\pi}_t^j}$ 达到最大值这一点。

□

3.3 定理 2.2 的证明

现在我们可以证明定理 2.2 了。

°定理 *2.2* 的证明。我们先考虑 (21)，并假设 $\mu, \nu \in \mathcal{S}^n$。将其完整展开可得

$$\begin{aligned}\mathrm{d}\Delta_\infty(t) &= \sum_{i\in\mathbf{N}}\sum_{k\in\mathbf{N}} \mathbf{1}_{\{(i^\star,k^\star)(t)=(i,k)\}} \frac{\mathrm{d}}{\mathrm{d}t}\left(\log\frac{\pi_t^i}{\pi_t^k} - \log\frac{\tilde\pi_t^i}{\tilde\pi_t^k}\right)\mathrm{d}t \\ &= \left[-\sum_{\substack{j=0\\ j\neq k^\star}}^{n} q_{jk^\star}\left(\frac{\pi_t^j}{\pi_t^{k^\star}} - \frac{\tilde\pi_t^j}{\tilde\pi_t^{k^\star}}\right) + \sum_{\substack{j=0\\ j\neq i^\star}}^{n} q_{ji^\star}\left(\frac{\pi_t^j}{\pi_t^{i^\star}} - \frac{\tilde\pi_t^j}{\tilde\pi_t^{i^\star}}\right)\right]\mathrm{d}t, \end{aligned} \tag{23}$$

为便于阅读，我们省略了 (t,ω) 对 $(i^\star,k^\star)$ 的依赖关系。

用引理 3.8 的记号重写 (23)，可得

$$\begin{aligned}\mathrm{d}\Delta_\infty(t) &= -\left[\sum_{\substack{j=0\\ j\neq k^\star}}^{n} q_{jk^\star}T_{jk^\star}(t) - \sum_{\substack{j=0\\ j\neq i^\star}}^{n} q_{ji^\star}T_{ji^\star}(t)\right]\mathrm{d}t \\ &= -\big(q_{i^\star k^\star}T_{i^\star k^\star}(t) - q_{k^\star i^\star}T_{k^\star i^\star}(t)\big)\,\mathrm{d}t - \sum_{\substack{j=0\\ j\neq i^\star,k^\star}}^{n}\left(q_{jk^\star}T_{jk^\star}(t) - q_{ji^\star}T_{ji^\star}(t)\right)\mathrm{d}t. \end{aligned} \tag{24}$$

由于按定义 Q-矩阵的非对角元素非负，且根据引理 3.8，对于 $j \neq i^\star, k^\star$，差值 $q_{i^\star k^\star}T_{i^\star k^\star}(t) - q_{k^\star i^\star}T_{k^\star i^\star}(t)$ 和 $q_{jk^\star}T_{jk^\star}(t) - q_{ji^\star}T_{ji^\star}(t)$ 也都非负，因此右端非正。我们希望求出 $[0,\infty)$ 上斯蒂尔切斯测度 $\mathrm{d}\Delta_\infty(t)$ 的上界。去掉求和中的非正项后，可简化为

$$\begin{aligned}\mathrm{d}\Delta_\infty(t) &\le -\big(q_{i^\star k^\star}T_{i^\star k^\star}(t) - q_{k^\star i^\star}T_{k^\star i^\star}(t)\big)\,\mathrm{d}t \\ &= -\left[q_{i^\star k^\star}\left(\frac{\pi_t^{i^\star}}{\pi_t^{k^\star}} - \frac{\tilde\pi_t^{i^\star}}{\tilde\pi_t^{k^\star}}\right) + q_{k^\star i^\star}\left(\frac{\tilde\pi_t^{k^\star}}{\tilde\pi_t^{i^\star}} - \frac{\pi_t^{k^\star}}{\pi_t^{i^\star}}\right)\right]\mathrm{d}t \\ &= -\left[q_{i^\star k^\star}\left(M_t - \frac{1}{m_t}\right)\frac{\tilde\pi_t^{i^\star}}{\pi_t^{k^\star}} + q_{k^\star i^\star}\left(m_t - \frac{1}{M_t}\right)\frac{\pi_t^{k^\star}}{\tilde\pi_t^{i^\star}}\right]\mathrm{d}t \\ &\le -2\sqrt{q_{i^\star k^\star}q_{k^\star i^\star}}\left(\frac{M_t m_t - 1}{\sqrt{M_t m_t}}\right)\mathrm{d}t, \end{aligned}$$

其中，$M_t := \frac{\pi_t^{i^\star}}{\tilde\pi_t^{i^\star}}$ 和 $m_t := \frac{\tilde\pi_t^{k^\star}}{\pi_t^{k^\star}}$ 如引理 3.7 所示，并且我们使用了不等式 $a + b \ge 2\sqrt{ab}$，其中 $a,\ b \ge 0$。回顾一下，由于根据引理 3.7 有 $M_t m_t \ge 1$，因此 $\Delta_\infty(t) = \Delta_{i^\star k^\star}(t) = \log(M_t m_t) \ge 0$。于是，我们可以将上述不等式改写为

$$\mathrm{d}\Delta_\infty(t) \le -4\sqrt{q_{i^\star k^\star}q_{k^\star i^\star}}\sinh\left(\frac{\Delta_\infty(t)}{2}\right)\mathrm{d}t$$

现在，如果 $\Delta_\infty(t) = 0$，则该定理显然成立，因此我们可以假设 $\Delta_\infty(t) > 0$。由于对于 $x > 0$，$\sinh(x) > 0$，我们可以将等式两边除以 $\sinh\left(\Delta_\infty(t)/2\right)$。对 $[s,t]$ 积分可得

$$\log\tanh\left(\frac{\Delta_\infty(t)}{4}\right) - \log\tanh\left(\frac{\Delta_\infty(s)}{4}\right) \le -\lambda(t-s),$$

其中我们定义了 $\lambda := 2\min_{i\neq j}\sqrt{q_{ij}q_{ji}}$，于是可得

$$\tanh\left(\frac{\Delta_\infty(t)}{4}\right) \le \tanh\left(\frac{\Delta_\infty(0)}{4}\right)e^{-\lambda t}.$$

对于 $x \ge 0$，双曲正切函数 $\tanh(x)$ 的凹性和单调性意味着 $\tanh(x)e^{-\lambda t} \le \tanh(xe^{-\lambda t})$，因此

$$\Delta_\infty(t) \le \Delta_\infty(0)e^{-\lambda t},$$

并且，回想式（19），证毕。

最后，我们去掉假设 $\mu,\nu \in \mathcal{S}^n$。根据引理 2.1，我们可以选取序列 $\mu_m,\nu_m \in \mathcal{S}^n$，使得 $\mathcal{H}(\mu_m,\nu_m) \to \mathcal{H}(\mu,\nu)$，且 $(\mu_m,\nu_m) \to (\mu,\nu)$ 在欧几里得范数下成立。考虑相应的滤波器 $\pi_t(\mu_m)$ $\tilde{\pi}_t(\nu_m)$（分别为以 μ_m 和 ν_m 为初始条件的方程 (4) 和 (5) 的解）。由于 Kushner–Stratonovich 方程在 $\mathcal{S}^n$ 上满足 Lipschitz 条件，根据随机微分方程的标准稳定性结果（例如 [15, 定理 16.4.3]），我们知道，当 $m \to \infty$ 时，$\pi_t(\mu_m) \to \pi_t$ 和 $\tilde{\pi}_t(\nu_m) \to \tilde{\pi}_t$ 依概率收敛。

由引理 3.2 可知 $\pi_t,\ \tilde{\pi}_t \in \mathcal{S}^n$。利用引理 2.1 给出的连续性，并应用上述结果，我们得到

$$\begin{aligned}\tanh\Big(\frac{\mathcal{H}(\pi_t,\tilde{\pi}_t)}{4}\Big) &= \lim_{m\to\infty}\Big[\tanh\Big(\frac{\mathcal{H}(\pi_t(\mu_m),\tilde{\pi}_t(\nu_m))}{4}\Big)\Big] \\ &\le \lim_{m\to\infty}\Big[\tanh\Big(\frac{\mathcal{H}(\mu_m,\nu_m)}{4}\Big)\Big]e^{-\lambda t} = \tanh\Big(\frac{\mathcal{H}(\mu,\nu)}{4}\Big)e^{-\lambda t}\end{aligned}$$

如所需。双曲正切函数的单调性和凹性再次完成了证明。 □

3.4 收缩率的最优性

我们刚刚证明的确定性收缩率 $\lambda = 2\min_{i\neq j}\sqrt{q_{ij}q_{ji}}$ 对于情形 $\pi_t,\ \tilde{\pi}_t \in \mathcal{S}^1$ 在 $\mu,\ \nu \in \mathring{\mathcal{S}}^1$ 上一致成立，且是尖锐的；也就是说，若存在 $\rho \in \mathbb{R}$ 使得 $\Delta_\infty(t) \le \Delta_\infty(s)e^{-\rho(t-s)}\ a.s.$ 对所有 $s < t$ 都成立，则可知 $\rho \le \lambda$。在这个基本情形下，最大过程就是 $\Delta_\infty = |\theta_0^1 - \tilde{\theta}_0^1|$。考虑 Q 为对称矩阵的具体情形，此时对角元素为 $q_{00} = q_{11} = -q$，非对角元素为 $q_{01} = q_{10} = q$。于是 $\lambda = 2q$。我们计算

$$\begin{aligned}\mathrm{d}|\theta_0^1(t) - \tilde{\theta}_0^1(t)| &= \operatorname{sign}\big(\theta_0^1(t) - \tilde{\theta}_0^1(t)\big)\Big[-q\big(e^{\theta_0^1(t)} - e^{\tilde{\theta}_0^1(t)} + e^{-\tilde{\theta}_0^1(t)} - e^{-\theta_0^1(t)}\big)\Big]\,\mathrm{d}t \\ &= \operatorname{sign}\big(\theta_0^1(t) - \tilde{\theta}_0^1(t)\big)\Big[-2q\big[(\theta_0^1(t) - \tilde{\theta}_0^1(t)) + \frac{1}{3!}((\theta_0^1(t))^3 - (\tilde{\theta}_0^1(t))^3) + \dots\big]\Big]\,\mathrm{d}t \\ &= -2q|\theta_0^1(t) - \tilde{\theta}_0^1(t)|\Big(1 + \frac{1}{3!}\big((\theta_0^1(t))^2 + (\tilde{\theta}_0^1(t))^2 + \theta_0^1(t)\tilde{\theta}_0^1(t)\big) + \dots\Big)\,\mathrm{d}t \\ &=: -2q|\theta_0^1(t) - \tilde{\theta}_0^1(t)|R_t\,\mathrm{d}t,\end{aligned}$$

由此我们推得

$$|\theta_0^1(t) - \tilde{\theta}_0^1(t)| = e^{-2q\int_s^t R_u\,\mathrm{d}u}|\theta_0^1(s) - \tilde{\theta}_0^1(s)|, \quad \text{for all } 0 \le s \le t. \tag{25}$$

注意，R_t 接近 1 当且仅当 $\theta_0^1(t) \approx \tilde{\theta}_0^1(t) \approx 0$，这在 π_t 和 $\tilde{\pi}_t$ 接近单纯形中心时发生，即 $\pi_t \approx \tilde{\pi}_t \approx \big(\frac{1}{2}, \frac{1}{2}\big)$。令 $\tau < \infty$ 为满足 $R_\tau \approx 1$ 的时间。注意，这样的 τ 以正概率存在，因为布朗运动动力学（在测度变换下）

对于 $\theta_0^1(t)$，确保 $\theta_0^1(t)$ 必须无限多次访问 0；那么根据定理 2.2，对于所有足够大的 τ，存在 $\varepsilon \ll 1$ 使得 $|\tilde{\theta}_0^1(\tau) - \theta_0^1(\tau)| \leq \varepsilon$。

由于 R_t 随时间连续，对于所有 $\delta > 0$，存在 t_δ 使得 $R_s \in (1,\ 1+\delta]$ 对所有$s \in [\tau, \tau + t_\delta]$。然后由 (25) 可知，对于所有 $s \in [\tau, \tau + t_\delta]$,

$$|\theta_0^1(\tau + \tau_\delta) - \tilde{\theta}_0^1(\tau + \tau_\delta)| \geq e^{-2q(1+\delta)t_\delta} \left|\theta_0^1(\tau) - \tilde{\theta}_0^1(\tau)\right|.$$

由于 δ 是任意选取的，我们可以看到界 $\lambda = 2q$ 得以达到。我们通过一个模拟示例对此进行说明，见图1（左）。

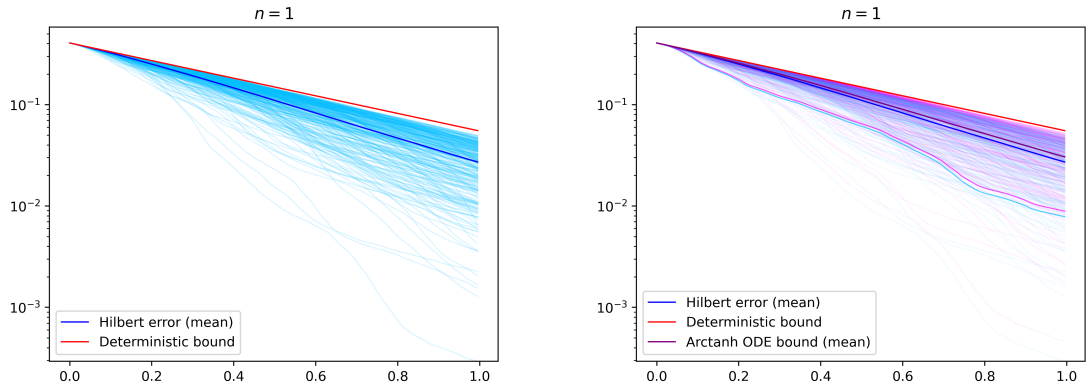

图 1：左侧以对数刻度绘制了希尔伯特射影误差$\mathcal{H}(\pi_t, \tilde{\pi}_t)$ 的 300 次实现（浅蓝色），其中 π_t, $\tilde{\pi}_t \in \mathring{\mathcal{S}}^1 \subset \mathbb{R}^2$ 分别是以 $\mu, \nu \in \mathring{\mathcal{S}}^1$为初始条件的方程 (4) 的解。蓝色表示样本均值，红色表示来自定理 2.2 的确定性逐路径界 $\mathcal{H}(\mu,\nu)e^{-\lambda t}$。可以看到，该界确实能够达到。在此模拟中，信号 X 是一个具有对称速率矩阵 Q 和跳跃率 $q = 1$的二状态马尔可夫链，其初始分布为其不变分布 $\mu = \left(\frac{1}{2}, \frac{1}{2}\right)$。这给出了 Wonham滤波器的初始条件 $\pi_0 = \mu$。‘错误’滤波器 $\tilde{\pi}_t$以 $\nu = \left(\frac{2}{5}, \frac{3}{5}\right)$为初始条件。固定这些参数，并将传感器函数设为 $h = (-1,1)$，我们为信号和观测过程模拟 300 条路径，计算这些路径对应的两个滤波器，并绘制它们的希尔伯特误差。右侧再次绘制希尔伯特误差的实现（极浅蓝色），并在同一图中加入通过对 $\tilde{\pi}_t$的每次实现数值求解命题 3.10 中的常微分方程 (27) 所得到的逐路径常微分方程界（品红色）。紫色表示常微分方程界的均值；在 300 对希尔伯特误差和常微分方程界中，我们随机突出显示其中一对。

另一方面，随着状态空间维数的增加，数值实验表明，λ 的最优性会降低。回顾上一小节中的记号，并重新审视我们的证明，很容易看出，这种次优性的原因在于我们舍弃了方程 (24) 右端形如 $q_{jk^\star}T_{jk^\star} - q_{ji^\star}T_{ji^\star}$ 的负项之和。具体而言，这类负项共有 $n-2$ 个，其取值可以落在 $(0,\ \infty)$ 内；这表明，随着 n 增大，Δ_∞ 的导数会变得更加负（而且可能大幅变小），因此 Δ_∞ 实际上应当比我们的界所表明的更快趋于 0。

遗憾的是，我们未能找到如下形式的统一下界$K_q\Delta_\infty$，适用于 $\sum_j(q_{jk^\star}T_{jk^\star} - q_{ji^\star}T_{ji^\star})$，其中 K_q是一个仅依赖于 Q的常数。不过，

假设可以观测到初始条件错误的滤波器 $\tilde{\pi}_t$的路径，我们便能给出一个更精确、依赖于 $\tilde{\pi}_t$的衰减率界。

我们将需要以下经典结果，现将其列出以求完整。

引理 3.9（比较原理）。 设 $X_t \in \mathbb{R}$ 是一个绝对连续过程，其几乎处处导数满足

$$\mathrm{d}X_t \le \alpha(t, X_t)\,\mathrm{d}t, \qquad X_0 = x_0,$$

在 $[0,\infty)$ 上，其中 $x \mapsto \alpha$（t、 x） 是局部利普希茨连续的。设 u_t 是常微分方程的唯一解（直到其首次爆炸时刻 $T > 0$）

$$\frac{\mathrm{d}u_t}{\mathrm{d}t} = \alpha(t, u_t), \qquad u_0 = x_0. \tag{26}$$

则对所有 $t < T$，都有 $X_t \le u_t$ 。

证明。 首先，回顾常微分方程理论中的标准结果（例如参见 [33, 定理 2.5]），可知 (26) 存在唯一解 u_t ，直到其首次爆炸时刻 $T > 0$。对于 $t < T$，考虑$H_t = X_t - u_t$。注意到 $H_0 = 0$，且 H_t 是绝对连续的，其几乎处处导数满足

$$\frac{\mathrm{d}H_t}{\mathrm{d}t} \le \alpha(t, X_t) - \alpha(t, u_t).$$

反设存在 $\tau < T$ 使得 $H_\tau > 0$。由 H_t 的连续性，存在 $t_0 \in [0,\tau)$ ，使得对所有 $s \in [t_0,\tau]$ 都有 $H_{t_0} = 0$ 和 $H_s \ge 0$。此外，由于 X_t 和 u_t 的连续性，存在 $R \in \mathbb{R}$ 使得对所有 $s \in [t_0,\tau]$ 都有 X_s, $u_s \in$ （$-R$， R）。于是

$$H_\tau = \int_{t_0}^{\tau} \frac{\mathrm{d}H_s}{\mathrm{d}s}\,\mathrm{d}s \le \int_{t_0}^{\tau} \big(\alpha(s, X_s) - \alpha(s, u_s)\big)\,\mathrm{d}s \le \int_{t_0}^{\tau} C_R(s)|X_s - u_s|\,\mathrm{d}s = \int_{t_0}^{\tau} C_R(s) H_s\,\mathrm{d}s,$$

其中 $C_R(t) \ge 0$ 是 $x_t \in (-R, R)$ 关于 $\alpha(t, x_t)$ 的利普希茨常数，并且根据假设，我们已用到 $H_s = X_s - u_s > 0$ 在 $[t_0,\tau]$ 上成立。于是，Gr¨onwall 不等式推出 $H_\tau \le 0$，这与假设矛盾。因此，对所有 $t < T$， $X_t \le u_t$ 。 □

命题 3.10。 假设对所有 $i \in$ **自然数集**， $\mu^i, \nu^i > 0$ 。对于所有 $t \ge 0$，令 $u_t \in (0,1)$ 为具有随机系数的常微分方程的唯一解，其定义为

$$\frac{\mathrm{d}u_t}{\mathrm{d}t} = -\tilde{\lambda}^\star(t, u_t) u_t, \qquad u_0 = \tanh\left(\frac{\mathcal{H}(\mu,\nu)}{4}\right), \tag{27}$$

