



上海交通大学硕士学位论文

# 现代数学观点下的信息论若干关键问题研究

姓 名：邹嘉阳

学 号：123034910057

导 师：王嘉教授

院 系：集成电路学院（信息与电子工程学院）

学科/专业：信息与通信工程

申请学位：工学硕士

2026年1月12日



**A Dissertation Submitted to  
Shanghai Jiao Tong University for the Degree of Master**

**A STUDY ON SELECTED KEY PROBLEMS IN  
INFORMATION THEORY VIA A MODERN  
MATHEMATICAL PERSPECTIVE**

**Author:** Jiayang Zou

**Supervisor:** Prof. Jia Wang

School of Integrated Circuits (School of Information Science and  
Electronic Engineering)

Shanghai Jiao Tong University

Shanghai, P.R. China

January 12<sup>th</sup>, 2026



# 上海交通大学

## 学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全知晓本声明的法律后果由本人承担。

学位论文作者签名：

日期： 年 月 日

# 上海交通大学

## 学位论文使用授权书

本人同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。

本学位论文属于：

- 公开论文
- 内部论文，保密  1年 /  2年 /  3年，过保密期后适用本授权书。
- 秘密论文，保密 \_\_\_\_ 年（不超过 10 年），过保密期后适用本授权书。
- 机密论文，保密 \_\_\_\_ 年（不超过 20 年），过保密期后适用本授权书。

（请在以上方框内选择打“√”）

学位论文作者签名：

指导教师签名：

日期： 年 月 日

日期： 年 月 日



## 摘要

香农信息论自诞生以来，便作为现代通信与数据压缩的基石深刻影响了数字时代的发展。然而，随着研究对象从经典的离散或高斯信源向一般化的随机过程、非紧度量空间以及复杂动力系统演进，传统的基于组合数学与概率论的分析框架逐渐显露出局限性。近年来，随机微分方程、最优传输理论以及非线性泛函分析等现代数学工具的引入，为重构信息论的基础提供了全新的几何与分析视角。本文立足于现代数学观点，围绕信息的动态演化与静态变分两大核心维度，对信息论中的若干关键问题展开了研究。本文的主要工作与创新点概括如下：

首先，本文在随机动力系统的框架下，深入研究了 Fokker-Planck 流的几何性质及其对信息量演化的影响，其中 FP 流可以将信道视为概率测度在 Wasserstein 空间中的梯度流，本文证明了在势函数满足特定凸性条件下，Fokker-Planck 方程解的适定性以及对数凹 (Log-concavity) 性质在时间演化中的保持性。在此基础上，本文系统推导了互信息沿 Fokker-Planck 流的二阶时间导数，建立了一般扩散过程下互信息凸性的几何判据，成功将热流 (Heat Flow) 和 Ornstein-Uhlenbeck (OU) 流中的经典结果推广至更一般的势函数情形。作为动力学视角的延伸，本文还简要探讨了引入非连续跳跃项后的跳扩散 (Jump-diffusion) 信道模型，给出了包含非高斯项的信息微分恒等式，完善了信道演化理论的边界。

其次，鉴于相对 Fisher 信息在刻画概率分布几何结构上的优势及其在采样算法分析中的潜力，本文在 Fokker-Planck 流的框架下构建了关于广义相对 Fisher 信息的强数据处理不等式 (SDPI) 理论。尽管针对 KL 散度与 Rényi 散度的 SDPI 研究已较为成熟，但针对更强的 Fisher 信息类泛函的研究仍相对匮乏。本文将视角拓展至更广泛的相对 Rényi-Fisher 信息与相对  $\Phi$ -Fisher 信息。通过构建精细的演化演算，本文推导了这些广义信息量沿一般 Fokker-Planck 流的时间导数解析式。在此基础上，本文引入了“Log-Hessian 比较原理” (The Log-Hessian Comparison Principle)。这一原理作为连接势函数几何曲率与信息量收缩系数的关键桥梁，提供了一种通过矩阵不等式判定 SDPI 成立与否的新视角。结合扩散半群的显式 Log-Hessian 公式与曲率界估计，本文建立了一套可验证的判别准则，不仅证明了在一般情形下的最终收缩性 (Eventual Contraction)，还确立了高斯初始化下的永恒收缩性 (Perpetual Contraction)。在理论应用层面，本文给出了 Proximal Sampler 算法在  $\Phi$ -Fisher 信息意义下的混合时

间上界，为理解采样算法在更广义统计量意义下的收敛行为提供了新工具。

最后，本文将视角从动态演化转向静态压缩，利用非线性分析方法与最优传输理论重构了率失真（Rate-Distortion, RD）理论。本文首先在最优弱传输（Optimal Weak Transport, OWT）的框架下，揭示了率失真函数与 Schrödinger Bridge 问题之间的对偶联系，导出了 RD 函数在一般波兰空间上的参数化表示。进一步地，针对抽象非紧空间中率失真函数最优重构分布的存在性这一悬而未决的数学难题，本文引入了非线性泛函分析中的集中紧性原理（Concentration-Compactness Principle）。通过对极小化序列进行“紧性（Compactness）、消散（Vanishing）、二分（Dichotomy）”的三态分析，本文排除了序列质量逃逸与分裂的可能性，从而严格证明了在极其宽泛的条件下（如无界空间、非连续失真度量），率失真泛函的最优解的存在性。这一结果从根本上弥补了经典信息论在处理非紧空间信源时的理论缺口。

综上所述，本文利用随机分析、偏微分方程与变分分析等现代数学工具，打通了信息论与几何、分析之间的壁垒。研究成果不仅深化了对信息量几何结构与演化规律的理解，也为复杂信源编码与高维采样算法提供了坚实的数学理论支撑。

**关键词：**Fokker-Planck 方程；互信息凸性；强数据处理不等式；率失真理论；最优传输理论

## Abstract

Since its inception, Shannon's information theory has served as the bedrock of modern communication and data compression. However, as research targets evolve from classical discrete or Gaussian sources to generalized stochastic processes, non-compact metric spaces, and complex dynamical systems, the traditional analytical framework based on combinatorics and probability theory has gradually revealed its limitations. In recent years, the introduction of modern mathematical tools—such as Stochastic Differential Equations (SDEs), Optimal Transport theory, and Nonlinear Functional Analysis—has provided a novel geometric and analytical perspective for reconstructing the foundations of information theory. Standing on the vantage point of modern mathematics, this thesis conducts a comprehensive study on several key issues in information theory, revolving around two core dimensions: Information Dynamics and Variational Principles. The main contributions and innovations of this thesis are summarized as follows:

First, within the framework of stochastic dynamical systems, this thesis investigates the geometric properties of Fokker-Planck flows and their impact on the evolution of information measures. Viewing the channel as a gradient flow of probability measures in the Wasserstein space, we prove the well-posedness of the solutions to the Fokker-Planck equation and the preservation of the log-concavity property during temporal evolution under specific convexity conditions of the potential function. On this basis, we systematically derive the second-order time derivative of mutual information along Fokker-Planck flows, establishing a geometric criterion for the convexity of mutual information. This generalizes classical results found in Heat flows and Ornstein-Uhlenbeck (OU) flows to scenarios with more general potential functions. As an extension of the dynamical perspective, we also briefly discuss the Jump-diffusion channel model, incorporating non-continuous jump terms, and present differential identities of information involving non-Gaussian terms, thereby completing the boundary of channel evolution theory.

Second, recognizing the strength of relative Fisher information in characterizing geometric structures of distributions and its potential in analyzing sampling algorithms, this thesis constructs a theoretical framework for Strong Data Processing Inequalities (SDPI) concern-

ing generalized relative Fisher information under Fokker-Planck flows. While SDPIs for KL and Rényi divergences are well-studied, research on the stronger Fisher information-type functionals remains relatively undeveloped. This thesis extends the scope to the broader classes of Relative Rényi-Fisher Information and Relative  $\Phi$ -Fisher Information. Through a rigorous calculus of evolution, we derive analytical expressions for the time derivatives of these generalized information measures along general Fokker-Planck flows. On this basis, we introduce and formalize the "Log-Hessian Comparison Principle." This principle serves as a crucial bridge connecting the geometric curvature of the potential function to the contraction coefficients of SDPI, offering a new perspective to verify SDPIs via matrix inequalities. By combining explicit Log-Hessian formulas for diffusion semigroups with curvature bound estimates, we establish a verifiable criterion that proves both the eventual contraction in general settings and the perpetual contraction under Gaussian initialization. As a theoretical application, we provide upper bounds on the mixing time of the Proximal Sampler algorithm in the sense of  $\Phi$ -Fisher information, offering new tools for understanding the convergence behavior of sampling algorithms under more general statistical quantities.

Finally, shifting the perspective from dynamic evolution to static compression, this thesis reconstructs Rate-Distortion (RD) theory using nonlinear analysis methods and Optimal Transport theory. We first reveal the dual connection between the RD function and the Schrödinger Bridge problem within the framework of Optimal Weak Transport (OWT), deriving the parametric representation of the RD function on general Polish spaces. Furthermore, addressing the mathematical difficulty regarding the existence of optimal reconstruction distributions for RD functions in abstract non-compact spaces, we introduce the Concentration-Compactness Principle from nonlinear functional analysis. By analyzing the minimizing sequence through the trichotomy of "Compactness, Vanishing, and Dichotomy," we rule out the possibilities of mass escape and splitting. Consequently, we rigorously prove that under extremely broad conditions (e.g., unbounded spaces, non-absolutely continuous marginals), the optimal solution to the rate-distortion functional exists and possesses good analytical properties. This result fundamentally fills the theoretical gap in classical information theory regarding sources in non-compact spaces.

In summary, this thesis bridges the gap between information theory and geometry/analysis by employing modern mathematical tools such as stochastic analysis, partial differential

---

equations, and variational analysis. The research findings not only deepen the understanding of the geometric structure and evolutionary laws of information measures but also provide a solid mathematical theoretical foundation for complex source coding and high-dimensional sampling algorithms.

**Key words:** Fokker-Planck Equation; Convexity of Mutual Information; Strong Data Processing Inequalities; Rate-Distortion Theory; Optimal Transport Theory



## 目 录

<b>第 1 章 绪论</b> .....	<b>1</b>
1.1 研究背景及意义.....	1
1.2 研究现状及难点.....	4
1.2.1 基于随机流的信息演化与几何不等式 .....	4
1.2.2 率失真理论的现代分析方法 .....	5
1.3 全文结构以及主要创新点.....	5
<b>第 2 章 背景知识</b> .....	<b>9</b>
2.1 符号说明.....	9
2.2 统计量与信息度量.....	10
2.3 对数凹性质与泛函不等式.....	13
2.4 随机分析基础与 Fokker-Planck 方程.....	15
2.4.1 伊藤积分 (The Itô Integral).....	15
2.4.2 伊藤公式 (Itô's Formula) .....	19
2.4.3 随机微分方程的存在唯一性 .....	21
2.4.4 马尔可夫半群 (Markov Semigroup) .....	22
2.4.5 Fokker-Planck 方程 .....	24
2.5 最优传输理论基础.....	25
2.5.1 最优传输问题描述与性质 .....	25
2.5.2 Wasserstein 距离与空间 .....	30
2.6 率失真理论简介.....	32
<b>第 3 章 Fokker-Planck 流的几何性质与互信息凸性研究</b> .....	<b>37</b>
3.1 本章引言.....	37
3.2 Fokker-Planck 流的适定性理论 .....	39
3.3 互信息沿 FP 流的凸性分析.....	41
3.3.1 抛物型系统解的对数凹性 .....	41
3.3.2 互信息沿 FP 流的凸性 .....	43

3.4 扩展模型：跳扩散信道下的信息微分性质.....	45
3.4.1 跳-扩散信道模型与演化方程 .....	46
3.4.2 熵与互信息的微分性质 .....	47
3.4.3 KL 散度的演化与强数据处理不等式.....	48
3.5 本章定理证明.....	49
3.5.1 定理 3.1 的证明 .....	49
3.5.2 引理 3.2 的证明 .....	55
3.5.3 定理 3.2 的证明 .....	56
3.5.4 引理 3.3 的证明 .....	60
3.5.5 引理 3.4 的证明 .....	60
3.5.6 定理 3.3 的证明 .....	63
3.5.7 定理 3.9 的证明 .....	64
3.6 本章小结.....	66
<b>第 4 章 广义相对 Fisher 信息的强数据处理不等式.....</b>	<b>67</b>
4.1 本章引言.....	67
4.2 相对 Rényi-Fisher 信息沿 Fokker-Planck 流的演化行为.....	71
4.2.1 相对 Fisher 信息的强数据处理不等式 .....	71
4.2.2 相对 Rényi-Fisher 信息沿 Fokker-Planck 流的时间导数 .....	74
4.2.3 相对 Rényi-Fisher 信息的强数据处理不等式 .....	76
4.3 相对 $\Phi$ -Fisher 信息沿 Fokker-Planck 流的演化行为.....	81
4.3.1 相对 $\Phi$ -Fisher 信息沿 Fokker-Planck 流的时间导数 .....	82
4.3.2 相对 $\Phi$ -Fisher 信息的强数据处理不等式 .....	83
4.3.3 相对 $\Phi$ -Fisher 信息下近端采样器的混合时间 .....	85
4.4 对数 Hessian 公式与对数 Hessian 比较原理.....	88
4.4.1 热流的对数 Hessian 比较原理 .....	88
4.4.2 OU 流的对数 Hessian 比较原理 .....	94
4.4.3 FP 流的对数 Hessian 比较原理 .....	99
4.5 本章定理证明与技术细节.....	105
4.5.1 定理 4.4 的证明 .....	105
4.5.2 定理 4.5 的证明 .....	113
4.5.3 定理 4.6 的证明 .....	114

4.5.4 计算时间导数的技术性论证 .....	121
4.5.5 定理 4.7 的证明 .....	128
4.5.6 定理 4.8 的证明 .....	131
4.5.7 定理 4.9 的证明 .....	134
4.5.8 定理 4.10 的证明.....	136
4.5.9 定理 4.11 的证明.....	144
4.5.10 定理 4.12 的证明 .....	146
4.5.11 定理 4.13 的证明 .....	148
4.5.12 定理 4.14 的证明 .....	150
4.5.13 推论 4.2 的证明.....	152
4.5.14 引理 4.5 的证明.....	153
4.5.15 定理 4.18 的证明 .....	155
4.5.16 推论 4.7 的证明.....	157
4.5.17 定理 4.21 中势函数的一个例子 .....	158
4.6 本章小结.....	160
<b>第 5 章 基于现代分析方法的率失真理论 .....</b>	<b>161</b>
5.1 本章引言.....	161
5.2 最优弱传输视角下的率失真函数.....	163
5.2.1 最优弱传输理论简介 .....	163
5.2.2 薛定谔桥问题简介 .....	165
5.2.3 最优弱传输视角下的率失真理论 .....	166
5.3 非紧空间下最优重构分布的存在性分析.....	170
5.4 本章定理证明.....	175
5.4.1 引理 5.2 的证明 .....	175
5.4.2 引理 5.3 的证明 .....	176
5.4.3 假设 5.2 不成立的例子 .....	178
5.4.4 定理 5.2 的证明 .....	178
5.4.5 命题 5.1 的证明 .....	179
5.4.6 定理 5.3 的证明 .....	179
5.4.7 定理 5.4 的证明 .....	183
5.4.8 定理 5.7 的证明 .....	184

---

---

5.4.9 引理 5.5 的证明 .....	187
5.4.10 定理 5.8 的证明.....	188
5.5 本章小结.....	193
<b>第 6 章 总结与展望 .....</b>	<b>195</b>
6.1 全文总结.....	195
6.2 未来工作展望.....	197
<b>参考文献.....</b>	<b>201</b>
<b>致 谢.....</b>	<b>213</b>
<b>学术论文和科研成果目录.....</b>	<b>217</b>

## 插图

- 图 4.1 对于  $p = 1$  的  $V(x) = \frac{x^2}{2} + (1 + x^2)^{p/2}$ ,  $\Xi$  和  $\Upsilon$  的示意图。.....159
- 图 4.2 对于  $p = 1$  且  $c = 1.8$  的  $V(x) = \frac{x^2}{2} + (1 + x^2)^{p/2}$ ,  $\tilde{h}$  的二阶导数。.....160



## 符号对照表

$a, \sigma$	模型参数或系数 (Model parameters)
$b(x, t)$	漂移向量 (Drift vector)
$B_n(x, t)$	传播子矩函数 (Propagator moment function)
$B_R(x)$	以 $x$ 为心 $R$ 为半径的球 (Ball)
$c(x)$	薛定谔方程中的势函数项
$c(x, y)$	传输代价函数 (Transport cost function)
$C(x, p)$	弱传输成本泛函 (Weak transport cost functional)
$C_b(\cdot)$	有界连续函数空间
$\text{Cov}(\cdot)$	协方差矩阵 (Covariance matrix)
$d, N$	空间维度 (Dimension)
$d_x, d_y$	空间上的度量 (Metric on spaces)
$D$	扩散张量或失真约束 (Diffusion tensor or Distortion constraint)
$D_{KL}(\cdot\ \cdot)$	相对熵 / KL 散度 (亦作 KL, $H_\nu$ )
$\mathbb{E}$	数学期望 (Expectation)
$\text{FI}(\cdot\ \cdot)$	相对 Fisher 信息 (Relative Fisher Information)
$\text{FI}_\Phi(\cdot\ \cdot)$	相对 $\Phi$ -Fisher 信息 (Relative $\Phi$ -Fisher Information)
$\mathcal{F}_t$	滤子 / 流 (Filtration)
$g(y)$	薛定谔系统中的对偶函数 (Dual potential in Schrödinger system)
$G_\nu(\mu)$	相对熵二阶导数中的势能海森项
$h(X)$	微分熵 (Differential entropy)
$h(x)$	辅助势函数 (Auxiliary potential function)
$H(\mu)$	微分熵 (Differential entropy)
$I$	单位矩阵
$I(X; Y)$	互信息 (Mutual information)
$J(\mu)$	Fisher 信息量
$J(\nu, \beta)$	率失真对偶泛函 (Dual functional in RD)
$K$	归一化常数 (Normalization constant)

$K(x, y, t)$	热核、传播子或转移核 (Kernel / Propagator)
$L, \mathcal{L}$	线性算子或无穷小生成元 (Linear operator / Generator)
$N_t$	泊松过程 (Poisson process)
$P_t, Q_t$	马尔可夫半群 (Markov semigroup)
$\mathcal{P}(\mathcal{X})$	概率测度空间
$\mathcal{P}_t(\mathcal{X})$	具有有限 $t$ 阶矩的概率测度空间
$Q$	OU 过程的漂移矩阵或通用概率测度
$Q(r)$	集中函数 (Concentration function)
$R(D)$	率失真函数 (Rate-Distortion function)
$R_q(\cdot\ \cdot)$	Rényi 散度 (Rényi divergence)
$\text{RFI}_q(\cdot\ \cdot)$	相对 Rényi-Fisher 信息 (Relative Rényi-Fisher Information)
$\mathbb{R}^d, \mathbb{R}^N$	欧几里得空间
$T$	传输映射、终止时刻或足够大的时间
$\text{Tr}(\cdot)$	矩阵的迹 (Trace)
$V(x)$	势函数 (Potential function)
$V_C(\mu, \nu)$	最优弱传输值 (Optimal Weak Transport value)
$\text{Var}(\cdot)$	方差 (Variance)
$W(x)$	变换后的势函数 (Transformed potential)
$W_t$	布朗运动 / 维纳过程 (Brownian motion)
$W_p(\cdot, \cdot)$	$p$ -阶 Wasserstein 距离
$X, Y, Z$	随机变量或随机过程
$\mathcal{X}, \mathcal{Y}$	波兰空间 / 字母表 (Polish spaces / Alphabets)
$\alpha$	强对数凹参数、强凸参数或归一化因子
$\beta$	庞加莱常数或拉格朗日乘子 (Lagrange multiplier)
$\gamma$	联合分布 / 参考测度 (Joint measure / Reference measure)
$\gamma_t$	OU 过程参数或变换后的密度 ( $e^{V/2}\mu_t$ )
$\delta$	狄拉克测度或微小量
$\Delta$	拉普拉斯算子 (Laplacian)
$\epsilon$	任意小的正数
$\lambda$	跳跃强度或集中紧性中的质量极限
$\mu, \nu$	概率测度 (源分布 / 重构分布)

$\Xi$	跳跃幅度随机变量 (Jump amplitude)
$\pi$	传输计划 / 耦合 (Transport plan / Coupling)
$\pi_x$	正则分解 / 条件概率 (Regular disintegration)
$\rho$	失真函数 (Distortion) 或密度比 (Density ratio)
$\sigma$	扩散系数 (Diffusion coefficient)
$\Sigma_t$	协方差矩阵 (Covariance matrix in OU process)
$\tau$	停时 (Stopping time)
$\phi, \psi$	势函数、对偶变量或测试函数
$\Phi(\cdot)$	$\Phi$ -散度生成函数
$\Psi(\cdot \cdot)$	后向/相互相对二阶 Fisher 信息
$\Omega$	空间区域 (Domain) 或样本空间
$\nabla$	梯度算子 (Gradient)
$\nabla^2$	海森矩阵 (Hessian matrix)
$\nabla \cdot$	散度算子 (Divergence)
$\langle \cdot, \cdot \rangle$	欧几里得内积
$\  \cdot \ $	欧几里得范数
$\  \cdot \ _{\text{HS}}$	希尔伯特-施密特范数 (Frobenius norm)
$*$	卷积运算 (Convolution)
$\#$	推送算子 (Pushforward), 如 $T_{\#}\mu$
$\ll$	绝对连续 (Absolutely continuous)
$\succeq$	矩阵半正定序 (Positive semi-definite order)
$\lesssim$	小于等于 (差一个常数因子) / Inequality up to a constant
$\oplus$	函数直和, 如 $(\phi \oplus \psi)(x, y) = \phi(x) + \psi(y)$
$\Pi(\mu, \nu)$	边缘分布为 $\mu, \nu$ 的耦合集合



# 第 1 章 绪论

## 1.1 研究背景及意义

自 1948 年克劳德·香农 (Claude E. Shannon) 发表那篇被誉为“数字时代大宪章”的论文《通信的数学理论》[1] 以来, 信息论便以其惊人的数学完备性和深刻的物理洞察力, 成为了现代通信、数据压缩、统计推断乃至机器学习的基石。香农创造性地将“信息”这一抽象概念量化为“熵 (Entropy)”, 并利用概率论中的大数定律和遍历理论 (Ergodic Theory), 在渐近意义下确立了信源编码定理和信道编码定理。这两大定理划定了通信系统性能的理论极限, 为之后七十余年的通信技术发展指明了方向。

在经典信息论的框架下, 研究对象通常被假设为离散符号序列、无记忆信源或性质良好的高斯过程。其所依赖的数学工具, 主要集中于组合数学、渐近统计学以及经典的概率极限理论。这种“离散与统计”的分析范式在处理传统通信问题时取得了巨大的成功。然而, 随着信息技术向更广阔的领域延伸, 特别是随着大数据分析、高维统计学习以及生成式人工智能的兴起, 信息处理的研究对象和应用场景正经历着一场深刻的变革。

首先, 数据分布的形式日益复杂。现代信息处理不再局限于低维、单模态或具有紧致支撑 (Compact Support) 的简单分布, 而是通过深度神经网络映射到高维流形上, 表现出高度的非线性、多模态以及非对数凹 (Non-log-concave) 特性。其次, 系统模型从静态转向动态。在朗之万动力学 (Langevin Dynamics) 采样、扩散模型生成以及随机控制等前沿领域, 信息不再是静态的存储量, 而是随着随机微分方程 (SDE) 驱动的时间演化而流动的物理量。最后, 度量空间的拓扑结构发生改变。在涉及连续信号的函数空间或非欧几里得几何空间中, 传统的基于欧氏距离或汉明距离的分析方法面临严峻挑战。

在这一全新的时代背景下, 经典信息论的分析工具逐渐显露出其边界与局限。例如, 经典的熵幂不等式 (Entropy Power Inequality, EPI) 虽然在加性高斯噪声信道中成立 [2], 但难以直接推广至一般的非线性扩散过程; 基于 KL 散度 (Kullback-Leibler Divergence) 以及 Wasserstein 距离的收敛性分析虽然成熟, 但在处理需要更强信息度量控制 (如 Rényi 散度或 Fisher 信息等) 的采样算法时仍具有局限性; 而在率失真理论中, 当信源定义在非紧波兰空间 (Unbounded Polish Space) 时, 经典的变分方法因

缺乏紧性而无法保证最优重构分布的存在，导致理论与实际应用的脱节，尤其是经典变分方法中的 Euler-Lagrange 方程依赖于最优解的存在性，为通过 Euler-Lagrange 方程研究最优解的性质，首要的便是证明其存在性。因此，重构信息论的数学基础，将其从传统的组合与概率框架，推进到更加深邃的几何与分析框架，已成为当前学术界的迫切需求。

为了突破上述局限，现代数学观点主张将信息量的演化视为概率测度空间上的动力学行为。这一视角的转变，最早可以追溯到物理学家对 Boltzmann 方程和热力学第二定律的研究。1959 年，Stam 引入了著名的 de Bruijn 恒等式 [2]，揭示了在热流 (Heat Flow) ——即加性高斯噪声干扰下，熵的导数与 Fisher 信息之间的等式关系。这一发现打通了信息论、统计物理与偏微分方程之间的壁垒。

近年来，随着随机分析理论的成熟，这种“动力学视角”被进一步深化。信道不再仅仅被视为一个输入到输出的条件概率核，而是被建模为由 Fokker-Planck 方程描述的随机流 (Stochastic Flow) [3]。在这种视角下，信息量的变化率直接对应于流形的几何曲率和势函数 (Potential Function) 的凸性。这一理论框架的优势在于，它允许我们利用偏微分方程中的强力工具 (如 Bakry, Émery 等人引入的 Gamma 演算 [4]) 来研究信息的耗散规律。

然而，现有的研究大多局限于线性漂移项 (Ornstein-Uhlenbeck 过程) 或简单的对数凹势函数 [5-6]。在更一般的非线性系统中，互信息的凸性变化规律如何？当引入非连续的跳跃项形成跳扩散 (Jump-diffusion) 过程时，信息的微分性质将发生怎样的突变？这些问题不仅具有重要的理论价值，更直接关系到非高斯噪声环境下通信系统的性能评估。

此外，有关强数据处理不等式 (SDPI) 的研究 [7]，也是信息论中极为重要的研究主题。SDPI 旨在量化信息通过信道后的严格收缩系数，这对于证明郎之万蒙特卡罗算法 (Langevin MCMC) [8] 等算法的混合时间 (Mixing Time) 至关重要。虽然针对 KL 散度与 Rényi 散度的 SDPI 已有较多成果 [9-11]，但针对更强的度量——广义相对 Fisher 信息 (如 Rényi-Fisher 和  $\Phi$ -Fisher 信息) 的研究仍相对匮乏。由于广义 Fisher 信息涉及概率密度的高阶导数，其演化方程极其复杂，缺乏有效的几何判据 (如 Log-Hessian 比较原理) 来刻画其收缩行为。填补这一空白，将为高维空间中的高效采样提供坚实的理论保证。