其中

$$\tilde{\lambda}^\star(t, u_t) = \min_{i \ne k}\left\{\left(q_{ik}\frac{\tilde{\pi}_t^i}{\tilde{\pi}_t^k} + \sum_{\substack{j\ne i,k,\\ j\notin \tilde{\mathcal{J}}_k^i(t,u_t)}} q_{jk}\frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^k}\right)\frac{1+u_t}{1-u_t} + \left(q_{ki}\frac{\tilde{\pi}_t^k}{\tilde{\pi}_t^i} + \sum_{\substack{j\ne i,k,\\ j\in \tilde{\mathcal{J}}_k^i(t,u_t)}} q_{ji}\frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^i}\right)\frac{1-u_t}{1+u_t}\right\}, \tag{28}$$

and

$$\tilde{\mathcal{J}}_k^i(t, u_t) := \left\{ j \in \mathbf{N} \,:\, \frac{q_{jk}}{\tilde{\pi}_t^k} \ge \frac{q_{ji}}{\tilde{\pi}_t^i}\Big(\frac{1-u_t}{1+u_t}\Big)^2 \right\}.$$

那么，对于所有 $t < \infty$,

$$\tanh\left(\frac{\mathcal{H}(\pi_t, \tilde{\pi}_t)}{4}\right) \le u_t. \tag{29}$$

~ ~特别地，$\lambda^\star(t,u_t)\geq\lambda_t^\star$，其中

$$\tilde{\lambda}_t^\star := 2\min_{i\neq k}\left\{\min_{S\subseteq\mathbf{N}}\sqrt{q_{ik}q_{ki}+\sum_{j\in S,\,j\neq i,k}q_{ik}q_{ji}\frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^k}+\sum_{l\notin S,\,l\neq i,k}q_{ki}q_{lk}\frac{\tilde{\pi}_t^l}{\tilde{\pi}_t^i}+\sum_{\substack{j\in S,\\ j\neq i,k}}\sum_{\substack{l\notin S,\\ l\neq i,k}}q_{ji}q_{lk}\frac{\tilde{\pi}_t^j\tilde{\pi}_t^l}{\tilde{\pi}_t^i\tilde{\pi}_t^k}}\right\},$$

$$\geq 2\min_{i\neq k}\left\{\sqrt{q_{ik}q_{ki}+\sum_{j\neq i,k}\min\left\{\frac{q_{ji}q_{ik}}{\tilde{\pi}_t^k},\frac{q_{jk}q_{ki}}{\tilde{\pi}_t^i}\right\}\tilde{\pi}_t^j}\right\}=:\tilde{\lambda}_t, \tag{30}$$

这意味着对于所有 $t<\infty$,

$$\tanh\left(\frac{\mathcal{H}(\pi_t,\tilde{\pi}_t)}{4}\right)\leq\tanh\left(\frac{\mathcal{H}(\mu,\nu)}{4}\right)e^{-\int_0^t\tilde{\lambda}_s^\star\,\mathrm{d}s}. \tag{31}$$

证明。回顾定理 2.2 的证明中使用的记号。考虑 (24)，并应用链式法则推导 $\Delta_\infty(t)/4$ 的双曲正切函数的动态，从而得到

$$\begin{aligned}\mathrm{d}\tanh\left(\frac{\Delta_\infty(t)}{4}\right)=&-\frac{1}{4}\cosh^{-2}\left(\frac{\Delta_\infty(t)}{4}\right)\left(q_{i^\star k^\star}T_{i^\star k^\star}(t)-q_{k^\star i^\star}T_{k^\star i^\star}(t)\right)\mathrm{d}t\\&-\frac{1}{4}\cosh^{-2}\left(\frac{\Delta_\infty(t)}{4}\right)\sum_{\substack{j=0\\ j\neq i^\star,k^\star}}^{n}\left(q_{jk^\star}T_{jk^\star}(t)-q_{ji^\star}T_{ji^\star}(t)\right)\mathrm{d}t.\end{aligned} \tag{32}$$

我们逐一考察上述等式右端的各项。首先考虑 $T_{i^\star k^\star}(t)$，并注意到

$$\begin{aligned}T_{i^\star k^\star}(t)&=e^{\tilde{\theta}_{k^\star}^{i^\star}(t)}\left(e^{\theta_{k^\star}^{i^\star}(t)-\tilde{\theta}_{k^\star}^{i^\star}(t)}-1\right)=e^{\tilde{\theta}_{k^\star}^{i^\star}(t)}\left(e^{\Delta_\infty(t)}-1\right)=e^{\tilde{\theta}_{k^\star}^{i^\star}(t)+\frac{\Delta_\infty(t)}{2}}\left(e^{\frac{\Delta_\infty(t)}{2}}-e^{-\frac{\Delta_\infty(t)}{2}}\right)\\&=2e^{\tilde{\theta}_{k^\star}^{i^\star}(t)+\frac{\Delta_\infty(t)}{2}}\sinh\left(\frac{\Delta_\infty(t)}{2}\right).\end{aligned}$$

回顾恒等式 $\sinh(2x)=2\sinh(x)\cosh(x)$，可得

$$\frac{T_{i^\star k^\star}(t)}{\cosh^2\left(\frac{\Delta_\infty(t)}{4}\right)}=4e^{\tilde{\theta}_{k^\star}^{i^\star}(t)+\frac{\Delta_\infty(t)}{2}}\tanh\left(\frac{\Delta_\infty(t)}{4}\right)=4\frac{\tilde{\pi}_t^{i^\star}}{\tilde{\pi}_t^{k^\star}}e^{\frac{\Delta_\infty(t)}{2}}\tanh\left(\frac{\Delta_\infty(t)}{4}\right).$$

同理,

$$-\frac{T_{k^\star i^\star}(t)}{\cosh^2\left(\frac{\Delta_\infty(t)}{4}\right)}=4e^{-\tilde{\theta}_{k^\star}^{i^\star}(t)-\frac{\Delta_\infty(t)}{2}}\tanh\left(\frac{\Delta_\infty(t)}{4}\right)=4\frac{\tilde{\pi}_t^{k^\star}}{\tilde{\pi}_t^{i^\star}}e^{-\frac{\Delta_\infty(t)}{2}}\tanh\left(\frac{\Delta_\infty(t)}{4}\right).$$

现在，对于 $j\neq i^\star,\ k^\star$，考虑 $q_{jk^\star}T_{jk^\star}(t)-q_{ji^\star}T_{ji^\star}(t)$。注意到

$$\Delta_{jk^\star}(t)=\theta_{k^\star}^j(t)-\tilde{\theta}_{k^\star}^j(t)=\left(\theta_{k^\star}^{i^\star}(t)-\tilde{\theta}_{k^\star}^{i^\star}(t)\right)+\left(\theta_{i^\star}^j(t)-\tilde{\theta}_{i^\star}^j(t)\right)=\Delta_\infty(t)+\Delta_{ji^\star}.$$

由引理 3.8，并回想 log 是递增的，我们有 $\Delta_{jk^\star}(t)\geq 0$ 和 $\Delta_{ji^\star}\leq 0$。此外，根据 $\Delta_\infty(t)$的定义，我们有 $\Delta_{jk^\star}(t)\leq\Delta_\infty(t)$，且 $|\Delta_{ji^\star}(t)|\leq\Delta_\infty(t)$。于是可以写成

$$\begin{aligned}q_{jk^\star}T_{jk^\star}(t)-q_{ji^\star}T_{ji^\star}(t)&=q_{jk^\star}e^{\tilde{\theta}_{k^\star}^j}(e^{\Delta_{jk^\star}(t)}-1)+q_{ji^\star}e^{\tilde{\theta}_{i^\star}^j}(1-e^{\Delta_{jk^\star}(t)-\Delta_\infty(t)})\\&=\tilde{\pi}_t^j\left[\frac{q_{jk^\star}}{\tilde{\pi}_t^{k^\star}}(e^{\Delta_{jk^\star}(t)}-1)+\frac{q_{ji^\star}}{\tilde{\pi}_t^{i^\star}}(1-e^{\Delta_{jk^\star}(t)-\Delta_\infty(t)})\right],\end{aligned}$$

特别地，若 $\frac{q_{jk^\star}}{\tilde\pi_t^{k^\star}} \geq \frac{q_{ji^\star}}{\tilde\pi_t^{i^\star}} e^{-\Delta_\infty(t)}$，则 $q_{jk^\star}T_{jk^\star}(t) - q_{ji^\star}T_{ji^\star}(t)$ 关于 $\Delta_{jk^\star}(t)$ 单调递增；否则单调递减。因此

$$\frac{q_{jk^\star}}{\tilde\pi_t^{k^\star}} \geq \frac{q_{ji^\star}}{\tilde\pi_t^{i^\star}} e^{-\Delta_\infty(t)} \Longrightarrow \min_{0\leq\Delta_{jk^\star}(t)\leq\Delta_\infty(t)} \left\{q_{jk^\star}T_{jk^\star}(t) - q_{ji^\star}T_{ji^\star}(t)\right\} = q_{ji^\star}\frac{\tilde\pi_t^j}{\tilde\pi_t^{i^\star}}(1-e^{-\Delta_\infty(t)}),$$

$$\frac{q_{jk^\star}}{\tilde\pi_t^{k^\star}} < \frac{q_{ji^\star}}{\tilde\pi_t^{i^\star}} e^{-\Delta_\infty(t)} \Longrightarrow \min_{0\leq\Delta_{jk^\star}(t)\leq\Delta_\infty(t)} \left\{q_{jk^\star}T_{jk^\star}(t) - q_{ji^\star}T_{ji^\star}(t)\right\} = q_{jk^\star}\frac{\tilde\pi_t^j}{\tilde\pi_t^{k^\star}}(e^{\Delta_\infty(t)}-1).$$

对所有 $t<\infty$, 以及所有 $i, k \in$ **自然数集** $\times$ **自然数集**, let

$$\tilde J_k^i(t,\Delta_\infty(t)) := \left\{j\in\mathbf{N}:\ \frac{q_{jk}}{\tilde\pi_t^k} \geq \frac{q_{ji}}{\tilde\pi_t^i}e^{-\Delta_\infty(t)} \text{ and } j\neq i,k\right\},$$

以及 $^{\mathsf{c}}\tilde J_k^i(t,\Delta_\infty(t)) :=$ **自然数集** $\setminus\left(\tilde J_k^i(t,\Delta_\infty(t)) \cup \{i,k\}\right)$。综合以上所有估计，我们可以对 (32) 的右端（按勒贝格–斯蒂尔切斯测度的意义）给出如下界

$$\begin{aligned}
\mathrm{d}\tanh\left(\frac{\Delta_\infty(t)}{4}\right) \leq & -\left[q_{i^\star k^\star}\frac{\tilde\pi_t^{i^\star}}{\tilde\pi_t^{k^\star}}e^{\frac{\Delta_\infty(t)}{2}} + q_{k^\star i^\star}\frac{\tilde\pi_t^{k^\star}}{\tilde\pi_t^{i^\star}}e^{-\frac{\Delta_\infty(t)}{2}}\right]\tanh\left(\frac{\Delta_\infty(t)}{4}\right)\mathrm{d}t \\
& -\left[\sum_{j\in\tilde J_{k^\star}^{i^\star}(t,\Delta_\infty(t))} q_{ji^\star}\frac{\tilde\pi_t^j}{\tilde\pi_t^{i^\star}}e^{-\frac{\Delta_\infty(t)}{2}} + \sum_{j\in{}^{\mathsf{c}}\tilde J_k^i(t,\Delta_\infty(t))} q_{jk^\star}\frac{\tilde\pi_t^j}{\tilde\pi_t^{k^\star}}e^{\frac{\Delta_\infty(t)}{2}}\right]\tanh\left(\frac{\Delta_\infty(t)}{4}\right)\mathrm{d}t \\
\leq & -\tilde\lambda(t,\Delta_\infty(t))\tanh\left(\frac{\Delta_\infty(t)}{4}\right)\mathrm{d}t, \qquad (33)
\end{aligned}$$

其中我们定义

$$\tilde\lambda(t,\Delta_\infty(t)) := \min_{i\neq k}\left\{q_{ik}\frac{\tilde\pi_t^i}{\tilde\pi_t^k}e^{\frac{\Delta_\infty(t)}{2}} + q_{ki}\frac{\tilde\pi_t^k}{\tilde\pi_t^i}e^{-\frac{\Delta_\infty(t)}{2}} + \sum_{j\in\tilde J_k^i(t,\Delta_\infty(t))} q_{ji}\frac{\tilde\pi_t^j}{\tilde\pi_t^i}e^{-\frac{\Delta_\infty(t)}{2}} + \sum_{j\in{}^{\mathsf{c}}\tilde J_k^i(t,\Delta_\infty(t))} q_{jk}\frac{\tilde\pi_t^j}{\tilde\pi_t^k}e^{\frac{\Delta_\infty(t)}{2}}\right\}.$$

利用 t 对于不等式 $a+b\geq 2\sqrt{ab}$，当 $a, b\geq 0$ 时，我们有 $\forall(i,\ k) \in$ **自然数集** $\times$ **自然数集**，并且 $\forall t<\infty$ ，

$$\begin{aligned}
& q_{ik}\frac{\tilde\pi_t^i}{\tilde\pi_t^k}e^{\frac{\Delta_\infty(t)}{2}} + q_{ki}\frac{\tilde\pi_t^k}{\tilde\pi_t^i}e^{-\frac{\Delta_\infty(t)}{2}} + \sum_{j\in\tilde J_k^i(t,\Delta_\infty(t))} q_{ji}\frac{\tilde\pi_t^j}{\tilde\pi_t^i}e^{-\frac{\Delta_\infty(t)}{2}} + \sum_{j\in{}^{\mathsf{c}}\tilde J_k^i(t,\Delta_\infty(t))} q_{jk}\frac{\tilde\pi_t^j}{\tilde\pi_t^k}e^{\frac{\Delta_\infty(t)}{2}} \\
\geq & 2\left[\frac{e^{\frac{\Delta_\infty(t)}{2}}}{\tilde\pi_t^k}\left(q_{ik}\tilde\pi_t^i + \sum_{j\in{}^{\mathsf{c}}\tilde J_k^i(t,\Delta_\infty(t))} q_{jk}\tilde\pi_t^j\right)\right]^{\frac12}\left[\frac{e^{\frac{-\Delta_\infty(t)}{2}}}{\tilde\pi_t^i}\left(q_{ki}\tilde\pi_t^k + \sum_{j\in\tilde J_k^i(t,\Delta_\infty(t))} q_{ji}\tilde\pi_t^j\right)\right]^{\frac12} \\
\geq & 2\left(q_{ik}q_{ki} + \sum_{j\in\tilde J_k^i(t,\Delta_\infty(t))} q_{ik}q_{ji}\frac{\tilde\pi_t^j}{\tilde\pi_t^k} + \sum_{l\in{}^{\mathsf{c}}\tilde J_k^i(t,\Delta_\infty(t))} q_{ki}q_{lk}\frac{\tilde\pi_t^l}{\tilde\pi_t^i} + \sum_{\substack{j\in\tilde J_k^i(t,\Delta_\infty(t)),\\ l\in{}^{\mathsf{c}}\tilde J_k^j(t,\Delta_\infty(t))}} q_{ji}q_{lk}\frac{\tilde\pi_t^j\tilde\pi_t^l}{\tilde\pi_t^i\tilde\pi_t^k}\right)^{\frac12},
\end{aligned}$$

由此得到

$$\tilde\lambda(t,\Delta_\infty(t)) \geq 2\min_{i\neq k}\min_{S\subseteq\mathbf{N}}\left(q_{ik}q_{ki} + \sum_{\substack{j\in S,\\ j\neq i,k}} q_{ik}q_{ji}\frac{\tilde\pi_t^j}{\tilde\pi_t^k} + \sum_{\substack{l\notin S,\\ l\neq i,k}} q_{ki}q_{lk}\frac{\tilde\pi_t^l}{\tilde\pi_t^i} + \sum_{\substack{j\in S,\\ j\neq i,k}}\sum_{\substack{l\notin S,\\ l\neq i,k}} q_{ji}q_{lk}\frac{\tilde\pi_t^j\tilde\pi_t^l}{\tilde\pi_t^i\tilde\pi_t^k}\right)^{\frac12}. \qquad (34)$$

接着，对（33）的右端进行估计，并应用 Gr¨onwall 论证（针对绝对连续过程），即可得到（31）。不等式（30）可通过进一步最小化（34）的右端立即得出，特别是

$$\tilde{\lambda}(\tilde{\pi}_t, \Delta_\infty(t)) \geq \tilde{\lambda}_t := 2 \min_{i \neq k} \left\{ \sqrt{q_{ik}q_{ki} + \sum_{j \neq i,k} \min\left\{ \frac{q_{ji}q_{ik}}{\tilde{\pi}_t^k}, \frac{q_{jk}q_{ki}}{\tilde{\pi}_t^i} \right\} \tilde{\pi}_t^j} \right\} > 0, \tag{35}$$

°因为根据假设， Q 的非对角元素严格为正，并且由引理 3.1 可知，对所有 $t < \infty$， $\tilde{\pi}_t \in \mathcal{S}^n$。

现在令 X_t :=双曲正切函数($\Delta_\infty(t)/4$)。那么我们可以将 (33) 改写为

$$\mathrm{d}X_t \leq -\tilde{\lambda}^\star(t, X_t) X_t \, \mathrm{d}t,$$

其中

$$\begin{aligned}\tilde{\lambda}^\star(t, X_t) &:= \tilde{\lambda}(t, 4 \operatorname{arctanh}(X_t)) \\ &= \min_{i \neq k} \left\{ \left(q_{ik} \frac{\tilde{\pi}_t^i}{\tilde{\pi}_t^k} + \sum_{j \in \mathfrak{c}\tilde{\mathcal{J}}_k^i(t, X_t)} q_{jk} \frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^k} \right) \frac{1 + X_t}{1 - X_t} + \left(q_{ki} \frac{\tilde{\pi}_t^k}{\tilde{\pi}_t^i} + \sum_{j \in \tilde{\mathcal{J}}_k^i(t, X_t)} q_{jk} \frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^k} \right) \frac{1 - X_t}{1 + X_t} \right\},\end{aligned}$$

并且 $\tilde{\mathcal{J}}_k^i(t,\ X_t) := \tilde{J}_k^i(t,$ 4 反双曲正切$(X_t))$，且 $\mathfrak{c}\tilde{\mathcal{J}}_k^i(t,\ X_t)$ = **自然数集** $\setminus \left(\tilde{\mathcal{J}}_k^i(t,\right.$ $\left.X_t\right) \cup \{i, k\})$。我们希望将 X_t 与 (27) 的解进行比较。接下来，我们论证 (27) 对所有 $t < \infty$ 都有良定义的解。于是，该命题可由直接应用引理 3.9 得出。

°令 u_t 为 (27) 的一个解。注意到 $\mathcal{H}(\mu, \nu) \in [0, \infty)$，因为根据假设， $\mu, \nu \in \mathcal{S}^n$，所以 $u_0 \in [0, 1)$。如果 $\mathcal{H}(\mu, \nu) = 0$，那么 $u_0 = 0$，并且对所有 $t \geq 0$， $u_t = 0$，因此该命题˜显然成立。故从现在起，假设 $u_0 \in (0, 1)$。现在，系数 $\lambda^\star(\tilde{\pi}_t, u_t)$ 依赖于过程 $\tilde{\pi}_t$，而该过程由 ω-ω 固定。注意到当 $u_t \uparrow 1$ 时， $\lambda^\star(\tilde{\pi}_t, u_t)$ 会发散。由于 $u_0 < 1$，爆炸时刻 T 满足 $u_{T-} = 1$ 严格为正。回顾引理 (3.1)，对所有 $t < \infty$， $\tilde{\pi}_t \in \mathcal{S}^n$。于是，在区间 $[0, T)$ 上， $x \mapsto \lambda^\star(\tilde{\pi}_t, x)x$ 局部利普希茨连续（其利普希茨常数依赖于 t、ω 和 x），常微分方程理论中的标准结果（例如 [33, 定理 2.5]）表明，(27) 在 $[0, T)$ 上有唯一解 u_t。˜ ˜另一方面，对于 $u_t \in (-1, 1)$， $\lambda^\star(\tilde{\pi}_t, u_t) \geq \lambda_t \geq 2 \min_{i \neq k} \sqrt{q_{ik}q_{ki}}$ 严格为正，因此当 $u_t \in (0, 1)$ 时，$-\lambda^\star(\tilde{\pi}_t, u_t)u_t$ 严格为负。于是对所有 $t \geq 0$， $u_t \leq u_0 < 1$，因此实际上爆炸时刻 $T = \infty$，且 (27) 对所有 $t \geq 0$ 都有唯一解。此外，当 u_t 趋近于 0 时， $-\lambda^\star(\tilde{\pi}_t, u_t)u_t$ 趋于 0，因此很容易推出，对所有 $t \geq 0$， $u_t \in (0, u_0]$。

□

根据对称性，命题 3.10 的界也可以用真实滤波器 π_t 表示。

推论 3.10.1。 假设对于所有 $i \in$ **自然数集**， $\mu^i, \nu^i > 0$，且已观测到 π_t。对于所有 $t \geq 0$，令$u_t \in (0,\ 1)$ 为由随机系数常微分方程给出的唯一解

$$\frac{\mathrm{d}u_t}{\mathrm{d}t} = -\lambda^\star(t, u_t)u_t, \qquad u_0 = \tanh\left(\frac{\mathcal{H}(\mu, \nu)}{4}\right), \tag{36}$$

其中

$$\lambda^\star(t, u_t) = \min_{i \neq k} \left\{ \left(q_{ik} \frac{\pi_t^i}{\pi_t^k} + \sum_{\substack{j \neq i,k, \\ j \in \mathcal{J}_k^i(t, u_t)}} q_{jk} \frac{\pi_t^j}{\pi_t^k} \right) \frac{1 - u_t}{1 + u_t} + \left(q_{ki} \frac{\pi_t^k}{\pi_t^i} + \sum_{\substack{j \neq i,k, \\ j \in \mathcal{J}_k^i(t, u_t)}} q_{ji} \frac{\pi_t^j}{\pi_t^i} \right) \frac{1 + u_t}{1 - u_t} \right\}, \tag{37}$$

且 $\mathcal{J}_k^i(t,u_t) := \left\{ j \in \text{自然数集} : \frac{q_{jk}}{\pi_t^k} \le \frac{q_{ji}}{\pi_t^i}\left(\frac{1+u_t}{1-u_t}\right)^2 \right\}$。那么对于所有 $t<\infty$,

$$\tanh\left(\frac{\mathcal{H}(\pi_t,\tilde{\pi}_t)}{4}\right) \le u_t.$$

特别地，对于所有 $t<\infty$，双曲正切函数 $\left(\frac{\mathcal{H}(\pi_t,\tilde{\pi}_t)}{4}\right) \le$ 双曲正切函数 $\left(\frac{\mathcal{H}(\mu,\nu)}{4}\right) e^{-\int_0^t \lambda_s^\star, } \mathrm{d}^s$，其中 $\lambda_t^\star \ge \lambda_t$，且 $\lambda_t^\star$ 以及 λ_t 的定义分别与 (30) 中的 $\lambda_t^\star$ 和 λ_t 等价，只是用 π_t 替代了 $\tilde{\pi}_t$。证明。证明与命题 3.10 的证明类似。

□

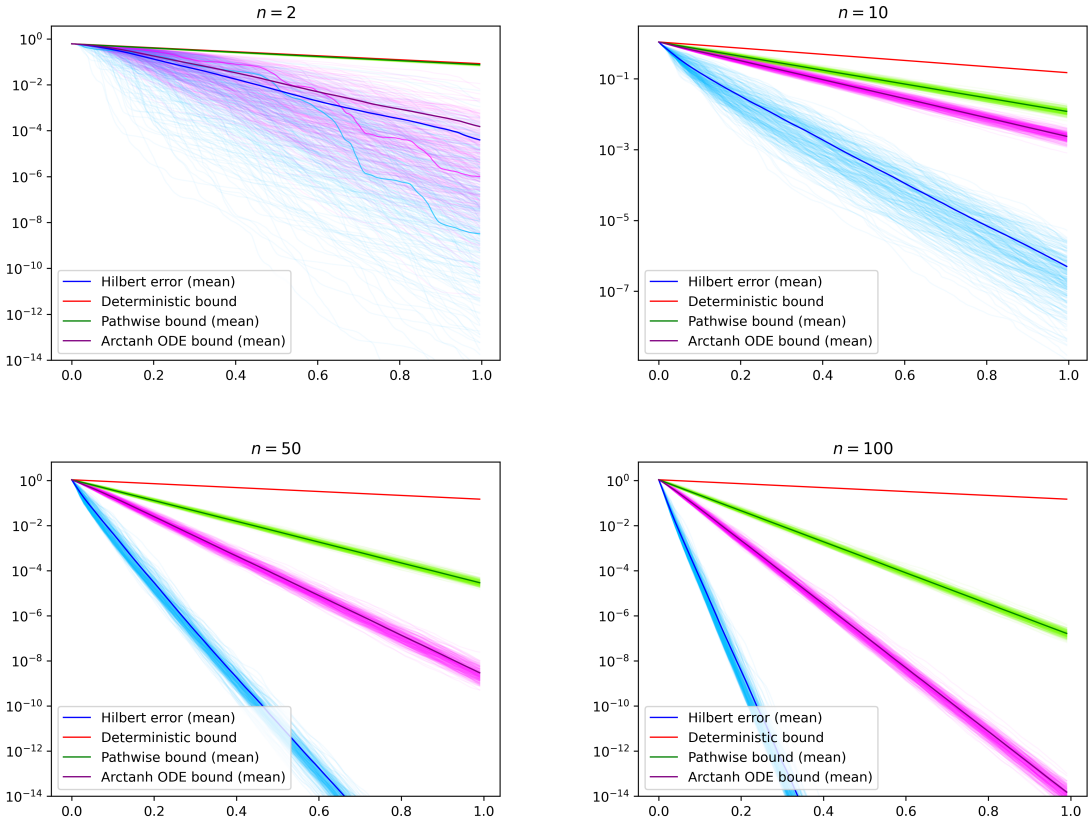

图 2：对于维度为 $n = 2,20,50,100$ 和 $\pi_t, \tilde{\pi}_t \in \mathcal{S}^n$ 的情形，分别以 $\mu, \nu \in \mathcal{S}^n$ 为初始值，我们绘制了 300 次希尔伯特射影误差的实现 $\mathcal{H}(\pi_t, \tilde{\pi}_t)$（浅蓝色）、逐路径界 $\mathcal{H}(\mu,\nu)e^{-\int_0^t \tilde{\lambda}_s \mathrm{d}^s}$（浅绿色）以及命题 3.10 中常微分方程界（洋红色），均采用对数刻度，$t \in [0,1]$。蓝色、深绿色和紫色曲线分别表示相应的样本均值。红色曲线表示定理 2.2 中的确定性界 $\mathcal{H}(\mu,\nu)e^{-\lambda t}$。对于 $n=2$ 的情形，我们特别标出一次希尔伯特误差实现（随机选取）及其对应的常微分方程界。在这次模拟中，我们在所有维度下保持速率矩阵的结构不变，因此确定性速率 λ 也保持不变，始终等于 2。可以看到，随着维度增加，确定性收缩率越来越不理想，而改为计算逐路径速率或常微分方程界所带来的改进十分显著（尽管该界仍远不够紧）。有关这次模拟的更多详情，请参见附录 B。

~ ~ ~在 (30) 中，我们给出了 $\lambda_t^\star$ 的下界 λ_t，因为从数值计算的角度来看，对于每个 t，通过遍历**自然数集** 的所有可能子集来求得 $\lambda_t^\star$，代价可能很高，尤其是在高维情况下。另一方面， λ_t 很容易计算。我们在图 2 中展示命题 3.10 所给界的表现：我们绘制了逐路径界 $\mathcal{H}(\mu,\nu)e^{-2\int_0^t \tilde{\lambda}_s \,\mathrm{d}s}$，它~可直接由 (31) 得出，使用速率 λ_t；以及常微分方程界 4反双曲正切(u_t)，其中 u_t 是 (27) 的数值解，并将它们与定理 2.2 给出的确定性速率进行比较。

从图 2 的曲线可以看出，即使使用逐路径收缩率或命题 3.10 中的常微分方程界，我们在维度 $n \geq 2$下得到的界仍不够紧。这也影响了我们在第 4.2.1 节中对误差界所做的模拟。鉴于迄今为止尝试的那类进一步代数运算似乎不太可能得到更好的界，我们可以考虑改进估计方法，转而考察期望希尔伯特误差的衰减率；从模拟结果来看，该误差的表现似乎相当良好。或者，我们也可以考察期望收缩率。若要沿任一方向继续推进，就需要找到一种方法，估计 π_t 和 $\tilde{\pi}_t$各分量之比的最大值点和极小值点的期望。

我们的数值实验还表明，至少对于所考察的例子，在高维情形下会出现测度集中现象；此时，以压倒性概率成立的收敛率将远快于我们已证明的收敛率。我们将这一问题留待未来研究。

4 鲁棒性与误差界限

第 3 节的压缩结果使我们得以考察采用近似滤波器而非最优滤波器时的误差行为。

4.1 Wonham滤波器关于模型参数的连续性

在本节中，我们重新证明 Chigansky 和 Van Handel 关于 Wonham滤波器对模型参数具有鲁棒性的结果（见 [14]）。请注意，我们采用的方法与 [14], 完全不同，而且我们通过希尔伯特误差而非 ℓ^1-范数来获得鲁棒性。由于希尔伯特度量强于 ℓ^1 （见 [3]中的引理 1），我们在此得到的误差估计比 [14]中的更紧。

考虑模型参数不正确的近似 Wonham 滤波器

$$\mathrm{d}\tilde{\pi}_t = \tilde{Q}^T\tilde{\pi}_t\,\mathrm{d}t + \left(\tilde{H} - \tilde{\pi}_t^\top \tilde{h}\,\mathbb{I}_{n+1}\right)\tilde{\pi}_t\left(\mathrm{d}Y_t - \tilde{\pi}_t^\top\tilde{h}\,\mathrm{d}t\right), \quad \tilde{\pi}_0 = \nu, \tag{38}$$

~ ~其中 $Q = (\tilde{q}_{ij})$ 和 h 分别是一个转移强度矩阵和一个有界传感器函数，且不同于 Q 和 h； $\tilde{H} = \mathrm{diag}(\tilde{h})$ 是对角矩阵，其对角元为 $(\tilde{H})_{ii} = \tilde{h}^i$。我们关注希尔伯特误差 $\mathcal{H}(\pi_t, \tilde{\pi}_t)$。

记号。 与前几节相比，$(\tilde{\pi}_t)_{t\geq 0}$ 现在表示方程 (38) 的解，而 $(\pi_t)_{t\geq 0}$ 仍然是方程 (4) 的解。

注。 再次提醒，所有结果均可推广到多维观测 ($d \neq 1$) 和可逆的 σ 情形。尤其要注意，本文给出的论证很容易适用于 (1) 和 (4) 中的 $\sigma \neq 1$ （这在更高维情形中对应于可逆的 $\sigma \neq \mathbb{I}_d$）。这将在 (38) 中再增加一个“设定错误”的参数 $\tilde{\sigma}$ 。在这段内容中。

˜我们在下文给出的证明、 σ 和 $\tilde{\sigma}$ 可以分别直接纳入方程（4）和（38）中的 h 和 h ，从而得到“正确的”和“错误的”Wonham滤波器。由于其中的论证需要了解观测 Y 的二次变差，因此无法在 [14] 中处理这种情况（另见 [14,备注4.2]）。

本节余下部分将用于证明以下定理。

定理 4.1 (模型稳健性**)**。 设 π_t 为方程 (4) 的解， $\tilde{\pi}_t$ 为方程 (38)的解。假设对所有 $i \in \mathbf{N}$ ，μ^i、 $\nu^i > 0$ 成立；并且对所有 $i \neq j$， q_{ij}、 $\tilde{q}_{ij} > 0$ 成立。对所有 $t < \infty$，

$$\mathbf{E}\left[\mathcal{H}(\pi_t, \tilde{\pi}_t)\right] \leq \mathcal{H}(\mu, \nu)e^{-\lambda t} + K_q \int_0^t e^{-\lambda(t-s)} \mathbf{E}\left[\frac{1}{\min_k \tilde{\pi}_s^k}\right] \mathrm{d}s + K_h \int_0^t e^{-\lambda(t-s)}\, \mathrm{d}s$$
$$+ \frac{1}{4} \sum_{(i,k)} \sum_{(j,l)\neq(i,k)} \mathbf{E}\left[\int_0^t e^{-\lambda(t-s)}\, \mathrm{d}L_s^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot))\right],$$

其中， $\lambda = 2$为$_{i\neq j} \sqrt{q_{ij} q_{ji}}$的最小值， $K_q = 2$ 为$_j$的最大值，$_k |\tilde{q}_{jk} - q_{jk}|$ ，而 $K_h = 2$ 为$_j |h^j|$ 、$_i |h^i - \tilde{h}^i| +$、$_i |(h^i)^2 - (\tilde{h}^i)^2|$的最大值； $L_t^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot))$表示过程 $(\Delta_{ik}(t) - \Delta_{jl}(t))_{t\geq 0}$在 0 处的局部时，其中 Δ_{ik} 对所有 $(i,k) \in N \times N$ 均按下文的 (39) 演化。

此外，˜对固定的紧集内 $\tilde{h}$，即若在常数项都会消失，得因为 $\tilde{h} \to h$ 对于 ˜

$$\lim_{\tilde{h}\to h} \mathbf{E}\left[\mathcal{H}(\pi_t, \tilde{\pi}_t)\right] \leq \mathcal{H}(\mu, \nu)e^{-\lambda t} + K_q \tilde{C}(1 - e^{-\lambda}).$$

˜即使在 h 保持不变的情况下，这一结果也使我们能够很好地控制 $\mathcal{H}(\pi_t, \tilde{\pi}_t)$，正如下一个命题所示。

命题 4.2. 定理 *4.1* 中的误差项在 $t \to \infty$ 时保持有限。具体而言，

$$\sup_{t\geq 0} \mathbf{E}\left[\int_0^t e^{-\lambda(t-s)}\, \mathrm{d}L_s^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot))\right] < \infty, \quad \forall (i,k),(j,l) \in \mathbf{N} \times \mathbf{N}, (i,k) \neq (j,l).$$

备注。 我们在上文陈述定理 4.1 时，将衰减率 λ 设为 Q的函数，其中 Q是马尔可夫链的“真实”动力学；并将（第一项）误差项设为“模型设定错误”过程 $\tilde{\pi}_t$的函数。然而，如果我们愿意，证明中的任何内容都不妨碍我们反过来处理：若将 λ替换为 $\tilde{\lambda} = 2\min_{i\neq j} \sqrt{\tilde{q}_{ij}\tilde{q}_{ji}}$，并将 $\tilde{\pi}_t$ 替换为 π_t，该定理仍然成立。请注意，由 h 的模型设定错误导致的误差项（以及局部时间项）保持不变。

现在我们开始证明这些结果。与第 3 节一样，我们先将 $\pi_t, \tilde{\pi}_t$ 变换为$\theta_t, \tilde{\theta}_t$，并推导出 $\Delta_{ik}(t) = \theta_k^i(t) - \tilde{\theta}_k^i(t) = \log \frac{\pi^i}{\pi^k}(t) - \log \frac{\tilde{\pi}^i}{\tilde{\pi}^k}(t)$ 在 $(i,k) \in$ **自然数集 $\times$ 自然数集</style} 中的动力学（回忆一下）。请注意，对于有限的 , 、 和 (,** $\tilde{\pi}_t$) 几乎处处都有定义（因为当滤波器 $\tilde{\pi}$ 的动力学由参数 ˜ ˜Q 和 h 给出时，引理 3.1 同样成立）。

对（4）和（38）应用伊藤公式，我们推导出差分过程$\Delta_{ik}(t)$的动态方程如下

$$\mathrm{d}\Delta_{ik}(t) = -\sum_{\substack{j=0\\ j\neq k}}^{n} \left(q_{jk}\frac{\pi_t^j}{\pi_t^k} - \tilde{q}_{jk}\frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^k}\right) \mathrm{d}t + \sum_{\substack{j=0\\ j\neq i}}^{n} \left(q_{ji}\frac{\pi_t^j}{\pi_t^i} - \tilde{q}_{ji}\frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^i}\right) \mathrm{d}t + (q_{ii} - \tilde{q}_{ii} - q_{kk} + \tilde{q}_{kk})\, \mathrm{d}t$$
$$+ (h^i - \tilde{h}^i - h^k + \tilde{h}^k)(\, \mathrm{d}B_t + \pi_t^\top h\, \mathrm{d}t) + \frac{1}{2}\Big((h^k)^2 - (\tilde{h}^k)^2 - (h^i)^2 + (\tilde{h}^i)^2\Big)\, \mathrm{d}t,$$
$$\Delta_{ik}(0) = \log\frac{\mu^i}{\mu^j} - \log\frac{\nu^i}{\nu^j}, \tag{39}$$

其中，我们再次引入了创新过程 $B_t = Y_t - \int_0^t \pi_s^\top h \mathrm{d}s$。

~我们马上注意到，如果 $h = h$，随机项以及所有涉及 h 和 $\tilde{h}$的漂移项都会消失。因此，在这个简单情形下，我们再次得到 C^1 对于$\Delta_{ik}(t)$ 的动力学，适用于第 3 节中类似的论证，从而得到对所有 $t < \infty$成立的以下估计

$$\begin{aligned}\mathcal{H}(\pi_t, \tilde{\pi}_t) &\le \mathcal{H}(\mu,\nu)e^{-\int_0^t \tilde{\lambda}_s \,\mathrm{d}s} + \int_0^t e^{-\int_s^t \tilde{\lambda}_r \,\mathrm{d}r} \max_{i,k}\left\{\frac{(\delta Q^\top \tilde{\pi}_s)^i}{\tilde{\pi}_s^i} - \frac{(\delta Q^\top \tilde{\pi}_s)^k}{\tilde{\pi}_s^k}\right\}\mathrm{d}s \\ &\le e^{-\lambda t}\mathcal{H}(\mu,\nu) + 2\max_{i,k}|\tilde{q}_{ik} - q_{ik}| \int_0^t e^{-\lambda(t-s)} \frac{1}{\min_j \tilde{\pi}^j(s)}\,\mathrm{d}s \\ &\le e^{-\lambda t}\mathcal{H}(\mu,\nu) + \frac{2}{\lambda}\max_{i,k}|\tilde{q}_{ik} - q_{ik}| \,\frac{1}{\min_{j\in\mathbf{N},\, s\in[0,t]} \tilde{\pi}^j(s)}(1 - e^{-\lambda t}), \end{aligned} \tag{40}$$

其中 $\delta Q := \tilde{Q} - Q$ 且 $\min_{j\in\mathbf{N},\ s\in[0,t]} \tilde{\pi}^j(s) \neq 0$ 几乎必然成立（由引理 3.1 可知），而 $\tilde{\lambda}_t$ 和 λ 分别是命题 3.10 中式（30）给出的逐路径收缩率，以及定理 2.2 中的确定性速率。类似命题 3.10 的思路，可以得到这个误差的更紧确界；不过，我们将在第 4.2 节讨论一般近似滤波器的误差时再给出这些界（模型参数设定错误的滤波器就是其中一个特例）。

在这种情况~第 3 节中提出的证明策略无法直接应用。通过仔细修改论证以考虑随机项，例如反复应用田中公式，有可能推导出类似于 (21) 的$\Delta_\infty(t) = \max_{(i,k)\in\mathbf{N}\times\mathbf{N}} \Delta_{ik}(t)$的动力学。然而，我们在此采用另一种策略。我们首先引入几个定义。

回顾最大值和最大值点的以下光滑近似 .