在信息的动态演化之外，信息的静态压缩，即率失真理论，同样面临着现代数学的重构。率失真理论作为有损数据压缩的基石，旨在刻画给定失真约束下的最小编码

速率。自 Shannon 于 1948 年和 1959 年 [12] 提出该理论以来，其研究重心长期集中于离散或连续无记忆信源。直到 1974 年，Csiszár [13] 利用精细的测度论工具与凸分析方法，奠定了抽象字母表上 RD 理论的数学基础，并给出了 RD 函数的参数化表示，为 Blahut-Arimoto [14-15] 等数值算法提供了理论支撑。此后，Rezaei [16] 等学者进一步利用 Stone-Čech 紧化和拓扑向量空间中的不动点定理，放松了对再生空间的紧性假设，将理论推广至更一般的波兰空间。进入 21 世纪，随着 Villani [17] 等人对最优传输理论 (Optimal Transport, OT) 的推动，学术界开始尝试从几何视角重新审视 RD 问题。2013 年，Robert Gray [18] 系统阐述了 OT 与信息论的联系，但这两种理论在数学形式上存在微妙的错位：传统 OT 旨在最小化两个固定边缘分布之间的传输成本，而 RD 理论则是在固定源分布的前提下，寻找最优的条件转移概率 (Disintegration)。为了弥合这一鸿沟，Gozlan [19] 等人提出的最优弱传输 (Optimal Weak Transport, OWT) 理论应运而生。OWT 作为一个更广泛的框架，天然地包含了 RD 问题，并揭示了 RD 函数与随机控制中的薛定谔桥 (Schrödinger Bridge) [20] 问题之间的深刻对偶联系。这种视角的引入，让我们能够利用 Wasserstein 空间中的几何性质来直观理解信息的压缩与表示。

然而，非紧空间中最优解的存在性仍是悬而未决的难题。在经典的紧致空间中，概率测度的弱收敛足以保证变分问题的极值存在。但在一般的非紧波兰空间 (如  $\mathbb{R}^n$  或无穷维函数空间) 中，极小化序列可能会出现病态行为：序列的质量可能会“逃逸”到无穷远处 (Vanishing)，或者分裂成相距甚远的碎片 (Dichotomy)。这种“紧性缺失 (Loss of Compactness)”现象，使得经典的变分证明失效。现有的文献往往通过引入强硬的紧支撑假设来规避这一问题，这极大地限制了理论的普适性。而此前最普适的结论为 Rezaei 等人于 2006 年提出的工作，他们通过弱星拓扑 (Weak-\* Topology) 和不动点方法，在失真函数 (Distortion Function) 为连续函数的情况下，证明了非紧波兰空间中 RD 函数最优解的存在性。然而，对于更一般的非连续型可测失真函数 (如阶梯型增长、存在截断等)，非紧波兰空间中 RD 函数最优解的存在性依然悬而未决。

为解决这一问题，本文引入了非线性泛函分析中的重要工具：集中紧性原理 (Concentration-Compactness Principle) [21-24]。这一由 P.-L. Lions 发展用于解决无界区域上椭圆偏微分方程极值问题的深刻理论，为处理非紧空间中的变分问题提供了一套完整的分析框架。作为完善现代信息论数学基础的关键一步，本文将集中紧性原理引入信息论，通过精细的“三态分析” (Compactness, Vanishing, Dichotomy) 来严格排除病态情形，从而在普适的条件下证明了率失真函数最优解的存在性。

综上所述，本研究立足于现代数学观点，利用随机分析方法、最优传输理论的几何观点以及非线性泛函分析的集中紧性理论，对信息论中的关键问题进行重构与深化。本研究的意义主要体现在以下三个方面：

第一，理论深度的拓展。通过研究 Fokker-Planck 流下的信息几何性质，特别是建立广义 Fisher 信息的 SDPI 理论和 Log-Hessian 比较原理，本文将信息演化理论从线性高斯领域推广至一般的非线性、非高斯乃至跳跃扩散领域，深化了对“信息流”本质的理解。

第二，分析框架的革新。通过引入最优弱传输与集中紧性原理，本文推进了非紧空间中率失真函数存在性这一长期存在的分析难题，为一般化信源的编码理论提供了一套严密、通用的数学分析框架，弥补了经典理论的缺口。

第三，应用价值的支撑。本文的理论成果直接服务于高维数据分析与算法设计。SDPI 的收缩性界为 Proximal Sampler 等前沿采样算法提供了全新的收敛速度估计；而率失真理论在非紧空间下的重构，则攻克了连续变量最优编码策略的存在性难题，为复杂信源的压缩理论补全了关键的数学拼图。

这不仅是一次对信息论基础问题的数学求解，更是一次将现代分析数学的深层力量注入信息科学的尝试。

## 1.2 研究现状及难点

### 1.2.1 基于随机流的信息演化与几何不等式

将信道视为随机动力系统是近年来连接信息论、统计物理与偏微分方程的重要桥梁。Fisher 信息 (Fisher Information, FI) 最早由 R. A. Fisher 引入统计学 [25]，随后被推广为定义在概率测度空间上的非参数泛函。McKean、Villani 等学者 [26-30] 的开创性工作揭示了 Fisher 信息与 Boltzmann 方程熵产生率的深刻联系。而在信息论领域，Stam 引入的 de Bruijn 恒等式与 Guo, Wibisono 等人推广的 I-MMSE 恒等式 [3, 31]，建立了沿着 Fokker-Planck (FP) 流，熵导数与 Fisher 信息的关系。

尽管已有诸多成果，但现有研究仍存在难点：首先，关于互信息凸性的几何判据尚不完善。现有结论多集中于热流或 Ornstein-Uhlenbeck (OU) 流，对于更一般形式的 Fokker-Planck 流，特别是当势函数具有一般几何结构时，互信息随时间演化的二阶性质尚不清晰。此外，如何将信息微分性质推广至包含非连续跳跃项的跳扩散 (Jump-diffusion) 信道，也是完善演化理论边界的重要课题。

其次，针对广义 Fisher 信息的强数据处理不等式 (SDPI) 研究相对滞后。SDPI

是分析采样算法（如 Proximal Sampler）收敛速度的关键工具 [9, 32-34]。虽然针对 KL 散度与 Rényi 散度的 SDPI 已有较多研究，但针对更强的相对 Fisher 信息（及其推广形式如 Rényi-Fisher 和  $\Phi$ -Fisher）的研究仍相对匮乏。如何在一般势函数下建立全新的广义 Fisher 信息的 SDPI，并给出其在统计学等领域中的应用，是当前该领域的主要挑战。

## 1.2.2 率失真理论的现代分析方法

率失真理论关注给定失真约束下的最小编码速率。Csiszár 奠定了抽象字母表上的 RD 理论基础。近年来，Rezaei 等人利用弱星拓扑（Weak-\* Topology）和不动点方法，将 RD 理论推广至波兰空间，解决了紧性假设的部分限制。与此同时，Gozlan 等人提出的最优弱传输（Optimal Weak Transport, OWT）理论，通过优化测度的正则分解（Regular Disintegration），为 RD 函数提供了新的对偶几何视角。

然而，将理论应用于非紧致空间时，仍面临核心的分析学难题：最优重构分布的存在性。在无穷维或无界空间中，极小化序列可能会发生“质量逃逸（Mass Escape/Vanishing）”或“分裂（Dichotomy）”，导致传统的紧性论证失效。现有的工作往往依赖强硬的矩约束或紧支撑假设，缺乏一种通用的分析框架来处理这些病态情形。如何引入非线性泛函分析中的“集中紧性原理”（Concentration-Compactness Principle）等现代分析方法，通过精细分析极小化序列的渐近行为来严格证明最优解的存在性，是完善现代率失真理论数学基础的关键难点。

## 1.3 全文结构以及主要创新点

本文围绕现代数学视角下的信息论关键问题，从动态演化与静态变分两个核心层面展开深入研究。全文共分为六章，结构安排如下：

**第一章：绪论。**首先系统回顾信息论从经典离散与高斯框架向一般化随机过程及抽象空间拓展的历史脉络，指出传统组合与概率方法在处理高维非线性系统与非紧空间时所面临的理论局限。随后，简要介绍本文所采用的现代数学工具——随机分析、最优传输与非线性泛函分析，并阐明本研究在拓展信息演化理论的适用范围、以及解决率失真函数存在性难题方面的科学意义。

**第二章：背景知识。**本章为后续的理论推导奠定数学基础。首先系统地定义本文中使用的信息度量（熵、Fisher 信息、KL 散度等），并回顾对数凹（Log-concavity）性质及其相关的泛函不等式（如 Poincaré 不等式与 Log-Sobolev 不等式）。接着，引

入 Fokker-Planck 方程作为描述概率测度动态演化的核心工具，并简述其与随机微分方程的联系。最后，简要介绍最优传输理论中的 Wasserstein 空间几何结构以及率失真理论的基本定义，为后文的理论分析提供必要的预备知识。

**第三章：Fokker-Planck 流的几何性质与互信息凸性研究。**本章作为动态视角的基石，将深入探讨信息量在随机动力系统演化中的规律。首先，利用 Ishige 等人引入的抛物型偏微分方程中对数凹保持性质的理论，在势函数满足特定凸性条件下，严格证明 Fokker-Planck 方程解的适定性以及对数凹性质在时间演化中的保持性。这一结论为后续研究信息量的单调性提供了几何保障。在此基础上，本章将系统推导互信息沿 Fokker-Planck 流的二阶时间导数解析式，并建立具有一般非线性势函数的扩散过程中互信息凸性的几何判据。该判据不仅将经典的热流与 OU 流情形作为特例包含在内，还可推广至更广泛的强凸势函数情形，揭示出势函数的 Hessian 曲率是决定信息耗散速率的关键因素。此外，为进一步拓展理论的适用范围，本章还将专门探讨引入非连续跳跃项后的跳扩散信道模型，推导出包含非高斯跳跃项的信息微分恒等式，揭示了当随机路径失去连续性时，信息演化规律如何发生改变。

**第四章：广义相对 Fisher 信息的强数据处理不等式 (SDPI)。**本章聚焦于一类比 KL 散度更能反映概率分布高阶几何结构的广义信息度量，旨在解决高维采样算法收敛性分析中的核心难题。首先，在传统的 Shannon 熵框架外，引入并研究了相对 Rényi-Fisher 信息与相对  $\Phi$ -Fisher 信息。通过建立一套精细的演化分析工具，本章推导了这些广义信息度量沿 Fokker-Planck 流的时间导数，并引入了“Log-Hessian 比较原理” (The Log-Hessian Comparison Principle)。该原理以一个矩阵不等式的形式，刻画了目标分布 Hessian (即二阶几何结构) 与信息量收缩速率之间的定量关系。基于扩散半群的显式 Log-Hessian 公式与曲率界估计，本章建立了一套可验证的判别准则，不仅证明了 SDPI 在有限时间后的最终收缩性 (Eventual Contraction)，还进一步证明：当初始分布为高斯分布时，该收缩性在整个时间轴上恒成立 (即“永恒收缩性”)。作为理论应用，本章利用该框架分析了 Proximal Sampler 算法，给出了其在  $\Phi$ -Fisher 信息意义下的混合时间上界，显著优化了现有文献中的收敛速率估计，为采样算法在强范数下的收敛行为提供了新的理论解释。

**第五章：基于现代分析方法的率失真理论。**本章将研究视角由动态演化转向静态压缩，借助变分法与最优传输理论，为率失真 (Rate-Distortion, RD) 理论建立新的分析框架。首先，在最优弱传输理论的框架下，揭示了 RD 函数与随机控制中的薛定谔桥问题之间的对偶联系，导出了 RD 函数在一般波兰空间上的参数化表示。随后，针

对抽象非紧空间中最优重构分布的存在性难题，本章创新性地引入了非线性泛函分析中的“集中紧性原理”（Concentration-Compactness Principle）。传统的变分方法往往因极小化序列的“质量逃逸（Vanishing）”或“分裂（Dichotomy）”而失效（即紧性缺失），而本章通过对序列的三态分析（即集中紧性原理中的三类渐近情形），严格排除了上述病态情形，从而在极其宽泛的条件下（如无界空间、非连续失真度量）证明了最优解的存在性。这一结果有效克服了经典率失真理论在处理非紧致或高维连续信源时的局限，为现代连续信号压缩的数学分析提供了关键的存在性保障。

**第六章：总结与展望。**全面总结全文的理论贡献与创新点，并对非对数凹势函数下的 SDPI、McKean 完全单调猜想等未来研究方向进行展望。

本文的主要创新点归纳如下：

### 1. 第三章：

- **建立了 Fokker-Planck 流下互信息凸性的几何判据与几何保持性理论：**针对信息量在非线性随机动力系统演化中的问题，本文利用抛物型方程理论，严格证明了在势函数满足特定凸性条件下，Fokker-Planck 方程解的稳定性以及对数凹（Log-concavity）几何性质在时间演化中的不变性。在此基础上，推导了互信息沿流的二阶时间导数解析式，突破了经典热流与 Ornstein-Uhlenbeck 流的限制，建立了一般强凸势函数下互信息凸性的几何判据。此外，还将研究边界拓展至跳扩散模型，推导了包含非高斯跳跃项的信息微分恒等式，完善了信道演化理论。

### 2. 第四章：

- **引入了“Log-Hessian 比较原理”：**针对高维概率分布收敛性分析的难题，本文提出了一种基于矩阵不等式的新方法——Log-Hessian 比较原理。该原理深刻揭示了目标分布的二阶几何结构（Hessian）与信息量收缩系数之间的内蕴联系。结合扩散半群的显式公式与曲率界估计，本文建立了一套可验证的判别准则，成功证明了在一般情形下的最终收缩性（Eventual Contraction）以及高斯初始化下的永恒收缩性（Perpetual Contraction），为研究强数据处理不等式（SDPI）提供了全新的数学工具。
- **构建了广义相对 Fisher 信息的 SDPI 框架并优化了采样算法分析：**填补了相对 Fisher 信息类泛函在 SDPI 研究中的空白，将研究对象从传统的 KL 散度拓展至更强的相对 Rényi-Fisher 信息与相对  $\Phi$ -Fisher 信息。通过构建精细的演化演算体系，本文导出了这些广义泛函沿 Fokker-Planck 流的演化

行为。作为理论应用，利用该框架对 Proximal Sampler 算法进行了深入分析，给出了其在  $\Phi$ -Fisher 信息意义下的混合时间上界，证明了算法在更广义信息度量下的收敛性质。

### 3. 第五章:

- **基于最优弱传输理论重构了率失真函数的几何结构与对偶表示:** 从最优传输的几何视角重新审视经典的信源编码问题，在最优弱传输的框架下，揭示了率失真函数与随机控制中的薛定谔桥问题之间的深刻对偶联系。本文推导了 RD 函数在一般波兰空间上的参数化表示，明确了其变分结构。这一工作打通了信息论、最优传输与随机控制之间的壁垒，为理解信息的压缩与表示提供了新的几何视窗。
- **利用集中紧性原理解决了抽象空间率失真理论的存在性难题:** 针对传统变分方法在处理非紧波兰空间（如  $\mathbb{R}^d$  或函数空间）时因“紧性缺失”而失效的问题，本文创新性地引入了非线性泛函分析中的集中紧性原理。通过对率失真泛函的极小化序列进行“紧性、消散、二分”的三态分析，严格排除了质量逃逸与分裂的病态情形。从而在无需强硬紧支撑约束、且允许失真函数非连续（如下半连续、强制增长）的极宽泛条件下，严格证明了最优重构分布的存在性，弥补了经典信息论的理论缺口。

## 第 2 章 背景知识

### 2.1 符号说明

本文主要在  $d$  维欧几里得空间  $\mathbb{R}^d$  ( $d \geq 1$ ) 中展开讨论。对于向量  $u = (u_1, \dots, u_d)$  和  $v = (v_1, \dots, v_d) \in \mathbb{R}^d$ , 我们定义欧几里得内积与范数如下:

$$\langle u, v \rangle = u^\top v = \sum_{i=1}^d u_i v_i, \quad \|u\| = \sqrt{\langle u, u \rangle}. \quad (2.1)$$

对于对称矩阵  $A \in \mathbb{R}^{d \times d}$ , 我们记  $A \succeq 0$  表示  $A$  为半正定矩阵。对于两个对称矩阵  $A, B \in \mathbb{R}^{d \times d}$ , 记号  $A \succeq B$  意味着  $A - B \succeq 0$ 。此外, 我们用  $I$  表示  $d \times d$  维的单位矩阵。

对于任意矩阵  $A \in \mathbb{R}^{d \times d}$ ,  $\text{Tr}(A)$  表示矩阵  $A$  的迹 (trace),  $\|A\|_{\text{HS}}$  表示  $A$  的希尔伯特-施密特范数 (Hilbert-Schmidt norm, 亦称 Frobenius 范数)。相应的希尔伯特-施密特内积则定义为:

$$\langle A, B \rangle_{\text{HS}} := \text{Tr}(A^\top B) \equiv \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} A_{ij} B_{ij}, \quad (2.2)$$

其中  $A, B \in \mathbb{R}^{d_1 \times d_2}$  是两个大小相同的矩阵。给定向量  $u \in \mathbb{R}^d$  和矩阵  $A \in \mathbb{R}^{d \times d}$ , 定义 (伪) 二次型为:

$$\|u\|_A^2 := u^\top A u. \quad (2.3)$$

若  $A \succeq 0$ , 则对任意  $u$  均有  $\|u\|_A^2 \geq 0$ ; 若  $A$  非半正定, 则该量对某些  $u$  可能取负值。值得注意的是,  $\|u\|_A^2$  仅依赖于矩阵  $A$  的对称部分

$$A_{\text{sym}} := \frac{1}{2}(A + A^\top), \quad (2.4)$$

即满足  $\|u\|_A^2 = \|u\|_{A_{\text{sym}}}^2$ 。

对于可微函数  $f: \mathbb{R}^d \rightarrow \mathbb{R}$ , 我们分别用  $\nabla f$  表示  $f$  的梯度 (gradient),  $\nabla^2 f$  表示  $f$  的海森矩阵 (Hessian), 并用  $\Delta f = \text{Tr}(\nabla^2 f)$  表示  $f$  的拉普拉斯算子 (Laplacian)。

对于 (充分光滑的) 向量场  $v: \mathbb{R}^d \rightarrow \mathbb{R}^d$ , 其中  $v(x) = (v_1(x), \dots, v_d(x))$ , 其雅可比矩阵 (Jacobian matrix) 定义为:

$$\nabla v(x) \in \mathbb{R}^{d \times d}, \quad (\nabla v(x))_{ij} = \frac{\partial v_i(x)}{\partial x_j}. \quad (2.5)$$

同时,  $\nabla \cdot v = \text{Tr}(\nabla v)$  表示向量场  $v$  的散度 (divergence)。

对于两个可测函数  $f$  和  $g$ , 若存在常数  $C > 0$  使得对于几乎处处的  $x \in \mathbb{R}^d$  都有  $f(x) \leq Cg(x)$ , 我们记作  $f \lesssim g$ 。

## 2.2 统计量与信息度量

令  $\mathcal{P}_2(\mathbb{R}^d)$  表示  $\mathbb{R}^d$  上关于勒贝格测度绝对连续且具有有限二阶矩的概率测度集合, 即:

$$\mathcal{P}_2(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \|x\|^2 \mu(x) dx < \infty \right\}, \quad (2.6)$$

其中  $\mathcal{P}(\mathbb{R}^d)$  表示  $\mathbb{R}^d$  上所有概率测度构成的空间。我们将测度  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$  与其 (关于勒贝格测度的) 密度函数视为同一对象, 不加区分, 并将其数学期望记为  $\mathbb{E}_\mu[\cdot] = \int_{\mathbb{R}^d} (\cdot) \mu(x) dx$ 。对于服从分布  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$  的随机变量  $X$ , 我们分别用  $\text{Var}(X)$  和  $\text{Cov}(X)$  表示  $X$  的方差和协方差矩阵, 定义如下:

$$\begin{aligned} \text{Var}(X) &:= \mathbb{E}_\mu[\|X - \mathbb{E}_\mu[X]\|^2] = \mathbb{E}_\mu[X^\top X] - \mathbb{E}_\mu[X]^\top \mathbb{E}_\mu[X], \\ \text{Cov}(X) &:= \mathbb{E}_\mu[(X - \mathbb{E}_\mu[X])(X - \mathbb{E}_\mu[X])^\top]. \end{aligned} \quad (2.7)$$

此外, 对于函数  $\phi \in \mathbb{L}^2(\mu)$ , 定义其方差为:

$$\text{Var}_\mu(\phi) := \int \|\phi\|^2 d\mu - \left\| \int \phi d\mu \right\|^2. \quad (2.8)$$

对于两个测度  $\mu$  和  $\nu$ , 记号  $\mu \ll \nu$  表示  $\mu$  关于  $\nu$  绝对连续 (即对于任意可测集  $A$ , 若  $\nu(A) = 0$  则必有  $\mu(A) = 0$ )。(香农) 微分熵泛函  $H : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R} \cup \{-\infty\}$  定义为:

$$H(\mu) := -\mathbb{E}_\mu[\log \mu] = -\int_{\mathbb{R}^d} \mu(x) \log \mu(x) dx, \quad (2.9)$$

前提是该积分定义良好。

(非参数) Fisher 信息量 (Fisher information) 泛函  $J : \mathcal{P}(\mathbb{R}^d) \rightarrow [0, \infty]$  定义为:

$$J(\mu) := \mathbb{E}_\mu[\|\nabla \log \mu\|^2] = \int_{\mathbb{R}^d} \mu(x) \|\nabla \log \mu(x)\|^2 dx. \quad (2.10)$$

依照惯例, 若  $\mu$  不存在 (充分正则的) 密度函数, 则规定  $J(\mu) = +\infty$ 。当  $\mu$  光滑且在无穷远处衰减足够快时, 利用分部积分法可得如下等价公式:

$$J(\mu) = -\mathbb{E}_\mu[\Delta \log \mu]. \quad (2.11)$$

对于两个具有密度函数的测度  $\mu \ll \nu$ , Kullback-Leibler (KL) 散度定义为:

$$H_\nu(\mu) \equiv \text{KL}(\mu \parallel \nu) \equiv D_{KL}(\mu \parallel \nu) := \mathbb{E}_\mu \left[ \log \frac{\mu}{\nu} \right] = \int_{\mathbb{R}^d} \mu(x) \log \frac{\mu(x)}{\nu(x)} dx. \quad (2.12)$$

若  $\mu \not\ll \nu$ , 则按惯例定义  $\text{KL}(\mu \parallel \nu) = +\infty$ 。

当密度函数可微时,  $\mu$  相对于  $\nu$  的相对 Fisher 信息量定义为:

$$J_\nu(\mu) \equiv \text{FI}(\mu \parallel \nu) := \mathbb{E}_\mu \left[ \left\| \nabla \log \frac{\mu}{\nu} \right\|^2 \right] = \int_{\mathbb{R}^d} \mu(x) \left\| \nabla \log \frac{\mu(x)}{\nu(x)} \right\|^2 dx. \quad (2.13)$$

若上述积分在通常意义下无定义, 则令其为  $+\infty$ 。

此外我们定义  $\mu$  相对于  $\nu$  的相对二阶 Fisher 信息量 (Relative Second-order Fisher Information) 为:

$$K_\nu(\mu) = \int_{\mathbb{R}^d} \mu \left\| \nabla^2 \log \frac{\mu}{\nu} \right\|_{\text{HS}}^2 dx. \quad (2.14)$$

进一步地, 基于上述信息量, 我们可以定义它们的相互形式 (Mutual Version) [6]:

**定义 2.1** 给定一个关于随机变量  $Y \sim \mu_Y$  的泛函  $F(Y) \equiv F(\mu_Y)$ , 我们可以为联合随机变量  $(X, Y) \sim \mu_{XY}$  定义其相互形式  $F(X; Y)$  如下:

$$F(X; Y) = F(Y|X) - F(Y) \quad (2.15)$$

其中  $F(Y|X) = \int_{\mathbb{R}^d} \mu_X(x) F(\mu_{Y|X}(\cdot|x)) dx$  表示泛函  $F$  在条件随机变量  $Y|\{X=x\} \sim \mu_{Y|X}(\cdot|x)$  上的期望, 该期望是对  $X \sim \mu_X$  进行平均得到的。

值得注意的是, 相互形式仅捕捉了泛函中的非线性成分。因此, 若两个泛函仅相差一个线性项, 它们将拥有相同的相互形式。

**定义 2.2** 互信息 (Mutual Information) 等于负熵  $-H(\mu)$  或相对熵  $H_\nu(\mu)$ , 的相互形式, 即:

$$I(X; Y) = H(Y) - H(Y|X) = H_\nu(X; Y). \quad (2.16)$$

相互相对 Fisher 信息 (Mutual Relative Fisher Information) 定义为:

$$J_\nu(X; Y) = J_\nu(Y|X) - J_\nu(Y) \quad (2.17)$$

相互相对二阶 Fisher 信息 (Mutual Relative Second-order Fisher Information) 定义为:

$$K_\nu(X; Y) = K_\nu(Y|X) - K_\nu(Y). \quad (2.18)$$

需要指出的是,  $J_\nu(X; Y)$  等价于后向 Fisher 信息 (Backward Fisher Information), 或称为统计 Fisher 信息 (Statistical Fisher Information)  $\Phi(X|Y)$ , 其定义为 [3]:

$$\Phi(X|Y) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mu_{XY} \left\| \nabla_y \log \mu_{X|Y} \right\|^2 dx dy. \quad (2.19)$$

因此,  $J_\nu(X; Y)$  始终是非负的。然而, 相互相对二阶 Fisher 信息  $K_\nu(X; Y) = K_\nu(Y|X) - K_\nu(Y)$  与  $J_\nu(X; Y)$  不同, 它有可能取到负值。

此外,  $\text{KL}(\cdot \| \nu)$  实际上是与泛函  $-H(\cdot)$  相关的 Bregman 散度; 类似地,  $\text{FI}(\cdot \| \nu)$  在形式上也可被解释为与 Fisher 信息量泛函  $J(\cdot)$  相关的类 Bregman 散度。

设  $\mu$  和  $\nu$  为两个概率测度且  $\mu \ll \nu$ , 定义  $\mu$  相对于  $\nu$  的  $q$  阶 Rényi 散度 (其中  $q \geq 0$ ) 为:

$$R_q(\mu \| \nu) := \begin{cases} \frac{1}{q-1} \log \int \rho^q d\nu, & q \neq 1, q \in (0, \infty), \\ \text{KL}(\mu \| \nu), & q = 1, \\ \log \|\rho\|_{\mathbb{L}^\infty(\nu)}, & q = +\infty \\ -\log \nu(\text{supp}(\mu)), & q = 0, \end{cases} \quad (2.20)$$

其中  $\rho = \frac{d\mu}{d\nu}$  为 Radon-Nikodym 导数。若  $\mu \not\ll \nu$ , 对于  $q \geq 1$  定义  $R_q(\mu \| \nu)$  为  $+\infty$ ; 而对于  $q < 1$ , 定义

$$R_q(\mu \| \nu) := \frac{1}{q-1} \log \int \left( \frac{d\mu}{d\lambda} \right)^q \left( \frac{d\nu}{d\lambda} \right)^{1-q} d\lambda, \quad (2.21)$$

其中  $\lambda$  是  $\mu$  和  $\nu$  的公共控制测度 (Common Dominating Measure) (例如  $\lambda = \frac{\mu + \nu}{2}$ )。

若  $\mu$  和  $\nu$  具有可微的密度函数, 则  $\mu$  相对于  $\nu$  的相对 Rényi-Fisher 信息量定义为:

$$\text{RFI}_q(\mu \| \nu) := q \left( \int \rho^{q-1} d\mu \right)^{-1} \left( \int \rho^{q-1} \|\nabla \log \rho\|^2 d\mu \right). \quad (2.22)$$

令  $\mathbb{R}_+ := [0, +\infty)$  及  $\mathbb{R}_{++} := (0, +\infty)$ 。对于任意满足  $\Phi(1) = 0$  的凸函数  $\Phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ , 若  $\mu \ll \nu$ , 定义概率测度  $\mu$  和  $\nu$  之间的  $\Phi$ -散度为:

$$D_\Phi(\mu \| \nu) := \mathbb{E}_\nu [\Phi(\rho)], \quad (2.23)$$

若  $\mu \not\ll \nu$  则取值为  $+\infty$ 。此外, 定义相对  $\Phi$ -Fisher 信息量  $\text{FI}_\Phi(\mu \| \nu)$  为:

$$\text{FI}_\Phi(\mu \| \nu) := \mathbb{E}_\nu \left[ \|\nabla \rho\|^2 \Phi''(\rho) \right] = \mathbb{E}_\mu \left[ \rho \|\nabla \log \rho\|^2 \Phi''(\rho) \right]. \quad (2.24)$$

设  $U$  为  $\mathbb{R}$  或  $\mathbb{R}_{++}$ 。若  $\Phi \in C(\bar{U}) \cap C^4(U)$  且满足以下条件:

$$\begin{aligned} \Phi'' &\neq 0, \quad \Phi'' \geq 0 \quad \text{在 } U \text{ 上,} \\ (\Phi''')^2 &\leq \frac{\theta}{2} \Phi'' \Phi^{(4)} \quad \text{在 } U \text{ 上,} \end{aligned} \quad (2.25)$$

其中  $\theta > 0$ ，则称函数  $\Phi$  为对应  $\Phi$ -散度的  $\theta$ -容许生成函数 ( $\theta$ -admissible generating function)。在  $\theta = 1$  的特例中，我们简称  $\Phi$  为容许生成函数，并参照 [35] 的术语，将导出的散度称为容许相对熵 (admissible relative entropy)。

**例 2.1** ( $\theta$ -容许生成函数示例) 考虑函数  $f(x) := \Phi''(x) = (1+x)^{-m}, x > 0, m > 0$ ，我们有：

$$\frac{(f')^2}{ff''} = \frac{m^2(1+x)^{-2m-2}}{m(m+1)(1+x)^{-2m-2}} = \frac{m}{m+1} \leq \frac{\theta}{2}. \quad (2.26)$$

由此可导出一类  $\theta$ -容许生成函数：

- 若  $\theta \in (0, 2)$ ，可取  $m \in (0, \frac{\theta}{2-\theta}]$ 。
- 若  $\theta \geq 2$ ，则对任意  $m > 0$  均成立。

通过对  $f$  进行两次积分，我们可以得到  $\Phi$  的表达式：

- 若  $m \neq 1, 2$ ，则  $\Phi(x) = \frac{(1+x)^{2-m}}{(1-m)(2-m)} + C_1x + C_0$ ，且满足  $\Phi(1) = 0$ 。
- 若  $m = 1$ ，则  $\Phi(x) = (x+1)\log(1+x) + C_1x + C_0$ ，且满足  $\Phi(1) = 0$ 。
- 若  $m = 2$ ，则  $\Phi(x) = -\log(1+x) + C_1x + C_0$ ，且满足  $\Phi(1) = 0$ 。

一个典型的例子是 KL 散度，其对应于参数  $\theta = 1$  的生成函数  $\Phi(x) = x \log x$ 。

### 2.3 对数凹性质与泛函不等式

若对于所有的  $x, y \in \mathbb{R}^d$  和  $t \in [0, 1]$ ，函数  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  满足

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\alpha}{2}t(1-t)\|x-y\|^2, \quad (2.27)$$

则称  $f$  是  $\alpha$ -强凸的 ( $\alpha$ -strongly convex, 简称  $\alpha$ -SC)。若进一步假设  $f \in C^2(\mathbb{R}^d)$ ，则该条件等价于海森矩阵的下界条件： $\nabla^2 f(x) \succeq \alpha I, \forall x \in \mathbb{R}^d$ 。

对于函数  $f \in C^2(\mathbb{R}^d)$ ，若存在常数  $M > 0$  使得  $-MI \preceq \nabla^2 f(x) \preceq MI, \forall x \in \mathbb{R}^d$ ，则称  $f$  是  $M$ -光滑的 ( $M$ -smooth)。

若概率测度  $\nu \in \mathcal{P}(\mathbb{R}^d)$  的负对数密度函数  $-\log \nu$  是  $\alpha$ -强凸的 (其中  $\alpha > 0$ )，则称  $\nu$  是  $\alpha$ -强对数凹的 ( $\alpha$ -strongly log-concave, 简称  $\alpha$ -SLC)。等价地，若  $\log \nu$  二阶可微，这意味着：

$$-\nabla^2 \log \nu(x) \succeq \alpha I, \quad \forall x \in \mathbb{R}^d. \quad (2.28)$$

若  $-\log \nu$  是凸函数，即对所有  $x \in \mathbb{R}^d$  有  $-\nabla^2 \log \nu(x) \succeq 0$ ，则称  $\nu$  是 (弱) 对数凹的。此外，若存在  $0 < L < \infty$  使得

$$-LI \preceq -\nabla^2 \log \nu(x) \preceq LI, \quad \forall x \in \mathbb{R}^d, \quad (2.29)$$

则称  $\nu$  是  $L$ -对数光滑的 ( $L$ -log-smooth)。进一步, 若对于严格非负函数  $\lambda(x)$  满足  $-\nabla^2 \log \mu(x) \succeq \nabla^2 \lambda(x)$ , 则称  $X \sim \mu$  是  $\lambda$ -相对对数凹的。

此处列举本文用到的 (强) 对数凹性质的一些重要特性:

1. 正态分布及多元正态分布均属于对数凹分布。
2. 卷积操作保持对数凹性。即, 若  $f$  和  $g$  是对数凹的, 则其卷积

$$(f * g)(x) = \int f(x - y)g(y)dy \quad (2.30)$$

也是对数凹的。

3. 对数凹性质在依分布收敛下保持不变 [36]。
4. 若随机变量  $Z = X + Y$ , 其中  $X$  是  $\alpha$ -SLC 的,  $Y$  是  $\beta$ -SLC 的, 且  $X$  与  $Y$  相互独立 (假设两者维数相同), 则  $Z$  是  $(\alpha^{-1} + \beta^{-1})^{-1}$ -SLC 的。
5. 若随机变量  $X$  是  $\alpha$ -SLC 的, 则对于任意常数  $k > 0$ , 随机变量  $kX$  是  $k^{-2}\alpha$ -SLC 的。

关于 (强) 对数凹性更详细的保持性质可参见文献 [37]。

若对于所有的  $\mu \ll \nu$ , 测度  $\nu$  满足如下不等式:

$$\text{FI}(\mu \parallel \nu) \geq 2\alpha \text{KL}(\mu \parallel \nu), \quad (2.31)$$

则称  $\nu$  满足  $\alpha$ -对数索伯列夫不等式 ( $\alpha$ -log-Sobolev inequality, 简称  $\alpha$ -LSI)。根据 Bakry-Émery 准则 [38],  $\alpha$ -SLC 性质蕴含  $\alpha$ -LSI 性质。值得注意的是, 即便在 SLC 条件下, KL 散度与相对 Fisher 信息量之间的差距也可能任意大, 具体反例可见 [10] 附录 D 中的 Lemma 9。

若对于所有光滑函数  $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}$ , 测度  $\nu$  满足

$$\text{Var}_\nu(\varphi) \leq \frac{1}{\beta} \mathbb{E}_\nu[\|\nabla \varphi\|^2], \quad (2.32)$$

则称  $\nu$  满足  $\beta$ -庞加莱不等式 ( $\beta$ -Poincaré inequality, 简称  $\beta$ -PI)。等价地, 对于光滑向量场  $\psi: \mathbb{R}^d \rightarrow \mathbb{R}^d$ , 该不等式表述为:

$$\text{Var}_\nu(\psi) \leq \frac{1}{\beta} \mathbb{E}_\nu[\|\nabla \psi\|_{\text{HS}}^2]. \quad (2.33)$$

值得注意的是,  $\alpha$ -LSI 蕴含  $\alpha$ -PI [17]。

## 2.4 随机分析基础与 Fokker-Planck 方程

本章将系统地回顾随机分析的核心理论，为后续章节的研究奠定数学基础。我们将从布朗运动的定义出发，构建伊藤 (Itô) 随机积分理论，进而讨论随机微分方程及其解的性质，并最终推导描述概率密度演化的 Fokker-Planck 方程。本章的主要理论框架参考自 [39-43]。

### 2.4.1 伊藤积分 (The Itô Integral)

1828 年，植物学家 Robert Brown 在显微镜下观察悬浮在水中的花粉微粒时，发现它们呈现出一种奇特的、持续的无规则运动 [44]。随后的物理学研究，特别是 Albert Einstein 在 1905 年发表的开创性论文 [45]，从理论上阐明了这种所谓的“布朗运动”源于流体分子对微粒的频繁随机碰撞。在数学领域，关于这一现象的理论研究最早由法国数学家 Louis Bachelier 开启 [46]。自诞生以来，布朗运动已成为概率论的核心基石，被广泛应用于热传导建模、金融衍生品定价、偏微分方程求解、流形几何性质探索，以及复杂概率分布的采样算法中。

尽管布朗运动作为数学对象的存在性并非显而易见<sup>①</sup>，但为了聚焦于随机微积分的构建，我们在此直接给出其标准定义。

**定义 2.3** (标准布朗运动) 称一个取值于  $\mathbb{R}^d$  的随机过程  $\{W_t\}_{t \geq 0}$  为布朗运动，如果它满足以下四条性质：

- $W_0 = 0$ ;
- (增量独立性) 对于任意时间序列  $0 < t_1 < \dots < t_k$ ，随机变量组  $(W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_k} - W_{t_{k-1}})$  是相互独立的；
- (增量分布) 对于任意  $0 \leq s < t$ ，增量  $W_t - W_s$  服从均值为 0、协方差矩阵为  $(t-s)I_d$  的高斯分布  $\mathcal{N}(0, (t-s)I_d)$ ；
- (路径连续性) 几乎必然地 (Almost surely, a.s.)，样本路径  $t \mapsto W_t$  是连续函数。

我们本节的首要目标是建立随机微积分的语言，特别是赋予涉及布朗运动的积分以严格的数学意义：给定  $\mathbb{R}$  上的随机过程  $(\eta_t)_{t \geq 0}$ ，我们要如何理解形如  $\int_0^T \eta_t dW_t$  的表达式？一旦掌握了随机积分这一工具，我们便可以形式化并求解随机微分方程。

定义随机积分的主要技术障碍在于布朗运动路径的极度不规则性：对于微小的  $t > 0$ ，根据定义有  $W_t \sim \mathcal{N}(0, t)$ ，这意味着  $|W_t|$  的典型量级是  $\sqrt{t}$ 。这种特性导致布

<sup>①</sup> 布朗运动的数学存在性由 Norbert Wiener 严格建立 [47]，因此该过程常被称为维纳过程 (Wiener process)。但在本节中，我们将略去其具体的构造过程，直接给出其公理化定义。

朗运动在  $t = 0$  处（事实上是在任何处）都不可微，且其路径具有无穷大的全变差。因此，经典的黎曼-斯蒂尔杰斯（Riemann-Stieltjes）积分理论在此失效。尽管如此，我们依然可以定义有意义的随机积分，其主要步骤概述如下。

我们在一个完备的、带流的（Filtered）概率空间  $(\Omega, \mathcal{F}, \mathbb{P})$  上进行讨论。假设流（也可译为滤子） $(\mathcal{F}_t)_{t \geq 0}$  满足右连续性，即对于所有  $s \geq 0$ ，有  $\bigcap_{t > s} \mathcal{F}_t = \mathcal{F}_s$ ，且  $\mathcal{F}_0$  包含所有零测集的子集。同时，我们假设布朗运动是关于该流适应的（Adapted），即对于每个  $t \geq 0$ ， $W_t$  是  $\mathcal{F}_t$ -可测的。

定义伊藤积分的标准范式遵循“由简入繁”的逼近思想：首先针对简单过程定义积分，然后将其推广到更一般的过程类。最简单的过程是分段常数过程，我们称之为简单过程（Elementary Process），其形式如下：

$$\eta_t = \sum_{i=0}^{k-1} H_i \mathbf{1}_{[t_i, t_{i+1})}(t) \quad (2.34)$$

其中  $0 \leq t_0 < t_1 < \dots < t_k$ ，且系数  $H_i$  是有界随机变量并关于  $\mathcal{F}_{t_i}$  可测。对于此类过程，随机积分最合理的定义无疑是黎曼和形式：

$$\int_0^T \eta_t dW_t := \sum_{i=0}^{k-1} H_i (W_{t_{i+1} \wedge T} - W_{t_i \wedge T}). \quad (2.35)$$

虽然这是我们要采用的定义，但为了避免混淆，暂时我们不使用积分符号，而是将其记为算子  $\mathcal{I}_{[0, T]}(\eta)$ 。我们需要记录该随机积分的两个核心性质。

首先，过程  $t \mapsto \mathcal{I}_{[0, t]}(\eta)$  是一个连续鞅（Martingale）。也就是它不仅连续，还满足以下定义。

**定义 2.4** 称过程  $(M_t)_{t \geq 0}$  为关于滤子  $(\mathcal{F}_t)_{t \geq 0}$  的鞅，如果对于所有  $0 \leq s < t$ ，满足  $M_s$  是  $\mathcal{F}_s$ -可测且可积的，并且满足鞅性质：

$$\mathbb{E}[M_t | \mathcal{F}_s] = M_s.$$

事实上，我们可以从以下两个事实推导出  $(\mathcal{I}_{[0, t]}(\eta))_{t \geq 0}$  是鞅：(1) 对于每个  $i$ ， $H_i$  是  $\mathcal{F}_{t_i}$ -可测的；(2) 布朗运动  $(W_t)_{t \geq 0}$  本身是一个鞅。

第二个核心性质是我们可以显式计算其方差。这是伊藤积分最重要的等距性质：

$$\begin{aligned} \mathbb{E}[\mathcal{I}_{[0, T]}(\eta)^2] &= \mathbb{E} \left[ \left[ \sum_{i=0}^{k-1} H_i (W_{t_{i+1} \wedge T} - W_{t_i \wedge T}) \right]^2 \right] \\ &= \mathbb{E} \left[ \sum_{i=0}^{k-1} H_i^2 (t_{i+1} \wedge T - t_i \wedge T) \right] \\ &= \mathbb{E} \left[ \int_0^T \eta_t^2 dt \right]. \end{aligned} \quad (2.36)$$

在此推导中，我们利用了布朗运动定义中的基本性质，如增量的独立性与方差。该方程表明：如果定义积测度  $\mathbb{P}_T := \mathbb{P} \otimes \mathcal{L}^1|_{[0,T]}$ （其中  $\mathcal{L}^1|_{[0,T]}$  是  $[0, T]$  上的 Lebesgue 测度），那么映射  $\eta \mapsto \mathcal{I}_{[0,T]}(\eta)$  是从  $\mathbb{L}^2(\mathbb{P}_T)$  到  $\mathbb{L}^2(\mathbb{P})$  的等距同构（Isometry），即满足  $\|\mathcal{I}_{0,T}(\eta)\|_{\mathbb{L}^2(\mathbb{P})} = \|\eta\|_{\mathbb{L}^2(\mathbb{P}_T)}$ 。

正是通过这一等距同构关系，我们可以将随机积分的定义从简单过程推广到简单过程的闭包。

这个闭包由循序可测（Progressive）<sup>①</sup> 且满足如下可积性条件的过程  $(\eta_t)_{t \geq 0}$  组成：

$$\|\eta\|_{\mathbb{L}^2(\mathbb{P}_T)}^2 = \mathbb{E} \int_0^T \eta_t^2 dt < \infty. \quad (2.37)$$

这意味着，对于任意满足上述条件的  $(\eta_t)_{t \geq 0}$ ，都可以被一系列形如 (2.34) 的简单过程  $\{(\eta_t^{(k)})_{t \geq 0} : k \in \mathbb{N}\}$  在  $\mathbb{L}^2(\mathbb{P}_T)$  范数下逼近。对于每个  $k$ ，通过 (2.35) 定义随机积分  $\mathcal{I}_{[0,T]}(\eta^{(k)})$ ，得益于等距性质，序列  $\mathcal{I}_{[0,T]}(\eta^{(k)})$  在  $\mathbb{L}^2(\mathbb{P})$  中存在极限。我们将此极限定义为随机积分  $\mathcal{I}_{[0,T]}(\eta)$ 。

上述过程虽然成功定义了固定时刻  $t$  的积分  $\mathcal{I}_{[0,t]}(\eta)$ ，但并未保证不同时刻  $t$  之间的连贯性。问题在于  $\mathcal{I}_{[0,t]}(\eta)$  是作为极限定义的，该极限仅在忽略零测集的意义下唯一确定，而不同时刻的零测集可能会累积成非零测集。这是不可接受的，因为随机微积分的真正威力在于将随机积分视为一个以时间为索引的随机过程。

解决这一问题的关键见解（源于伊藤清 [48]）是回到逼近序列  $\{(\eta_t^{(k)})_{t \geq 0}\}$ 。对于每个  $k$ ，积分  $t \mapsto \mathcal{I}_{[0,t]}(\eta^{(k)})$  定义了一个完整的随机过程，且该过程是连续鞅。应用强大的鞅收敛定理，我们可以得到如下定理：

**定理 2.1** 假设  $(\eta_t)_{t \geq 0}$  是循序可测过程且满足  $\mathbb{E} \int_0^T \eta_t^2 dt < \infty$ 。则存在一个适应于  $(\mathcal{F}_t)_{t \geq 0}$  的连续鞅，记为  $(\int_0^t \eta_s dW_s)_{t \geq 0}$ ，满足：

$$\mathbb{E} \left[ \left| \int_0^t \eta_s dW_s \right|^2 \right] = \mathbb{E} \int_0^t \eta_s^2 ds, \quad \forall t \in [0, T]. \quad (2.38)$$

公式 (2.38) 被称为伊藤等距（Itô isometry）。此外，对于每个  $t \in [0, T]$ ，几乎必然成立：

$$\int_0^t \eta_s dW_s = \mathcal{I}_{[0,t]}(\eta). \quad (2.39)$$

为了构建完备的理论，通常还需要最后一步扩展，即允许被积函数  $\eta$  属于更广泛的循序可测过程类，仅需满足几乎必然（a.s.）可积条件：

$$\int_0^T \eta_t^2 dt < \infty, \quad a.s. \quad (2.40)$$

① 若过程  $(\eta_t)_{t \geq 0}$  满足：对任意  $T \geq 0$ ，映射  $(\omega, t) \mapsto \eta_t(\omega)$  关于  $\mathcal{F}_T \otimes \mathcal{B}_{[0,T]}$  是可测的，则称其为循序可测的。其中  $\mathcal{B}_{[0,T]}$  是  $[0, T]$  上的 Borel  $\sigma$ -代数。

注意，此条件比  $\mathbb{E} \int_0^T \eta_t^2 dt < \infty$  要弱得多。这种扩展在数学上极具意义，因为我们希望定义的适用范围尽可能广。但同样重要的是，它引入了局部化（Localization）这一技术手段。总体而言，局部化实际上减少了技术上的繁琐：一旦引入该机制，我们总是可以利用停时（Stopping Time）在过程表现良好的时间段内（例如有界时）进行操作。此外，叙述定理2.3中最实用的伊藤公式的版本也需要局部化技术。

**定义 2.5** 关于滤子  $(\mathcal{F}_t)_{t \geq 0}$  的停时是一个随机变量  $\tau : \Omega \rightarrow [0, \infty]$ ，满足对于所有  $t \geq 0$ ，事件  $\{\tau \leq t\}$  属于  $\mathcal{F}_t$ 。

**定义 2.6** 称递增的停时序列  $(\tau_n)_{n \geq 1}$  为  $\eta$  在  $[0, T]$  上的局部化序列，如果满足：

- 对于所有  $n \in \mathbb{N}$ ，被停止的过程  $(\eta_t \mathbf{1}_{[0, \tau_n]}(t))_{t \geq 0}$  具有有限的  $\mathbb{L}^2(\mathbb{P}_T)$  范数；
- 当  $n \rightarrow \infty$  时， $\tau_n \rightarrow T$  几乎必然成立。

事实上，上述局部化序列通常很容易构造。

**命题 2.1** 假设  $(\eta_t)_{t \geq 0}$  是循序可测的，且满足  $\int_0^T \eta_t^2 dt < \infty$  a.s.。则由下式定义的序列  $(\tau_n)_{n \geq 1}$  构成了  $\eta$  在  $[0, T]$  上的局部化序列：

$$\tau_n := \inf \left\{ t \geq 0 : \int_0^t \eta_s^2 ds \geq n \right\} \wedge T. \quad (2.41)$$

有了局部化序列，我们可以如下定义  $\eta$  的伊藤积分。对于满足  $\int_0^T \eta_t^2 dt < \infty$  a.s. 的循序可测过程  $\eta$ ，设  $(\tau_n)_{n \geq 1}$  为其局部化序列。对于每个  $n$ ，根据定义，我们可以对停止过程  $(\eta_{t \wedge \tau_n})_{t \geq 0}$  应用现有的伊藤积分定义，从而得到连续鞅：

$$t \mapsto \int_0^t \eta_s \mathbf{1}_{[0, \tau_n]}(s) dW_s. \quad (2.42)$$

随之，我们将  $\eta$  的伊藤积分定义为过程 (2.42) 的极限。此外，我们还可以利用局部化序列定义鞅的类比概念：这些过程几乎是鞅，但缺乏所需的可积性，称为局部鞅。

**定义 2.7** 称过程  $(M_t)_{t \geq 0}$  为局部鞅（Local Martingale），如果它适应于滤子  $(\mathcal{F}_t)_{t \geq 0}$ ，且存在递增停时序列  $(\tau_n)_{n \geq 1}$  使得  $\tau_n \rightarrow \infty$ ，并且对每个  $n$ ，停止过程  $t \mapsto M_{t \wedge \tau_n} - M_0$  是关于  $(\mathcal{F}_t)_{t \geq 0}$  的鞅。

**命题 2.2** 对于任意满足  $\int_0^T \eta_t^2 dt < \infty$  a.s. 的循序可测过程  $(\eta_t)_{t \geq 0}$ ，通过局部化定义的伊藤积分  $(\int_0^t \eta_s dW_s)_{t \geq 0}$  是一个连续局部鞅。

作为一个具体的例子，考虑  $t \geq 0$  时的过程  $\eta_t = \exp(W_t^{2025})$ 。对于较大的  $T \geq 0$ ，该过程并不满足  $\mathbb{E} \int_0^T \eta_t^2 dt < \infty$  的条件，但我们依然可以使用局部化序列定义其伊藤积分  $\int_0^T \eta_t dW_t$ 。通常情况下，局部化的主要优势在于我们在进行计算时，无需预先验证繁琐的  $\mathbb{L}^2$  条件。

## 2.4.2 伊藤公式 (Itô's Formula)

掌握了伊藤积分后, 我们进一步考虑以下一类重要的随机过程。

**定义 2.8** 称  $\mathbb{R}^d$  上的随机过程  $(X_t)_{t \geq 0}$  为伊藤过程 (Itô process), 如果它可以表示为:

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \quad (2.43)$$

其中漂移系数  $(b_t)_{t \geq 0}$  是  $\mathbb{R}^d$  值循序可测过程, 满足  $\int_0^T \|b_t\| dt < \infty$  a.s.; 扩散系数  $(\sigma_t)_{t \geq 0}$  是  $\mathbb{R}^{d \times N}$  值循序可测过程, 满足  $\int_0^T \|\sigma_t\|_{\text{HS}}^2 dt < \infty$  a.s.;  $(W_t)_{t \geq 0}$  是  $N$  维布朗运动。此外, 随机变量  $X_0$  需为  $\mathcal{F}_0$ -可测, 此时过程  $(X_t)_{t \geq 0}$  也是循序可测的。

当漂移系数为零时,  $(X_t)_{t \geq 0}$  简化为伊藤积分, 因此是连续局部鞅。若漂移系数非零, 该过程不再必然是局部鞅, 而是成为半鞅 (Semi-martingale)。为了定义半鞅, 我们需要引入有限变差和二次变差的概念。

**定义 2.9** 称连续过程  $(A_t)_{t \in [0, T]}$  具有有限变差 (Finite Variation), 如果对于任意划分  $0 = t_0 < t_1 < \dots < t_n = T$ , 定义划分的模为  $\text{mesh}(t_i) := \max_{1 \leq i \leq n} |t_i - t_{i-1}|$ , 则全变差满足:

$$\lim_{\text{mesh}(t_i) \rightarrow 0} \sum_{i=1}^n \|A_{t_i} - A_{t_{i-1}}\| < \infty. \quad (2.44)$$

**定义 2.10** 称连续过程  $(A_t)_{t \in [0, T]}$  具有有限二次变差 (Finite Quadratic Variation), 如果对于任意划分, 当划分的模趋于 0 时, 满足:

$$\lim_{\text{mesh}(t_i) \rightarrow 0} \sum_{i=1}^n \|A_{t_i} - A_{t_{i-1}}\|^2 < \infty. \quad (2.45)$$

**定理 2.2** 设  $(M_t)_{t \in [0, T]}$  为连续局部鞅。则存在几乎必然唯一的递增过程  $t \mapsto [M, M]_t$ , 使得  $t \mapsto M_t^2 - [M, M]_t$  为连续局部鞅。此外, 若对于每个  $n \in \mathbb{N}^+$ ,  $(t_i)$  为  $[0, T]$  的划分且其模随  $n \rightarrow \infty$  趋于零, 则有:

$$[M, M]_t = \lim_{\text{mesh}(t_i) \rightarrow 0} \sum_{i=1}^n (M_{t_i \wedge t} - M_{t_{i-1} \wedge t})^2, \quad \forall t \in [0, T], \quad \text{依概率收敛。} \quad (2.46)$$

过程  $([M, M]_t)_{t \in [0, T]}$  称为  $(M_t)_{t \in [0, T]}$  的二次变差。

**定义 2.11** 称过程  $(X_t)_{t \in [0, T]}$  为半鞅 (Semi-martingale), 如果它可以分解为:

$$X_t = M_t + A_t, \quad (2.47)$$

其中  $(M_t)_{t \in [0, T]}$  是连续局部鞅,  $(A_t)_{t \in [0, T]}$  是初始值为 0 的有限变差过程。此分解几乎必然唯一。我们可以定义半鞅的二次变差为  $[X, X]_t := [M, M]_t$ 。

作为简写，我们通常将伊藤过程写成微分形式：

$$dX_t = b_t dt + \sigma_t dW_t. \quad (2.48)$$

我们的核心目标是研究当将一个 Itô 过程与光滑函数  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  复合时该过程如何变换；这正是伊藤公式的内容，也是随机微积分的基本工具。尽管 (2.48) 处的记号较为非正式，但它传达了主要的直觉。对于足够小的  $h > 0$ ，可近似有

$$X_{t+h} \approx X_t + h b_t + \sqrt{h} \sigma_t \xi, \quad (2.49)$$

其中  $\xi \sim \mathcal{N}(0, I_N)$ 。注意到  $\sqrt{h}$  来自布朗运动增量  $W_{t+h} - W_t \sim \mathcal{N}(0, hI_N)$ 。若  $f \in C^2(\mathbb{R}^d)$ ，为了计算  $f(X_{t+h}) - f(X_t)$  到  $o(h)$  量级，仅用一阶泰勒展开会遗漏由布朗运动产生的重要项：布朗增量的二阶项为  $O(h)$ ，因此不能忽略。这一额外的二次项即构成伊藤校正项，从而导出完整的伊藤公式 [49]。

**定理 2.3** 设  $X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s$  为  $\mathbb{R}^d$  上的伊藤过程，且  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  为二次连续可微函数。则过程  $(f(X_t))_{t \geq 0}$  也是伊藤过程，并满足：

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t \nabla f(X_s)^\top b_s ds + \int_0^t \nabla f(X_s)^\top \sigma_s dW_s \\ &\quad + \frac{1}{2} \int_0^t \langle \nabla^2 f(X_s), \sigma_s \sigma_s^\top \rangle ds. \end{aligned} \quad (2.50)$$

由于伊藤积分（在适当的可积条件下）是连续鞅，伊藤公式中最后一项的期望通常为零。因此，取期望后得到：

$$\mathbb{E}f(X_t) - \mathbb{E}f(X_0) = \int_0^t \mathbb{E} \left[ \langle \nabla f(X_s), b_s \rangle + \frac{1}{2} \langle \nabla^2 f(X_s), \sigma_s \sigma_s^\top \rangle \right] ds, \quad (2.51)$$