定义 4.3. 令 $\alpha \in (0,\infty)$，并令 $\mathbf{X}_t = \{X_t^0, \ldots, X_t^n\}$ 为一族取实值的连续随机变量。定义 *LogSumExp* 函数 $LSE_\alpha(\mathbf{X}_\cdot)(t)$为

$$LSE_\alpha(\mathbf{X}_\cdot)(t) = \frac{1}{\alpha}\log\sum_i e^{\alpha X_t^i}. \tag{41}$$

对于$\mathbf{g}_t = \{g_t^i(x)\}_{i=0}^n$ 这一族取实值的函数，定义*SoftArgMax* （也称为*SoftMax*）函数 $S_\alpha^{arg}(\mathbf{X}_\cdot, \mathbf{g}_\cdot)(t)$为

$$S_\alpha^{arg}(\mathbf{X}_\cdot, \mathbf{g}_\cdot)(t) = \frac{\sum_i g_t^i(X_t^i)e^{\alpha X_t^i}}{\sum_k e^{\alpha X_t^k}}. \tag{42}$$

注意，对于每个 ω， $LSE_\alpha(\mathbf{X}_\cdot)$ 都逐点收敛到 $\max_i X^i$，而 $S_\alpha^{arg}(\mathbf{X}_\cdot, \mathbf{g}_\cdot(\mathbf{X}_\cdot))(t)$ 则逐点收敛到 $\frac{1}{|\mathcal{I}|}\sum_{j\in\mathcal{I}} g_t^j(X_t^j)$，其中$\mathcal{I}$ =最大值点$_i$ X^i，当 $\alpha \to \infty$ 对每个 t 均成立时（见附录 A）。

在接下来的论证中，我们还需要连续半鞅的局部时 L_a^t 的定义。特别是，我们只关注有限变差部分绝对连续的连续半鞅的局部时（例如 Itô 过程）。在这种情况下，对于具有绝对连续有限变差部分的半鞅 Z，我们可以采用以下局部时定义（改编自 Revuz 和 Yor [30,第 6 章推论 1.9]，并注意到如果 Z 的有限变差部分绝对连续，那么 [30, 第 6 章定理 1.7] 的证明表明， L_t^a 在 a 中有一个几乎处处双连续的修正版本，且 t）。

定义 4.4。 设 $Z_t = M_t + A_t$ 为实值连续半鞅，其中 M 是局部鞅， A 是绝对连续有限变差过程。我们将 Z 在 $a \in \mathbb{R}$处、时间 t的局部时定义为过程 L_a^t ，它在 $t \in \mathbb{R}^+$ 中连续，并由 a 给出

$$L_t^a = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{(a-\varepsilon, a+\varepsilon)}(Z_s)\, \mathrm{d}\langle Z \rangle_s \quad a.s.$$

4.1.1 定理 4.1 的证明

定理 4.1 的证明策略如下。我们利用 (41) 定义最大过程 Δ_∞ 的光滑近似，并通过伊藤公式推导其动力学。随后，在令 $\alpha \to \infty$ 趋于极限时，需要稍加留意，因为当最大过程在同一时刻于多个指标处取到时，一些被积函数会收敛到狄拉克质量。在附录 A 中，我们说明如何处理这些项；这些项决定了估计中局部时的出现。

定理 *4.1* 的证明。 考虑过程族 $\boldsymbol{\Delta}_t = \{\Delta_{ik}(t)\}_{i \in \mathbf{N},\ k \in \mathbf{N}}$ 其演化方程由 (39) 给出。我们应用伊藤引理推导 $LSE_\alpha(\boldsymbol{\Delta}_\cdot)(t)$ 的动力学

$$\begin{aligned}
\mathrm{d}LSE_\alpha(\boldsymbol{\Delta}_\cdot)(t) &= \sum_{(j,l)} \frac{e^{\alpha \Delta_{jl}(t)}}{\sum_{(i,k)} e^{\alpha \Delta_{ik}(t)}}\, \mathrm{d}\Delta_{jl}(t) \\
&\quad + \frac{1}{2} \sum_{(j,l)} \sum_{(u,v)} \alpha e^{\alpha \Delta_{jl}(t)} \left(\frac{\mathbb{1}_{\{(u,v)=(j,l)\}}}{\sum_{(i,k)} e^{\alpha \Delta_{ik}(t)}} - \frac{e^{\alpha \Delta_{uv}(t)}}{(\sum_{(i,k)} e^{\alpha \Delta_{ik}(t)})^2} \right) \mathrm{d}\langle \Delta_{jl}(\cdot), \Delta_{uv}(\cdot) \rangle_t,
\end{aligned}$$

其中，所有求和均遍历双重指标集合 **自然数集** $\times$**自然数集**。

回顾我们在第 3 节中的记号 $T_{ik}(t) := \frac{\pi_t^i}{\pi_t^k} = \frac{\tilde{\pi}_t^i}{\tilde{\pi}_t^k}$ ，我们通过适当地添加和减去各项，对 Δ_{ik} 的动力学进行整理，得到

$$\begin{aligned}
\mathrm{d}\Delta_{ik}(t) &= \left(-\sum_{\substack{j=0 \\ j \neq k}}^{n} q_{jk} T_{jk}(t) + \sum_{\substack{j=0 \\ j \neq i}}^{n} q_{ji} T_{ji}(t) \right) \mathrm{d}t + \left(\frac{(\delta Q^\top \tilde{\pi}_t)^k}{\tilde{\pi}_t^k} - \frac{(\delta Q^\top \tilde{\pi}_t)^i}{\tilde{\pi}_t^i} \right) \mathrm{d}t \\
&\quad + (h^i - \tilde{h}^i - h^k + \tilde{h}^k)\, \mathrm{d}B_t + \big(h^i - \tilde{h}^i - h^k + \tilde{h}^k\big) \pi_t^\top h \, \mathrm{d}t \\
&\quad + \frac{1}{2} \Big((h^k)^2 - (\tilde{h}^k)^2 - (h^i)^2 + (\tilde{h}^i)^2 \Big)\, \mathrm{d}t \\
&=: C_{ik}^{q,1}(t)\, \mathrm{d}t + C_{ik}^{q,2}(t)\, \mathrm{d}t + C_{ik}^{h,1}\, \mathrm{d}B_t + C_{ik}^{h,1} \pi_t^\top h \, \mathrm{d}t + \frac{1}{2} C_{ik}^{h,2}\, \mathrm{d}t,
\end{aligned}$$

˜其中再次有 $\delta Q = Q - Q$。

那么，对于所有 $s \le t < \infty$，我们有

$$
\begin{aligned}
LSE_\alpha(\boldsymbol{\Delta}_\cdot)(t) = LSE_\alpha(\boldsymbol{\Delta}_\cdot)(s)
&+ \int_s^t \sum_{(j,l)} \frac{e^{\alpha\Delta_{jl}(r)}}{\sum_{(i,k)} e^{\alpha\Delta_{ik}(r)}} \left(C_{jl}^{q,1}(r) + C_{jl}^{q,2}(r) + C_{jl}^{h,1}\, \pi_r^\top h + \frac{1}{2} C_{jl}^{h,2} \right) \mathrm{d}r \\
&+ \int_s^t \sum_{(j,l)} \frac{e^{\alpha\Delta_{jl}(r)}}{\sum_{(i,k)} e^{\alpha\Delta_{ik}(r)}}\, C_{jl}^{h,1}\, \mathrm{d}B_r \\
&+ \frac{1}{2} \int_s^t \sum_{(j,l)} \sum_{(u,v)\neq(j,l)} \frac{\alpha e^{\alpha(\Delta_{jl}(r)+\Delta_{uv}(r))}}{(\sum_{(i,k)} e^{\alpha\Delta_{ik}(r)})^2} \left((C_{jl}^{h,1})^2 - C_{jl}^{h,1} C_{uv}^{h,1} \right) \mathrm{d}r \\
=: LSE_\alpha(\boldsymbol{\Delta}_\cdot)(s) &+ I_1 + I_2 + \frac{1}{2} I_3.
\end{aligned}
$$

我们希望在两边取 $\alpha \to \infty$ 的极限。

记作

$$\Delta_\infty(t) = \max_{(j,l)\in \mathbf{N}\times\mathbf{N}} \Delta_{jl}(t),$$

并且我们立即得到，当 $\alpha \to \infty$ 时， $LSE_\alpha(\boldsymbol{\Delta}_\cdot)(t)$ 收敛到 $\Delta_\infty(t)$。

我们令 $\lambda = 2\min_{i\neq k} \sqrt{q_{ik}\, q_{ki}}$，并对 $j \in$ **自然数集** 定义误差项

$$\mathcal{E}_t^{1,j} = \frac{(\delta Q^\top \tilde{\pi}_s)^j}{\tilde{\pi}_s^j}, \quad \mathcal{E}^{2,j} = (\tilde{h}^j)^2 - (h^j)^2, \quad \mathcal{E}^{3,j} = \tilde{h}^j - h^j.$$

对于每个时间 t，定义 $\mathcal{I}_t \subset$ **自然数集** $\times$ **自然数集** 为 $\boldsymbol{\Delta}_t$ 的最大值点，即满足对于所有 $(i,k) \in \mathcal{I}_t$，均有 $\Delta_{ik}(t) = \Delta_\infty(t)$ 的双重指标 (i,k) 的集合。令 $|\mathcal{I}_t|$ 表示 $\mathcal{I}_t$ 的大小。我们逐一考虑 I_1、I_2 和 I_3。

从 I_1 开始。我们将 $S_\alpha^{arg}(\boldsymbol{\Delta}_\cdot, C^q{,}1(\cdot))(r)$、$S_\alpha^{arg}(\boldsymbol{\Delta}_\cdot, C^q{,}2(\cdot))(r)$ 和 $S_\alpha^{arg}(\boldsymbol{\Delta}_\cdot, C^h{,}1)(r)\ \pi_r^\top h$ 识别为被积函数，此外还有 $S_\alpha^{arg}(\boldsymbol{\Delta}_\cdot, C^h{,}2)(r)$。它们分别由 $\max_{i,j\in\mathbf{N}\times\mathbf{N}} C^q{}_{,1}$ 和 $\max_{i,j\in\mathbf{N}\times\mathbf{N}} C^q{}_{,ij}{}^2$ 控制；这些函数关于时间连续，因此在$[0,t]$上可积。此外，它们还由 $\max_k |h^k| \max_{i,j\in\mathbf{N}\times\mathbf{N}} C^h{}_{,ij}{}^1 < \infty$ 和 $\max_{i,j\in\mathbf{N}\times\mathbf{N}} C^h{}_{,}{}^2 < \infty$ 控制，而根据关于 h 的假设，这些量有界，因此可积。于是，我们可以应用控制收敛定理，将极限移入积分号内；由引理 A.2 可得，对所有 $s \le t$,

$$
\begin{aligned}
\lim_{\alpha\to\infty} I_1 &= \int_s^t \frac{1}{|\mathcal{I}_r|} \sum_{(i,k)\in\mathcal{I}_r} \left(C_{ik}^{q,1}(r) + C_{ik}^{q,2}(r) + C_{ik}^{h,1}\, \pi_r^\top h + \frac{1}{2} C_{ik}^{h,3} \right) \mathrm{d}r \\
&\le -\int_s^t \frac{4}{|\mathcal{I}_r|} \sum_{(i,k)\in\mathcal{I}_r} \sqrt{q_{ik}\, q_{ki}}\, \sinh\left(\frac{\Delta_{ik}(r)}{2} \right) \mathrm{d}r + \int_s^t \frac{1}{|\mathcal{I}_r|} \sum_{(i,k)\in\mathcal{I}_r} \left(\frac{(\delta Q^\top \tilde{\pi}_r)^k}{\tilde{\pi}_r^k} - \frac{(\delta Q^\top \tilde{\pi}_r)^i}{\tilde{\pi}_r^i} \right) \mathrm{d}r \\
&\quad + \max_j |h^j| \int_s^t \frac{1}{|\mathcal{I}_r|} \sum_{(i,k)\in\mathcal{I}_r} \left(h^i - \tilde{h}^i - h^k + \tilde{h}^k \right) \mathrm{d}r \\
&\quad + \int_s^t \frac{1}{2|\mathcal{I}_r|} \sum_{(i,k)\in\mathcal{I}_r} \left((h^k)^2 - (\tilde{h}^k)^2 - (h^i)^2 + (\tilde{h}^i)^2 \right) \mathrm{d}r \\
&\le -\lambda \int_s^t \Delta_\infty(r)\, \mathrm{d}r + \int_s^t \max_{i,k} \left\{ \mathcal{E}_r^{1,i} - \mathcal{E}_r^{1,k} + \frac{1}{2}\left(\mathcal{E}^{2,k} - \mathcal{E}^{2,i} \right) \right\} \mathrm{d}r \\
&\quad + \max_j |h^j| \max_{i,k} \left\{ \mathcal{E}^{3,i} - \mathcal{E}^{3,k} \right\}(t-s),
\end{aligned}
$$

其中，我们按照定理 2.2 的证明对 $C^{q}{}_{,ik}{}^{1}(r)$ 进行了有界估计，并注意到当 $x \geq 0$ 时，$2\sinh(x/2) \geq x$。

同样地，对于随机积分，我们可以利用 I_2 控制收敛交换极限与积分的顺序，从而得到

$$\lim_{\alpha\to\infty} I_2 = \int_s^t \frac{1}{|\mathcal{I}_r|} \sum_{(i,k)\in\mathcal{I}_r} (h^i - \tilde{h}^i - h^k + \tilde{h}^k)\,\mathrm{d}B_r.$$

最后，回顾 (39)，并注意到过程 $\{\Delta_{ik}\}$ 是附录 A 中所考虑的形如 (55) 的连续半鞅，命题 A.6 适用，因此有

$$\lim_{\alpha\to\infty} I_3 \leq \frac{1}{2}\sum_{(i,k)} \sum_{(j,l)\neq(i,k)} \Big(L_t^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot)) - L_s^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot)) \Big) \quad \text{a.s.},$$

其中，$L_t^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot))$ 表示差分过程 $(\Delta_{ik}(r) - \Delta_{jl}(r))_{r\geq 0}$ 在时间 t 于 0 处的局部时。

综合以上所有估计可得，对于所有 $s \leq t$,

$$\begin{aligned}
\Delta_\infty(t) \leq \Delta_\infty(s) &- \lambda \int_s^t \Delta_\infty(r)\,\mathrm{d}r + \int_s^t \max_{i,k} \Big\{ \mathcal{E}_r^{1,i} - \mathcal{E}_r^{1,k} + \frac{1}{2}\big(\mathcal{E}^{2,k} - \mathcal{E}^{2,i}\big) \Big\}\,\mathrm{d}r \\
&+ \max_j |h^j| \max_{i,k} \big\{ \mathcal{E}^{3,i} - \mathcal{E}^{3,k} \big\}(t-s) + \int_s^t \frac{1}{|\mathcal{I}_r|} \sum_{(i,k)\in\mathcal{I}_r} (h^i - \tilde{h}^i - h^k + \tilde{h}^k)\,\mathrm{d}B_r \\
&+ \frac{1}{4}\sum_{(i,k)} \sum_{(j,l)\neq(i,k)} \Big(L_t^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot)) - L_s^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot)) \Big).
\end{aligned}$$

对两边关于参考测度 $\mathbb{P}$ 取期望，由于被积函数有界，随机积分项消失，因此有

$$\begin{aligned}
\mathrm{d}\mathbf{E}\left[\Delta_\infty(t)\right] \leq -\lambda \mathbf{E}\left[\Delta_\infty(t)\right]\,\mathrm{d}t &+ \mathbf{E}\left[\max_{i,k} \Big\{ \mathcal{E}_t^{1,i} - \mathcal{E}_t^{1,k} + \frac{1}{2}\big(\mathcal{E}^{2,k} - \mathcal{E}^{2,i}\big) \Big\} \right] \mathrm{d}t \\
&+ \max_j |h^j| \max_{i,k} \big\{ \mathcal{E}^{3,i} - \mathcal{E}^{3,k} \big\}\,\mathrm{d}t + \frac{1}{4}\sum_{(i,k)} \sum_{(j,l)\neq(i,k)} \mathrm{d}\mathbf{E}\left[L_t^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot)) \right],
\end{aligned}$$

其中，左端和右端最后一项应理解为勒贝格–斯蒂尔切斯测度。利用链式法则求出 $e^{\lambda t}\mathbf{E}\left[\Delta_\infty(t)\right]$，的动力学方程，并对所得微分不等式两边积分可得，对于所有 $s \leq t < \infty$,

$$\begin{aligned}
\mathbf{E}\left[\Delta_\infty(t)\right] \leq \mathbf{E}\left[\Delta_\infty(s)\right] e^{-\lambda(t-s)} &+ \int_s^t e^{-\lambda(t-r)} \mathbf{E}\left[\max_{i,k} \Big\{ \mathcal{E}_r^{1,i} - \mathcal{E}_r^{1,k} + \frac{1}{2}\big(\mathcal{E}^{2,k} - \mathcal{E}^{2,i}\big) \Big\} \right] \mathrm{d}r \\
&+ \max_j |h^j| \max_{i,k} \big\{ \mathcal{E}^{3,i} - \mathcal{E}^{3,k} \big\} \int_s^t e^{-\lambda(t-r)}\,\mathrm{d}r \\
&+ \frac{1}{4}\sum_{(i,k)} \sum_{(j,l)\neq(i,k)} \int_s^t e^{-\lambda(t-r)}\,\mathrm{d}\mathbf{E}\left[L_r^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot)) \right].
\end{aligned}$$

对误差项进行界定时，由于局部时具有有限变差（其期望也如此），我们可以对斯蒂尔切斯积分两次应用分部积分，并对.....使用富比尼–托内利定理

右端的最后一项，得到

$$\begin{aligned}\mathbf{E}\left[\Delta_\infty(t)\right] \le \mathbf{E}\left[\Delta_\infty(s)\right] e^{-\lambda(t-s)} &+ 2\max_{i,k}|\tilde{q}_{ik}-q_{ik}|\int_s^t e^{-\lambda(t-r)}\mathbf{E}\left[\frac{1}{\min_j \tilde{\pi}_r^j}\right]\mathrm{d}r \\ &+ \Big(2\max_j|h^j|\max_i|h^i-\tilde{h}^i| + \max_i|(h^i)^2-(\tilde{h}^i)^2|\Big)\int_s^t e^{-\lambda(t-r)}\,\mathrm{d}r \\ &+ \frac{1}{4}\sum_{(i,k)}\sum_{(j,l)\neq(i,k)}\mathbf{E}\bigg[\int_s^t e^{-\lambda(t-r)}\,\mathrm{d}L_r^0(\Delta_{ik}(\cdot)-\Delta_{jl}(\cdot))\bigg], \end{aligned} \tag{43}$$

对所有 $s \le t < \infty$ 均成立，这正是我们要证明的结论。

现在我们继续讨论定理的第二部分。首先，类似于 [14, 引理 3.6], ，可以得到 $\mathbf{E}[(\min_j \tilde{\pi}_t^j)^{-1}]$ 的一个显式界，它关于参数 $(\tilde{\ } \tilde{\ } \ ^\circ, Q, h)$ 连续依赖于 $\nu \in \mathcal{S}^n$。特别地，我们有

$$\mathbf{E}\left[\frac{1}{\min_j \tilde{\pi}_t^j}\right] \le \max_j\left\{\frac{1}{\nu^j}\exp\Big\{-\tilde{q}_{jj}t + \max_k(\tilde{h}^j-\tilde{h}^k)^2 t\Big\}\right\}$$

对所有 $t < \infty$ 均成立；当我们令 $h \to h$ 趋于˜ ˜ ˜极限时，(43) 式右端的第一个积分在 h 周围的某个紧集内对 h 一致有界。第二项显然在 $h \to h$ 时趋于零。