或者写成微分形式：

$$\partial_t \mathbb{E}f(X_t) = \mathbb{E} \left[ \langle \nabla f(X_t), b_t \rangle + \frac{1}{2} \langle \nabla^2 f(X_t), \sigma_t \sigma_t^\top \rangle \right]. \quad (2.52)$$

实际上，伊藤公式也可以推广到显含时间的函数  $f(t, x)$ ：

$$\begin{aligned} f(t, X_t) - f(0, X_0) &= \int_0^t \left\{ \partial_s f(s, X_s) + \nabla f(s, X_s)^\top b_s + \frac{1}{2} \langle \nabla^2 f(s, X_s), \sigma_s \sigma_s^\top \rangle \right\} ds \\ &\quad + \int_0^t \langle \sigma_s^\top \nabla f(s, X_s), dW_s \rangle. \end{aligned} \quad (2.53)$$

此外，我们也可以利用连续局部鞅的二次变差过程来重新表述伊藤公式，这对于处理一般半鞅尤为方便。

**定理 2.4** 设  $X$  为  $\mathbb{R}^d$  值半鞅, 记  $X = (X^1, \dots, X^d)$ 。设  $f \in C^2(\mathbb{R}^d)$ 。则  $f(X)$  也是半鞅, 且满足:

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \partial_i f(X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_{ij}^2 f(X_s) d[X^i, X^j]_s, \quad (2.54)$$

其中协变差定义为  $[X, Y] := \frac{1}{2}([X + Y, X + Y] - [X, X] - [Y, Y])$ 。

### 2.4.3 随机微分方程的存在唯一性

设  $b: \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  和  $\sigma: \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times N}$  为给定函数。现在我们考虑随机微分方程 (SDE):

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x \in \mathbb{R}^d, \quad (2.55)$$

假设给定一个完备的带流概率空间  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ , 其上支撑着一个标准  $N$  维适应布朗运动  $W$ 。该 SDE 的强解 (strong solution) 是指一个适应的  $\mathbb{R}^d$  值过程  $X$ , 满足:

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \quad (2.56)$$

本节我们将陈述该方程存在性与唯一性的基本结果。定理指出, 如果系数  $b, \sigma$  关于空间变量满足一致 Lipschitz 条件, 则 SDE 存在唯一解。所谓的唯一性是指, 如果在同一概率空间上、由同一布朗运动驱动的两个解  $X$  和  $\tilde{X}$ , 则必然有  $X = \tilde{X}$ 。关于此定理的详细证明可参见 [39, 50]。

**定理 2.5** 假设  $b$  和  $\sigma$  是连续的, 且存在常数  $K > 0$  使得对于所有  $t \geq 0$  和  $x, y \in \mathbb{R}^d$ , 满足:

$$\|b(t, x) - b(t, y)\| + \|\sigma(t, x) - \sigma(t, y)\|_{\text{HS}} \leq K\|x - y\|. \quad (2.57)$$

那么, 对于任意完备带流概率空间  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  和初始条件  $X_0 = x \in \mathbb{R}^d$ , SDE (2.55) 拥有唯一的强解  $(X_t)_{t \geq 0}$ 。此外, 该解  $(X_t)_{t \in [0, T]}$  是一个 Markov 过程。

如果在较弱的假设下, 即  $b$  和  $\sigma$  仅满足局部 Lipschitz 条件, 我们也可以建立类似于定理 2.5 的结论 [51]。但在这种情况下, SDE 的解并不保证在所有时间内都存在。问题在于, 当系数增长速度超过线性时, 可能会在有限的随机时刻  $\tau_e$  发生爆破 (explosion), 即当  $t \rightarrow \tau_e$  时  $\|X_t\| \rightarrow \infty$ , a.s. 这种现象在常微分方程 (ODE) 中也同样存在。在此情形下, 我们仍然可以在某个区间  $[0, \tau_e - \delta]$  (其中  $\delta < \tau_e$ ) 定义唯一的强解。

### 2.4.4 马尔可夫半群 (Markov Semigroup)

Markov 半群理论的核心思想是通过作用在函数上的算子来编码 Markov 过程  $(X_t)_{t \geq 0}$  的行为。我们将建立一套针对这些算子的演算规则，并通过泛函分析的方法对其进行研究。这类似于通过研究离散时间 Markov 链的转移矩阵来揭示链的遍历性与收敛性 [52]。

**定义 2.12** 对于  $\mathbb{R}^d$  上的时齐 (time-homogeneous) <sup>①</sup> Markov 过程  $(X_t)_{t \geq 0}$ ，其关联的 Markov 半群  $(P_t)_{t \geq 0}$  定义为：

$$P_t f(x) := \mathbb{E}[f(X_t) | X_0 = x], \quad \forall f \in \mathcal{B}_b(\mathbb{R}^d), \quad (2.58)$$

其中  $\mathcal{B}_b(\mathbb{R}^d)$  是  $\mathbb{R}^d$  上有界可测函数的空间。

利用 Markov 性和迭代条件期望，可以得到如下引理：

**引理 2.1** 算子族  $(P_t)_{t \geq 0}$  满足半群性质：

$$P_{t+s} = P_t P_s = P_s P_t, \quad \forall t, s \geq 0, \quad (2.59)$$

且  $P_0 = Id$  为恒等算子。

构建演算规则的起点是对半群  $t \mapsto P_t$  进行微分，这通过以下定义实现。

**定义 2.13** Markov 半群  $(P_t)_{t \geq 0}$  的无穷小生成元 (Infinitesimal Generator)  $\mathcal{L}$  定义为：

$$\mathcal{L}f := \lim_{t \rightarrow 0^+} \frac{P_t f - f}{t}, \quad (2.60)$$

该定义适用于所有使得上述极限在最大范数意义下存在的函数  $f$ 。生成元  $\mathcal{L}$  的定义域记为  $\mathcal{D}(\mathcal{L})$ 。

正如预期的那样，Markov 半群捕捉了原始 Markov 过程的所有信息。一种证明方法是展示 Markov 过程可以完全从其半群中恢复。另一种我们现在采取的方法，是展示 Markov 过程的动力学特征可以通过涉及半群的计算规则来描述。特别地，我们要介绍以下命题。

<sup>①</sup> 若 Markov 过程  $(X_t)_{t \geq 0}$  的转移概率仅依赖于时间差，而与绝对时间无关，即对任意可测集  $A$  和  $s > 0$ ，有  $\mathbb{P}(X_{t+s} \in A | X_t = x) = \mathbb{P}(X_s \in A | X_0 = x)$ ，则称该过程为时齐的。

**命题 2.3** (Kolmogorov 后向方程) 设  $(X_t)_{t \geq 0}$  为时齐 Markov 过程, 具有半群  $(P_t)_{t \geq 0}$  和生成元  $\mathcal{L}$ 。则对于任意  $f \in \mathcal{D}(\mathcal{L})$ , 成立:

$$\partial_t P_t f = \mathcal{L} P_t f = P_t \mathcal{L} f. \quad (2.61)$$

特别地,  $\mathcal{L}$  与半群  $(P_t)_{t \geq 0}$  可交换。

该方程有一个对偶形式: 设  $\pi_0$  为  $X_0$  的密度函数。形式上, 我们可以写成  $\mathbb{E}f(X_t) = \int P_t f d\pi_0 = \int f dP_t^* \pi_0$ , 其中  $P_t^*$  是  $P_t$  的伴随算子 (adjoint)。这意味着  $X_t$  的分布规律  $\pi_t$  形式上由  $P_t^* \pi_0$  给定。此外, 利用 Kolmogorov 后向方程, 推导如下:

$$\begin{aligned} \partial_t \int f dP_t^* \pi_0 &= \int \partial_t P_t f d\pi_0 = \int \mathcal{L} P_t f d\pi_0 = \int P_t f d\mathcal{L}^* \pi_0 = \int f dP_t^* \mathcal{L}^* \pi_0 \\ &= \int P_t \mathcal{L} f d\pi_0 = \int \mathcal{L} f dP_t^* \pi_0 = \int f d\mathcal{L}^* P_t^* \pi_0. \end{aligned} \quad (2.62)$$

由于上述等式必须对所有函数  $f$  成立, 我们得出以下结论。

**命题 2.4** (Kolmogorov 前向方程) 设  $(X_t)_{t \geq 0}$  为时齐 Markov 过程, 具有半群  $(P_t)_{t \geq 0}$  和生成元  $\mathcal{L}$ 。则对于任意初始密度  $\pi_0$ ,  $X_t$  的概率密度  $\pi_t$  满足:

$$\partial_t \pi_t = \mathcal{L}^* \pi_t, \quad \pi_t = P_t^* \pi_0. \quad (2.63)$$

我们将描述密度演化的方程 (即 Kolmogorov 前向方程) 称为 Fokker-Planck 方程。关于 Fokker-Planck 方程的详细讨论可参考 [53]。

**例 2.2** 考虑 Langevin 扩散 (或过阻尼 Langevin 动力学) 的生成元:

$$dZ_t = -\nabla V(Z_t) dt + \sqrt{2} dW_t. \quad (2.64)$$

假设  $f \in C^2(\mathbb{R}^d)$  且导数有界, 利用伊藤公式, 我们有:

$$\begin{aligned} \mathcal{L} f(x) &= \lim_{t \rightarrow 0^+} \frac{\mathbb{E}[f(Z_t) | Z_0 = x] - f(x)}{t} \\ &= \lim_{t \rightarrow 0^+} \frac{1}{t} \mathbb{E} \left[ \int_0^t \langle \nabla f(Z_s), -\nabla V(Z_s) \rangle + \Delta f(Z_s) ds \middle| Z_0 = x \right] \\ &= \langle \nabla f(x), -\nabla V(x) \rangle + \Delta f(x). \end{aligned} \quad (2.65)$$

对于函数  $f, g: \mathbb{R}^d \rightarrow \mathbb{R}$ , 分部积分给出:

$$\begin{aligned} \int f \mathcal{L} g dx &= \int f \langle \nabla g, -\nabla V \rangle + f \Delta g dx \\ &= \int g \{ \Delta f + \operatorname{div}(f \nabla V) \} dx, \end{aligned} \quad (2.66)$$

这意味着：

$$\mathcal{L}^* f = \operatorname{div}(f \nabla V) + \Delta f. \quad (2.67)$$

因此， $Z_t$  的密度  $\pi_t$  对应的 Fokker-Planck 方程为：

$$\partial_t \pi_t = \operatorname{div}(\pi_t \nabla V) + \Delta \pi_t. \quad (2.68)$$

由此，我们可以通过令方程右端为零来求解平稳分布，得到 Gibbs 测度  $\pi(x) \propto e^{-V(x)}$  [54]。

### 2.4.5 Fokker-Planck 方程

考虑一个通用的随机信道，其输出为实值随机过程  $(X_t)_{t \geq 0}$ ，由如下随机微分方程控制：

$$dX_t = a(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x \in \mathbb{R}^d, \quad (2.69)$$

其中  $a: \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  和  $\sigma: \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times N}$  是可测函数，且满足足够正则的条件以保证 SDE 存在唯一的强解。设  $\pi_t$  表示  $X_t$  的概率密度。由命题 2.4 可知， $\pi_t$  满足如下 Fokker-Planck 方程：

$$\partial_t \pi_t = -\operatorname{div}(a(t, \cdot) \pi_t) + \sum_{i,j=1}^d \partial_{ij}^2 (D_{ij}(t, \cdot) \pi_t), \quad (2.70)$$

其中  $D = \frac{1}{2} \sigma \sigma^\top$ ， $D_{ij} = \frac{1}{2} \sum_{k=1}^N \sigma_{ik} \sigma_{jk}$ 。该方程描述了随机过程  $(X_t)_{t \geq 0}$  的概率密度函数  $\pi_t$  随时间的演化规律。方程右端第一项代表漂移分量，第二项则反映了由随机噪声引起的扩散效应。它有如下常见的特殊形式：

1. 布朗运动: 当  $a(t, x) = 0, \sigma(t, x) = I_d$  时，Fokker-Planck 方程退化为经典的热方程：

$$\partial_t \pi_t = \frac{1}{2} \Delta \pi_t. \quad (2.71)$$

这等价于连续时间高斯信道：

$$X_t = X_0 + W_t \stackrel{d}{=} X_0 + \sqrt{t}Z \sim \mathcal{N}(X_0, t). \quad (2.72)$$

2. Ornstein-Uhlenbeck (OU) 过程: 当  $a(t, x) = -\alpha x, \sigma(t, x) = 1, d = 1$  时，Fokker-Planck 方程化为：

$$\partial_t \pi_t = \alpha \partial_x (x \pi_t) + \frac{1}{2} \partial_{xx}^2 \pi_t. \quad (2.73)$$

其对应的 SDE 为：

$$dX_t = -\alpha X_t dt + dW_t. \quad (2.74)$$

该随机微分方程可以显式求解：

$$X_t = e^{-\alpha t} \left( X_0 + \int_0^t e^{\alpha s} dW_s \right) \stackrel{d}{=} e^{-\alpha t} X_0 + \sqrt{\frac{1}{2\alpha} (1 - e^{-2\alpha t})} Z, \quad (2.75)$$

因此当  $t \rightarrow \infty$  时,  $X_t \rightarrow \mathcal{N}(0, \frac{1}{2\alpha})$ 。

3. Langevin 扩散: 当  $a(t, x) = -\nabla V(x), \sigma(t, x) = \sqrt{2}I_d$  时, Fokker-Planck 方程化为:

$$\partial_t \pi_t = \operatorname{div}(\pi_t \nabla V) + \Delta \pi_t. \quad (2.76)$$

其对应的 SDE 为:

$$dX_t = -\nabla V(X_t)dt + \sqrt{2}dW_t. \quad (2.77)$$

并且该随机微分方程的平稳分布为 Gibbs 分布  $\pi(x) \propto e^{-V(x)}$ 。当  $V$  满足适当的条件时, 在 Wasserstein 度量意义下 (见章节 2.5)  $\pi_t$  会以指数速度收敛到  $\pi$ 。然而对于这种一般的 Langevin 扩散,  $\pi_t$  通常没有显式的表达式。

## 2.5 最优传输理论基础

### 2.5.1 最优传输问题描述与性质

1781 年, Monge 在他的著作中 [55], 从以下实际问题开始引入了最优传输问题的概念: 假设有人从地面上提取土壤来建造防御工事。运输土壤成本最低的方式是什么? 为了严谨地表述这个问题, 我们需要明确传输成本, 也就是把一个单位质量的物体从  $x$  点移动到  $y$  点要花多少钱。在当时 Monge 的观点与实际问题上, 考察的空间为 3 维欧式空间, 代价函数  $c(x, y) = |x - y|$  为欧几里得距离。而这个问题可以推广到更高维及更抽象的空间与分布中, 问题描述如下:

**定义 2.14** (Monge Problem) 给定两个概率分布  $\mu \in \mathcal{P}(\mathcal{X})$  以及  $\nu \in \mathcal{P}(\mathcal{Y})$ , 以及代价函数  $c: \mathcal{X} \times \mathcal{Y} \rightarrow [0, +\infty]$ , 求解:

$$(MP) \quad \inf \left\{ M(T) := \int c(x, T(x)) d\mu : T_{\#}\mu = \nu \right\} \quad (2.78)$$

其中测度  $T_{\#}\mu(A) := \mu(T^{-1}(A))$  对任意的可测集  $A$  均成立, 称为测度  $\mu$  关于映射  $T$  的像测度或 push-forward 测度。

事实上, 这个问题是难以求解的, 因为其限制条件过于苛刻, 也无法利用离散化逼近的手法去逼近所求得的最优传输映射  $T$ , 并且可以找到这样的反例: 在弱收敛

意义下, 满足  $T_{\#}\mu = \nu$  这样的传输映射全体构成的集合, 并不是封闭的。此外, 根据 Villani 构造的反例 [17], 很多情况下 Monge 问题的解不存在, 也就是不存在一个最优传输映射, 尤其是对于非绝对连续测度间的最优传输问题。因此需要对 Monge 问题进行松弛与推广, 由此引出了 Kantorovich 问题 [56]:

**定义 2.15** (Kantorovich Problem) 给定两个概率分布  $\mu \in \mathcal{P}(\mathcal{X})$  以及  $\nu \in \mathcal{P}(\mathcal{Y})$ , 以及代价函数  $c: \mathcal{X} \times \mathcal{Y} \rightarrow [0, +\infty]$ , 求解:

$$(KP) \quad \inf \left\{ K(\pi) := \int_{\mathcal{X} \times \mathcal{Y}} c d\pi : \pi \in \Pi(\mu, \nu) \right\} \quad (2.79)$$

其中  $\Pi(\mu, \nu)$  被称为传输方案, 它表示满足如下条件的集合:

$$\Pi(\mu, \nu) = \{ \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : (Proj_x)_{\#}\pi = \mu, (Proj_y)_{\#}\pi = \nu \} \quad (2.80)$$

其中  $Proj_x$  是  $\mathcal{X} \times \mathcal{Y}$  到  $\mathcal{X}$  的投影映射,  $Proj_y$  是  $\mathcal{X} \times \mathcal{Y}$  到  $\mathcal{Y}$  的投影映射, 显然在这个问题描述中, 允许单个点传输到多个点甚至无穷多点的情况存在, 而不是仅限于映射。而式子 (2.2) 对应的解即为  $\mu$  和  $\nu$  之间的最优传输方案。特别地, 如果存在最优传输映射, 记为  $T$ , 则 Kantorovich 问题对应的解为  $\pi = (Id, T)_{\#}\mu$ 。

Kantorovich 提出的这个更广义的问题比 Monge 提出的问题更容易求解。例如, 在 Monge 情况下, 我们需要至少存在一个满足约束条件的映射  $T$ , 而如果初始测度  $\mu = \delta_0$  为单个点的狄拉克测度, 而目标测度  $\nu$  并不是单个点的狄拉克测度, 则不存在任何一个满足限制条件的映射, 但存在满足 KP 限制条件的传输方案。此外, 我们可以说 KP 是 Monge 提出的原始问题的松弛。并且可以证明 Kantorovich 问题的解是存在的, 即可以取到最小值。我们有如下一系列定理:

**定理 2.6** 设  $\mathcal{X}, \mathcal{Y}$  为紧的度量空间,  $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ , 代价函数  $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  为连续函数, 则 Kantorovich 问题存在一个解。

**定理 2.7** 设  $\mathcal{X}, \mathcal{Y}$  为紧的度量空间,  $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ , 代价函数  $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  为下半连续函数并且有下界, 则 Kantorovich 问题存在一个解。

**定理 2.8** 若函数  $f: \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$  为下半连续函数, 且有下界,  $\mathcal{X}$  为度量空间, 则泛函  $J: \mathcal{M}_+ \rightarrow \mathbb{R} \cup \{+\infty\}$  以  $\mathcal{X}$  上的正测度为自变量, 并且  $J(\mu) := \int f d\mu$  是下半连续的。

**定理 2.9** 设  $\mathcal{X}, \mathcal{Y}$  为波兰空间, 即完备的可分度量空间,  $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ , 代价函数  $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  为下半连续函数, 则 Kantorovich 问题存在一个解。

Kantorovich 问题是一个具备凸限制条件的线性优化问题, 由线性的等式或者不等式进行限制, 因此一种重要的求解工具就是对偶理论, 它常用于处理凸优化问题。记对偶问题为 DP。

首先将限制条件  $\gamma \in \Pi(\mu, \nu)$  改写, 对任意的  $\gamma \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$ , 我们有:

$$\sup_{\phi, \psi} \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} (\phi(x) + \psi(y)) d\pi = \begin{cases} 0, & \text{if } \pi \in \Pi(\mu, \nu) \\ +\infty, & \text{otherwise} \end{cases} \quad (2.81)$$

其中  $\phi, \psi$  为有界连续函数, 因此我们添加这个取上确界的过程, 便可以去掉  $\gamma$  原有的限制条件。一旦满足限制条件, 则不会加上任何余项, 如果不满足这一限制条件, 则最终会得到  $+\infty$ , 与 Kantorovich 问题解的存在性矛盾。因此我们可以将 Kantorovich 问题转化为:

$$\min_{\pi} \int_{\mathcal{X} \times \mathcal{Y}} c d\pi + \sup_{\phi, \psi} \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} (\phi(x) + \psi(y)) d\pi \quad (2.82)$$

并更换 sup 与 inf 的位置, 可以得到:

$$\sup_{\phi, \psi} \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu + \inf_{\pi \in \mathcal{M}_+} \int_{\mathcal{X} \times \mathcal{Y}} (c(x, y) - (\phi(x) + \psi(y))) d\pi \quad (2.83)$$

事实上, 对 sup 与 inf 交换顺序, 两个优化式子并不一定相等, 这里用到了 Rockafellar 的一个定理[57], 它要求一个变量是凹的, 另一个是凸的, 以及一些关于紧性的假设。此外容易得到:

$$\inf_{\pi \in \mathcal{M}_+} \int_{\mathcal{X} \times \mathcal{Y}} (c - (\phi \oplus \psi)) d\pi = \begin{cases} 0, & \text{if } \phi \oplus \psi \leq c \text{ on } \mathcal{X} \times \mathcal{Y} \\ -\infty, & \text{otherwise} \end{cases} \quad (2.84)$$

其中  $\phi \oplus \psi$  表示:  $(\phi \oplus \psi)(x, y) := \phi(x) + \psi(y)$ . 如果存在集合  $E$ , 在  $E$  上有  $\phi \oplus \psi > c$  成立, 则取  $\gamma$  完全集中在这个集合  $E$  上, 并使得  $\int_E d\pi$  趋于  $+\infty$ , 从而  $\inf_{\pi \in \mathcal{M}_+} \int_{\mathcal{X} \times \mathcal{Y}} (c - (\phi \oplus \psi)) d\pi = -\infty$ . 由此引出了如下对偶最优传输问题:

**定义 2.16** (对偶最优传输问题) 给定两个概率分布  $\mu \in \mathcal{P}(\mathcal{X})$  以及  $\nu \in \mathcal{P}(\mathcal{Y})$ , 以及代价函数  $c: \mathcal{X} \times \mathcal{Y} \rightarrow [0, +\infty)$ , 考虑求解下述问题:

$$(DP) \quad \max \left\{ \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu : \phi \in C_b(\mathcal{X}), \psi \in C_b(\mathcal{Y}), \phi \oplus \psi \leq c \right\} \quad (2.85)$$

其中  $C_b(\mathcal{X})$  为  $\mathcal{X}$  上的有界连续函数空间。注意到:  $\sup(DP) \leq \min(KP)$ , 因为根据限制条件:  $\phi \oplus \psi \leq c$ , 对任意的  $\pi \in \Pi(\mu, \nu)$ , 有:

$$\int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \psi d\nu = \int_{\mathcal{X} \times \mathcal{Y}} (\phi \oplus \psi) d\pi \leq \int_{\mathcal{X} \times \mathcal{Y}} c d\pi \quad (2.86)$$

然而, 对偶问题的解并不一定存在, 因为满足限制条件的函数空间并不是紧的, 需要假定其他的条件来保证这个问题的解存在。

**定义 2.17** ( $c$ -变换) 给定函数  $\xi : \mathcal{X} \rightarrow \bar{\mathbb{R}}$ , 我们定义  $c$ -变换 (或称为  $c$ -共轭函数)  $\xi^c : \mathcal{Y} \rightarrow \bar{\mathbb{R}}$ , 其中:

$$\xi^c(y) = \inf_{x \in \mathcal{X}} c(x, y) - \xi(x) \quad (2.87)$$

此外我们还可以定义  $\bar{c}$ -变换, 给定函数  $\zeta : \mathcal{Y} \rightarrow \bar{\mathbb{R}}$ , 则有:

$$\zeta^{\bar{c}}(x) = \inf_{y \in \mathcal{Y}} c(x, y) - \zeta(y) \quad (2.88)$$

并且, 我们称  $\mathcal{Y}$  上的函数  $\psi$  是  $\bar{c}$ -concave 的, 如果存在一个函数  $\xi$ , 使得  $\psi = \xi^c$ . 相应地, 我们称  $\mathcal{X}$  上的函数  $\phi$  是  $c$ -concave 的, 如果存在一个函数  $\zeta$ , 使得  $\psi = \zeta^{\bar{c}}$ .

如果代价函数  $c(x, y)$  是连续的, 并且在一个紧集上取值有限, 那么代价函数也是一致连续的, 这说明存在一个单调递增的连续函数  $\omega : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ ,  $\omega(0) = 0$ , 使得:

$$|c(x, y) - c(x', y')| \leq \omega(d(x, x') + d(y, y')) \quad (2.89)$$

因此对任意给的  $x \in \mathcal{X}$ , 我们有  $|(c(x, y) - \xi(x)) - (c(x, y') - \xi(x))| \leq \omega(d(y, y'))$ , 由  $c$ -变换的定义,  $\xi^c(y) = \inf_{x \in \mathcal{X}} c(x, y) - \xi(x)$ , 由于连续模对下确界封闭, 可得  $|\xi^c(y) - \xi^c(y')| \leq \omega(d(y, y'))$ . 此外, 我们可以将  $c$ -变换应用到对偶最优传输问题当中:

**命题 2.5** 假定  $\mathcal{X}, \mathcal{Y}$  都是紧的, 代价函数  $c$  为连续函数, 则对偶最优传输问题存在一组解  $(\phi, \psi)$ , 并且  $\psi = \phi^c$ . 特别地:

$$\max(DP) = \max_{\phi \in c\text{-conc}(\mathcal{X})} \int_{\mathcal{X}} \phi d\mu + \int_{\mathcal{Y}} \phi^c d\nu \quad (2.90)$$

其中  $c\text{-conc}(\mathcal{X})$  表示  $\mathcal{X}$  上的  $c$ -concave 函数构成的集合。

事实上, 对偶最优传输问题与 Kantorovich 问题在一定假设下是等价的, 我们有如下定理:

**定理 2.1.5** (对偶最优传输问题与 Kantorovich 问题的等价性) 假定  $\mathcal{X}$  与  $\mathcal{Y}$  均为波兰空间,  $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  是一致连续且有界的, 则对偶最优传输问题存在一组解  $(\phi, \phi^c)$ , 并且  $\max(DP) = \min(KP)$ .

在本节中, 我们将探讨 Monge 问题解存在的判据, 即在什么条件下我们可以求得最优传输映射, 从而给出 Monge 问题与 Kantorovich 问题等价的一个充分条件, 这个等价性证明最初由 Brenier 提出[58]:

**定理 2.10** (Brenier 定理) 若两个概率空间均是  $d$  维欧氏空间, 即  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ , 并且代价函数  $c(x, y) = \|x - y\|^2$ , 如果两个输入测度 (用  $\mu$  表示) 中至少有一个具有关于勒贝格测度的概率密度, 那么 Kantorovich 问题中的最优传输方案  $\pi$  是唯一的, 并且  $(x, T(x))$  为其支撑, 其中  $T$  是 Monge 问题下的最优传输映射, 即  $\pi = (Id, T)_\# \mu$ , 从而有

$$\forall h \in C(\mathcal{X} \times \mathcal{Y}), \int_{\mathcal{X} \times \mathcal{Y}} h(x, y) d\pi = \int_{\mathcal{X}} h(x, T(x)) d\mu \quad (2.91)$$

此外, 这个最优传输映射被一个凸函数  $\varphi$  的梯度唯一确定, 即  $T(x) = \nabla\varphi(x)$ .

这个结果表明, 在这个具有非奇异密度的问题中, Monge 问题和它的 Kantorovich 松弛是等价的 (即松弛是紧的)。

此外, Brenier 定理也有很多变体以及加强, 有如下定理:

**定理 2.11** 若两个概率空间均是  $d$  维欧氏空间, 即  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ , 并且代价函数  $c(x, y) = \frac{1}{2} \|x - y\|^2$ . 假定

$$\int_{\mathbb{R}^d} |x|^2 d\mu + \int_{\mathbb{R}^d} |y|^2 d\nu < +\infty \quad (2.92)$$

并且初始测度  $\mu$  相对勒贝格测度来说是绝对连续测度, 则存在唯一的最优传输映射  $T$ , 并且存在一个凸函数  $\varphi$ , 使得  $T = \nabla\varphi$ .

**定理 2.12** (Pratelli 定理[59]) 假定初始测度是没有原子的测度 (为绝对连续测度), 代价函数  $c$  是连续函数, 则 Monge 问题与 Kantorovich 问题等价, 即  $\min(KP) = \inf(MP)$ 。

### 2.5.2 Wasserstein 距离与空间

如前文所述，对于两个概率分布之间的最优传输问题（在实际情境下可以视为负责生产者和消费者之间的货物运输，他们各自的空间分布由概率度量建模），初始测度上一点与目标测度上一点之间的度量越大（即生产者和消费者之间的距离越远），则相应的最优传输总代价就越大，需要一个量来概括这个最优传输总代价，事实上这在前文中已经提及，记：

$$C(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int c(x, y) d\pi \quad (2.93)$$

其中  $c(x, y)$  为代价函数，是从  $x$  处运输单位质量到  $y$  处所需要的代价（或成本）。在这里我们不需要考察最优传输具体的传输形式，而只关注最终的最优传输代价， $C(\mu, \nu)$  可以看做是两个测度  $\mu$  与  $\nu$  之间的某种“距离”，但很多时候它并不满足度量，即“距离”所要满足的三条公理。然而，当代价函数本身是一个度量函数时，这个最优传输代价  $C(\mu, \nu)$  也会成为两个测度之间的度量函数。我们给出如下定义：

**定义 2.18** （Wasserstein 距离）设  $(\mathcal{X}, d)$  为波兰空间，即完备的可分度量空间，令  $p \in [1, +\infty)$ 。对于任意两个  $\mathcal{X}$  上的概率测度  $\mu, \nu$ ，它们之间的以  $p$  为参数的 Wasserstein 距离定义为：

$$\begin{aligned} W_p(\mu, \nu) &= \left( \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X}} d(x, y)^p d\pi \right)^{1/p} \\ &= \inf \left\{ [\mathbb{E}d(X, Y)^p]^{\frac{1}{p}}, \text{law}(X) = \mu, \text{law}(Y) = \nu \right\} \end{aligned} \quad (2.94)$$

事实上，通常来说  $W_p$  在严格意义上不算是一个“距离”，因为它可能会取值为  $+\infty$ ，并不是所有的情况下最优传输代价都是有限的，但它确实满足度量所要求的三个条件，即对称性、正定性、三角不等式。

**命题 2.6** Wasserstein 距离是一个度量。

为了使得 Wasserstein 距离在严格意义下仍为一个度量，即使得它不会取到正无穷大，需要对它进行一些限制，由此定义出 Wasserstein 空间：

**定义 2.19** （Wasserstein 空间）设  $(\mathcal{X}, d)$  为波兰空间，即完备的可分度量空间，令  $p \in [1, +\infty)$ 。以  $p$  为参数的 Wasserstein 空间定义为：

$$\mathcal{P}_p(\mathcal{X}) := \left\{ \mu \in \mathcal{P}(\mathcal{X}); \int_{\mathcal{X}} d(x_0, x)^p d\mu < +\infty \right\} \quad (2.95)$$

其中  $x_0 \in \mathcal{X}$  的选取是任意的, 并且这个空间也不依赖于  $x_0 \in \mathcal{X}$  的选取。则 Wasserstein 距离  $W_p$  在 Wasserstein 空间  $\mathcal{P}_p$  上是一个有限的度量。也即 Wasserstein 空间。关于 Wasserstein 距离在 Wasserstein 空间中是有限的证明也很容易:

设两个概率测度  $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$ , 则有:

$$d(x, y)^p \leq 2^{p-1}[d(x, x_0)^p + d(x_0, y)^p] \quad (2.96)$$

对两边积分可以得到  $\int d(x, y)^p d\pi < +\infty$ .

**命题 2.7** 对任意的  $1 \leq p \leq q < \infty$ , 有  $W_p \leq W_q$ .

一般来说,  $W_1$  距离更灵活, 更容易被限制住, 因为它的值是最小的, 而  $W_2$  距离可以更好地反映几何特征, 对于高斯测度可以赋予黎曼流形结构[60], 它还可以更好地根据维度进行缩放。然而  $W_2$  距离的结果通常比  $W_1$  距离更难以确定。

在平常意义下, 测度的弱收敛定义为: 一系列测度  $\mu_k$  弱收敛到  $\mu$ , 如果对任意的有界连续函数  $\varphi$ , 有  $\int \varphi d\mu_k \rightarrow \int \varphi d\mu$ , 记为  $\mu_k \rightharpoonup \mu$ . 下面将引出 Wasserstein 空间中的弱收敛:

**定义 2.20** ( $\mathcal{P}_p$  空间上的弱收敛) 对于  $p \in [1, +\infty)$  以及波兰空间  $(\mathcal{X}, d)$ , 设  $\mu$  以及一系列测度  $\{\mu_k\}_{k \in \mathbb{N}}$  是  $\mathcal{P}_p(\mathcal{X})$  上的概率测度, 则  $\mu_k$  在  $\mathcal{P}_p(\mathcal{X})$  意义下弱收敛到  $\mu$ , 如果对任意  $x_0 \in \mathcal{X}$  满足如下四个等价条件中的一条:

- (i)  $\mu_k \rightharpoonup \mu$  并且  $\int d(x_0, x)^p d\mu_k \rightarrow \int d(x_0, x)^p d\mu$ ;
- (ii)  $\mu_k \rightharpoonup \mu$  并且  $\limsup_{k \rightarrow \infty} \int d(x_0, x)^p d\mu_k \leq \int d(x_0, x)^p d\mu$ ;
- (iii)  $\mu_k \rightharpoonup \mu$  并且  $\lim_{R \rightarrow \infty} \limsup_{k \rightarrow \infty} \int_{d(x_0, x) \geq R} d(x_0, x)^p d\mu_k = 0$ ;
- (iv) 对于所有满足  $|\varphi(x)| \leq C(1 + d(x_0, x)^p)$ ,  $C \in \mathbb{R}$  的连续函数  $\varphi$ , 有下式成立:

$$\int \varphi(x) d\mu_k \rightarrow \int \varphi(x) d\mu \quad (2.97)$$

**定理 2.13** 设  $(\mathcal{X}, d)$  为波兰空间, 即完备的可分度量空间, 令  $p \in [1, +\infty)$ , 则 Wasserstein 距离  $W_p$  刻画了 Wasserstein 空间中的弱收敛, 即如果  $\{\mu_k\}_{k \in \mathbb{N}}$  是 Wasserstein 空间  $\mathcal{P}_p(\mathcal{X})$  中的一系列测度,  $\mu \in \mathcal{P}(\mathcal{X})$  中的一个概率测度, 则有  $\mu_k$  在  $\mathcal{P}_p(\mathcal{X})$  空间意义下弱收敛到  $\mu$ , 等价于  $W_p(\mu_k, \mu) \rightarrow 0$ .

**推论 2.1** 设  $(\mathcal{X}, d)$  为波兰空间, 即完备的可分度量空间, 令  $p \in [1, +\infty)$ ,  $\{\mu_k\}_{k \in \mathbb{N}}$  是 Wasserstein 空间  $\mathcal{P}_p(\mathcal{X})$  中的一系列测度,  $\mu \in \mathcal{P}(\mathcal{X})$  中的一个概率测度, 则以下两条论述等价:

1.  $\mu_k \rightharpoonup \mu$  并且  $\int_{\mathcal{X}} d(x_0, x)^p d\mu_k \rightarrow \int_{\mathcal{X}} d(x_0, x)^p d\mu$ .
2.  $W_p(\mu_k, \mu) \rightarrow 0$

**推论 2.2** (Wasserstein 距离的连续性定理) 设  $(\mathcal{X}, d)$  为波兰空间, 即完备的可分度量空间, 令  $p \in [1, +\infty)$ ,  $\{\mu_k\}_{k \in \mathbb{N}}, \{\nu_k\}_{k \in \mathbb{N}}$  是 Wasserstein 空间  $\mathcal{P}_p(\mathcal{X})$  中的一系列测度,  $\mu, \nu \in \mathcal{P}(\mathcal{X})$  中的一个概率测度, 则 Wasserstein 距离是连续的, 即如果  $\mu_k$  与  $\nu_k$  分别在 Wasserstein 空间意义下弱收敛到  $\mu$  与  $\nu$ , 则有:

$$W_p(\mu_k, \nu_k) \rightarrow W_p(\mu, \nu) \quad (2.98)$$

事实上, 如果  $\mu_k$  与  $\nu_k$  分别弱收敛到  $\mu$  与  $\nu$ , 但并不一定在 Wasserstein 空间意义下弱收敛到  $\mu$  与  $\nu$ , 则会有如下类似于 Fatou 引理的定理:

$$W_p(\mu, \nu) \leq \liminf_{k \rightarrow \infty} W_p(\mu_k, \nu_k) \quad (2.99)$$

并且 Wasserstein 距离本身在  $\mathcal{P}(\mathcal{X})$  上是一个下半连续函数。

## 2.6 率失真理论简介

率失真理论 (Rate-Distortion Theory) 是信息论的核心分支, 旨在研究在允许一定失真的前提下, 信源压缩的理论极限。给定一个概率分布为  $p(x)$  的信源  $X$  以及一个失真度量函数  $\rho(x, y)$ , 率失真函数  $R(D)$  定义为在平均失真不超过  $D$  的约束下所需的最小信息速率:

$$R(D) = \inf_{p(y|x)} I(X; Y) \quad (2.100)$$

其中下确界是在所有满足以下失真约束的条件概率分布 (即重构映射) 上选取的:

$$\mathbb{E}[\rho(X, Y)] \leq D. \quad (2.101)$$

这里  $I(X; Y)$  表示输入  $X$  与重构  $Y$  之间的互信息。

为了给率失真理论建立严格的数学基础, I. Csiszár 在其经典文献 [13] 中, 在极弱的假设条件下证明了  $R(D)$  的基本性质及其参数化表示。这些工作为处理一般抽象

空间上的信源编码问题提供了坚实的理论支撑。首先，我们需要对失真度量（即损失函数）本身施加最基本的正则性假设。

**假设 2.1** 令  $\rho: \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty]$  为给定的可测函数，称为损失函数。假设其满足：

$$\inf_{y \in \mathcal{Y}} \rho(x, y) = 0, \quad \forall x \in \mathcal{X}. \quad (2.102)$$

该假设直观上意味着，对于任意信源符号  $x$ ，我们在重构空间中总能找到使其失真  $\rho(x, y)$  任意小的元素（即理论上不存在不可消除的固有失真底限）。在此基础上，Csiszár 首先刻画了  $R(D)$  的几何性质。

**引理 2.2** [Lemma 1.1 in [13]] 在假设 2.1 下，率失真函数  $R(D)$  是关于  $D$  的非增凸函数。此外，我们可以将其表示为：

$$R(D) = \inf_{\mathbb{E}\rho(X; Y) = D} I(X; Y) \quad (2.103)$$

其中  $D \leq D_1$ 。这里  $D_1$  定义为使得  $R(D)$  在区间  $(D_1, +\infty)$  上为常数的最小非负数（通常对应于  $R(D) = 0$  的临界失真值）。

上述引理虽然刻画了  $R(D)$  的形态，但并未保证  $R(D)$  在有限值处有定义。为了排除  $R(D)$  恒为无穷大这种病态情形，我们需要引入如下的随机变量存在性假设。

**假设 2.2** 假设存在定义在概率空间上的随机变量  $Y$ ，其中  $Y \sim \nu$ ，满足：

$$I(X; Y) < \infty, \quad \text{且} \quad \mathbb{E}\rho(X, Y) < \infty, \quad (2.104)$$

其中  $X \sim \mu$  是一个固定的信源分布。

此假设保证了至少存在一种编码方案，能在有限的互信息代价下实现有限的平均失真。进一步地，为了处理更一般的拓扑结构，我们通常还需要如下关于有限逼近的假设。

**假设 2.3** 假设存在一个有限集合  $B \subset \mathcal{Y}$ ，使得

$$\int \rho(x, B) \mu(dx) < \infty \quad \text{其中} \quad \rho(x, B) := \min_{y \in B} \rho(x, y). \quad (2.105)$$

定义  $D_{\min} := \inf\{D \geq 0 \mid R(D) < \infty\}$  为率失真函数有限的起始点。接下来的定理给出了  $R(D)$  在所有正失真水平上均有限（即  $D_{\min} = 0$ ）的充分条件。

**定理 2.14** [Theorem 2.1 in [13]] 在假设 2.1 和 2.2 下, 若对于任意  $\epsilon > 0$ , 存在  $\mathcal{X}$  的可测子集序列  $A_1 \subset A_2 \subset \dots$  以及  $\mathcal{Y}$  的有限子集序列  $B_1 \subset B_2 \subset \dots$ , 使得当  $n \rightarrow \infty$  时  $\mu(A_n) \rightarrow 1$ , 且满足

$$\min_{y \in B_n} \rho(x, y) < \epsilon, \quad \forall x \in A_n, n = 1, 2, \dots \quad (2.106)$$

则对于所有  $D > 0$ , 都有  $R(D)$  有限, 即  $D_{\min} = 0$ 。

值得注意的是, 如果重构字母表  $\mathcal{Y}$  是至多可数的, 则上述定理的假设条件自然满足。更重要的是, 如果  $\mathcal{X}$  是 Polish 空间 (完备可分度量空间), 且截断失真函数  $\rho^{(K)}(x, y) := \min\{\rho(x, y), K\}$  对于某个固定的  $K > 0$  是上半连续的, 该假设同样成立。在这种情况下, 由于 Polish 空间上的任意概率测度都是胎紧的 (tight), 我们可以选取  $A_i$  为  $\mathcal{X}$  中的一列递增紧集序列 [13]。

在讨论了  $R(D)$  的有限性后, 我们需要考察其在两端的渐进性质。下面的定理描述了当允许失真趋于无穷大时,  $R(D)$  的收敛行为。

**定理 2.15** [Theorem 2.1 in [13]] 在假设 2.1 和 2.2 下, 若存在如定理 2.14 所述的集合序列  $A_n$ , 且存在  $y_n \in \mathcal{Y}$  使得

$$\max_{x \in A_n} \rho(x, y_n) < \infty, \quad n = 1, 2, \dots \quad (2.107)$$

则当  $D \rightarrow \infty$  时,  $R(D) \rightarrow 0$ 。进一步地, 定义最大失真  $D_{\max}$  为:

$$D_{\max} = \lim_{K \rightarrow \infty} \left( \inf_y \int \rho^{(K)}(x, y) d\mu \right). \quad (2.108)$$

则对于  $D < D_{\max}$ ,  $R(D)$  为正值; 若  $D_{\max} < \infty$ , 则对于  $D > D_{\max}$ ,  $R(D) = 0$ 。

除了定性的存在性与渐进分析, 如何计算  $R(D)$  也是理论研究的核心。利用拉格朗日乘子法, Csiszár 给出了  $R(D)$  的参数化表示, 这一结果揭示了率失真问题与凸优化的对偶联系。

**定理 2.16** [Theorem 2.3 in [13]] 在假设 2.1, 2.2 和 2.3 下, 对于任意  $D > D_{\min}$ , 成立

$$R(D) = \max_{\beta, \alpha(x)} \left\{ \int \log \alpha(x) d\mu - \beta D \right\} \quad (2.109)$$

其中最大化是在约束条件  $\alpha(x) \geq 1, \beta \geq 0$  以及

$$\int \alpha(x) e^{-\beta \rho(x, y)} d\mu \leq 1, \quad \forall y \in \mathcal{Y} \quad (2.110)$$

之下进行的。使目标函数达到最大值的  $\beta$  正是通过如下斜率关系与  $D$  相关联的  $\beta$  值:

$$R(D') + \beta D' \geq R(D) + \beta D, \quad \forall D'. \quad (2.111)$$

上述 Csiszár 的一系列工作奠定了率失真函数的理论框架。在此基础上, R. M. Gray 在其专著中对这些结果进行了系统的梳理与推广, 特别是将  $R(D)$  的计算归结为更为紧凑的变分形式。这一形式不仅概括了上述对偶定理, 也直接构成了著名的 Blahut-Arimoto 算法及其推广形式的理论基础。

**定理 2.17** [Theorem 9.2 in [61]] 若  $R(D) < \infty$ , 则率失真函数可表示为如下关于参数  $\beta$  的最大化问题:

$$\begin{aligned} R(D) &= \max_{\beta \geq 0} \{F(\beta) - \beta D\} \\ F(\beta) &= \inf_{\nu \in \mathcal{P}(\mathcal{Y})} \int \log \frac{1}{\int e^{-\beta \rho(x,y)} d\nu} d\mu. \end{aligned} \tag{2.112}$$

其中  $\nu$  遍历重构空间  $\mathcal{Y}$  上的所有概率测度。这一表示将  $R(D)$  的求解转化为寻找最优输出分布  $\nu$  和最优拉格朗日乘子  $\beta$  的双重优化问题。



## 第 3 章 Fokker-Planck 流的几何性质与互信息凸性研究

### 3.1 本章引言

Fokker-Planck 方程是一类重要的偏微分方程，主要用于描述随机过程概率密度函数的演化，在统计物理、机器学习以及其他诸多领域中扮演着核心角色。针对该方程，我们可以从不同的理论视角和方法论出发，对其进行深入的理解与求解。

Fokker-Planck 方程可以被视为与伊藤随机微分方程相关联的随机过程概率密度的演化方程 [63]。这意味着，对于满足伊藤随机微分方程的马尔可夫过程而言，在给定初始状态下，其条件概率密度随时间的演化过程均可由 Fokker-Planck 方程精确描述 [64]。因此，将这一现象称为“Fokker-Planck 流”（Fokker-Planck flow, FP flow）是非常恰当的，这一术语凸显了概率密度函数严格遵循 Fokker-Planck 方程进行的动态演进与变换过程。此外，Fokker-Planck 方程还被广泛应用于非线性系统的分析中，以解决诸如锁相频率解调等多样化的实际问题 [65]。

在统计物理学与机器学习领域，Fokker-Planck 方程的重要性在于其能够描述密度演化的连续性方程；在此过程中，密度的变化完全由一个时变的速度场决定，而该速度场反过来又依赖于当前的密度函数本身 [66]。这种自洽性（self-consistency）使得 Fokker-Planck 方程成为了设计隐函数（latent functions）以及神经网络参数化模型的基础，进而可以通过生成完整的密度轨迹来逼近 Fokker-Planck 方程的解 [66]。在具体的应用领域中，例如等离子体物理和非线性滤波，Fokker-Planck 方程同样被用于描述特定的物理现象，诸如射频波与等离子体之间的相互作用，以及基于观测结果所给出的状态概率密度等 [67]。这些广泛的应用充分展示了 Fokker-Planck 方程在解决实际物理问题方面的灵活性与有效性。

近年来，A. Wibisono 和 V. Jog 提出了利用 Fokker-Planck 方程进行信道建模的新思路，并将原本基于离散马尔可夫链建模的信道扩展为由连续马尔可夫链进行建模 [3]。这一范式的转变将信道概念化为一个随时间动态演化的系统，从而使得我们能够利用随机微分方程的方法对信道进行深入分析。特别地，沿着由 SDE 定义的轨迹，概率密度函数的演化过程可以由 Fokker-Planck 方程来精确表征。因此，符合这一描述特征的信道被统称为 Fokker-Planck 信道。此外，Wibisono 和 Jog 还成功推广了 I-MMSE（互信息-最小均方误差）恒等式 [31, 68-69]，或者等价地称为 De Bruijn 恒等

---

本章内容已发表在 IEEE International Symposium on Information Theory (ISIT 2025) [62].

式 [2, 70]; 这些恒等式最初是针对加性高斯白噪声信道场景提出的, 用于描绘布朗运动的演进过程且可用简单的 SDE 表示, 而现在已被拓展至更一般的 Fokker-Planck 信道情形。

在传统信息论的框架下, 众所周知, 受数据处理不等式的约束, 互信息在马尔可夫链的传输过程中会逐渐减少 [71]。并且, 对于固定的初始概率密度函数而言, 它通常是一个凸函数。然而, 当我们转换至 Fokker-Planck 信道的视角时, 互信息被视为一个随时间变化的函数。我们已知互信息沿着 Fokker-Planck 过程是单调递减的, 这也就意味着其关于时间的一阶导数为负 [3]。此外, 现有的研究已经证明, 如果初始分布是对数凹 (log-concave) 的, 那么互信息沿着热流 (Heat flow, 即 Fokker-Planck 流的一种特例) 始终是时间的凸函数 [5]。进一步地, 如果初始分布相对于目标高斯测度具有足够强的对数凹性, 那么互信息沿着 OU 流也始终保持为时间的凸函数 [6]。然而, 目前对于互信息沿着一般 Fokker-Planck 流 (即使是最简单的线性 FPE) 演化的性质仍缺乏相关的研究, 主要原因在于一般的 Fokker-Planck 流并不像热流或 OU 流那样拥有显式解。因此, 我们需要探索一种能够规避对显式解依赖的新方法, 以深入研究互信息的凸性问题。

在本文中, 我们在满足适当假设的前提下, 深入研究了互信息沿着 Fokker-Planck 流演化的凸性, 旨在将热流情形下的相关结论推广至更一般的情形。众所周知, Fokker-Planck 流的性质与熵之间存在着紧密的联系。根据最优传输理论, Fokker-Planck 流可以被视为 Wasserstein 空间 (即赋予了 Wasserstein 度量结构的概率分布空间) 中的梯度流, 其所关联的能量泛函取为相对于稳态分布的相对熵的形式 [17, 72-75]。此外本章节还进一步拓展了信道模型, 探讨了信息量沿着跳扩散信道的信息微分特性 [76]。综上所述, 本章节的主要贡献总结如下:

### 1. 建立了 Fokker-Planck 流经典解的适定性理论框架:

针对一般非线性势函数下的 Fokker-Planck 方程, 本文并未依赖现有的显式解 (如热核或 Mehler 公式), 而是利用随机分析中的 Feynman-Kac 公式与抛物型偏微分方程理论, 严格证明了在势函数满足特定凸性假设及初始分布有界的情况下, Fokker-Planck 方程在全空间  $\mathbb{R}^N$  上经典解的存在性与唯一性。此外, 本文还进一步确立了解的光滑性 ( $C^{1,\infty}$  正则性), 从而为后续对互信息进行高阶时间求导提供了坚实的数学基础。

### 2. 推导了互信息沿 Fokker-Planck 流演化的二阶时间导数公式:

本文在 Wasserstein 空间梯度流的框架下, 通过构建精细的演化演算, 系统地

推导了相对熵与互信息沿 Fokker-Planck 流的一阶及二阶时间导数的解析表达式。该推导揭示了互信息二阶导数的符号直接取决于变换后概率密度函数（即  $\gamma_t = e^{V/2}\mu_t$ ）的 Log-Hessian 矩阵的半正定性。这一结果建立了信息量演化与概率分布微观几何结构之间的内蕴联系，为分析互信息的凸性提供了明确的代数判据。

### 3. 证明了互信息凸性的几何判据与保持性定理：

结合 Ishige 等人关于抛物型方程解的空间对数凹性（Spatial Log-concavity）理论，本文证明了互信息凸性的充分条件。具体而言，当势函数  $V(x)$  满足几何约束  $\Delta V \leq \frac{1}{2}\|\nabla V\|^2$  时，如果初始分布相对于稳态分布表现出足够强的对数凹性（即  $V/2$ -相对对数凹），则这种几何性质在随后的时间演化中将保持不变，进而保证了互信息始终是时间的凸函数。这一结论成功将热流（Heat flow）和 Ornstein-Uhlenbeck 流中的经典结果推广至了更广泛的 Fokker-Planck 流情形。

### 4. 完善了理论边界：最终凸性与无界区域逼近：

本文进一步探讨了初始条件不满足强对数凹假设时的情形，提出了“最终凸性”（Eventual Convexity）的结论：即只要存在某个时刻  $T > 0$ ，使得系统演化至该时刻后的分布满足  $V/2$ -相对对数凹条件，互信息将在  $t > T$  之后永久保持凸性。此外，针对全空间  $\mathbb{R}^N$  上的分析难题，本文采用了一系列有界光滑凸区域逼近全空间的方法，并通过比较原理和  $L^1$  收敛性分析，严格去除了对边界条件的依赖，从而证明了上述凸性定理在无界空间中的普适性。

### 5. 进一步拓展了信道模型，探讨了跳扩散信道中的信息微分特性：

基于跳扩散过程的随机微分方程模型，本文引入了更为复杂的信道模型，涵盖了连续扩散与跳跃过程的混合动力学。通过构建相应的偏微分-积分方程，系统分析了信息量在跳扩散信道中的演化规律，揭示了跳跃机制对信息微分性质的影响，为信道建模与信息理论的交叉研究提供了新的视角和工具。

## 3.2 Fokker-Planck 流的适定性理论

尽管在第二章背景知识中我们已经对 Fokker-Planck 方程进行了介绍，但为了保证本章内容的完整性与自洽性，此处我们将重新回顾其定义，并引入本章所需的特定数学符号与模型假设。考虑一个通用的 Fokker-Planck 信道，其输出为实值随机过程  $(X_t)_{t \geq 0}$ ，取值于  $\mathbb{R}^N$ ，并由如下随机微分方程（SDE）控制：

$$dX_t = a(X_t, t)dt + \sigma(X_t, t)dW_t \quad (3.1)$$

其中  $(W_t)_{t \geq 0}$  是  $\mathbb{R}^N$  上的标准布朗运动,  $a(x, t)$  是漂移系数,  $\sigma(x, t) \succ 0$  是扩散系数, 且  $\sigma(x, t) \in \mathbb{R}^{N \times N}$ . 我们记  $X_t$  在固定时刻  $t \geq 0$  的概率密度函数为  $\mu(x, t)$  或  $\mu_t(x)$  (其中空间变量  $x \in \mathbb{R}^N$ ), 则该密度函数满足 Fokker-Planck 方程:

$$\partial_t \mu = \frac{\partial \mu}{\partial t} = \Delta(D\mu) - \nabla \cdot (a\mu). \quad (3.2)$$

这里  $D(x, t) = \sigma(x, t)^\top \sigma(x, t)/2$ ,  $\nabla \cdot = \sum_{i=1}^n \frac{\partial}{\partial x_i}$  为散度算子,  $\Delta = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}$  为拉普拉斯算子。

当选择  $a \equiv 0$  且  $\sigma \equiv \sqrt{2}$  时, 方程即为热流 (Heat Flow), 对应于高斯信道。此外, 若选择  $a(x, t) = -\alpha x$  (其中  $\alpha > 0$ ) 且  $\sigma \equiv 1$ , 该方程即为 OU 流。在本章中, 我们考虑  $a(x, t)$  独立于时间  $t$  且  $\sigma \equiv \sqrt{2}$  的情形。定义势函数  $V(x)$ , 使其负梯度等于漂移系数, 即  $\nabla V = -a(x)$ , 且势函数  $V = V(x)$  满足  $e^{-V} \in \mathbb{L}^1(\mathbb{R}^N)$ 。此时, 我们得到如下形式的 Fokker-Planck 流:

$$\frac{\partial \mu}{\partial t} = \Delta \mu + \nabla \cdot (\mu \nabla V). \quad (3.3)$$

这一结果要求  $a(x, t)$  和  $\sigma(x, t)$  满足适当的正则性与增长条件, 例如光滑性和 Lipschitz 性质 [77]。在本章节中, 如无特殊说明, 文中所述的 Fokker-Planck 流特指上述形式的 Langevin 扩散。

若令  $\gamma_t = e^{\frac{V}{2}} \mu_t$ , 则 Fokker-Planck 流 (3.3) 可简化为如下形式:

$$\frac{\partial \gamma}{\partial t} = \left( \frac{1}{2} \Delta V - \frac{1}{4} \|\nabla V\|^2 \right) \gamma + \Delta \gamma \quad (3.4)$$