接下来，我们讨论关于局部时的积分。令 $\tilde{\pi}_t(v)$ 表示以 $v \in \mathbb{R}^{n+1}$ 替代 $\tilde{h}$ 后，方程 (38) 的唯一解。注意，对于 $(i,k),(j,l) \in$ **自然数集** $\times$ **自然数集** 以及 $(i,k) \neq (j,l)$，各过程 $\Delta_{ik}-\Delta_{jl}$ 的漂移项由 $b^{ik,\ jl}(\tilde{h})$ 给出，其中

$$\begin{aligned} b_t^{ik,jl}(v) := &-\sum_{r=0}^n\left(q_{rk}\frac{\pi_t^r}{\pi_t^k} - \tilde{q}_{rk}\frac{\tilde{\pi}_t^r(v)}{\tilde{\pi}_t^k(v)}\right) + \sum_{r=0}^n\left(q_{ri}\frac{\pi_t^r}{\pi_t^i} - \tilde{q}_{ri}\frac{\tilde{\pi}_t^r(v)}{\tilde{\pi}_t^i(v)}\right) \\ &+ \sum_{r=0}^n\left(q_{rl}\frac{\pi_t^r}{\pi_t^l} - \tilde{q}_{rl}\frac{\tilde{\pi}_t^r(v)}{\tilde{\pi}_t^l(v)}\right) - \sum_{r=0}^n\left(q_{rj}\frac{\pi_t^r}{\pi_t^j} - \tilde{q}_{rj}\frac{\tilde{\pi}_t^r(v)}{\tilde{\pi}_t^j(v)}\right) \\ &+ (h^i - v^i - h^k + v^k - h^j + v^j + h^l - v^l)\pi_t^\top h \\ &+ \frac{1}{2}\Big((h^k)^2 - (v^k)^2 - (h^i)^2 + (v^i)^2 - (h^l)^2 + (v^l)^2 + (h^j)^2 - (v^j)^2\Big). \end{aligned} \tag{44}$$

考虑 $b^{ik,jl}(\tilde{h})$ 和 $b^{ik,jl}(h)$ 在 $[0,t]$ 上的差。利用 $\tilde{\pi}(h)$ 和 $\tilde{\pi}(\tilde{h})$ 均位于单纯形中这一事实，可得

$$\begin{aligned}&\mathbf{E}\bigg[\sup_{s\le t}\left|b_s^{ik,jl}(h) - b_s^{ik,jl}(\tilde{h})\right|\bigg]^2 \\ &\quad\le \sum_{u\in\{i,k,j,l\}}\sum_{r\neq u}\tilde{q}_{ru}\mathbf{E}\left[\sup_{s\le t}\left(\frac{1}{\tilde{\pi}_s^u(h)\tilde{\pi}_s^u(\tilde{h})}\right)^2\right]\mathbf{E}\left[\sup_{s\le t}\Big(\left|\tilde{\pi}_s^u(\tilde{h}) - \tilde{\pi}_s^u(h)\right| + \left|\tilde{\pi}_s^r(\tilde{h}) - \tilde{\pi}_s^r(h)\right|\Big)^2\right].\end{aligned}$$

对于所有 $u \in$ **自然数集**，引理 3.6] 的一个简单推广 [14, 可得，第一个期望关于 h h 在 h附近的一个紧集内一致有界。由于 $\tilde{\pi}_t$ 位于单纯形中，随机微分方程 (38) 具有利普希茨系数，因此我们可以应用标准的稳定性论证（例如˜ [15, 定理 16.4.3]），从而得知第二个期望当 $h \to h$ 时趋于 0。因此，我们在 $[0,t]$ 上有依概率一致收敛 $b^{ik,\ jl}(\tilde{h}) \to b^{ik,\ jl}(h)$，当 $\tilde{h} \to h$时。

˜ ˜现在，固定一个任意序列 $\{h_n\}_{n\in\mathbb{N}}$，使得 $h_n \to h$。根据上述结果，我们可以选取一个子序列 $\{\tilde{h}_{n_r}\}_{r\in\mathbb{N}}$，使得 $b^{ik}{}_{,s}{}^{jl}(\tilde{h}_{n_r})$ 在 $[0,t]$ 上几乎必然一致收敛于 $b^{ik}{}_{,s}{}^{jl}(h)$。从现在起，当我们写 $h \to h$ 时，均指沿该子序列得到的极限。记 ˜$(\Delta_{ik}-\Delta_{jl})^\star_t$ 为当 $h \to h$ 时 $(\Delta_{ik}-\Delta_{jl})_t$ 的极限。利用这一一致收敛性，我们可以得到，几乎必然地，对所有 $s\in[0,t]$,

$$\begin{aligned}(\Delta_{ik}-\Delta_{jl})^\star_s &= (\Delta_{ik}-\Delta_{jl})_0 + \lim_{\tilde{h}\to h}\int_0^s b_r^{ik,jl}(\tilde{h})\,\mathrm{d}r \\ &\quad + \lim_{\tilde{h}\to h}(h^i-\tilde{h}^i-h^k+\tilde{h}^k-h^j+\tilde{h}^j+h^l-\tilde{h}^l)B_s \\ &= (\Delta_{ik}-\Delta_{jl})_0 + \int_0^s b_r^{ik,jl}(h)\,\mathrm{d}r,\end{aligned}$$

并且 $(\Delta_{ik}-\Delta_{jl})^\star_s$ 是绝对连续的，其导数为 $b^{ik}{}_{,\ s}{}^{jl}(h)$。

现在，根据田中公式，我们有

$$L_t^0(\Delta_{ik}-\Delta_{jl}) = |(\Delta_{ik}-\Delta_{jl})_t| - |(\Delta_{ik}-\Delta_{jl})_0| + \int_0^t \operatorname{sign}((\Delta_{ik}-\Delta_{jl})_s)\,\mathrm{d}(\Delta_{ik}-\Delta_{jl})_s,$$

˜，其中约定 $\operatorname{sign}(0) = -1$。令 $h\to h$ 趋于无穷，对上述等式两边取极限，随机积分消失，并对涉及 $b^{ik}{}_{,\ jl}$的积分应用控制收敛，我们得到

$$\lim_{\tilde{h}\to h} L_t^0(\Delta_{ik}-\Delta_{jl}) = |(\Delta_{ik}-\Delta_{jl})^\star_t| - |(\Delta_{ik}-\Delta_{jl})_0| + \int_0^t \lim_{\tilde{h}\to h}\operatorname{sign}((\Delta_{ik}-\Delta_{jl})_s) b_s^{ik,jl}(\tilde{h})\,\mathrm{d}s. \quad (45)$$

考虑积分内的极限。注意，对于所有满足 $b^{ik}{}_{,\ s}{}^{jl}(h) \neq 0$ 且 $(\Delta_{ik}-\Delta_{jl})^\star_s \neq 0$的 $s\le t$，我们几乎处处有

$$\lim_{\tilde{h}\to h}\operatorname{sign}((\Delta_{ik}-\Delta_{jl})_s) b_s^{ik,jl}(\tilde{h}) = \operatorname{sign}((\Delta_{ik}-\Delta_{jl})^\star_s) b_s^{ik,jl}(h).$$

现在考虑满足 $b^{ik}{}_{,\ s}{}^{jl}(h) = 0$的 $s\le t$。则有

$$\lim_{\tilde{h}\to h}\operatorname{sign}((\Delta_{ik}-\Delta_{jl})_s) b_s^{ik,jl}(\tilde{h}) = 0 = \operatorname{sign}((\Delta_{ik}-\Delta_{jl})^\star_s) b_s^{ik,jl}(h),$$

对于所有这样的 s均成立。最后，考虑满足 $b^{ik}{}_{,\ s}{}^{jl}(h) \neq 0$ 但 $(\Delta_{ik}-\Delta_{jl})^\star_s = 0$的时间 $s\le t$。那么，当 $\tilde{h}$ 趋于 h时，我们可能有 $\operatorname{sign}((\Delta_{ik}-\Delta_{jl})_s) b^{ik}{}_{,\ s}{}^{jl}(\tilde{h}) \nrightarrow \operatorname{sign}((\Delta_{ik}-\Delta_{jl})^\star_s) b^{ik}{}_{,\ s}{}^{jl}(h)$。然而，根据引理 3.4，该集合

$$\left\{ s : (\Delta_{ik}-\Delta_{jl})^\star_s = 0, \frac{\mathrm{d}}{\mathrm{d}s}(\Delta_{ik}-\Delta_{jl})^\star_s = b_s^{ik,jl}(h) \neq 0 \right\}$$

具有勒贝格测度零。因此，最后我们可以得出结论：几乎处处成立。

$$\lim_{\tilde{h}\to h}\operatorname{sign}((\Delta_{ik}-\Delta_{jl})_s) b_s^{ik,jl} = \operatorname{sign}((\Delta_{ik}-\Delta_{jl})^\star_s) b_s^{ik,jl}(h), \quad \text{for a.a. } s\le t,$$

因此，根据 $(\Delta_{ik}-\Delta_{jl})^\star_t$ 的绝对连续性，(45) 的右端为 0。由此我们证明了 $L_t^0(\Delta_{ik}-\Delta_{jl}) \to 0$ as $\tilde{h} \to h$ 沿子序列 ˜$\{h_{n_r}\}_{r\in\mathbb{N}}$ 几乎必然收敛，这意味着它沿同一子序列依概率收敛。在

˜另一方面，原序列 $\{h_n\}_{n\in\mathbb{N}}$ 是任意的，因此我们可以沿任意序列重复上述论证，并总能找到一个子序列，使得 $L_t^0(\Delta_{ik}-\Delta_{jl})$ 依概率收敛到 0。由此可知，当 $\tilde{h}\to h$ 时， $L_t^0(\Delta_{ik}-\Delta_{jl})$ 依概率趋于 0。根据田中公式，我们也可以像推导依概率一致收敛那样，验证$\mathbf{E}[L_t^0(\Delta_{ik}-\Delta_{jl})^2]$ 关于 $\tilde{h}$ 一致有界，其中 $\tilde{h}$ 位于 h 周围的一个紧集内，因此由维塔利收敛定理可得 $\mathbf{E}[L_t^0(\Delta_{ik}-\Delta_{jl})]\to 0$ as $\tilde{h}\to h$。这就得到了该定理。 □

命题 *4.2* 的证明。 我们重点讨论局部时间项，因为通过与 [14,命题 3.7], 类似的论证，立即可得 $\sup_{t>0}\mathbf{E}[(\min_k\tilde{\pi}_t^k)^{-1}]<\infty$。 令 $(i,k),(j,l)\in$**自然数集** $\times$ **自然数集**，其中 $(i,k)\neq(j,l)$。回忆由田中公式，我们可以将 X_t 在 0 处的局部时写为：$=(\Delta_{ik}-\Delta_{jl})_t$ 为

$$L_t^0(X)=|X_t|-|X_0|+\int_0^t \operatorname{sign}(X_s)\,\mathrm{d}X_s.$$

于是有

$$\begin{aligned}\mathbf{E}\Big[\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}L_s^0(X_\cdot)\Big]&=\mathbf{E}\Big[\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}|X_s|\Big]+\mathbf{E}\Big[\int_0^t e^{-\lambda(t-s)}\operatorname{sign}(X_s)\,\mathrm{d}X_s\Big]\\&=\mathbf{E}\Big[\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}|X_s|\Big]+\mathbf{E}\Big[\int_0^t e^{-\lambda(t-s)}\operatorname{sign}(X_s)b_s^{ik,jl}(\tilde{h})\,\mathrm{d}s\Big]\\&\le\mathbf{E}\Big[\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}|X_s|\Big]+\sup_{s\le t}\mathbf{E}\Big[\big|b_s^{ik,jl}(\tilde{h})\big|\Big]\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}s,\end{aligned}\tag{46}$$

其中 $b^{ik}{}_{,s}{}^{jl}(\tilde{h})$ 是 X_t 的漂移项，定义见式 **(44)**。由于

$$|b_t^{ik,jl}|\le K_q\frac{1}{\min_i\pi_t^i}+K_{\tilde{q}}\frac{1}{\min_i\tilde{\pi}_t^i}+K_h,$$

˜ ˜其中 K_q、 $K_{\tilde{q}}$ 和 K_h 是仅依赖于 Q、 Q、 h 和 h 的常数，因此当我们对所有 $t>0$ 取上确界时，**(46)** 中的第二项是有限的。对于第一项，两次分部积分可得

$$\mathbf{E}\Big[\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}|X_s|\Big]=\mathbf{E}\big[|X_t|\big]-|X_0|e^{-\lambda t}-\lambda\mathbf{E}\Big[\int_0^t|X_s|e^{-\lambda(t-s)}\,\mathrm{d}s\Big],$$

并且由于 $X_t=\Delta_{ik}-\Delta_{jl}$, $\Delta_{ik}=\log\pi_i/\pi_k-\log\tilde{\pi}_i/\tilde{\pi}_k$and $|\log(x)|\le 1/x$，我们可以再次将右端界定为 $\mathbf{E}\big[1/\min_i\pi_t^i\big]$ 和 $\mathbf{E}\big[1/\min_i\tilde{\pi}_t^i\big]$ 的倍数；当我们对 $t>0$ 取上确界时，这些量仍然有限。 □

4.2 近似滤波器的误差界

上一小节介绍的方法可以得到一个更一般的结果。我们可以完全按照之前的步骤，计算一般近似滤波器的误差，而不只是修改了 Q 和 h （若允许 $\sigma\neq 1$，还包括 σ）的滤波器。本节的讨论将给出定理 2.3 和定理 2.4 的证明。

考虑由 **(6)** 给出的一般近似滤波模型，即

$$\mathrm{d}\tilde{\pi}_t=\tilde{f}_t\,\mathrm{d}t+\tilde{g}_t\,\mathrm{d}Y_t,\quad \tilde{\pi}_0=\nu,\tag{47}$$

其中， $\tilde{f}_t,\tilde{g}_t$ 是取值于 $\mathbb{R}^{n+1}$ 的可预测过程。我们还需要以下假设：

。

˜**(A4).** 以概率 *1*， $\tilde{\pi}_t \in \mathcal{S}^n$ 对所有 $t < \infty$均成立。此外，f_t 和 $\tilde{g}_t$ 局部有界，并满足可积性条件

$$\mathbf{E}\left[\int_0^t \max_i \frac{|\tilde{f}_s^i|}{\pi_s^i}\,\mathrm{d}s + \left(\int_0^t \max_i \left(\frac{\tilde{g}_s^i}{\pi_s^i}\right)^2 \mathrm{d}s\right)^{1/2}\right] < \infty$$

对所有 $t < \infty$。

请注意，根据引理 3.1 和（对） [14, 引理 3.6]（的简单推广），Wonham滤波器随机微分方程 (4) 或由 (38) 给出的模型参数设定错误的 Wonham滤波器，立即满足假设 (A4)。

我们首先证明一个中间结果。

命题4.5 （近似滤波器的希尔伯特误差动力学）。 设 π_t 为 *(4)* 的解， $\tilde{\pi}_t$ 为 *(47)* 的解。假设对于所有 $i \neq j$， μ^i、 $\nu^i > 0 \quad \forall i$ 和 $q_{ij} > 0$ 成立。在假设 *(A4)* 下，对于所有 $s \leq t < \infty$，我们有

$$\begin{aligned}\mathcal{H}(\pi_t, \tilde{\pi}_t) \leq \mathcal{H}(\pi_s, \tilde{\pi}_s) &- 2\int_s^t \kappa_r \sinh\left(\frac{\mathcal{H}(\pi_r, \tilde{\pi}_r)}{2}\right)\mathrm{d}r + \int_s^t \max_{i,k}\left\{\mathcal{E}_r^{1,i} - \mathcal{E}_r^{1,k} + \frac{1}{2}\left(\mathcal{E}_r^{2,k} - \mathcal{E}_r^{2,i}\right)\right\}\mathrm{d}r \\ &+ \max_j |h^j| \int_s^t \max_{i,k}\left\{\mathcal{E}_r^{3,i} - \mathcal{E}_r^{3,k}\right\}\mathrm{d}r + \int_s^t \frac{1}{|\mathcal{I}_r|}\sum_{(i,k)\in\mathcal{I}_r}\left(\mathcal{E}_r^{3,i} - \mathcal{E}_r^{3,k}\right)\mathrm{d}B_r \\ &+ \frac{1}{4}\sum_{(i,k)}\sum_{(j,l)\neq(i,k)}\int_s^t \mathrm{d}L_r^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot)),\end{aligned} \tag{48}$$

其中 $B_t = Y_t - \int_0^t \pi_s^\top h\mathrm{d}s$ 是创新过程。对于 $j \in$ **自然数集**，误差项由下式给出

$$\mathcal{E}_t^{1,j} = \left(\sum_{m=0}^n q_{mj}\frac{\tilde{\pi}_t^m}{\tilde{\pi}_t^j}\right) - \frac{\tilde{f}_t^j}{\tilde{\pi}_t^j}, \quad \mathcal{E}_t^{2,j} = (h^j)^2 - \frac{(\tilde{g}_t^j)^2}{(\tilde{\pi}_t^j)^2}, \quad \mathcal{E}_t^{3,j} = h^j - \frac{\tilde{g}_t^j}{\tilde{\pi}_t^j},$$

以及过程 $(\Delta_{ik}(t))_{t\geq 0}$，对于 $(i,k) \in N \times N$，定义为 $\Delta_{ik}(t) = \log\frac{\pi_t^i}{\pi_t^k} - \log\frac{\tilde{\pi}_t^i}{\tilde{\pi}_t^k}$。集合 $\mathcal{I}_t = \{(i,k) : \Delta_{ik}(t) = \mathcal{H}(\pi_t, \tilde{\pi}_t)\}$ 是这些过程对所有 $t < \infty$ 的最大值点，并且 $L_t^0(\Delta_{ik}(\cdot) - \Delta_{jl}(\cdot))$ 表示差分过程 $(\Delta_{ik} - \Delta_{jl})$ 在 *0* 处的局部时，对于所有 (i,k)、$(j,l) \in N^{\times}N$。衰减系数 $\kappa_t > 0$ 可以取为以下任意值。

$$\kappa_t = \begin{cases} \lambda, \\ \tilde{\lambda}^\star\left(t, \tanh\left(\frac{\mathcal{H}(\pi_t,\tilde{\pi}_t)}{4}\right)\right), \\ \lambda^\star\left(t, \tanh\left(\frac{\mathcal{H}(\pi_t,\tilde{\pi}_t)}{4}\right)\right), \end{cases} \tag{49}$$

˜其中 λ 是定理 *2.2* 中的确定性速率，而 $\lambda^\star$ 和 $\lambda^\star$ 分别是命题 *3.10* 和推论 *3.10.1* 中定义的函数（分别见(28) 和(37) ）。

证明。 假设 (A4) 允许我们通过定义常用变换 $\theta_k^i : (0,1)^{\times 2} \to \mathbb{R}$，将分析从单纯形转移到 $\mathbb{R}^n$。$\tilde{\theta}_k^i = \log\frac{\tilde{\pi}_t^i}{\tilde{\pi}_t^k}$ 的动力学由下式给出

$$\mathrm{d}\log\frac{\tilde{\pi}^i}{\tilde{\pi}^k}(t) = \frac{1}{\tilde{\pi}_t^i}\left(\tilde{f}_t^i\,\mathrm{d}t + \tilde{g}_t^i\,\mathrm{d}Y_t\right) - \frac{1}{\tilde{\pi}_t^k}\left(\tilde{f}_t^k\,\mathrm{d}t + \tilde{g}_t^k\,\mathrm{d}Y_t\right) + \frac{1}{2}\left(\left(\frac{\tilde{g}_t^k}{\tilde{\pi}_t^k}\right)^2 - \left(\frac{\tilde{g}_t^i}{\tilde{\pi}_t^i}\right)^2\right)\mathrm{d}t,$$

因此，令 $\theta_k^i(t) = \log \frac{\pi_t^i}{\pi_t^k}$ and $\Delta_{ik}(t) = \theta_k^i(t) - \tilde{\theta}_k^i(t)$，定义创新过程$B_t = Y_t - \int_0^t \pi_s^\top h \mathrm{d}s$，并回顾式 (17) 中 $\theta_k^i(t)$ 的动力学，我们得到

$$\begin{aligned}\mathrm{d}\Delta_{ik}(t) = &\left[\frac{1}{\pi_t^i}\Big(\sum_{j=0}^n q_{ji}\pi_t^j\Big) - \frac{\tilde{f}_t^i}{\tilde{\pi}_t^i}\right]\mathrm{d}t + \left[\frac{\tilde{f}_t^k}{\tilde{\pi}_t^k} - \frac{1}{\pi_t^k}\Big(\sum_{j=0}^n q_{jk}\pi_t^j\Big)\right]\mathrm{d}t \\ &+ \frac{1}{2}\left((h^k)^2 - (h^i)^2 + \left(\frac{\tilde{g}_t^i}{\tilde{\pi}_t^i}\right)^2 - \left(\frac{\tilde{g}_t^k}{\tilde{\pi}_t^k}\right)^2\right)\mathrm{d}t + \Big(h^i - h^k - \frac{\tilde{g}_t^i}{\tilde{\pi}_t^i} + \frac{\tilde{g}_t^k}{\tilde{\pi}_t^k}\Big)(\,\mathrm{d}B_t + \pi_t^\top h\,\mathrm{d}t).\end{aligned} \tag{50}$$

这个方程乍看之下可能有些棘手，但我们可以完全按照处理模型设定错误的 Q 和 h时的方法来处理。适当地添加和减去各项后，我们可以将 (50) 改写为

$$\begin{aligned}\mathrm{d}\Delta_{ik}(t) = &\left(\sum_{\substack{j=0\\ j\neq i}}^n q_{ji}\left(\frac{\pi_t^j}{\pi_t^i} - \frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^i}\right) - \sum_{\substack{j=0\\ j\neq k}}^n q_{jk}\left(\frac{\pi_t^j}{\pi_t^k} - \frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^k}\right)\right)\mathrm{d}t + \big(\mathcal{E}_t^{1,i} - \mathcal{E}_t^{1,k}\big)\,\mathrm{d}t \\ &+ \frac{1}{2}\big(\mathcal{E}_t^{2,k} - \mathcal{E}_t^{2,i}\big)\,\mathrm{d}t + \big(\mathcal{E}_t^{3,i} - \mathcal{E}_t^{3,k}\big)(\,\mathrm{d}B_t + \pi_t^\top h\,\mathrm{d}t),\end{aligned} \tag{51}$$

其中，对所有 $j \in$**自然数集**，我们定义误差项

$$\mathcal{E}_t^{1,j} = \left(\sum_{m=0}^n q_{mj}\frac{\tilde{\pi}_t^m}{\tilde{\pi}_t^j}\right) - \frac{\tilde{f}_t^j}{\tilde{\pi}_t^j}, \qquad \mathcal{E}_t^{2,j} = (h^j)^2 - \left(\frac{\tilde{g}_t^j}{\tilde{\pi}_t^j}\right)^2, \qquad \mathcal{E}_t^{3,j} = h^j - \frac{\tilde{g}_t^j}{\tilde{\pi}_t^j}. \tag{52}$$

现在，我们按照定理 4.1 的证明继续进行：令 $\mathbf{\Delta}_t = \{\Delta_{ik}(t)\}_{i\in\mathbf{N},\ k\in\mathbf{N}}$ 为演化方程由 (51) 给出的过程族，定义过程 $LSE_\alpha(\Delta_\cdot)(t)$及其动力学，最后取 $\alpha\to\infty$，从而得到误差估计。

再次令 $T_{ik}(t) = \frac{\pi_t^i}{\pi_t^k} - \frac{\tilde{\pi}_t^i}{\tilde{\pi}_t^k}$对所有 $(i,k) \in$ **自然数集** $\times$ **自然数集**成立，则对所有 $s \le t < \infty$,

$$\begin{aligned}LSE_\alpha(\mathbf{\Delta}_\cdot)(t) = &LSE_\alpha(\mathbf{\Delta}_\cdot)(s) + \int_s^t S_\alpha^{arg}\Big(\mathbf{\Delta}_\cdot, \sum_{j\neq i}^n q_{ji}T_{ji}(\cdot) - \sum_{j\neq k}^n q_{jk}T_{jk}(\cdot)\Big)(r)\,\mathrm{d}r \\ &+ \int_s^t S_\alpha^{arg}\big(\mathbf{\Delta}_\cdot, \mathcal{E}_\cdot^{1,i} - \mathcal{E}_\cdot^{1,k}\big)(r)\,\mathrm{d}r + \frac{1}{2}\int_s^t S_\alpha^{arg}\big(\mathbf{\Delta}_\cdot, \mathcal{E}_\cdot^{2,k} - \mathcal{E}_\cdot^{2,i}\big)(r)\,\mathrm{d}r \\ &+ \int_s^t S_\alpha^{arg}\big(\mathbf{\Delta}_\cdot, \mathcal{E}_\cdot^{3,i} - \mathcal{E}_\cdot^{3,k}\big)(r)(\,\mathrm{d}B_r + \pi_s^\top h\,\mathrm{d}r) \\ &+ \frac{1}{2}\int_s^t \sum_{(j,l)}\sum_{(u,v)\neq(j,l)} \frac{\alpha e^{\alpha(\Delta_{jl}(r)+\Delta_{uv}(r))}}{(\sum_{(i,k)} e^{\alpha\Delta_{ik}(r)})^2}\left(\big(\mathcal{E}_r^{3,j} - \mathcal{E}_r^{3,l}\big)^2 - \big(\mathcal{E}_r^{3,j} - \mathcal{E}_r^{3,l}\big)\big(\mathcal{E}_r^{3,u} - \mathcal{E}_r^{3,v}\big)\right)\mathrm{d}r.\end{aligned}$$

注意，由引理 3.1 和假设 (A4) 可知，右端前四个被积函数具有足够的正则性，因此在令 $\alpha\to\infty$ 时，可以对勒贝格积分或随机积分应用控制收敛定理。对于最后一项，我们再次使用命题 A.6；假设 (A4) 保证了该命题适用，从而将积分界定为差分过程的局部

时的大小 $\Delta_{ik}-\Delta_{jl}$，同时令 $\alpha\to\infty$。由此可得，对所有 $s\le t<\infty$,

$$\begin{aligned}\Delta_\infty(t)\le\Delta_\infty(s)&+\int_s^t\frac{1}{|\mathcal{I}_r|}\sum_{(i,k)\in\mathcal{I}_r}\left(\sum_{j\neq i}q_{ji}T_{ji}(r)-\sum_{j\neq k}q_{jk}T_{jk}(r)\right)\mathrm{d}r\\&+\int_s^t\max_{i,k}\left\{\mathcal{E}_r^{1,i}-\mathcal{E}_r^{1,k}+\frac{1}{2}\big(\mathcal{E}_r^{2,k}-\mathcal{E}_r^{2,i}\big)\right\}\mathrm{d}r+\max_j|h^j|\int_s^t\max_{i,k}\left\{\mathcal{E}_r^{3,i}-\mathcal{E}_r^{3,k}\right\}\mathrm{d}r\\&+\int_s^t\frac{1}{|\mathcal{I}_r|}\sum_{(i,k)\in\mathcal{I}_r}\big(\mathcal{E}_r^{3,i}-\mathcal{E}_r^{3,k}\big)\,\mathrm{d}B_r+\frac{1}{4}\sum_{(i,k)}\sum_{(j,l)\neq(i,k)}\int_s^t\mathrm{d}L_r^0(\Delta_{ik}(\cdot)-\Delta_{jl}(\cdot)),\quad(53)\end{aligned}$$

其中，对于所有 $r\in[s,t]$，我们定义 $\mathcal{I}_r=\{(i,\ k):\ \Delta_{ik}(r)=\Delta_\infty(r)\}\subset$ **自然数集** $\times$ **自然数集** 为 $\mathbf{\Delta}_r$ 的最大值点，并令 $|\mathcal{I}_r|$ 表示其大小。

最后，考虑右端第一个积分中的漂移项。根据引理 3.7，对于所有 $r\in[s,t]$，我们有

$$\Delta_{ik}(r)=\Delta_\infty(r)=\mathcal{H}(\pi_r,\tilde{\pi}_r)=\log\max_j\frac{\pi_r^j}{\tilde{\pi}_r^j}-\log\min_j\frac{\pi_r^j}{\tilde{\pi}_r^j}=:\log M_r-\log\frac{1}{m_r},\quad\forall(i,k)\in\mathcal{I}_r,$$

其中，$M_r\ge1$ 和 $1/m_r\le1$ 分别是 π_r 与 $\tilde{\pi}_r$各分量之比的逐点最大值和最小值。根据引理 3.8，对于所有 $r\in[s,t]$，以及所有 $(i,\ k)\in\mathcal{I}_r$、$T_{ji}(r)\le0$ 和 $T_{jk}(r)\ge0$，对所有 $j\in$ **自然数集**，均有此结论。因此，右端的第一个积分为负，特别地

$$\int_s^t\frac{1}{|\mathcal{I}_r|}\sum_{(i,k)\in\mathcal{I}_r}\left(\sum_{j\neq i}^n q_{ji}T_{ji}(r)-\sum_{j\neq k}^n q_{jk}T_{jk}(r)\right)\mathrm{d}r\le-\int_s^t\min_{(i,k)\in\mathcal{I}_r}\left(\sum_{j\neq k}^n q_{jk}T_{jk}(r)-\sum_{j\neq i}^n q_{ji}T_{ji}(r)\right)\mathrm{d}r.$$

现在，我们可以像定理 2.2、命题 3.10 或推论 3.10.1 的证明中那样，通过代数计算将被积函数最小化，从而得到

$$\int_s^t\frac{1}{|\mathcal{I}_r|}\sum_{(i,k)\in\mathcal{I}_r}\left(\sum_{j\neq i}^n q_{ji}T_{ji}(r)-\sum_{j\neq k}^n q_{jk}T_{jk}(r)\right)\mathrm{d}r\le-2\int_s^t\kappa_r\sinh\left(\frac{\Delta_\infty(r)}{2}\right)\mathrm{d}r,$$

其中，衰减率 κ_r 可以取自定理 2.2 中的任意系数 λ、命题 3.10 中公式 (28) 的 $\lambda^\star(t,$双曲正切函数$(\Delta_\infty(t)/4)$，或推论 3.10.1 中公式 (37) 的 $\lambda^\star(t,$双曲正切函数$(\Delta_\infty(t)/4)$。

□

由上述命题可直接推出定理 2.3 和定理 2.4。

定理 *2.3* 的证明。从（48）出发。对于所有 $t<\infty$，我们将 κ_t 从下方估计为确定性速率 $\lambda=2\min_{i\neq k}\sqrt{q_{ik}q_{ki}}$。此外，注意到当 $x\ge0$ 时，$2\sinh(x/2)\ge x$。将这两个界代入（48）右端的第一个积分中。我们取期望，并注意到随机积分为零，因为它是一个鞅（根据假设 (A4)，被积函数是局部 L^2-可积的）。对标准 Gr¨onwall 论证作适当修改，以处理勒贝格–斯蒂尔切斯测度（如定理 4.1 的证明中所示），即可完成证明。

□

定理 *2.4* 的证明。 如果 $\mathcal{E}^{3,\,i}_{t}=0$ 对所有 $i\in$ **自然数集** 以及 $t<\infty$均成立，那么 $\mathcal{E}^{2,\,i}_{t}=0$ 也成立。于是（51）化为

$$\mathrm{d}\Delta_{ik}(t)=\Bigg(\sum_{\substack{j=0\\ j\neq i}}^{n}q_{ji}\Big(\frac{\pi_t^j}{\pi_t^i}-\frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^i}\Big)-\sum_{\substack{j=0\\ j\neq k}}^{n}q_{jk}\Big(\frac{\pi_t^j}{\pi_t^k}-\frac{\tilde{\pi}_t^j}{\tilde{\pi}_t^k}\Big)\Bigg)\,\mathrm{d}t+\big(\mathcal{E}_t^{1,i}-\mathcal{E}_t^{1,k}\big)\,\mathrm{d}t,$$

对于所有 $(i,k)\in\mathbf{N}\times\mathbf{N}$，因此我们恢复了差分过程 $\Delta_{ik}(t)$ 的 C^1 动力学。 C^1过程不会产生局部时，因此（48）可简化为

$$\Delta_\infty(t)\le\Delta_\infty(s)-2\int_s^t\kappa_r\sinh\Big(\frac{\Delta_\infty(r)}{2}\Big)\,\mathrm{d}r+\int_s^t\max_{i,k}\big\{\mathcal{E}_r^{1,i}-\mathcal{E}_r^{1,k}\big\}\,\mathrm{d}r,$$

对于所有 $s\le t$，其中 $\kappa_t>0$ 是由（49）中的任一速率给出的系数。现在，定理的第二部分很容易得出：首先将 κ_r 从下方界定为一个由以下之一给出的正的（可测）过程 γ_t

$$\gamma_t=\begin{cases}\lambda,\\ \tilde{\lambda}_t^\star\le\tilde{\lambda}^\star\Big(t,\tanh\Big(\frac{\mathcal{H}(\pi_t,\tilde{\pi}_t)}{4}\Big)\Big),\\ \lambda_t^\star\le\lambda^\star\Big(t,\tanh\Big(\frac{\mathcal{H}(\pi_t,\tilde{\pi}_t)}{4}\Big)\Big),\end{cases}$$

˜其中 $\lambda_t^\star$ 和 $\lambda_t^\star$ 与命题 3.10 和推论 3.10.1 中的相同。然后，再次回忆到对于 $x\ge0$，有 2 $\sinh(x/2)\ge x$，常用的 Gr¨onwall 论证可得

$$\Delta_\infty(t)\le\Delta_\infty(0)e^{-\int_0^t\gamma_s\,\mathrm{d}s}+\int_0^t e^{-\int_s^t\gamma_r\,\mathrm{d}r}\max_{i,k}\big\{\mathcal{E}_s^{1,i}-\mathcal{E}_s^{1,k}\big\}\,\mathrm{d}s,$$

˜即为 $\gamma_t=\lambda_t^\star$对应的式（11）。

现在我们来寻找一个更紧的界。考虑过程 $X_t=$ 双曲正切函数 $\big(\Delta_\infty(t)/4\big)$。应用链式法则可得

$$\begin{aligned}\mathrm{d}X_t=\frac{1}{4}\cosh^{-2}\Big(\frac{\Delta_\infty(t)}{4}\Big)\,\mathrm{d}\Delta_\infty(t)&\le-\kappa_t\frac{\sinh\big(\frac{\Delta_\infty(t)}{2}\big)}{2\cosh^2\big(\frac{\Delta_\infty(t)}{4}\big)}\,\mathrm{d}t+\frac{1}{4}\frac{\max_{i,k}\big\{\mathcal{E}_r^{1,i}-\mathcal{E}_r^{1,k}\big\}}{\cosh^2\big(\frac{\Delta_\infty(t)}{4}\big)}\\&\le-\kappa_tX_t\,\mathrm{d}t+\frac{1}{2}\max_{i,k}\big\{\mathcal{E}_t^{1,i}-\mathcal{E}_t^{1,k}\big\}\frac{X_t}{\sinh\big(2\,\mathrm{arctanh}(X_t)\big)}\,\mathrm{d}t,\end{aligned}$$

其中我们使用了恒等式 $\sinh(2x)=2\sinh(x)\cosh(x)$。由于 $\sinh(2\,\text{反双曲正切}(x))=\frac{2x}{1-x^2}$，因此可以将上式改写为

$$\mathrm{d}X_t\le\alpha(t,X_t)\,\mathrm{d}t,\quad\text{where}\quad\alpha(t,X_t)=-\tilde{\lambda}^\star(t,X_t)X_t+\frac{1}{4}\max_{i,k}\big\{\mathcal{E}_t^{1,i}-\mathcal{E}_t^{1,k}\big\}\big(1-X_t^2\big),$$

˜为便于下文表述，我们用 $\lambda^\star(t,\ X_t)$ 替代 κ_t （不过，无论 κ_t 是定理 2.2 中的确定性速率 λ 、命题 3.10 中的逐路径˜速率 $\lambda_t^\star$，还是系数 $\lambda^\star(t,\ X_t)$ 或推论 3.10.1 中的逐路径速率 $\lambda_t^\star$，论证都类似）。将 $\tilde{\lambda}^\star(t,\ X_t)$ 下界估为 $\tilde{\lambda}_t^\star$，并将 $(1-X_t^2)$ 上界估为 1，再次应用 Gr¨onwall 不等式即可得到 (10)。

˜回顾 (28)，其中给出了 $\lambda^\star$ 的定义。注意，映射 $x \mapsto \alpha(t,\ x)$ 是局部˜利普希茨连续的（其利普希茨常数依赖于 $\omega,\ t$ 和 x），因为根据引理 3.1 和假设 (A4)，$x \mapsto \lambda^\star(t,\ x)x$ 局部利普希茨连续，且 $\max_{i,\ k}\{\mathcal{E}_t^{1,i} - \mathcal{E}_t^{1,k}\}$ 局部有界。现在，令 u_t 为具有随机系数的常微分方程的解，其方程为

$$\frac{\mathrm{d}u_t}{\mathrm{d}t} = \alpha(t, u_t), \qquad u_0 = X_0 = \tanh\left(\frac{\Delta_\infty(0)}{4}\right), \tag{54}$$

其中 α，或者具体而言，$\tilde{\lambda}^\star$ 和 $\max_{i,\ k}\{\mathcal{E}_t^{1,i} - \mathcal{E}_t^{1,k}\}$ 取决于过程 $\tilde{\pi}_t$，而该过程对于˚每个 ω都是固定的。回想一下，由于 $\mu,\ \nu \in \mathcal{S}^n$ 根据假设，$\mathcal{H}(\mu,\ \nu) < \infty$，因此 $u_0 \in (0,\ 1)$。由于右端是局部 Lipschitz 的，(54) 在其首次爆炸时刻 $T > 0$ 之前有唯一解 u_t （同样参见例如 [33, 定理 2.5]）。现在，如果 $T < \infty$，那么 T 是满足 $u_{T-} = 1$的第一个时间。由 u_t的连续性，对于所有 $\varepsilon > 0$，存在一个 $\delta > 0$，使得对于$s \in (T - \delta,\ T)$，我们有 $u_s \in (1 - \varepsilon,\ 1)$。然后

$$\begin{aligned} 1 &= u_{T-\delta} + \int_{T-\delta}^{T} \alpha(s, u_s)\,\mathrm{d}s \\ &\le u_{T-\delta} - \frac{\delta(2-\varepsilon)(1-\varepsilon)}{\varepsilon} \inf_{s\in[T-\delta,T]} \min_{i\neq k}\left\{q_{ik}\frac{\tilde{\pi}_s^i}{\tilde{\pi}_s^k}\right\} + \frac{\delta\varepsilon(2-\varepsilon)}{4} \sup_{s\in[T-\delta,T]} \max_{i,k}\left\{\mathcal{E}_s^{1,i} - \mathcal{E}_s^{1,k}\right\}, \end{aligned}$$

利用 $\lambda^\star$ 的（严格）正性以及 $\max_{i,\ k}\{\mathcal{E}_t^{1,i} - \mathcal{E}_t^{1,k}\}$，并且 $\frac{1+u_t}{1-u_t} \geq \frac{(2-\varepsilon)\ \ (1-\varepsilon)\ \varepsilon}{}$ 且 $(1-u_s^2) \le \varepsilon(2-\varepsilon)$ 对于 $s \in (T-\delta,\ T)$ 成立。由于当 ε 足够小时，负项会超过正项，这意味着 $1 < u_{T-\delta}$，而它严格小于 1，因此产生矛盾。于是$T = \infty$，且对于所有 $t \geq 0$，(54) 都有唯一解。

类似的论证可证明，对于所有 $t \geq 0,\ \ u_t > 0$ 成立；最后，由引理 3.9 可得该定理。 □

推论 *2.4.1* 的证明。证明与定理 2.4 的证明类似。 □

4.2.1 一个数值示例

我们得出结 本节以及本文通过几组模拟检验我们的误差界 .