这即为一个 (虚时) 薛定谔方程 (Imaginary-time Schrödinger Equation)。令  $c(x) = \frac{1}{4} \|\nabla V\|^2 - \frac{1}{2} \Delta V$ 。在本文中, 我们始终假设满足以下条件:

**假设 3.1**  $V(x)$  是非负且光滑的, 并且  $c(x)$  是具有下界的凸函数。

若我们考虑 Ornstein-Uhlenbeck 过程, 则此假设成立。

**假设 3.2**  $\gamma_0(x)$  是有界的, 即  $\mu_0(x) \lesssim e^{-\frac{V(x)}{2}}$ 。

**注** 通过考虑变换  $\tilde{\gamma}_t := \gamma_t e^{-\beta t}$ , 其中  $\beta > \inf_{x \in \mathbb{R}^N} c(x)$ , 我们有:

$$\frac{\partial \tilde{\gamma}}{\partial t} + (c(x) + \beta) \tilde{\gamma} = \Delta \tilde{\gamma}. \quad (3.5)$$

由于方程 (3.4) 与 (3.5) 的存在性与唯一性是等价的, 且  $\tilde{\gamma}$  具有与  $\gamma$  相同的对数凹性质, 因此我们只需考虑  $c(x)$  为非负的情形。

根据 Fokker-Planck 方程的长时行为理论 [53, 78-79], 我们已知 Fokker-Planck 流 (3.3) 具有唯一的归一化稳态  $\mu_\infty = e^{-V}$ , 并且  $\mu_t$  与  $\mu_\infty$  之间的相对熵随  $t \rightarrow \infty$  呈指数级收敛于 0 [80]。此外,  $\mu_t - \mu_\infty$  的  $\mathbb{L}^1$  范数以及它们之间的 Wasserstein 距离也随  $t \rightarrow \infty$  呈指数级收敛于 0 [17]。

在本章中, 我们利用随机分析的方法, 结合 Feynman-Kac 公式与抛物型偏微分方程的经典解理论, 证明了在假设 3.1 和 3.2 成立的前提下, Fokker-Planck 流 (3.3) 在全空间  $\mathbb{R}^N$  上存在唯一的经典解。具体细节将在本章的证明部分 (第 3.5.1 小节) 给出。

**定理 3.1** 若假设 3.1 和 3.2 成立, 则全空间上的 Fokker-Planck 流 (3.3) 存在唯一的经典解:

$$\mu(x, t) = e^{-\frac{V(x)}{2}} \mathbb{E}_{x/\sqrt{2}} \left[ \gamma_0(\sqrt{2}W_t) e^{-\int_0^t c(\sqrt{2}W_r) dr} \right] \quad (3.6)$$

其中  $c(x) = \frac{1}{4} \|\nabla V\|^2 - \frac{1}{2} \Delta V$ ,  $\gamma_0 = e^{\frac{V}{2}} \mu_0$ 。此外, 该解满足  $\mu(x, t) \in C^{1,\infty}((0, \infty) \times \mathbb{R}^N)$ 。

进而, 该解可以表示为核函数与  $\mu_0$  的卷积形式:

$$\begin{aligned} \mu(x, t) &= e^{-\frac{V(x)}{2}} \int_{\mathbb{R}^N} K(x, y, t) \gamma_0(y) dy \\ &= \int_{\mathbb{R}^N} K(x, y, t) e^{\frac{V(y)-V(x)}{2}} \mu_0(y) dy. \end{aligned} \quad (3.7)$$

**证明** 证明见第 3.5.1 小节。 ■

此外, 我们也可以通过算子半群来表示该解 [81]:

$$\mu(x, t) = Q_t \mu_0 := e^{-V} P_t (e^V \mu_0) \quad (3.8)$$

其中  $P_t = e^{tL}$  是由 Witten-Laplace 算子  $L = \Delta - \nabla V \cdot \nabla$  生成的热半群。

事实上, 如果该 Fokker-Planck 流仅存在于有界区域内, 根据文献 [82] 的第 2.4.4 小节, 其解也可以表示为如下形式:

$$\mu(x, t) = \int_{\Omega} K_{\Omega}(x, y, t) e^{\frac{V(y)-V(x)}{2}} \mu_0(y) dy \quad (3.9)$$

其中  $K_{\Omega}$  是具有光滑边界的有界区域  $\Omega$  中对应于该方程的核函数。

### 3.3 互信息沿 FP 流的凸性分析

#### 3.3.1 抛物型系统解的对数凹性

为保证本章内容的完整性与自洽性, 此处我们将重新回顾对数凹性质的定义, 下面的表述与第 2.3 介绍的定义内容等价:

**定义 3.1** 称  $\mathbb{R}^N$  上的非负函数  $u$  是对数凹 (log-concave) 的, 如果对于任意  $t \in [0, 1]$  和使得  $u(x)u(y) > 0$  的  $x, y \in \mathbb{R}^N$ , 满足

$$u((1-t)x + ty) \geq u(x)^{1-t}u(y)^t \quad (3.10)$$

这等价于以下条件:

1. 集合  $S_u := \{x \in \mathbb{R}^N : u(x) > 0\}$  是凸集, 且  $\log u$  在  $S_u$  上是凹函数。
2.  $u$  可以表示为  $u = e^{-\phi}$ , 其中  $\phi$  是一个凸函数。

通过研究高斯核的对数凹性以及 Prekópa-Leindler 不等式, Brascamp 和 Lieb 证明了热流保持对数凹性 [83]。此外, Lee 和 Vázquez 证明了如果初始分布是  $\mathbb{R}^N$  中具有紧支集的有界非负函数, 那么它沿热流最终会变为对数凹的 [84]。

另外, Ishige 等人引入了更广义的凹性概念, 包括空间凹性 (spatial concavity) [85] 和 F-凹性 (F-concavity) [86], 旨在检验热流过程中所保持的最强或最弱凹性性质。他们对抛物型系统解的对数凸性进行了深入研究。

**引理 3.1** (Theorem 1.1 in [85]) 设  $\Omega$  为  $\mathbb{R}^N$  中的有界凸区域, 且  $d_1, d_2 > 0$ 。令  $D := \Omega \times (0, \infty)$ , 设  $(u, v) \in C^{2,1}(D; \mathbb{R}^2) \cap C(\bar{D}; \mathbb{R}^2)$  满足

$$\begin{cases} \partial_t u - d_1 \Delta u + f(x, t, u, v, \nabla u) = 0 & \text{in } D \\ \partial_t v - d_2 \Delta v + g(x, t, u, v, \nabla v) = 0 & \text{in } D \\ u, v \geq 0 & \text{in } D \\ u(x, t) = v(x, t) = 0 & \text{on } \partial\Omega \times [0, \infty) \\ u(x, 0) = u_0(x), v(x, 0) = v_0(x) & \text{in } D \end{cases} \quad (3.11)$$

其中  $f, g$  是  $D \times [0, \infty)^2 \times \mathbb{R}^N$  上的非负连续函数。假设满足以下条件:

1. 上述系统满足粘性比较原理 (viscosity comparison principle)。
2. 对于任意固定的  $t > 0$  和  $\theta \in \mathbb{R}^N$ , 函数

$$\mathbf{f}_{t,\theta}(x, r, s) := e^{-r} f(x, t, e^r, e^s, e^r \theta) \quad (3.12)$$

和

$$\mathbf{g}_{t,\theta}(x, r, s) := e^{-s} g(x, t, e^r, e^s, e^s \theta) \quad (3.13)$$

在  $\Omega \times (0, +\infty)^2$  上是凸的。

如果  $\log u_0$  和  $\log v_0$  在  $\Omega$  中是凹的, 则对于任意固定的  $t > 0$ ,  $\log u(\cdot, t)$  和  $\log v(\cdot, t)$  在  $\Omega$  中也是凹的。

考虑虚时薛定谔方程：

$$\frac{\partial \gamma}{\partial t} + c(x)\gamma - \Delta \gamma = 0 \quad (3.14)$$

其中  $c(x)$  是非负且足够光滑的。基于假设 3.1 和 3.2，已确立方程 (3.14) 拥有唯一的经典解。此外，由于极大值原理适用于此类抛物型偏微分方程——特别是经典的比较原理——因此可以证明粘性比较原理同样适用于方程 (3.14) [87]。

**推论 3.1** 设  $\Omega$  为  $\mathbb{R}^N$  中的有界凸区域，并设  $u \in C^2(D) \cap C(\bar{D})$ ,  $D := \Omega \times (0, \infty)$  满足

$$\begin{cases} \partial_t u + c(x)u - \Delta u = 0 & \text{in } D \\ u(x, t) = 0 & \text{on } \partial\Omega \times [0, \infty) \\ u(x, 0) = u_0(x) & \text{in } \Omega \end{cases} \quad (3.15)$$

其中  $u_0$  是  $\bar{\Omega}$  上的非负连续函数，且假设 3.1 成立。若  $u_0$  在  $\Omega$  中是对数凹的，则对于任意固定的  $t > 0$ ， $u_t$  在  $\Omega$  中也是对数凹的。

### 3.3.2 互信息沿 FP 流的凸性

在本小节中，我们将前述基本量的定义应用于联合随机变量  $(X, Y) = (X_0, X_t)$ ，其中  $X_0 \sim \mu_{X_0}$  代表初始分布， $X_t \sim \mu_{X_t}$  代表沿 Fokker-Planck 流从  $\mu_{X_0}$  演化到给定时间  $t \geq 0$  时的分布。为简化记号，我们将  $\mu_{X_0}$  和  $\mu_{X_t}$  分别简记为  $\mu_0$  和  $\mu_t$ 。此外，联合随机变量  $(X_0, X_t)$  的分布记为  $\mu_{0t}$  或  $\mu_{t0}$ 。同时，条件随机变量  $X_t | \{X_0 = x_0\}$  的分布记为  $\mu_{t|0}(\cdot | x_0)$ 。

**引理 3.2** 沿 Fokker-Planck 流，对于参考测度  $\nu = \mu_\infty = e^{-V}$ ，我们有：

$$\begin{aligned} \frac{d}{dt} H_\nu(\mu_t) &= -J_\nu(\mu_t) = \int_{\mathbb{R}^N} \mu_t \left\| \nabla \log \frac{\mu_t}{\nu} \right\|^2 dx \\ \frac{d^2}{dt^2} H_\nu(\mu_t) &= 2K_\nu(\mu_t) + 2G_\nu(\mu_t) \\ \frac{d^2}{dt^2} H_\nu(\mu_t | \mu_0) &= 2K_\nu(\mu_t | \mu_0) + 2G_\nu(\mu_t | \mu_0) \end{aligned} \quad (3.16)$$

其中

$$G_\nu(\mu) = \int_{\mathbb{R}^N} \mu \left\langle (\nabla^2 V) \nabla \log \frac{\mu}{\nu}, \nabla \log \frac{\mu}{\nu} \right\rangle dx \quad (3.17)$$

注意  $\nu$  和  $V$  的变量由  $x_t$  表示。这里  $\langle \cdot, \cdot \rangle$  表示内积。

证明见第 3.5.2 小节。

如果我们选择函数  $U(x) = x \log x$ ，此引理即为文献 [17] 中 *Theorem 24.2* 的直接推论。此时能量泛函  $U_\nu$  退化为相对熵  $H_\nu$ 。因此，我们可以计算互信息沿 Fokker-Planck 流的导数。

**定理 3.2** 沿 Fokker-Planck 流, 对于参考测度  $\nu = \mu_\infty = e^{-V}$ , 我们有:

$$\begin{aligned} \frac{d}{dt} I(X_0; X_t) &= -J_\nu(\mu_0; \mu_t) \leq 0 \\ \frac{d^2}{dt^2} I(X_0; X_t) &= 2\Psi(\mu_0|\mu_t) + 4 \int_{\mathbb{R}^{2N}} \mu_{0,t} \langle \nabla^2 \log \gamma_t, \nabla^2 \log \mu_{0|t} \rangle_{\text{HS}} dx_0 dx_t \\ &= 2\Psi(\mu_0|\mu_t) - 4 \int_{\mathbb{R}^{2N}} \mu_{0,t} (\nabla \log \mu_{0|t})^\top (\nabla^2 \log \gamma_t) (\nabla \log \mu_{0|t}) dx_0 dx_t. \end{aligned} \quad (3.18)$$

此处  $\gamma_t := e^{\frac{V}{2}} \mu_t$ , 且给定  $\mu_t$  时  $\mu_0$  的后向二阶 Fisher 信息定义为:

$$\Psi(\mu_0|\mu_t) = \int_{\mathbb{R}^{2N}} \mu_{t,0} \left\| \nabla^2 \log \mu_{0|t} \right\|_{\text{HS}}^2 dx_0 dx_t. \quad (3.19)$$

有时我们将  $\mathbb{R}^{2N}$  替换为  $\mathbb{R}_0^N \times \mathbb{R}_t^N$  以区分两个变量  $x_0, x_t$  的空间。

证明见第 3.5.3 小节。

接下来, 我们阐明关于互信息沿 Fokker-Planck 流凸性性质的主要结果。在整个小节中, 我们指定  $X_t \sim \mu_t$  表示从初始状态  $X_0 \sim \mu_0$  演化的 FP 过程。证明思路遵循 Ishige 等人的工作 [88]。

**引理 3.3** 令假设 3.1, 3.2 成立。设  $\Omega$  为有界光滑凸区域,  $\mu_0 \in C(\bar{\Omega})$  且在  $\partial\Omega$  上  $\mu_0 = 0$ 。此外, 假设  $X_0 \sim \mu_0$  是  $\frac{V}{2}$ -相对对数凹的, 其中  $V(x)$  是 FP 流 (3.3) 的势函数。那么互信息  $I(X_0; X_t)$  关于  $t$  恒为凸函数。

证明见第 3.5.4 小节。

若对于某个严格非负函数  $\lambda(x)$ , 满足  $-\nabla^2 \log \mu(x) \succeq \nabla^2 \lambda(x)$ , 我们称  $X \sim \mu$  是  $\lambda$ -相对对数凹的。进一步地, 我们证明  $\partial\Omega$  上  $\mu_0 = 0$  这一条件实际上是多余的。

**引理 3.4** 令假设 3.1, 3.2 成立。设  $\Omega$  为有界光滑凸区域,  $\mu_0 \in C(\bar{\Omega})$ , 但不假设在  $\partial\Omega$  上  $\mu_0 = 0$ 。此外, 假设  $\mu_0$  是  $\frac{V}{2}$ -相对对数凹的, 其中  $V(x)$  是 FP 流 (3.3) 的势函数。那么互信息  $I(X_0; X_t)$  关于  $t$  恒为凸函数。

证明见第 3.5.5 小节。

对于  $\mathbb{R}^N$  中的无界凸区域  $\Omega$ , 我们考虑使用一系列有界光滑凸区域来逼近它。进而我们推导出  $\Omega$  的有界性也是多余的。在接下来的讨论中, 我们仅考虑  $\Omega = \mathbb{R}^N$  的情形。

**定理 3.3** 令假设 3.1, 3.2 成立。设  $\mu_0 \in C(\mathbb{R}^N)$ 。此外, 假设  $\mu_0$  是  $\frac{V}{2}$ -相对对数凹的, 其中  $V(x)$  是 FP 流 (3.3) 的势函数。那么互信息  $I(X_0; X_t)$  关于  $t$  恒为凸函数。

证明见第 3.5.6 小节。

特别地，如果存在一个足够大的时刻，使得该时刻的分布是  $\frac{V}{2}$ -相对对数凹的，那么在此之后互信息将保持凸性。

**推论 3.2** 令假设 3.1, 3.2 成立。设  $\mu_T \in C(\mathbb{R}^N), T \geq 0$ 。如果  $\mu_T$  是  $\frac{V}{2}$ -相对对数凹的，其中  $V(x)$  是 FP 流 (3.3) 的势函数。那么互信息  $I(X_0; X_t)$  对于  $t > T$  保持凸性。

上述结论的证明均基于对数凹性质在 Fokker-Planck 流(3.3)下的保持性。实际上，若  $\Delta$  前的系数不为 1，类似的结论同样成立，证明过程几乎等同。

**定理 3.4** 设概率测度  $\mu_t$  沿 FP 流演化， $c > 0$  为常数：

$$\begin{cases} \partial_t \mu_t = \frac{c}{2} \Delta \mu + \nabla \cdot (\mu \nabla V) & \text{in } \mathbb{R}^d \\ \mu(x, 0) = \mu_0(x) & \text{in } \mathbb{R}^d \end{cases} \quad (3.20)$$

如果假设 3.1 与 3.2 成立，且进一步假设  $\mu_0$  是  $\frac{V}{c}$ -相对对数凹的，那么对于所有  $t > 0$ ， $\mu_t$  也是  $\frac{V}{c}$ -相对对数凹的。

### 3.4 扩展模型：跳扩散信道下的信息微分性质

在上一节中，我们深入探讨了 Fokker-Planck 流及其相关信道的几何性质。这类模型本质上是由布朗运动驱动的随机微分方程（SDE）所描述的，其概率密度函数的演化遵循 Fokker-Planck 方程（亦称为前向 Kolmogorov 方程）。这类过程的一个显著特征在于其样本轨迹的连续性，即系统状态随时间的演化是平滑且无突变的。

然而，在诸如移动通信、复杂网络以及金融市场等实际应用场景中，信道特性往往更为复杂。系统不仅受到背景噪声的连续扰动（扩散项），还经常遭受突发性的、大幅度的脉冲干扰或状态跳变。例如，无线通信中的深衰落、金融资产价格的剧烈波动或神经网络中的脉冲信号，均表现出显著的非连续性特征 [89-90]。对于此类包含“突变”特性的动力学行为，仅依靠连续的 Fokker-Planck 扩散模型已不足以进行精确刻画。

为此，本节将研究范围从单纯的扩散过程拓展至更一般的跳-扩散过程（Jump-diffusion Process）。在物理描述上，该模型由确定性的漂移（Drift）、连续的扩散（Diffusion）以及突发的跳跃（Jumps）三部分构成。我们将基于 Kramers-Moyal 展开和 Kolmogorov 积分-微分方程，研究此类信道下熵与互信息的演化规律，并将 I-MMSE 恒等式及 de Bruijn 恒等式推广至这一更广泛的框架中。

本节内容已发表在 IEEE International Symposium on Information Theory (ISIT 2025) [76].

### 3.4.1 跳-扩散信道模型与演化方程

考虑一个跳-扩散信道，其输出为一个实值随机过程  $(X_t)_{t \geq 0}$ ，该过程服从如下随机微分方程（SDE）：

$$dX_t = a(X_t, t) dt + \sqrt{b(X_t, t)} dW_t + \Xi(X_t, t) dN_t \quad (3.21)$$

其中  $(W_t)_{t \geq 0}$  是标准布朗运动； $(N_t)_{t \geq 0}$  是强度为  $\lambda(X_t, t)$  的泊松过程（Poisson process）[91]； $\Xi \sim w(\xi|X_t, t)$  是一个实值随机变量，表示跳跃的幅度。为分析简便，我们假设系数函数  $a(x, t)$ 、 $b(x, t) \geq 0$ 、 $\lambda(x, t) \geq 0$  以及跳跃分布  $w(\xi|x, t) \geq 0$  关于所有变量均是光滑的。在物理意义上，漂移项  $a(X_t, t) dt$  描述确定性演化，扩散项  $\sqrt{b(X_t, t)} dW_t$  引入连续的小尺度噪声，而跳跃项  $\Xi(X_t, t) dN_t$  则刻画突发的大尺度畸变。

对于微小的时间间隔  $\Delta t > 0$ ，我们定义传播子（Propagator） $\Delta X_t$  为：

$$\begin{aligned} \Delta X_t &:= X_{t+\Delta t} - X_t \\ &= a(X_t, t)\Delta t + \sqrt{b(X_t, t)\Delta t}Z + \Xi(X_t, t)Y + \tilde{\mathcal{O}}(\Delta t^2) \end{aligned} \quad (3.22)$$

其中  $Z \sim \mathcal{N}(0, 1)$  服从标准正态分布， $Y \sim \text{B}(1, \lambda(X_t, t)\Delta t)$  服从伯努利分布（以概率  $\lambda(X_t, t)\Delta t$  取值为 1）。符号  $\tilde{\mathcal{O}}(\Delta t^2)$  表示该项在概率意义下是  $\Delta t^2$  的高阶无穷小，即：如果随机变量  $U = \tilde{\mathcal{O}}(\Delta t^2)$ ，意味着  $1 - \mathbb{P}(U \neq \mathcal{O}(\Delta t^2)) = \mathcal{O}(\Delta t^2)$ 。

跳-扩散过程具有极强的建模能力，除某些具有无限跳跃率或非线性时间依赖系数的 Lévy 驱动过程外，它几乎可以覆盖广泛的马尔可夫过程 [89]。以下引理给出了传播子的概率密度函数。

**引理 3.5** (引理 1, [76]) 方程 (3.22) 中传播子  $\Delta X_t$  的概率密度函数（PDF）为：

$$\begin{aligned} p_{\Delta X}(\xi|\Delta t; x, t) &= w(\xi|x, t)\lambda(x, t)\Delta t \\ &+ \frac{1 - \lambda(x, t)\Delta t}{\sqrt{2\pi b(x, t)\Delta t}} \exp\left(-\frac{(\xi - a(x, t)\Delta t)^2}{2b(x, t)\Delta t}\right) + \mathcal{O}(\Delta t^2). \end{aligned} \quad (3.23)$$

进一步地，过程  $X_t$  的概率密度函数的演化遵循 Kramers-Moyal 展开。

**引理 3.6** (引理 4, [76])  $X_t$  的概率密度函数  $p(x, t)$  满足 Kramers-Moyal 方程 [92]：

$$\frac{\partial p(x, t)}{\partial t} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} [B_n(x, t)p(x, t)] \quad (3.24)$$

其中  $B_n(x, t)$  为随机过程  $(X_t)_{t \geq 0}$  的第  $n$  阶传播子矩函数（Propagator Moment Function），定义为：

$$B_n(x, t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[\Delta X_t^n | X_t = x]}{\Delta t}. \quad (3.25)$$

$B_n(x, t)$  的值不仅可以通过有限采样间隔的时间序列数据进行估计 [93], 也可以直接由 SDE (3.21) 中的系数函数导出:

**定理 3.5** (定理 1, [76]) 传播子矩函数具有如下形式:

$$B_n(x, t) = \begin{cases} \lambda(x, t)w_1(x, t) + a(x, t), & n = 1 \\ \lambda(x, t)w_2(x, t) + b(x, t), & n = 2 \\ \lambda(x, t)w_n(x, t), & n \geq 3 \end{cases} \quad (3.26)$$

其中  $w_n(x, t) = \int_{\mathbb{R}} \xi^n w(\xi|x, t) d\xi$  是跳跃分布  $w(\cdot|x, t)$  的第  $n$  阶矩。

该定理表明, 扩散部分的高阶矩是  $\Delta t$  的高阶无穷小, 因此在极限下消失; 而跳跃部分的高阶矩则得以保留, 这正是跳跃过程非连续性的数学体现。

此外, 由 Pawula 定理可知, 并非所有的传播子矩函数集合  $\{B_n(x, t)\}$  都能生成合法的马尔可夫过程 [94]。为了确保过程的存在性, 我们可以将概率密度的演化方程重写为 Kolmogorov 积分-微分方程的形式。

**定理 3.6** (定理 2, [76])  $X_t$  的概率密度函数  $p(x, t)$  满足如下 Kolmogorov 方程 [89]:

$$\begin{aligned} \frac{\partial p(x, t)}{\partial t} = & \int_{\mathbb{R}} p(x - \xi, t) \lambda(x - \xi, t) w(\xi|x - \xi, t) d\xi - p(x, t) \lambda(x, t) \\ & - \frac{\partial}{\partial x} (a(x, t) p(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (b(x, t) p(x, t)). \end{aligned} \quad (3.27)$$

方程 (3.27) 的右端清晰地展示了物理过程的分离: 前两项 (积分项) 对应于纯跳跃过程 (3.28) 的主方程 (Master Equation):

$$dX_t = \Xi(X_t, t) dN_t, \quad (3.28)$$

而后两项 (微分项) 则对应于纯扩散过程的 Fokker-Planck 方程 (2.70)。这表明在微小时间间隔内, 跳-扩散过程中的跳跃部分与扩散部分可以被独立研究。

### 3.4.2 熵与互信息的微分性质

我们将加性高斯信道和扩散信道中的 I-MMSE 恒等式推广至跳-扩散信道, 并将熵和互信息的时间导数表示为级数展开、Fisher 型信息以及 KL 散度的形式。

**定理 3.7** (定理 5, [76]) 信道 (3.21) 输出过程熵  $h(X_t)$  的时间导数为:

$$\begin{aligned} \frac{d}{dt} h(X_t) = & \frac{1}{2} J_{b_t}(X_t) + \mathbb{E} \left[ \lambda(X_t, t) \log \frac{p_{X_t}(X_t)}{p_{X_t}(X_t + \Xi)} \right] \\ & + \mathbb{E} \left[ \frac{\partial}{\partial x} a(X_t, t) - \frac{1}{2} \frac{\partial^2}{\partial x^2} b(X_t, t) \right]. \end{aligned} \quad (3.29)$$

需要注意的是，与高斯信道不同，由于 SDE (3.21) 中函数的任意性，输出熵随时间并不一定单调增加。

**定理 3.8** (定理 6, [76]) 信道 (3.21) 的输入  $X_0$  与输出  $X_t$  之间的互信息  $I(X_0; X_t)$  的时间导数为：

$$\frac{d}{dt}I(X_0; X_t) = -\mathbb{E} \left[ \lambda(X_t, t) D(p_{X_0|X_t} \parallel p_{X_0|X_t, \Xi} | \Xi, X_t) \right] - \frac{1}{2} J_{b_t}(X_0; X_t) \quad (3.30)$$

其中  $p_{X_0|X_t, \xi} := p_{X_0|X_t}(x_0|x_t + \xi)$  表示条件概率密度  $p_{X_0|X_t}$  在空间上的平移  $\xi$ ；而

$$D(p_{X_0|X_t} \parallel p_{X_0|X_t, \Xi} | \Xi = \xi, X_t = x_t) = \int_{\mathbb{R}} p_{X_0|X_t}(x_0|x_t) \log \frac{p_{X_0|X_t}(x_0|x_t)}{p_{X_0|X_t}(x_0|x_t + \xi)} dx_0 \quad (3.31)$$

是给定  $\Xi = \xi$  和  $X_t = x_t$  时，原分布与平移分布之间的 KL 散度。

公式 (3.30) 右端的第一项表示错配 KL 散度 (Mismatched KL Divergence) 的期望，这归因于跳跃过程；第二项则表示 Fisher 型信息，归因于扩散过程。由于这两项均非负，我们立即得到以下推论：

**推论 3.3** (推论 2, [76])

$$\frac{d}{dt}I(X_0; X_t) \leq -\frac{1}{2} J_{b_t}(X_0; X_t) = -\frac{1}{2} \text{mmse}_{b_t}(\varphi(X_0, X_t)|X_t). \quad (3.32)$$

这表明，跳-扩散信道的互信息总是随时间递减的。这一结论可以看作是连续时间情形下的数据处理不等式 (Data Processing Inequality) [71]。

### 3.4.3 KL 散度的演化与强数据处理不等式

KL 散度是度量两个概率分布差异的核心指标。我们进一步研究了它沿跳-扩散流演化的微分特性及其数据处理不等式。

**定理 3.9** 假设  $(\mu_t)_{t \geq 0}$  和  $(\nu_t)_{t \geq 0}$  均遵循跳-扩散信道的演化方程：

$$\begin{aligned} \partial_t \mu_t &= \int_{\mathbb{R}} \mu(x - \xi, t) \lambda(x - \xi, t) w(\xi|x - \xi, t) d\xi - \mu(x, t) \lambda(x, t) \\ &\quad - \frac{\partial}{\partial x} (a(x, t) \mu(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (b(x, t) \mu(x, t)), \\ \partial_t \nu_t &= \int_{\mathbb{R}} \nu(x - \xi, t) \lambda(x - \xi, t) w(\xi|x - \xi, t) d\xi - \nu(x, t) \lambda(x, t) \\ &\quad - \frac{\partial}{\partial x} (a(x, t) \nu(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (b(x, t) \nu(x, t)). \end{aligned} \quad (3.33)$$