我们考虑一个模型参数近似的 Wonham滤波器，其动力学由 (38) 给出˜。我们假设 h 已知，因此 $h = h_\circ$ 我们通过应用一种非负分解算法来近似 Q：从 Q中减去对角矩阵，使用 Python 软件包 sklearn.decomposition 中的 NMF 类对所得正矩阵进行近似，然后˜重构对角矩阵，以确保所有行之和为 0，从而得到 $Q_\circ$ 在这种设定下，不存在由 h的模型设定错误导致的误差，因此我们不必担心估计局部时间项。我们考虑定理 2.4 给出的误差界。

在下图中，我们比较了该近似滤波器与适用于三状态和六状态马尔可夫链的Wonham滤波器。在每种情况下，我们取 Q 矩阵为给定的

$$Q = \begin{pmatrix} -3 & 1 & 2 \\ 1 & -3 & 2 \\ 1.5 & 1.5 & -3 \end{pmatrix}, \qquad Q = \begin{pmatrix} -9 & 3 & 1 & 1.5 & 2.5 & 1 \\ 1 & -7.5 & 1 & 2 & 2.3 & 1.2 \\ 3 & 2 & -8 & 1 & 1 & 1 \\ 2 & 1.3 & 1 & -6 & 0.7 & 1 \\ 1.1 & 1 & 0.9 & 3 & -9 & 3 \\ 1 & 1 & 3 & 2 & 2.5 & -9.5 \end{pmatrix},$$

以及传感器函数 h 为

$$h = (-1, 0, 1)\,, \qquad h = (-3, -2, -1, 1, 2, 3)\,.$$

对于三状态马尔可夫链，我们取信号 X 的初始分布为其遍历分布，即 law$(X_0) = \big(0.3,$ $0.3, 0.4\big)$。这也是 Wonham滤波器 π_t 的初始条件。近似滤波器 $\tilde{\pi}_t$ 的近似速率矩阵 Q 通过对 Q 进行 2 通道非负矩阵分解近似得到。我们取 $\tilde{\pi}_t$ 的初始条件为 $\tilde{\pi}_0 = \big(0.2, 0.2, 0.6\big)$。在六状态情形下，我们令信号 X 的初始状态非常接近 $\mathcal{S}^5$ 的边界，其初始分布为 $\mu = \big(0.5,$ $0.04, 0.09, 0.2, 0.04, 0.13\big)$，这也是 π_t 的初始条件~。对于 $\tilde{\pi}_t$ 的速率矩阵 Q，我们采用对 Q 进行的 4 通道非负矩阵分解近似。我们还令 $\tilde{\pi}_t$ 相对接近 $\mathcal{S}^5$ 的边界，但靠近与 μ 不同的另一条边，并取 $\tilde{\pi}_0 = \big(0.25, 0.1, 0.06, 0.07, 0.22, 0.3\big)$。

~为体现透明性，我们在此列出每种情况下非负矩阵分解近似得到的矩阵 Q（四舍五入至两位有效数字）：

$$\tilde{Q} \approx \begin{pmatrix} -2.5 & 0.5 & 2 \\ 0.5 & -2.5 & 2 \\ 1.5 & 1.5 & -3 \end{pmatrix}, \qquad \tilde{Q} \approx \begin{pmatrix} -9 & 3.04 & 1.04 & 1.54 & 2.43 & 0.95 \\ 0.94 & -7.25 & 1.70 & 2.02 & 1.58 & 1.01 \\ 2.92 & 2.05 & -7.8 & 0.69 & 0.88 & 1.26 \\ 2.11 & 1.24 & 0.52 & -5.34 & 0.84 & 0.62 \\ 1.13 & 0.86 & 0.63 & 3.02 & -8.68 & 3.04 \\ 1.02 & 0.77 & 2.64 & 1.92 & 2.92 & -9.28 \end{pmatrix}.$$

在图3左侧，我们分别绘制了三状态和六状态非线性滤波器中，π_t 与 $\tilde{\pi}_t$ 之间的希尔伯特误差的100条实现轨迹（及其样本均值，蓝色），以及定理 2.4 给出的误差界（及其样本均值）。由于这些误差界是逐路径给出的，π_t 与 $\tilde{\pi}_t$ 之间误差的每条实现轨迹都对应三个误差界：洋红色曲线表示 4 反双曲正切(u_t)，其中 u_t 是常微分方程~(8)的（数值）解；绿色曲线表示式（11）中的界，其中衰减率由式（9）中定义的 λ_t 给出；红色曲线再次表示式（11）中的界，但采用的是定理 2.2 中的确定性衰减率 λ。在每个时间步上逐路径计算误差项（7）。右侧图中，对于相同的模拟，我们绘制了双曲正切函数$(\mathcal{H}(\pi_t,\ \tilde{\pi}_t)/4)$的100条实现轨迹（蓝色），以及式（8）的数值解 u_t（洋红色）。

在三状态情形中，滤波器位于 $\mathcal{S}^2$，可以看出，我们对误差的估计与其实际值非常接近（尤其是由方程 (8) 的解给出的常微分方程界）。在六状态情形中，滤波器为五维，我们的误差界就不那么精确了。事实上，可以合理地认为，随着状态空间维数增加，我们的误差界会逐渐变差。

为什么会这样？正如我们在第3.4节中提到的，我们的误差界的主要问题在于收缩率。关于稳定性估计的数值实验（见图 2）表明，误差的收缩速度远快于我们能够证明的速度。这是合理的，因为对 Q 的所有分量取最小值后，定理 2.2 和命题 3.10 中的衰减率给出的是针对“最坏情况”场景的逐路径界。类似的论证也适用于我们对近似误差项 (7) 的处理：为确保我们的界成立，我们需要对所有可能的指标取最大值，这意味着随着维数增加，我们高估误差的可能性也越来越大。

此时可以从几个方向着手，以收紧我们的误差估计。首先可以尝试利用对各个指标求平均，而不是

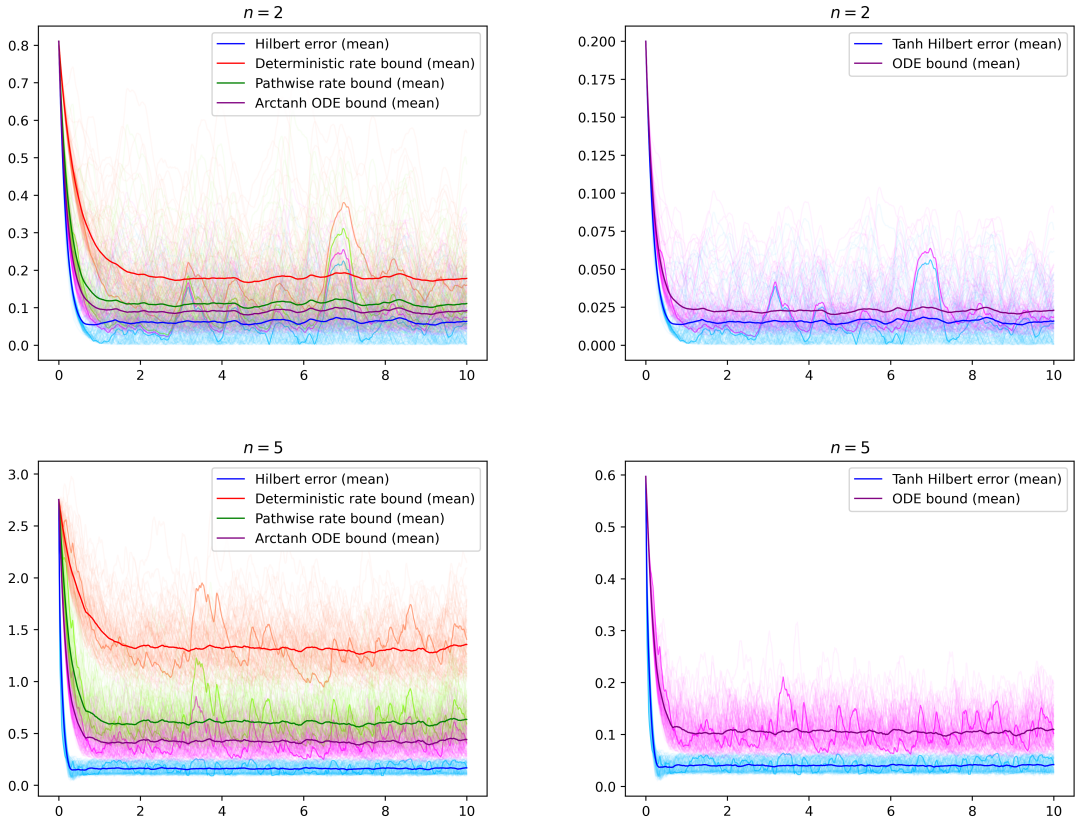

图3：对于维度 $n = 2,5$，我们将定理 2.4 中给出的误差界与 Wonham滤波器和近似滤波器之间的实际希尔伯特误差进行比较。在右侧，我们绘制了希尔伯特射影误差 $\mathcal{H}(\pi_t, \tilde{\pi}_t)$ 的100次实现（浅蓝色，淡化显示）、由 4 反双曲正切(u_t) 给出的常微分方程界（其中 u_t 满足方程 (8)，以淡紫红色显示）、逐路径界 (11)˜（采用逐路径衰减率 λ_t，以浅绿色显示），以及采用确定性衰减率 λ 的逐路径界（以橙色显示），其中 $t \in [0, 10]$。我们随机选取一条希尔伯特误差样本路径，并将其与对应的三个逐路径界一同突出显示。我们分别用蓝色、紫色、绿色和红色绘制误差及三个界的样本均值。在右侧，对于相同的模拟，我们绘制了量 $\tanh(\mathcal{H}(\pi_t, \tilde{\pi}_t)/4)$ 的100次实现，以及常微分方程界 u_t 和两者的样本均值。同样，我们随机突出显示一次双曲正切误差的实现及其对应的常微分方程界。

只需对它们取极小值/最大值，即可得到更紧的衰减率/误差项。如果不采用逐路径的界，而是寻求期望意义下的界，这或许可以实现。然而，这个问题似乎很难，因为它涉及估计 π_t 与 $\tilde{\pi}_t$ 的比值的对数差的极小值点和最大值点的期望。另一方面，如果不进行解析推导，而是能够用数值方法估计相关量，那么在高维情况下获得稳定性误差衰减率的良好数值估计应该相对容易，正如图 2 中的图示所示。随后，可以将估计得到的衰减率代入（11），从而为高维情况下近似滤波器的误差得到更紧的界（该界应以高概率成立）。高估误差项的问题仍然存在，但由于这些误差项受负指数项支配，提高衰减率将带来显著的整体改进。

对衰减率进行数值估计还可以带来其他可能性。本文所发展的论证适用于信号由任意遍历连续时间马尔可夫链给出的情形——Q 的非对角元素严格为正，仅用于保证我们推导出的衰减率非零。换言之，只要信号是遍历的，即使 Q-矩阵是稀疏的，Wonham滤波器的稳定性误差也会衰减（正如 [7] 所讨论的）。通过离散化，在紧状态空间上的非线性滤波器（由 Kushner–Stratonovich随机偏微分方程的解给出）通常可用具有大量状态的Wonham滤波器进行近似。然而，扩散算子对应的是一个非常稀疏的转移矩阵。给定离散化扩散算子的希尔伯特误差衰减率的数值估计，便可以使用我们的误差估计来理解无限维情况下近似滤波器的误差。

致谢 EF 的研究得到 EPSRC 项目 EP/L015811/1 的资助。SC 感谢 UKRI繁荣合作伙伴计划（FAIR）在 EPSRC 项目 EP/V056883/1 项下提供的支持，以及艾伦·图灵研究所、国家统计局（ONS）和牛津–曼氏定量金融研究所的支持。

A 半鞅族的最大值过程

在本附录中，我们使用适当的光滑近似，研究由共同布朗运动驱动的一族连续随机过程的最大值的动态。

回顾最大值和最大值点的以下光滑近似。设 $\alpha \in (0,\infty)$，并设 $\mathbf{x} = \{x_i\}_{i=0}^n$ 为实数序列。我们将 *LogSumExp* 函数 $LSE_\alpha(\mathbf{x})$ 定义为

$$LSE_\alpha(\mathbf{x}) = \frac{1}{\alpha} \log \sum_k e^{\alpha x_k},$$

并将 *SmoothMax* 函数 $S_\alpha(\mathbf{x})$ 定义为

$$S_\alpha(\mathbf{x}) = \frac{\sum_j x_j e^{\alpha x_j}}{\sum_k e^{\alpha x_k}}.$$

给定一族实值系数 $\mathbf{c} = \{c_i\}_{i=0}^n$，我们还将 *SoftArgMax*（或 *SoftMax*）函数 $S_\alpha^{arg}(\mathbf{x}, \mathbf{c})$ 定义为

$$S_\alpha^{arg}(\mathbf{x}, \mathbf{c}) = \frac{\sum_j c_j e^{\alpha x_j}}{\sum_k e^{\alpha x_k}}.$$

我们先证明几个简单的引理。

记号。设 $\mathcal{I}$ 为 $\mathbf{x}$ 的最大值点，即 $\mathcal{I} := \{j \in \text{自然数集} : x_j = \text{最大值}_{i \in \mathbf{N}}\, x_i\} \subset \text{自然数集}$。

引理 A.1（收敛到最大值）。

$$\lim_{\alpha\to\infty} LSE_\alpha(\mathbf{x}) = \lim_{\alpha\to\infty} S_\alpha(\mathbf{x}) = \max_{i\in\{0,\dots,n\}} x_i$$

证明。设 $M = \max^{i\in\{0,\dots,n\}} x_i$。我们有

$$M = \frac{1}{\alpha}\log e^{\alpha M} \le LSE_\alpha(\mathbf{x}) \le \frac{1}{\alpha}\log\left((n+1)e^{\alpha M}\right) = \frac{\log(n+1)}{\alpha} + M,$$

并令 $\alpha \to \infty$ 趋于极限，即得结论。对于 *SmoothMax* 函数，考虑 $\mathcal{I}$，即 $\mathbf{x}$ 的最大值点，并设其大小为 $|\mathcal{I}| = d \ge 1$。那么

$$\begin{aligned} S_\alpha(\mathbf{x}) &= \sum_{j\in\mathcal{I}} \frac{x_j}{d + \sum_{k\notin\mathcal{I}} e^{\alpha(x_k - x_j)}} + \sum_{j\notin\mathcal{I}} \frac{x_j}{1 + \sum_{k\neq j} e^{\alpha(x_k - x_j)}} \\ &= \frac{dM}{d + \sum_{k\notin\mathcal{I}} e^{-\alpha(M - x_k)}} + \sum_{j\notin\mathcal{I}} \frac{x_j}{1 + \sum_{\substack{k\neq j \\ k\in\mathcal{I}}} e^{\alpha(M - x_k)} + \sum_{\substack{k\neq j \\ k\notin\mathcal{I}}} e^{\alpha(x_k - x_j)}} \xrightarrow{\alpha\to\infty} M. \end{aligned}$$

□

引理 A.2。设 $\mathcal{I}$ 为 $\mathbf{x}$ 的最大值点，且 $|\mathcal{I}| = d \ge 1$ 为其大小。那么

$$\lim_{\alpha\to\infty} S_\alpha^{arg}(\mathbf{x}, \mathbf{c}) = \frac{1}{d}\sum_{j\in\mathcal{I}} c_j.$$

证明。与引理 A.1 类似。 □

引理 A.3 （ $LSE_\alpha(\mathbf{x})$ 的导数）。

$$\begin{aligned}\frac{\partial}{\partial x_i} LSE_\alpha(\mathbf{x}) &= \frac{e^{\alpha x_i}}{\sum_k e^{\alpha x_k}},\\ \frac{\partial^2}{\partial x_i^2} LSE_\alpha(\mathbf{x}) &= \alpha \sum_{j\neq i} \frac{e^{\alpha(x_i+x_j)}}{\left(\sum_k e^{\alpha x_k}\right)^2},\\ \frac{\partial^2}{\partial x_i \partial x_j} LSE_\alpha(\mathbf{x}) &= -\alpha \frac{e^{\alpha(x_i+x_j)}}{\left(\sum_k e^{\alpha x_k}\right)^2}.\end{aligned}$$

证明。简单计算即可得证。 □

现在考虑该函数

$$f_\alpha(\mathbf{x}) = \sum_i \frac{\partial^2}{\partial x_i^2} LSE_\alpha(\mathbf{x}) = -\sum_i \sum_{j\neq i} \frac{\partial^2}{\partial x_i \partial x_j} LSE_\alpha(\mathbf{x}) = \alpha \sum_i \sum_{j\neq i} \frac{e^{\alpha(x_i+x_j)}}{\left(\sum_k e^{\alpha x_k}\right)^2}.$$

引理 A.4. 如果最大值$_i\, x_i$ 是唯一的，也就是说，如果存在 $\exists!\, j^\star$ 使得最大值$_i\, x_i = x_{j^\star}$，那么

$$\lim_{\alpha\to\infty} f_\alpha(\mathbf{x}) = 0.$$

证明。 令 $x_{j^\star} := \max_i x_i$。由于 $x_{j^\star}$ 是唯一最大值点，存在 $\varepsilon_j > 0$ ，使得$x_j = x_{j^\star} - \varepsilon_j$ 对所有 $j \neq j^\star$都成立。于是我们有

$$\begin{aligned}f_\alpha(\mathbf{x}) &= \alpha\left[\sum_{j\neq j^\star} \frac{e^{\alpha(x_{j^\star}+x_j)}}{\left(\sum_k e^{\alpha x_k}\right)^2} + \sum_{i\neq j^\star} \frac{e^{\alpha(x_i+x_j^\star)}}{\left(\sum_k e^{\alpha x_k}\right)^2} + \sum_{i\neq j^\star} \sum_{\substack{j\neq i\\ j\neq j^\star}} \frac{e^{\alpha(x_i+x_j)}}{\left(\sum_k e^{\alpha x_k}\right)^2}\right]\\ &= \frac{\alpha}{\left(e^{\alpha x_{j^\star}} + \sum_{k\neq j^\star} e^{\alpha(x_{j^\star}-\varepsilon_k)}\right)^2}\left[2\sum_{j\neq j^\star} e^{\alpha(2x_{j^\star}-\varepsilon_j)} + \sum_{i\neq j^\star}\sum_{j\neq i} e^{\alpha(2x_{j^\star}-\varepsilon_i-\varepsilon_j)}\right]\\ &= \frac{\alpha}{\left(1 + \sum_{k\neq j^\star} e^{-\alpha\varepsilon_k}\right)^2}\left[2\sum_{j\neq j^\star} e^{-\alpha\varepsilon_j} + \sum_{i\neq j^\star}\sum_{j\neq i} e^{-\alpha(\varepsilon_i+\varepsilon_j)}\right],\end{aligned}$$

并且由于 ε_j 对所有 $j \neq j^\star$ 都严格为正，当 $\alpha \to \infty$ 时，在极限情况下，负指数项 $e^{-\alpha\varepsilon_j}$ 占主导地位，且 $f_\alpha \to 0$。 □

引理 A.5。 考虑函数

$$g_\alpha(x) = \alpha \frac{e^{\alpha x}}{(1+e^{\alpha x})^2}.$$

我们有 $g_\alpha(x)\mathrm{d}x \to \delta_0$ ，当 $\alpha \to \infty$ 时，在测度的弱收敛意义下，其中 δ_0 表示位于 *0* 处的狄拉克质量。

证明。 首先注意，对于所有 $\alpha > 0$

$$\int_{\mathbb{R}} g_\alpha(x)\,\mathrm{d}x = 1.$$

考虑任意连续有界函数 $\varphi(x) \in C_b(\mathbb{R})$。对于所有 $\varepsilon > 0$，存在一个 $\delta > 0$，使得

$$\begin{aligned}\Big|\int_{\mathbb{R}} \varphi(x) g_\alpha(x)\,\mathrm{d}x - \varphi(0)\Big| &\le \int_{\mathbb{R}} g_\alpha(x)\big|\varphi(x)-\varphi(0)\big|\,\mathrm{d}x \\ &\le \varepsilon \int_{-\delta}^{\delta} g_\alpha(x)\,\mathrm{d}x + \int_{-\infty}^{-\delta} \alpha e^{-\alpha\delta}\big|\varphi(x)-\varphi(0)\big|\,\mathrm{d}x \\ &\quad + \int_{\delta}^{\infty} \frac{\alpha}{1+e^{\alpha\delta}}\big|\varphi(x)-\varphi(0)\big|\,\mathrm{d}x \\ &\le \varepsilon + \int_{-\infty}^{-\delta} \alpha e^{-\alpha\delta}\big|\varphi(x)-\varphi(0)\big|\,\mathrm{d}x + \int_{\delta}^{\infty} \frac{\alpha}{1+e^{\alpha\delta}}\big|\varphi(x)-\varphi(0)\big|\,\mathrm{d}x.\end{aligned}$$