则对于任意  $t \geq 0$ , 我们有:

$$\begin{aligned} \frac{d}{dt} \text{KL}(\mu_t \| \nu_t) &= -\frac{1}{2} \int_{\mathbb{R}} b(x, t) \mu(x, t) \left\| \frac{\partial}{\partial x} \log \frac{\mu_t}{\nu_t} \right\|^2 dx \\ &\quad + \int_{\mathbb{R}} \mu(x, t) \lambda(x, t) \int_{\mathbb{R}} w(\xi | x, t) \left\{ \log \frac{\mu(x + \xi, t) \nu(x, t)}{\nu(x + \xi, t) \mu(x, t)} - \frac{\mu(x + \xi, t) \nu(x, t)}{\nu(x + \xi, t) \mu(x, t)} + 1 \right\} d\xi dx \\ &\leq -\frac{1}{2} \int_{\mathbb{R}} b(x, t) \mu(x, t) \left\| \frac{\partial}{\partial x} \log \frac{\mu_t}{\nu_t} \right\|^2 dx \leq 0. \end{aligned} \quad (3.34)$$

证明见第 3.5.7 小节。

基于上述定理, 我们可以得到关于 KL 散度的强数据处理不等式 (SDPI):

**推论 3.4** 如果扩散系数  $b(x, t)$  具有正下界, 即对于任意  $(x, t) \in \mathbb{R} \times \mathbb{R}_+$  均有  $b(x, t) \geq c > 0$ , 且概率测度  $\nu_t$  满足  $\alpha_t$ -对数索伯列夫不等式 (Log-Sobolev Inequality), 则 KL 散度满足如下指数衰减规律:

$$\text{KL}(\mu_t \| \nu_t) \leq \exp\left(-c \int_0^t \alpha_s ds\right) \text{KL}(\mu_0 \| \nu_0). \quad (3.35)$$

尽管我们推导了跳-扩散信道下 KL 散度的强数据处理不等式, 但仍存在诸多开放性问题的进一步探索:

- $\nu_t$  满足对数索伯列夫不等式的条件: 根据 [95] 的理论, 如果测度  $\nu_t$  在紧集之外是  $\beta_t$ -强对数凹的, 则其满足 Log-Sobolev 不等式。然而, 在跳-扩散演化下验证这一几何性质, 并给出常数  $\alpha_t$  的显式估计, 依然是一个极具挑战性的数学难题。
- Wasserstein 空间中的演化: 对于纯扩散过程, 概率密度的演化路径在 Wasserstein 空间中通常是绝对连续的, 便于利用 Otto 演算进行分析。然而, 跳跃项的引入可能会破坏路径的连续性, 导致概率测度可能出现奇异性。如何在包含跳跃的非连续框架下定义和计算 Wasserstein 距离的收缩性, 仍有待建立新的最优传输理论框架。

## 3.5 本章定理证明

### 3.5.1 定理 3.1 的证明

本节首先通过一个例子展示随机分析方法在求解偏微分方程 (PDE) 中的应用。

**例 3.1** 考虑如下热流方程:

$$\begin{cases} \partial_t u = \frac{1}{2} \Delta u, & (t, x) \in \mathbb{R}_+ \times \mathbb{R}^N \\ u(0, x) = u_0(x), & u_0 \in C_b(\mathbb{R}^N) \end{cases}$$

这里我们假设  $u \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^N)$ 。定义一个随机过程  $Z_s := u((t-s), W_s), 0 \leq s \leq t$ 。根据伊藤公式 (Itô's formula), 我们有:

$$\begin{aligned} dZ_s &= (-1) \partial_t u ds + \partial_i u dW_s^i + \frac{1}{2} \partial_{ij} u d\langle W_s^i, W_s^j \rangle \\ &= -\partial_t u + \partial_i u dW_s^i + \frac{1}{2} \Delta u ds \\ &= 0 + \partial_i u dW_s^i. \end{aligned} \tag{3.36}$$

因此,  $Z_s$  是一个鞅 (martingale)。此处我们使用了爱因斯坦求和约定,  $\partial_i u$  表示  $u$  关于  $x_i$  的偏导数,  $\langle X, Y \rangle$  表示两个过程  $X, Y$  的二次变差。由鞅的性质可知:

$$\begin{aligned} \mathbb{E}[Z_t] &= \mathbb{E}[Z_0] = \mathbb{E}[u(t, W_0)] \\ &= u(t, W_0) = u(t, x). \end{aligned} \tag{3.37}$$

注意, 我们要始终将  $W_0$  视为起始点  $x$ 。此外, 该期望可以通过积分表示为:

$$\begin{aligned} \mathbb{E}[Z_t] &= \mathbb{E}[u(0, W_t)] = \mathbb{E}[u_0(W_t)] \\ &= \int_{\mathbb{R}^N} u_0(y) (2\pi t)^{-\frac{N}{2}} e^{-\frac{\|y-x\|^2}{2t}} dy. \end{aligned} \tag{3.38}$$

因此, 我们得到解的显式表达:

$$u(t, x) = \int_{\mathbb{R}^N} u_0(y) (2\pi t)^{-\frac{N}{2}} e^{-\frac{\|y-x\|^2}{2t}} dy. \tag{3.39}$$

在方程 (3.39) 中, 我们求解了给定初始分布的热方程。接下来, 对于薛定谔方程 (3.4) 的情形, 我们可以利用类似的方法进行证明。

**引理 3.7** 考虑如下薛定谔方程:

$$\begin{cases} \partial_t u = \Delta u - c(x)u, & (t, x) \in \mathbb{R}_+ \times \mathbb{R}^N \\ u(0, x) = u_0(x), & u_0 \in C_b(\mathbb{R}^N) \end{cases} \tag{3.40}$$

假设  $u \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^N)$ , 即  $u$  关于  $t$  是  $C^1$  的, 关于  $x$  是  $C^2$  的。那么该解具有如下表示形式:

$$\begin{aligned} u(t, x) &= \mathbb{E}_{x/2} \left[ u_0(2W_{t/2}) e^{-2 \int_0^{t/2} c(2W_r) dr} \right] \\ &= \mathbb{E}_{x/\sqrt{2}} \left[ u_0(\sqrt{2}W_t) e^{-\int_0^t c(\sqrt{2}W_r) dr} \right] \\ &\equiv \int_{\mathbb{R}^N} K(x, y, t) u_0(y) dy \end{aligned} \tag{3.41}$$

**证明** 定义如下过程:

$$\begin{aligned} Z_s &:= u(2(t-s), 2W_s) e^{-2 \int_0^s c(2W_r) dr}, 0 \leq s \leq t \\ Y_s &:= e^{-2 \int_0^s c(2W_r) dr}, 0 \leq s \leq t. \end{aligned} \quad (3.42)$$

根据伊藤公式, 我们有

$$\begin{aligned} dZ_s &= Y_s \left( -2\partial_t u ds + 2\partial_i u dW_s^i + \frac{1}{2} \partial_{ij} u d \langle 2W_s^i, 2W_s^j \rangle \right) \\ &\quad + u \cdot \left[ -2Y_s \cdot c(2W_s) ds + M_i(W_s^i) dW_s^i + \frac{1}{2} d \langle Y_s^i, Y_s^j \rangle \right] \\ &\quad + d \langle u(2(t-s), 2W_s), Y_s \rangle \\ &\stackrel{(i)}{=} 2Y_s (-\partial_t u + \Delta u + c(2W_s) u) ds + u \cdot M_i(W_s^i) dW_s^i + 0 \\ &= u \cdot M_i(W_s^i) dW_s^i \end{aligned} \quad (3.43)$$

其中等号 (i) 成立是因为  $Y_s$  是一个有界变差 (bounded variation) 过程, 而  $M_i$  表示  $Y_s$  关于  $x_i$  的偏导数项。因此  $Z_s$  是一个鞅。利用鞅的性质, 可得:

$$\begin{aligned} \mathbb{E}[Z_t] &= \mathbb{E}[Z_0] = \mathbb{E}[u(2t, 2W_0)] = u(2t, 2x) \\ \mathbb{E}[Z_t] &= \mathbb{E}_x \left[ u(0, 2W_t) e^{-2 \int_0^t c(2W_r) dr} \right] \\ &= \mathbb{E}_x \left[ u_0(2W_t) e^{-2 \int_0^t c(2W_r) dr} \right] \\ u(t, x) &= \mathbb{E}_{x/2} \left[ u_0(2W_{t/2}) e^{-2 \int_0^{t/2} c(2W_r) dr} \right] \\ &\equiv \int_{\mathbb{R}^N} K(x, y, t) u_0(y) dy. \end{aligned} \quad (3.44)$$

在此语境下,  $\mathbb{E}$  的下标表示布朗运动的起始点,  $K(x, y, t)$  是一个固定的核函数。该核函数的作用类似于热核在求解热流问题时的作用。此外, 如果我们取

$$Z_t := u(t-s, \sqrt{2}W_s) e^{-\int_0^s c(\sqrt{2}W_r) dr}$$

那么通过同样的方法, 我们可以得到

$$u(t, x) = \mathbb{E}_{x/\sqrt{2}} \left[ u_0(\sqrt{2}W_t) e^{-\int_0^t c(\sqrt{2}W_r) dr} \right]. \quad (3.45)$$

■

基于上述讨论, 我们考虑证明一类方程经典解的存在性和唯一性:

$$\begin{cases} -\partial_t u + \Delta u - c(x)u = 0, & (t, x) \in \mathbb{R}_+ \times \mathbb{R}^N \\ u(0, x) = u_0(x), & u_0 \in C_b(\mathbb{R}^N) \end{cases} \quad (3.46)$$

其中  $c(x)$  是非负凸光滑函数。首先，我们考虑具有光滑边界的紧区域上的柯西问题 (Cauchy problem):

$$\begin{cases} -\partial_t u + \Delta u - c(x)u = 0, & (t, x) \in \mathbb{R}_+ \times \Omega \\ u(0, x) = u_0(x), & \text{for } x \in \bar{\Omega} \end{cases} \quad (3.47)$$

根据文献 [96] 第一章中的 *Theorem 10*, 方程 (3.47) 存在唯一的经典解, 并且 (3.9) 中的核函数  $K_\Omega(x, y, t)$  在  $\Omega$  内是有界的, 这对应于取  $\tau = 0$  时的基本解。接着, 我们利用文献 [97] 中的方法, 证明全空间问题 (3.46) 经典解的存在性与唯一性。

**引理 3.8** (*Feynman-Kac 解的连续性*) 令  $v(t, x) := \mathbb{E}_{x/2} \left[ u_0(2W_{t/2}) e^{-2 \int_0^{t/2} c(2W_r) dr} \right]$  如引理 3.7 证明中所定义, 则  $v(t, x)$  在  $[0, \infty) \times \mathbb{R}^N$  上是连续的。

**证明** 在此定义  $W_x(r) \equiv W_r(x)$  为从  $x$  出发、时间为  $r$  的布朗运动。则可以将  $v(t, x)$  重写为:

$$v(t, x) = \mathbb{E} \left[ u_0(2W_{x/2}(t/2)) e^{-2 \int_0^{t/2} c(2W_{x/2}(r)) dr} \right]. \quad (3.48)$$

为了简化该函数, 我们转而考虑证明  $v_1(t, x) := v(2t, 2x) = \mathbb{E} \left[ u_0(2W_x(t)) e^{-2 \int_0^t c(2W_x(r)) dr} \right]$  的连续性。设序列  $\{(t_n, x_n)\}_{n \in \mathbb{N}}, (t, x) \subset (0, \infty) \times \mathbb{R}^N$ , 并假设  $(t_n, x_n) \xrightarrow{n \rightarrow \infty} (t, x)$ 。我们需要证明

$$v_1(t_n, x_n) \xrightarrow{n \rightarrow \infty} v_1(t, x) \quad (3.49)$$

即对于任意  $\epsilon > 0$ , 存在  $N_1 \in \mathbb{N}$  使得对任意  $n \geq N_1$ , 均有

$$|v_1(t_n, x_n) - v_1(t, x)| < \epsilon. \quad (3.50)$$

设  $\epsilon > 0, 0 < \alpha \ll 1$ , 并取  $N_1 \in \mathbb{N}$  使得

$$\|(t_n, x_n) - (t, x)\| < \alpha, \text{ 当 } n \geq N_1. \quad (3.51)$$

此时若  $n \geq N_1$ , 我们有  $t - \alpha < t_n < t + \alpha$  以及  $\|x\| - \alpha < \|x_n\| < \|x\| + \alpha$ 。注意  $v_1(t, x)$  中的  $W_x(t)$  是如下最简随机微分方程的解:

$$\begin{aligned} dX(t) &= dW(t), \\ X(0) &= x. \end{aligned} \quad (3.52)$$

由文献 [97] 中的 *Proposition 2.1* 可知,  $W_x(t) \equiv W(t, x)$  在  $(0, \infty) \times \mathbb{R}^N$  上几乎处处 (a.s.) 连续。为了证明连续性, 我们定义如下随机变量:

$$Y_n := e^{-2 \int_0^t c(2W_x(r)) dr} u_0(2W_x(t)) - e^{-2 \int_0^{t_n} c(2W_{x_n}(r)) dr} u_0(2W_{x_n}(t_n)). \quad (3.53)$$

这些随机变量是一致可积 (uniformly integrable) 的, 因为:

$$\begin{aligned}
\int_{\Omega} Y_n^2 d\mathbb{P} &\leq 2 \int_{\Omega} \left( e^{-2 \int_0^t c(2W_x(r)) dr} u_0(2W_x(t)) \right)^2 d\mathbb{P} \\
&\quad + 2 \int_{\Omega} \left( e^{-2 \int_0^{t_n} c(2W_{x_n}(r)) dr} u_0(2W_{x_n}(t_n)) \right)^2 d\mathbb{P} \\
&\leq 4 \sup_{x \in \mathbb{R}^N} |u_0(x)| \int_{\Omega} d\mathbb{P} \\
&= 4 \sup_{x \in \mathbb{R}^N} |u_0(x)| < \infty.
\end{aligned} \tag{3.54}$$

其中上述公式中的  $\Omega$  表示维纳空间, 即  $C([0, \infty), \mathbb{R}^N)$ . 根据文献 [98] 第 215 页的 *Theorem 4.2*,  $\{Y_n\}_{n \geq N_1}$  是一致可积的. 令  $0 < \eta < 1$  且  $M > 0$ . 定义  $M_1 := 1 + M$ ,  $A_1 := [0, t + \alpha] \times [-M_1, M_1]^N$ ,  $A_2 := [-M_1, M_1]^N$ , 并对函数  $\nu : [0, \infty) \rightarrow \mathbb{R}^N$ ,  $T > 0$  定义  $\|\nu\|_T := \sup_{0 \leq s \leq T} \|\nu(s)\|$ . 由于  $\{Y_n\}_{n \geq N_1}$  的一致可积性, 对于任意  $\epsilon > 0$ , 存在  $\delta(\epsilon) > 0$ , 使得对于任意满足  $\mathbb{P}[B] < \delta(\epsilon)$  的可测集  $B$ , 都有  $\sup_n \int_B |Y_n| d\mathbb{P} < \frac{\epsilon}{2}$ . 利用  $W(t, x)$  和  $c(x)$  的连续性以及几乎处处不爆破 (not exploding in finite time) 的性质, 我们可以选取  $M > 0$  使得

$$\mathbb{P}[\|W_x\|_{t+\alpha} > M] \leq \frac{\delta(\epsilon)}{2} \tag{3.55}$$

并选取  $N_2 \in \mathbb{N}$  使得

$$\mathbb{P}[\|W_{x_n} - W_x\|_{t+\alpha} > \eta] \leq \frac{\delta(\epsilon)}{2}, \quad \mathbb{P}[\|c(W_{x_n}) - c(W_x)\|_{t+\alpha} > \eta] \leq \frac{\delta(\epsilon)}{2}. \tag{3.56}$$

令  $B_1 := \{\|W_x\|_{t+\alpha} \leq M\}$ ,  $B_2 := \{\|W_{x_n} - W_x\|_{t+\alpha} \vee \|c(W_{x_n}) - c(W_x)\|_{t+\alpha} \leq \eta\}$ , 其中  $A \vee B := \max\{A, B\}$ ,  $A \wedge B := \min\{A, B\}$ . 于是有:

$$\begin{aligned}
|v_1(t_n, x_n) - v_1(t, x)| &\leq \int_{B_1 \cap B_2} |Y_n| d\mathbb{P} + \int_{\Omega \setminus (B_1 \cap B_2)} |Y_n| d\mathbb{P} \\
&\leq \int_{B_1 \cap B_2} |Y_n| d\mathbb{P} + \frac{\epsilon}{2} \\
&= \int_{B_1 \cap B_2} \left| e^{-2 \int_0^t c(2W_x(r)) dr} u_0(2W_x(t)) - e^{-2 \int_0^{t_n} c(2W_{x_n}(r)) dr} u_0(2W_{x_n}(t_n)) \right| d\mathbb{P} \\
&\quad + \frac{\epsilon}{2} \\
&\leq \int_{B_1 \cap B_2} e^{-2 \int_0^t c(2W_x(r)) dr} |u_0(2W_x(t)) - u_0(2W_{x_n}(t_n))| d\mathbb{P} \\
&\quad + \int_{B_1 \cap B_2} |u_0(2W_{x_n}(t_n))| \left| e^{-2 \int_0^t c(2W_x(r)) dr} - e^{-2 \int_0^{t_n} c(2W_{x_n}(r)) dr} \right| d\mathbb{P} + \frac{\epsilon}{2} \\
&:= I_n + J_n + \frac{\epsilon}{2}.
\end{aligned} \tag{3.57}$$

由勒贝格控制收敛定理 (LDCT) 及  $c(x)$  的非负性, 我们可以找到  $N_3 \in \mathbb{N}$  使得对于任意  $n \geq N_3$ , 都有  $I_n < \frac{\epsilon}{4}$ 。接着我们估计  $J_n$ :

$$J_n \leq \sup_{x \in \mathbb{R}^N} |u_0(x)| \int_{B_1 \cap B_2} \left| e^{2 \int_0^{t_n} c(2W_{x_n}(r)) dr} - 2 \int_0^t c(2W_x(r)) dr - 1 \right| d\mathbb{P}. \quad (3.58)$$

在  $B_1 \cap B_2$  中,

$$\begin{aligned} q_n &:= \int_0^{t_n} c(2W_{x_n}(r)) dr - \int_0^t c(2W_x(r)) dr \\ &= \int_0^{t_n} c(2W_{x_n}(r)) dr - \int_0^t c(2W_{x_n}(r)) dr + \int_0^t c(2W_{x_n}(r)) dr - \int_0^t c(2W_x(r)) dr \\ &\leq \int_{t_n \vee t}^{t_n \wedge t} c(2W_{x_n}(r)) dr + \int_0^t c(2W_{x_n}(r)) dr - \int_0^t c(2W_x(r)) dr \\ &\leq 2\alpha \sup_{x \in A_2} c(x) + \int_0^t (c(2W_{x_n}(r)) - c(2W_x(r))) dr. \end{aligned} \quad (3.59)$$

从而  $q_n \xrightarrow{n \rightarrow \infty} 0$ , 且由 LDCT 可知存在  $N_4 \in \mathbb{N}$  使得  $J_n < \frac{\epsilon}{4}$ 。因此我们证明了对于任意  $n \geq \max_{i=1,2,3,4} \{N_i\}$ , 都有  $|v_1(t_n, x_n) - v_1(t, x)| < \epsilon$ 。这意味着  $v_1(t, x)$  和  $v(t, x)$  在  $(0, \infty) \times \mathbb{R}^N$  上是连续的。对于  $t = 0$  处的连续性, 可按相同方式证明。 ■

**引理 3.9** (Feynman-Kac 解的可微性) 令  $v$  定义如方程 (3.48) 所示, 则  $v \in C^{1,\infty}((0, \infty) \times \mathbb{R}^N)$ 。

**证明** 设  $T > 0$ 。考虑在具有  $C^2$  边界的有界开区域  $A \subset \mathbb{R}^N$  上的如下抛物型微分方程:

$$\begin{cases} -\partial_t u + \Delta u - c(x)u = 0, & (t, x) \in [0, T] \times A \\ u(0, x) = v(0, x), & x \in A \\ u(t, x) = v(t, x), & (t, x) \in (0, T] \times A \end{cases} \quad (3.60)$$

由  $v$  的连续性以及文献 [96] 的结果可知, 方程 (3.60) 存在唯一的经典解。定义如下停时 (stopping time):

$$\tau := \inf\{s > 0 \mid W_x(s) \notin \bar{A}\}. \quad (3.61)$$

遵循文献 [97] 中 Theorem 2.3 以及文献 [99] 第 6 章第 5 节的论证, 我们可以证明 (3.60) 的经典解具有如下表示:

$$w(t, x) = \mathbb{E}_{x/\sqrt{2}} \left[ v(t - \tau, \sqrt{2}W(\tau)) e^{-\int_0^\tau c(\sqrt{2}W(r)) dr} \right]. \quad (3.62)$$

利用过程  $W_x(s)$  的强马尔可夫性 (strong Markov property), 我们来证明  $v$  与  $w$  的相等性。考虑滤子  $\mathcal{F}_\tau$ 。对于  $v(t, x)$ , 利用条件期望的性质可得:

$$\begin{aligned}
v(t, \sqrt{2}x) &= \mathbb{E}_x \left[ u_0(\sqrt{2}W(t)) e^{-\int_0^t c(\sqrt{2}W(r)) dr} \right] \\
&= \mathbb{E}_x \left[ \mathbb{E} \left[ u_0(\sqrt{2}W(t)) e^{-\int_0^t c(\sqrt{2}W(r)) dr} \middle| \mathcal{F}_\tau \right] \right] \\
&= \mathbb{E}_x \left[ e^{-\int_0^\tau c(\sqrt{2}W(r)) dr} \mathbb{E} \left[ e^{-\int_0^{t-\tau} c(\sqrt{2}W(r+\tau)) dr} u_0(\sqrt{2}W(t-\tau+\tau)) \middle| \mathcal{F}_\tau \right] \right] \\
&= \mathbb{E}_x \left[ e^{-\int_0^\tau c(\sqrt{2}W(r)) dr} \mathbb{E}_{\sqrt{2}W(\tau)} \left[ e^{-\int_0^{t-\tau} c(\sqrt{2}W(r)) dr} u_0(\sqrt{2}W(t-\tau)) \right] \right] \\
&= \mathbb{E}_x \left[ e^{-\int_0^\tau c(\sqrt{2}W(r)) dr} v(t-\tau, \sqrt{2}W(\tau)) \right] \\
&= w(t, \sqrt{2}x).
\end{aligned} \tag{3.63}$$

因此  $v(t, x) \in C^{1,2}([0, T] \times A)$  且满足抛物方程  $\partial_t v = \Delta v - c(x)v$ 。由于  $T > 0$  和集合  $A$  的任意性, 可得  $v(t, x) \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^N)$ 。进一步地, 由热方程的表达式我们有:

$$u(x, t) = \frac{1}{(2\sqrt{\pi t})^N} \int_{\mathbb{R}^N} u_0(y) e^{-\frac{\|y-x\|^2}{4t}} dy - \frac{1}{(2\sqrt{\pi})^N} \int_0^t \int_{\mathbb{R}^N} \frac{c(y)u(y, \tau)}{(\sqrt{t-\tau})^N} e^{-\frac{\|y-x\|^2}{4(t-\tau)}} dy d\tau \tag{3.64}$$

这意味着  $u \in C^{1,\infty}(\mathbb{R}_+ \times \mathbb{R}^N)$ 。 ■

结合引理 3.7、引理 3.8 和引理 3.9, 我们可以证明 Fokker-Planck 流 (3.3) 和柯西问题 (3.4) 解的存在性与唯一性。

### 3.5.2 引理 3.2 的证明

根据梯度流和最优传输理论, 我们可以将 FP 流解释为相对熵  $H_\nu$  的梯度流  $\dot{\mu} = -\text{grad}_\mu H_\nu$ , 其中

$$H_\nu(\mu) = \int_{\mathbb{R}^N} \mu \log \frac{\mu}{\nu} dx \tag{3.65}$$

参考测度为  $\nu = \mu_\infty = e^{-V}$ 。于是我们有:

$$\begin{aligned}
\frac{d}{dt} H_\nu(\mu_t) &= \left\langle \text{grad}_\mu H_\nu, \frac{\partial \mu_t}{\partial t} \right\rangle \\
&= - \left\| \text{grad}_\mu H_\nu \right\|^2 \\
&= - \int_{\mathbb{R}^N} \mu_t \left\| \nabla \log \frac{\mu_t}{\nu} \right\|^2 dx \\
&= -J_\nu(\mu_t).
\end{aligned} \tag{3.66}$$

类似地，相对熵沿梯度流的二阶导数可由 Hessian 算子给出：

$$\begin{aligned}\frac{d^2}{dt^2}H_\nu(\mu_t) &= -\frac{d}{dt}\|\text{grad}_\mu H_\nu\|^2 \\ &= 2(\text{Hess}_\mu H_\nu)(\text{grad}_\mu H_\nu).\end{aligned}\quad (3.67)$$

根据文献 [17] 中的 *Equation (15.7)* 或 *Theorem 24.2*，该二阶导数可表示为如下等式：

$$\frac{d^2}{dt^2}H_\nu(\mu_t) = 2K_\nu(\mu_t) + 2G_\nu(\mu_t)\quad (3.68)$$

其中

$$\begin{aligned}K_\nu(\mu) &= \int_{\mathbb{R}^N} \mu \left\| \nabla^2 \log \frac{\mu}{\nu} \right\|_{\text{HS}}^2 dx \\ G_\nu(\mu) &= \int_{\mathbb{R}^N} \mu \left\langle (\nabla^2 V) \nabla \log \frac{\mu}{\nu}, \nabla \log \frac{\mu}{\nu} \right\rangle dx.\end{aligned}\quad (3.69)$$

特别地，若  $\mu_0$  是位于  $x_0 \in \mathbb{R}^N$  的点质量 (point mass)，我们有：

$$\begin{aligned}\frac{d}{dt}H_\nu(\mu_t|\mu_0 = x_0) &= -J_\nu(\mu_t|\mu_0 = x_0) \\ \frac{d^2}{dt^2}H_\nu(\mu_t|\mu_0 = x_0) &= 2K_\nu(\mu_t|\mu_0 = x_0) + 2G_\nu(\mu_t|\mu_0 = x_0).\end{aligned}\quad (3.70)$$

因此，对于任意初始分布  $\mu_0$ ，对  $\mu_0$  积分即可得到：

$$\begin{aligned}\frac{d}{dt}H_\nu(\mu_t|\mu_0) &= -J_\nu(\mu_t|\mu_0) \\ \frac{d^2}{dt^2}H_\nu(\mu_t|\mu_0) &= 2K_\nu(\mu_t|\mu_0) + 2G_\nu(\mu_t|\mu_0).\end{aligned}\quad (3.71)$$

### 3.5.3 定理 3.2 的证明

定理 3.2 的证明基于引理 3.10 和 3.11。

**引理 3.10** 对于任意联合分布  $(\mu_0, \mu_t)$  和任意参考概率测度  $\nu$ ，成立：

$$K_\nu(\mu_0; \mu_t) = \Psi(\mu_0|\mu_t) + 2 \int_{\mathbb{R}_t^N} \mu_t \left\langle \nabla^2 \log \frac{\mu_t}{\nu}, \int_{\mathbb{R}_0^N} \mu_{0|t} \nabla^2 \log \mu_{0|t} dx_0 \right\rangle_{\text{HS}} dx_t \quad (3.72)$$