令 $\alpha \to \infty$ 取极限时，最后两个积分趋于 0。因此，对于任意 $\varepsilon > 0$，左侧的极限小于 ε，故结论得证。 □

现在我们转而研究由同一个布朗运动驱动的一族连续半鞅的最大值的动态。需要特别指出的是，我们处理的是绝对连续有限变差部分的半鞅，这意味着它们的局部时在 $t \in \mathbb{R}^+$ 和 $a \in \mathbb{R}$ 中具有双连续修改（见定义 4.4）。所有能够表示为 Itˆo 随机微分方程的解、且其可积漂移项和随机项由具有绝对连续有限变差的半鞅驱动的随机过程，都属于这种情形。

考虑 一族取值于 $\mathbb{R}$ 的连续半鞅 $\mathbf{X}_t = \{X_t^i\}_{i=0}^n$，其动态为

$$\mathrm{d}X_t^i = b_t^i\,\mathrm{d}t + \sigma_t^i\,\mathrm{d}B_t, \tag{55}$$

其中，对于所有 $i = 0, \dots, n$，b_t^i 和 σ_t^i 分别是（实值、可预测、随机可积的）漂移系数和扩散系数，而 B_t 是一个标准布朗运动。

我们应用伊藤引理，推导 $LSE_\alpha(\boldsymbol{X}.)(t)$的动力学:

$$\begin{aligned}&\mathrm{d}LSE_\alpha(\boldsymbol{X}.)(t) \qquad (56)\\ &= \sum_{i=0}^n \frac{e^{\alpha X_t^i}}{\sum_k e^{\alpha X_t^k}}\,\mathrm{d}X_t^i + \frac{1}{2}\sum_{i=0}^n\sum_{j=0}^n \alpha e^{\alpha X_t^i}\left(\frac{\delta_{ij}}{\sum_k e^{\alpha X_t^k}} - \frac{e^{\alpha X_t^j}}{(\sum_k e^{\alpha X_t^k})^2}\right)\mathrm{d}\langle X_\cdot^i, X_\cdot^j\rangle_t \\ &= \sum_{i=0}^n \frac{e^{\alpha X_t^i}}{\sum_k e^{\alpha X_t^k}} b_t^i\,\mathrm{d}t + \sum_{i=0}^n \frac{e^{\alpha X_t^i}}{\sum_k e^{\alpha X_t^k}}\sigma_t^i\,\mathrm{d}B_t + \frac{1}{2} f_\alpha(\mathbf{X}., \boldsymbol{\sigma}.)(t)\,\mathrm{d}t, \qquad (57)\end{aligned}$$

其中，我们用 δ_{ij} 表示克罗内克 δ，并定义函数

$$f_\alpha(\mathbf{X}., \boldsymbol{\sigma}.)(t) := \alpha \sum_i \sum_{j\neq i} \frac{e^{\alpha(X_t^i + X_t^j)}}{\left(\sum_k e^{\alpha X_t^k}\right)^2}\left((\sigma_t^i)^2 - \sigma_t^i\sigma_t^j\right). \tag{58}$$

对于所有 $s \le t$，我们将 (57) 改写为如下积分形式，

$$\begin{aligned}LSE_\alpha(\boldsymbol{X}.)(t) = LSE_\alpha(\boldsymbol{X}.)(s) &+ \int_s^t S_\alpha^{arg}(\mathbf{X}., \mathbf{b}.)(r)\,\mathrm{d}r + \int_s^t S_\alpha^{arg}(\mathbf{X}., \boldsymbol{\sigma}.)(r)\,\mathrm{d}B_r \\ &+ \frac{1}{2}\int_s^t f_\alpha(\mathbf{X}., \boldsymbol{\sigma}.)(r)\,\mathrm{d}r. \qquad (59)\end{aligned}$$

我们关注令 α 趋于无穷时上述表达式的极限。对每个时间 t，定义 $\mathbf{X}_t$ 的最大值点为 $\mathcal{I}_t = \{j \in$ **自然数集**： $X_t^j \geq X_t^i \ \forall i \in$ **自然数集**$\}$。由于 $S_\alpha^{arg}(\mathbf{X}_\cdot, \mathbf{b}_\cdot)(r) \leq$ 最大值$_i\, b_t^i$，且根据假设，对所有 i，b^i 都可积，因此我们可以应用控制收敛定理，得到

$$\lim_{\alpha\to\infty} \int_0^t S_\alpha^{arg}(\mathbf{X}_\cdot, \mathbf{b}_\cdot)(s)\,\mathrm{d}s = \int_0^t \frac{1}{|\mathcal{I}_s|} \sum_{j\in\mathcal{I}_s} b_s^j\,\mathrm{d}s.$$

同样，$\max_i \sigma_t^i$ 对 B_t可积，因此我们可以对随机积分应用控制收敛定理，得到

$$\lim_{\alpha\to\infty} \int_s^t S_\alpha^{arg}(\mathbf{X}_\cdot, \boldsymbol{\sigma}_\cdot)(r)\,\mathrm{d}B_r = \int_s^t \frac{1}{|\mathcal{I}_r|} \sum_{j\in\mathcal{I}_r} \sigma_r^j\,\mathrm{d}B_r.$$

(59) 式右端的最后一个积分较难处理。

命题 A.6. 考虑一族连续半鞅$\mathbf{X}_t = \{X_t^i\}_{i=0}^n$，其动态由$(55)$给出。令 f_α 如(58)中所定义。那么，对所有 $s \leq t$

$$\lim_{\alpha\to\infty} \int_s^t f_\alpha(\mathbf{X}_\cdot, \boldsymbol{\sigma}_\cdot)(r)\,\mathrm{d}r \leq \sum_i \sum_{j>i} \Big(L_t^0(X_\cdot^i - X_\cdot^j) - L_s^0(X_\cdot^i - X_\cdot^j) \Big) \quad a.s.$$

证明。 利用对称性，我们首先将 $f_\alpha(\mathbf{X}_\cdot, \boldsymbol{\sigma}_\cdot)(t)$改写为

$$f_\alpha(\mathbf{X}_\cdot, \boldsymbol{\sigma}_\cdot)(t) = \frac{1}{2}\alpha \sum_i \sum_{j\neq i} \frac{e^{\alpha(X_t^i + X_t^j)}}{\left(\sum_k e^{\alpha X_t^k}\right)^2} \left(\sigma_t^i - \sigma_t^j\right)^2,$$

因此注意到，(59) 式右端的最后一个积分始终为正。此外，令 g_α 如引理 A.5 中所定义，

$$\begin{aligned}
\frac{\alpha e^{\alpha(X_t^i + X_t^j)}}{\left(\sum_k e^{\alpha X_t^k}\right)^2} &= \frac{\alpha}{2 + e^{\alpha(X_t^i - X_t^j)} + e^{\alpha(X_t^j - X_t^i)} + \sum_{k\neq i,j} \sum_{l\neq i,j} e^{\alpha(X_t^k + X_t^l - X_t^i - X_t^j)}} \\
&\leq \frac{\alpha e^{\alpha(X_t^i - X_t^j)}}{(1 + e^{\alpha(X_t^i - X_t^j)})^2} = g_\alpha(X_\cdot^i - X_\cdot^j)(t).
\end{aligned}$$

占用时间公式（例如参见 [30,第 6 章，推论 1.6]）给出

$$\begin{aligned}
\lim_{\alpha\to\infty} \int_s^t f_\alpha(\mathbf{X}_\cdot, \boldsymbol{\sigma}_\cdot)(r)\,\mathrm{d}r &\leq \lim_{\alpha\to\infty} \frac{1}{2} \int_s^t \sum_i \sum_{j\neq i} g_\alpha(X_\cdot^i - X_\cdot^j)(r) \left(\sigma_r^i - \sigma_r^j\right)^2 \mathrm{d}r \\
&= \sum_i \sum_{j\neq i} \lim_{\alpha\to\infty} \frac{1}{2} \int_s^t g_\alpha(X_\cdot^i - X_\cdot^j)(r)\,\mathrm{d}\langle X_\cdot^i - X_\cdot^j\rangle_r \\
&= \sum_i \sum_{j\neq i} \lim_{\alpha\to\infty} \frac{1}{2} \int_{\mathbb{R}} g_\alpha(z) \Big(L_t^z(X_\cdot^i - X_\cdot^j) - L_s^z(X_\cdot^i - X_\cdot^j) \Big)\,\mathrm{d}z \\
&= \frac{1}{2} \sum_i \sum_{j\neq i} \Big(L_t^0(X_\cdot^i - X_\cdot^j) - L_s^0(X_\cdot^i - X_\cdot^j) \Big)
\end{aligned}$$

几乎必然成立，其中最后一个等式依据引理 A.5 得出： $g_\alpha(z)\,\mathrm{d}z$ 弱收敛于位于 0 处的狄拉克质量。注意， $g_\alpha(z)$并非紧支撑（与定义 4.4 对比），但这没有影响，因为局部时 $L_t^z($ $X_\cdot^i - X_\cdot^j)$ 在 z a.s. 有界（参见 Barlow 和 Yor [6, 推论 5.2.2]）。 □

B 数值实验

本附录介绍图 2 中各图所用模拟的一些细节。为便于比较不同维度，我们为速率矩阵 Q 设定了固定结构，并使各维度下的压缩系数保持不变。

对于 $n = 2$，20，50，100，我们将信号过程 X 设定为一个具有 $n+1$ 个状态的马尔可夫链$\{0, \dots, n\}$，使得如果 X 在某个时间 t处于状态 i，那么它跳转到状态$i+1$ 或 $i-1$ 的概率相等，而跳转到状态 $j \neq i \pm 1$ 的概率则低得多。换句话说，该链在某个状态及其两个最近邻状态之间快速切换，但与距离较远的状态之间的混合速度较慢。我们令从状态 i 出发的跳转速率随链的维度增大而增大：对于 $n \geq 3$，我们将 $Q =(q_{ij})$的非三对角线元素设为 1，上下两条对角线元素设为 $n+1$，因此主对角线元素为 $-3n$，即

$$(q_{ij}) = \begin{cases} n+1, & \text{if } j \equiv i \pm 1 \pmod{n}, \\ -3n, & \text{if } j = i, \\ 1, & \text{otherwise.} \end{cases}$$

对于 $n = 2$，我们只需取 Q 为对角线上为 -2 、其余元素均为 1 的对称矩阵。对所有 n 均以这种方式固定 Q 后，定理 2.2 中的收缩率为 $\lambda = 2$，且在所有维度下保持不变。

链 X服从均匀平稳分布，我们将其记为 $\mu = (\frac{1}{n}, \dots, \frac{1}{n})$。这是概率单纯形 $\mathcal{S}^n$的中心点。我们取 law$(X_0) = \mu$。最后，我们将传感器函数 $h \in \mathbb{R}^{n+1}$ 设为一个随机生成的向量，使得对于每个$i \in$ **自然数集**， $h^i = z_i + x_i$，其中 z_i是 $\{-10, \dots, 10\}$中的一个随机整数，而 x_i 是 $[0, 1]$上均匀随机变量的一次实现。

最优滤波器 π_t 的初始条件为 $\pi_0 = \mu$。‘错误的’Wonham滤波器 $\tilde{\pi}_t$初始化为 $\nu \neq \mu$：为了确定 ν，我们通过根据 $n+1$ 个独立的伯努利随机变量，对 μ 的所有分量加上或减去 $\frac{1}{2} \min_i \mu_i$ 来扰动 μ，并重新归一化。

固定所有这些参数后，我们生成信号过程和观测过程的 300 条样本路径，并通过求解 Zakai 方程（例如参见 [5,备注 3.26]），使用简单的欧拉格式计算最优 Wonham滤波器和“错误的”Wonham滤波器，并在每一步之后重新归一化。我们绘制希尔伯特误差 $\mathcal{H}(\pi_t, \tilde{\pi}_t)$ 的实现，以及定理 2.2 和命题 3.10 给出的界。请注意，命题 3.10 给出的界依赖于样本路径（计算这些界需要观测 $\tilde{\pi}_t$），因此，对于希尔伯特误差的每个实现，都有与之对应的命题 3.10 中各界的实现。这些界还表示为 $\tanh(\mathcal{H}(\pi_t, \tilde{\pi}_t)/4)$ 的界（而非 $\mathcal{H}(\pi_t, \tilde{\pi}_t)$ 的界）。对式 (29) 两边取反双曲正切，再乘以 4，可得 $\mathcal{H}(\pi_t, \tilde{\pi}_t) \leq 4\operatorname{arctanh}(u_t)$，其中 u_t 是式 (27) 的解；鉴于 u 的动力学可能具有非常大的利普希茨系数，我们使用修正欧拉格式（例如参见 Hutzenthaler、Jentzen 和 Kloeden [19]）对常微分方程进行数值求解。由双曲正切函数的凹性和单调性，从式 (31) 可得 $\mathcal{H}(\pi_t, \tilde{\pi}_t) \leq \mathcal{H}(\mu, \nu) e^{-\int_0^t \tilde{\lambda}_s \, \mathrm{d}s}$；我们在每个时间步计算 $\tilde{\lambda}_t = 2 \min^{i \neq k} \left(q_{ik} q_{ki} + \sum_{j \neq i,k} \tilde{\pi}_t^j \min\{q_{ji} q_{ik} / \tilde{\pi}_t^k, q_{jk} q_{ki} / \tilde{\pi}_t^i\} \right)^{1/2}$，并进行数值积分以绘制该界。

参考文献

[1] S. Amari. 统计学中的微分几何方法，统计学讲义第 28 卷。施普林格出版社，纽约，1985 年。[2] S. Amari. 信息几何及其应用，应用数学科学第 194 卷。施普林格，东京，2016 年。[3] R. Atar 和 O. Zeitouni. 非线性滤波的指数稳定性。*Ann. Inst. H.Poincar´e Probab. Statist.*，33(6):697–725，1997 年。[4] R. Atar 和 O. Zeitouni. 有限状态非线性滤波的李雅普诺夫指数。*SIAM* 控制与优化杂志，35(1):36–55，1997 年。[5] A. Bain 和 D. Crisan. 随机滤波基础，随机建模与应用概率第 60 卷。施普林格，纽约，2009 年。[6] M. T. Barlow 和 M. Yor. 利用 Garsia-Rodemich-Rumsey 引理得到的半鞅不等式及其在局部时中的应用。《泛函分析杂志》，49(2):198–229，1982 年。[7] P. Baxendale、P. Chigansky 和 R. Liptser. Wonham滤波器的渐近稳定性：遍历信号与非遍历信号。*SIAM* 控制与优化杂志，43(2):643–669，2004 年。[8] G. Birkhoff. Jentzsch 定理的推广。*Trans. Amer. Math. Soc.*，85:219–227，1957 年。[9] G. Birkhoff. 格论。美国数学学会讨论会出版物，第 XXV 卷。美国数学学会，普罗维登斯，罗得岛州，第三版，1967 年。[10] A. Budhiraja. 非线性滤波器的渐近稳定性、遍历性及其他渐近性质。*Ann. Inst. H. Poincar´e Probab. Statist.*，39(6):919–941，2003 年。[11] A. Budhiraja 和 H. J. Kushner. 无限时间范围内非线性滤波器的鲁棒性。*SIAM* 控制与优化杂志，36(5):1618–1637，1998 年。[12] P. Chigansky. 非线性滤波器的稳定性：综述。讲义，巴西彼得罗波利斯，2006 年。[13] P. Chigansky、R. Liptser 和 R. Van Handel. 滤波器稳定性的内在方法。载于《牛津非线性滤波手册》，第 319–351 页。牛津大学出版社，牛津，2011 年。[14] P. Chigansky 和 R. Van Handel. 无限时间范围内有限状态非线性滤波的模型稳健性。《应用概率年刊》，17(2):688–715，2007 年。[15] S. N. Cohen 和 R. J. Elliott. 随机微积分及其应用。概率及其应用。施普林格，查姆，第二版，2015 年。[16] D. Crisan 和 B. Rozovskiˇı，编。《牛津非线性滤波手册》。牛津大学出版社，牛津，2011 年。

[17] B. Delyon 和 O. Zeitouni。滤波问题的李雅普诺夫指数。载于应用随机分析（伦敦，*1989*），第 5 卷随机学专著，第 511–521 页。Gordon and Breach，纽约，1991 年。[18] P. Dupuis 和 R. S. Ellis。大偏差理论的弱收敛方法。Wiley概率与统计丛书：概率与统计。John Wiley & Sons, Inc.，纽约，1997 年。Wiley-Interscience出版物。[19] M. Hutzenthaler、A. Jentzen 和 P. E. Kloeden。具有非全局利普希茨连续系数的 SDE 显式数值方法的强收敛。《应用概率年刊》，22(4):1611–1641，2012年。[20] R. E. K´alm´an。线性滤波与预测问题的新方法。《基础工程学杂志》，82(1):35–45，1960年。[21] R. E. K´alm´an 和 R. S. Bucy。线性滤波与预测理论的新成果。《基础工程学杂志》，83:95–108，1961年。[22] J. W. Kim 和 P. G. Mehta。Wonham滤波器稳定性的对偶刻画。载于*2021*年第*60*届*IEEE*决策与控制会议（*CDC*），第 1621–1628 页。IEEE，2021年。[23] J. W. Kim、P. G Mehta 和 S. Meyn。滤波器稳定性的条件 Poincar´e 不等式。载于*2021*年第*60*届*IEEE*决策与控制会议（*CDC*），第 1629–1636 页。IEEE，2021年。[24] H. Kunita。马尔可夫过程非线性滤波误差的渐近行为。《多元分析杂志》，1(4):365–393，1971年。[25] F. Le Gland 和 L. Mevel。射影乘积的基本性质及其在列允许非负矩阵乘积中的应用。《数学、控制、信号与系统》，13(1):41–62，2000年。[26] F. Le Gland 和 L. Mevel。隐马尔可夫模型中的指数遗忘与几何遍历性。《数学、控制、信号与系统》，13(1):63–93，2000年。[27] F. Le Gland 和 N. Oudjane。利用希尔伯特度量研究非线性滤波器的稳定性与一致逼近，并将其应用于粒子滤波器。《应用概率年刊》，14(1):144–187，2004年。[28] R. S. Liptser 和 A. N. Shiryayev。随机过程统计学。*I*。数学应用，第 5 卷。施普林格出版社，纽约—海德堡，1977年。一般理论，A. B. Aries 译。[29] D. Ocone 和 E. Pardoux。最优滤波器对其初始条件的渐近稳定性。《*SIAM*控制与优化杂志》，34(1):226–243，1996年。[30] D. Revuz 和 M. Yor。连续鞅与布朗运动，数学科学基础丛书第 293 卷 *[*数学科学基本原理*]*。施普林格出版社，柏林，第三版，1999年。

[31] E. Seneta。非负矩阵与马尔可夫链。施普林格统计学丛书。施普林格，纽约，2006年。1981年第二版修订重印本 [施普林格出版社，纽约；MR0719544]。[32] E. Seneta 和 S. Sheridan。非负矩阵乘积的强遍历性。《线性代数及其应用》，37:277–292，1981年。[33] G. Teschl。常微分方程与动力系统，第 140 卷数学研究生教材。美国数学学会，罗德岛州普罗维登斯，2012年。[34] R. Van Handel。滤波、稳定性与鲁棒性。博士论文，加州理工学院，2007年。[35] R. Van Handel。可观测性与非线性滤波。《概率论及相关领域》，145(1-2):35–74，2009年。[36] R. Van Handel。随机环境中条件马尔可夫过程与马尔可夫链的稳定性。《概率年刊》，37(5):1876–1925，2009年。[37] R. Van Handel。隐马尔可夫模型的一致可观测性与不稳定信号的滤波器稳定性。《应用概率年刊》，19(3):1172–1199，2009年。[38] W. M. Wonham。随机微分方程在最优非线性滤波中的若干应用。《*J. SIAM* 控制系列 *A*》，2:347–369（1965年），1965年。