**证明** 我们有如下分解：

$$\nabla^2 \log \frac{\mu_{t|0}}{\nu} = \nabla^2 \log \frac{\mu_t}{\nu} + \nabla^2 \log \mu_{0|t}.\quad (3.73)$$

于是可得:

$$\begin{aligned}
K_\nu(\mu_0; \mu_t) &= K_\nu(\mu_t | \mu_0) - K_\nu(\mu_t) \\
&= \int_{\mathbb{R}_0^N} \int_{\mathbb{R}_t^N} \mu_{t,0} \left( \left\| \nabla^2 \log \frac{\mu_{t,0}}{\nu} \right\|_{\text{HS}}^2 - \left\| \nabla^2 \log \frac{\mu_t}{\nu} \right\|_{\text{HS}}^2 \right) dx_t dx_0 \\
&= \int_{\mathbb{R}_0^N} \int_{\mathbb{R}_t^N} \mu_{t,0} \left\| \nabla^2 \log \mu_{0|t} \right\|_{\text{HS}}^2 dx_t dx_0 \\
&\quad + 2 \int_{\mathbb{R}_0^N} \int_{\mathbb{R}_t^N} \mu_{t,0} \left\langle \nabla^2 \log \frac{\mu_t}{\nu}, \nabla^2 \log \mu_{0|t} \right\rangle_{\text{HS}} dx_t dx_0 \\
&= \Psi(\mu_0 | \mu_t) + 2 \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \nabla^2 \log \frac{\mu_t}{\nu}, \nabla^2 \log \mu_{0|t} \right\rangle_{\text{HS}} dx_0 dx_t \\
&= \Psi(\mu_0 | \mu_t) + 2 \int_{\mathbb{R}_t^N} \mu_t \left\langle \nabla^2 \log \frac{\mu_t}{\nu}, \int_{\mathbb{R}_0^N} \mu_{0|t} \nabla^2 \log \mu_{0|t} dx_0 \right\rangle_{\text{HS}} dx_t
\end{aligned} \tag{3.74}$$

其中

$$\Psi(\mu_0 | \mu_t) := \int_{\mathbb{R}_0^N} \int_{\mathbb{R}_t^N} \mu_{t,0} \left\| \nabla^2 \log \mu_{0|t} \right\|_{\text{HS}}^2 dx_t dx_0.$$

**引理 3.11** 对于任意不依赖于  $x_0$  的光滑函数  $\rho_t$ , 任意联合分布  $(\mu_0, \mu_t)$  和任意参考概率测度  $\nu$ , 我们有如下分部积分公式:

$$\int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \rho_t, \nabla \log \mu_{0|t} (\nabla \log \mu_{0|t})^\top \right\rangle_{\text{HS}} dx_0 dx_t = - \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \rho_t, \nabla^2 \log \mu_{0|t} \right\rangle_{\text{HS}} dx_0 dx_t.$$

**证明**

$$\begin{aligned}
&\int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \rho_t, \nabla \log \mu_{0|t} (\nabla \log \mu_{0|t})^\top \right\rangle_{\text{HS}} dx_0 dx_t \\
&= \int_{\mathbb{R}_t^N} \mu_t \left\langle \rho_t, \int_{\mathbb{R}_0^N} \mu_{0|t} \nabla \log \mu_{0|t} (\nabla \log \mu_{0|t})^\top dx_0 \right\rangle_{\text{HS}} dx_t \\
&= \int_{\mathbb{R}_t^N} \mu_t \left\langle \rho_t, \int_{\mathbb{R}_0^N} \mu_{0|t} \left( \frac{\nabla^2 \mu_{0|t}}{\mu_{0|t}} - \nabla^2 \log \mu_{0|t} \right) dx_0 \right\rangle_{\text{HS}} dx_t \\
&= \int_{\mathbb{R}_t^N} \mu_t \left\langle \rho_t, \nabla^2 \left( \int_{\mathbb{R}_0^N} \mu_{0|t} dx_0 \right) \right\rangle_{\text{HS}} dx_t - \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \rho_t, \nabla^2 \log \mu_{0|t} \right\rangle_{\text{HS}} dx_0 dx_t \\
&= \int_{\mathbb{R}_t^N} \mu_t \left\langle \rho_t, \nabla^2(1) \right\rangle_{\text{HS}} dx_t - \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \rho_t, \nabla^2 \log \mu_{0|t} \right\rangle_{\text{HS}} dx_0 dx_t \\
&= 0 - \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \rho_t, \nabla^2 \log \mu_{0|t} \right\rangle_{\text{HS}} dx_0 dx_t \\
&= - \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \rho_t, \nabla^2 \log \mu_{0|t} \right\rangle_{\text{HS}} dx_0 dx_t.
\end{aligned}$$

结合引理 3.2、3.10 以及 3.11, 我们可以得到互信息关于  $t$  的一阶和二阶导数:

$$\begin{aligned}\frac{d}{dt}I(\mu_0; \mu_t) &= \frac{d}{dt}H_\nu(\mu_{t|0}) - \frac{d}{dt}H_\nu(\mu_t) \\ &= J_\nu(\mu_{t|0}) - J_\nu(\mu_t) \\ &= -J_\nu(\mu_0; \mu_t) \leq 0\end{aligned}$$

$$\begin{aligned}\frac{d^2}{dt^2}I(\mu_0; \mu_t) &= \frac{d^2}{dt^2}H_\nu(\mu_t|\mu_0) - \frac{d^2}{dt^2}H_\nu(\mu_t) \\ &= 2\Psi(\mu_0|\mu_t) + 4 \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle -\nabla^2 \log \frac{\mu_t}{\nu}, -\nabla^2 \log \mu_{0|t} \right\rangle_{\text{HS}} dx_t dx_0 \\ &\quad + 2 \int_{\mathbb{R}_0^N} \mu_0 dx_0 \int_{\mathbb{R}_t^N} \mu_{t|0} \left\langle \nabla^2 V, \nabla \log \frac{\mu_{t|0}}{\nu} \left( \nabla \log \frac{\mu_{t|0}}{\nu} \right)^\top \right\rangle_{\text{HS}} dx_t \\ &\quad - 2 \int_{\mathbb{R}^N} \mu_t \left\langle \nabla^2 V, \nabla \log \frac{\mu_t}{\nu} \left( \nabla \log \frac{\mu_t}{\nu} \right)^\top \right\rangle_{\text{HS}} dx_t \\ &= 2\Psi(\mu_0|\mu_t) + 4 \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle -\nabla^2 \log \frac{\mu_t}{\nu}, \nabla \log \mu_{0|t} \left( \nabla \log \mu_{0|t} \right)^\top \right\rangle_{\text{HS}} dx_t dx_0 \\ &\quad + 2 \int_{\mathbb{R}_0^N} \mu_0 dx_0 \int_{\mathbb{R}_t^N} \mu_{t|0} \left\langle \nabla^2 V, \nabla \log \frac{\mu_{t|0}}{\nu} \left( \nabla \log \frac{\mu_{t|0}}{\nu} \right)^\top \right\rangle_{\text{HS}} dx_t \\ &\quad - 2 \int_{\mathbb{R}^N} \mu_t \left\langle \nabla^2 V, \nabla \log \frac{\mu_t}{\nu} \left( \nabla \log \frac{\mu_t}{\nu} \right)^\top \right\rangle_{\text{HS}} dx_t \\ &= 2\Psi(\mu_0|\mu_t) + 4 \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle -\nabla^2 \log \mu_t, \nabla \log \mu_{0|t} \left( \nabla \log \mu_{0|t} \right)^\top \right\rangle_{\text{HS}} dx_t dx_0 \\ &\quad + 2 \int_{\mathbb{R}_0^N} \mu_0 dx_0 \int_{\mathbb{R}_t^N} \mu_{t|0} \left\langle \nabla^2 V, \nabla \log \frac{\mu_{t|0}}{\nu} \left( \nabla \log \frac{\mu_{t|0}}{\nu} \right)^\top \right\rangle_{\text{HS}} dx_t \\ &\quad - 2 \int_{\mathbb{R}^N} \mu_t \left\langle \nabla^2 V, \nabla \log \frac{\mu_t}{\nu} \left( \nabla \log \frac{\mu_t}{\nu} \right)^\top \right\rangle_{\text{HS}} dx_t \\ &\quad - 4 \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \nabla^2 V, \nabla \log \mu_{0|t} \left( \nabla \log \mu_{0|t} \right)^\top \right\rangle_{\text{HS}} dx_t dx_0 \\ &= 2\Psi(\mu_0|\mu_t) + 4 \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \nabla^2 \log \mu_t, \nabla^2 \log \mu_{0|t} \right\rangle_{\text{HS}} dx_t dx_0 \\ &\quad + 2 \int_{\mathbb{R}_0^N} \mu_0 dx_0 \int_{\mathbb{R}_t^N} \mu_{t|0} \left\langle \nabla^2 V, \nabla \log \mu_{t|0} \left( \nabla \log \mu_{t|0} \right)^\top + \nabla \log \nu \left( \nabla \log \nu \right)^\top \right. \\ &\quad \quad \left. - 2 \nabla \log \mu_{t|0} \left( \nabla \log \nu \right)^\top \right\rangle_{\text{HS}} dx_t \\ &\quad - 2 \int_{\mathbb{R}^N} \mu_t \left\langle \nabla^2 V, \nabla \log \mu_t \left( \nabla \log \mu_t \right)^\top + \nabla \log \nu \left( \nabla \log \nu \right)^\top \right. \\ &\quad \quad \left. - 2 \nabla \log \mu_t \left( \nabla \log \nu \right)^\top \right\rangle_{\text{HS}} dx_t \\ &\quad - 4 \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \left\langle \nabla^2 V, \nabla \log \mu_{0|t} \left( \nabla \log \mu_{0|t} \right)^\top \right\rangle_{\text{HS}} dx_t dx_0\end{aligned}$$

$$\begin{aligned}
&= 2\Psi(\mu_0|\mu_t) + 4 \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \langle \nabla^2 \log \mu_t, \nabla^2 \log \mu_{0|t} \rangle_{\text{HS}} dx_t dx_0 \\
&\quad - 2 \int_{\mathbb{R}_t^N} \mu_0 \int_{\mathbb{R}_0^N} \mu_{t|0} \langle \nabla^2 V, \nabla^2 \log \mu_{t|0} \rangle_{\text{HS}} dx_t dx_0 \\
&\quad + 2 \int_{\mathbb{R}_t^N} \mu_0 \int_{\mathbb{R}_0^N} \mu_{t|0} \langle \nabla^2 V, \nabla V (\nabla V)^\top \rangle_{\text{HS}} dx_t dx_0 \\
&\quad + 4 \int_{\mathbb{R}_t^N} \mu_0 \int_{\mathbb{R}_0^N} \mu_{t|0} \langle \nabla^2 V, \nabla \log \mu_{t|0} (\nabla V)^\top \rangle_{\text{HS}} dx_t dx_0 \\
&\quad + 2 \int_{\mathbb{R}_t^N} \mu_t \langle \nabla^2 V, \nabla^2 \log \mu_t \rangle_{\text{HS}} dx_t \\
&\quad - 2 \int_{\mathbb{R}_t^N} \mu_t \langle \nabla^2 V, \nabla V (\nabla V)^\top \rangle_{\text{HS}} dx_t - 4 \int_{\mathbb{R}_t^N} \mu_t \langle \nabla^2 V, \nabla \log \mu_t (\nabla V)^\top \rangle_{\text{HS}} dx_t \\
&\quad + 4 \int_{\mathbb{R}_t^N} \mu_t \int_{\mathbb{R}_0^N} \mu_{0|t} \langle \nabla^2 V, \nabla^2 \log \mu_{0|t} \rangle_{\text{HS}} dx_t dx_0.
\end{aligned}$$

进一步简化该公式，我们得到：

$$\begin{aligned}
\frac{d^2}{dt^2} I(\mu_0; \mu_t) &= 2\Psi(\mu_0|\mu_t) + 4 \int_{\mathbb{R}_0^N \times \mathbb{R}_t^N} \mu_{0,t} \langle \nabla^2 \log \mu_t, \nabla^2 \log \mu_{0|t} \rangle_{\text{HS}} dx_0 dx_t \\
&\quad + 4 \int_{\mathbb{R}_0^N \times \mathbb{R}_t^N} \mu_{0,t} \langle \nabla^2 V, \nabla^2 \log \mu_{0|t} (\nabla V)^\top \rangle_{\text{HS}} dx_0 dx_t \\
&\quad - 2 \int_{\mathbb{R}_0^N \times \mathbb{R}_t^N} \mu_{0,t} \langle \nabla^2 V, \nabla^2 \log \mu_{0|t} \rangle_{\text{HS}} dx_0 dx_t \\
&\quad + 4 \int_{\mathbb{R}_0^N \times \mathbb{R}_t^N} \mu_{0,t} \langle \nabla^2 V, \nabla^2 \log \mu_{0|t} \rangle_{\text{HS}} dx_0 dx_t \\
&= 2\Psi(\mu_0|\mu_t) + 2 \int_{\mathbb{R}_0^N \times \mathbb{R}_t^N} \mu_{0,t} \langle 2\nabla^2 \log \mu_t + \nabla^2 V, \nabla^2 \log \mu_{0|t} \rangle_{\text{HS}} dx_0 dx_t \\
&\quad + 4 \int_{\mathbb{R}_t^N} \mu_t \left\langle \nabla^2 V, (\nabla V)^\top \int_{\mathbb{R}_0^N} \mu_{0|t} \nabla \log \mu_{0|t} dx_0 \right\rangle_{\text{HS}} dx_t
\end{aligned}$$

其中

$$\begin{aligned}
&\int_{\mathbb{R}_t^N} \mu_t \left\langle \nabla^2 V, (\nabla V)^\top \int_{\mathbb{R}_0^N} \mu_{0|t} \nabla \log \mu_{0|t} dx_0 \right\rangle_{\text{HS}} dx_t \\
&\quad = \int_{\mathbb{R}_t^N} \mu_t \left\langle \nabla^2 V, (\nabla V)^\top \int_{\mathbb{R}_0^N} \nabla \mu_{0|t} dx_0 \right\rangle_{\text{HS}} dx_t \\
&\quad = \int_{\mathbb{R}_t^N} \mu_t \left\langle \nabla^2 V, (\nabla V)^\top \nabla \int_{\mathbb{R}_0^N} \mu_{0|t} dx_0 \right\rangle_{\text{HS}} dx_t \\
&\quad = 0.
\end{aligned} \tag{3.75}$$

最终，我们得到互信息关于  $t$  的二阶导数：

$$\begin{aligned} \frac{d^2}{dt^2} I(\mu_0; \mu_t) &= 2\Psi(\mu_0|\mu_t) + 2 \int_{\mathbb{R}_0^N \times \mathbb{R}_t^N} \mu_{0,t} \langle 2\nabla^2 \log \mu_t + \nabla^2 V, \nabla^2 \log \mu_{0|t} \rangle_{\text{HS}} dx_0 dx_t \\ &= 2\Psi(\mu_0|\mu_t) + 4 \int_{\mathbb{R}_0^N \times \mathbb{R}_t^N} \mu_{0,t} \langle \nabla^2 \log \gamma_t, \nabla^2 \log \mu_{0|t} \rangle_{\text{HS}} dx_0 dx_t \\ &= 2\Psi(\mu_0|\mu_t) - 4 \int_{\mathbb{R}_0^N \times \mathbb{R}_t^N} \mu_{0,t} (\nabla \log \mu_{0|t})^\top (\nabla^2 \log \gamma_t) (\nabla \log \mu_{0|t}) dx_0 dx_t. \end{aligned} \quad (3.76)$$

这里  $\gamma_t := e^{\frac{V}{2}} \mu_t$ 。因此，我们需要证明  $\gamma_t$  对于每个  $t \geq 0$  都是对数凹 (log-concave) 的。由此即可推出互信息  $I(\mu_0; \mu_t)$  关于  $t$  是凸的。

### 3.5.4 引理 3.3 的证明

由于  $\gamma_t = e^{\frac{V}{2}} \mu_t$  满足如下薛定谔方程：

$$\frac{\partial \gamma_t}{\partial t} + \left( \frac{1}{4} \|\nabla V\|^2 - \frac{1}{2} \Delta V \right) \gamma_t - \Delta \gamma_t = 0 \quad (3.77)$$

其中  $c(x) := \frac{1}{4} \|\nabla V\|^2 - \frac{1}{2} \Delta V \geq 0$ 。由推论 3.1 可知，如果  $\gamma_0$  在  $\Omega$  上是对数凹的且边界值为零，即  $X_0 \sim \mu_0$  在  $\Omega$  上是  $\frac{V}{2}$ -相对对数凹的，那么对于任意  $t > 0$ ， $\gamma_t$  在  $\Omega$  上始终保持对数凹性。因此，互信息关于  $t$  是凸的。

### 3.5.5 引理 3.4 的证明

引理 3.4 的证明基于引理 3.12 到 3.14。

**引理 3.12** 令  $u$  为如下方程的有界非负解：

$$\begin{cases} \partial_t u = \Delta u, & \text{在 } (0, \infty) \times \mathbb{R}^N \\ u(0, x) = u_0(x), & \text{在 } \mathbb{R}^N \end{cases} \quad (3.78)$$

其中  $u_0$  是  $\mathbb{R}^N$  上的有界非负函数。若  $u_0$  是对数凹的，则对于任意  $t > 0$ ， $u(t, \cdot)$  在  $\mathbb{R}^N$  上也是对数凹的。

**证明** 该引理直接作为对数凹性在卷积下保持性质的推论得出，因为方程 (3.78) 的解  $u(t, x)$  可以表示为热核与  $u_0$  的卷积，而热核本身是对数凹的。 ■

**引理 3.13** (文献 [100] 中的 *Theorem 4.1*) 令  $\eta$  为如下调和方程的解：

$$\begin{cases} -\Delta \eta = 1, & \text{在 } \Omega \\ \eta > 0, & \text{在 } \Omega \\ \eta = 0, & \text{在 } \partial\Omega \end{cases} \quad (3.79)$$

则  $\eta$  在  $\Omega$  上是  $1/2$ -凹的，这意味着  $\log \eta$  在  $\Omega$  上是凹的，即  $\eta$  在  $\Omega$  上是对数凹的。

此处我们称  $\mathbb{R}^N$  上的非负函数  $u$  是  $p$ -凹的, 如果对于  $s \in [0, 1]$  和  $x, y \in \mathbb{R}^N$ , 满足

$$u((1-s)x + sy) \geq M_p(u(x), u(y); s)$$

其中

$$M_p(a, b; \lambda) = \begin{cases} \max\{a, b\}, & p = +\infty \\ [(1-\lambda)a^p + \lambda b^p]^{1/p}, & \text{若 } p \neq -\infty, 0, +\infty \\ a^{1-\lambda}b^\lambda, & p = 0 \\ \min\{a, b\}, & p = -\infty \end{cases} \quad (3.80)$$

是  $a$  和  $b$  的 ( $\lambda$ -加权)  $p$ -均值。Jensen 不等式的一个简单推论是: 若  $p \leq q$ , 则

$$M_p(a, b; \lambda) \leq M_q(a, b; \lambda).$$

因此  $q$ -凹性可推导出  $p$ -凹性 (对于任意  $q \geq p$ ), 且我们注意到 0-凹性等价于对数凹性。

**引理 3.14** 设  $\Omega$  为有界光滑凸区域,  $u_0$  为  $\Omega$  上的有界非负对数凹函数。则存在一列对数凹函数  $\{u_{0,n}\}_{n \geq 1}$  几乎处处收敛到  $u_0$ , 且对于任意  $n$ ,  $u_{0,n}$  在  $\bar{\Omega}$  上非负且连续, 并具有零边界值, 即  $u_{0,n}|_{\partial\Omega} = 0$ 。

**证明** 考虑如下热方程:

$$\begin{cases} \partial_t v = \Delta v, & \text{在 } (0, \infty) \times \mathbb{R}^N \\ v(0, x) = v_0(x), & \text{在 } \mathbb{R}^N \end{cases} \quad (3.81)$$

其中

$$v_0(x) = \begin{cases} u_0(x), & \text{当 } x \in \Omega \\ 0, & \text{当 } x \notin \Omega \end{cases} \quad (3.82)$$

则对于  $x \in \mathbb{R}^N$  和  $t > 0$ , 解为  $v(t, x) = [e^{t\Delta} v_0](x)$ 。此处  $e^{t\Delta}$  为热半群 (heat semigroup), 且我们定义  $e^{-\infty} := 0$ 。显然  $v_0$  在  $\mathbb{R}^N$  上是对数凹的。由引理 3.12 可知, 对于任意  $t > 0$ ,  $v(t, \cdot)$  也是对数凹的。进一步, 由极大值原理 (maximum principle), 结合  $v_0 \in L^1(\mathbb{R}^N) \cap L^\infty(\mathbb{R}^N)$ , 可推知:

$$\text{对于任意 } t > 0, v(t, \cdot) \text{ 是 } \mathbb{R}^N \text{ 上的正连续函数,} \quad (3.83)$$

$$\text{对于任意 } t > 0, \|v(t, \cdot)\|_{L^\infty(\mathbb{R}^N)} < \|v_0\|_{L^\infty(\mathbb{R}^N)}, \quad (3.84)$$

$$\lim_{t \rightarrow 0^+} \|v(t, \cdot) - v_0\|_{L^1(\mathbb{R}^N)} = 0. \quad (3.85)$$

由 (3.85), 我们可以找到一个序列  $\{t_n\} \subset (0, \infty)$  满足  $\lim_{n \rightarrow \infty} t_n = 0$ , 使得对于几乎所有的  $x \in \mathbb{R}^N$ , 有

$$\lim_{n \rightarrow \infty} v(t_n, x) = v_0(x). \quad (3.86)$$

令  $\eta$  为如下调和方程的解:

$$\begin{cases} -\Delta \eta = 1, & \text{在 } \Omega \\ \eta > 0, & \text{在 } \Omega \\ \eta = 0, & \text{在 } \partial\Omega \end{cases}. \quad (3.87)$$

由引理 3.13 可知,  $\eta$  是对数凹的, 且当  $\text{dist}(x, \partial\Omega) \rightarrow 0$  时,  $\log \eta \rightarrow -\infty$ . 由 (3.83) 和 (3.84), 我们可以找到一个序列  $\{m_n\} \subset (1, +\infty)$  满足  $\lim_{n \rightarrow \infty} m_n = +\infty$ , 使得只要我们取

$$m_n \geq \left| \frac{\log \eta(x)}{\text{ess sup}_{x \in \Omega} \{\log v_0(x) - \log v(t_n, x)\}} \right| + 1 \quad (3.88)$$

函数

$$V_n(x) := \log v(t_n, x) + m_n^{-1} \log \eta(x) \quad (3.89)$$

在  $\Omega$  内就是连续且凹的, 并且

$$\sup_{x \in \Omega} V_n(x) \leq \text{ess sup}_{x \in \Omega} \log v_0. \quad (3.90)$$

此外, 由 (3.86) 我们有

$$\lim_{n \rightarrow \infty} V_n(x) = \log v_0(x) = \log u_0(x) \text{ 对几乎所有的 } x \in \Omega \text{ 成立,} \quad (3.91)$$

$$V_n(x) \rightarrow -\infty \text{ 当 } \text{dist}(x, \partial\Omega) \rightarrow 0 \text{ 时.} \quad (3.92)$$

那么函数  $u_{0,n}(x) := e^{V_n(x)}$  满足:

$$\lim_{n \rightarrow \infty} u_{0,n}(x) = u_0(x) \text{ 对几乎所有的 } x \in \Omega \text{ 成立,} \quad (3.93)$$

$$u_{0,n}(x) \text{ 在 } \Omega \text{ 上对数凹且在 } \bar{\Omega} \text{ 上连续 (对任意 } n), \quad (3.94)$$

$$u_{0,n} \geq 0 \text{ 在 } \Omega \text{ 内, 且 } u_{0,n} = 0 \text{ 在 } \partial\Omega \text{ 上.} \quad (3.95)$$

■

基于上述引理, 我们接着证明引理 3.4. 考虑薛定谔方程:

$$\begin{cases} \partial_t \gamma + c(x)\gamma - \Delta \gamma = 0 & \text{在 } (0, \infty) \times \Omega \\ \gamma(0, x) = \gamma_0(x) & \text{在 } \bar{\Omega} \end{cases} \quad (3.96)$$

此时我们不假设  $\gamma_0 = 0$  在  $\partial\Omega$  上。由引理 3.14, 我们可以找到一系列具有零边界值的对数凹函数  $\{\gamma_{0,n}\}$  几乎处处收敛到  $\gamma_0$ 。令

$$\gamma_n(t, x) := \int_{\Omega} K_{\Omega}(x, y, t) \gamma_{0,n}(y) dy, \quad (3.97)$$

其中  $K_{\Omega}(x, y, t)$  是该薛定谔方程的核函数 (根据文献 [82])。应用勒贝格控制收敛定理, 可得:

$$\lim_{n \rightarrow \infty} \gamma_n(t, x) = \int_{\Omega} K_{\Omega}(x, y, t) \gamma_0(y) dy = \gamma(t, x), \quad x \in \Omega, t > 0. \quad (3.98)$$

另一方面, 由引理 3.3 可知  $\gamma_n(t, \cdot)$  对于任意  $t > 0$  都在  $\Omega$  上是对数凹的。最后利用对数凹性在分布收敛下的保持性质, 即可证明引理 3.4。

### 3.5.6 定理 3.3 的证明

对于任意光滑凸区域  $\Omega$ , 根据文献 [101] 中的 *Theorem 2.7.1*, 存在一系列有界光滑凸区域  $\{\Omega_n\}_{n \geq 1}$  使得

$$\Omega_1 \subset \Omega_2 \subset \cdots \subset \Omega_n \subset \cdots, \quad \bigcup_{n=1}^{\infty} \Omega_n = \Omega. \quad (3.99)$$

如果  $\Omega = \mathbb{R}^N$ , 我们只需取  $\Omega_n = B_n(0)$ , 即以原点为中心半径为  $n$  的球。

考虑在  $\Omega_n$  上的方程序列:

$$\begin{cases} \partial_t \gamma_n + c(x) \gamma_n - \Delta \gamma_n = 0 & \text{在 } (0, \infty) \times \Omega_n \\ \gamma_n(t, x) = \gamma(t, x) \chi_{\overline{\Omega}_n}(x) & \text{在 } [0, \infty) \times \overline{\Omega}_n \end{cases} \quad (3.100)$$

其中  $\gamma(t, x)$  是全空间  $\mathbb{R}^N$  上薛定谔方程的解:

$$\begin{cases} \partial_t \gamma + c(x) \gamma - \Delta \gamma = 0 & \text{在 } (0, \infty) \times \mathbb{R}^N \\ \gamma(0, x) = \gamma_0(x) & \text{在 } \mathbb{R}^N \end{cases}. \quad (3.101)$$

由解的存在唯一性可知,  $\gamma(t, x) \chi_{\overline{\Omega}_n}(x)$  是方程 (3.100) 的唯一解。

由引理 3.4, 我们知道对于任意  $t > 0$ ,  $\gamma_n(t, \cdot)$  在  $\Omega_n$  上是对数凹的。进一步地,  $\gamma_n$  在  $L^1$  意义下收敛到  $\gamma$ , 这足以证明对于任意  $t > 0, x \in \mathbb{R}^N$ , 有

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^N \setminus \Omega_n} \gamma(t, x) dx = 0. \quad (3.102)$$

根据引理 3.7, 我们有

$$\begin{aligned}
\gamma(t, x) &= \mathbb{E} \left[ \gamma_0(2W_{t/2}) e^{-2 \int_0^{t/2} c(2W_r) dr} \right] \left( \frac{x}{2} \right) \\
&\leq \mathbb{E} \left[ \gamma_0(2W_{t/2}) \right] \left( \frac{x}{2} \right) \\
&= \int_{\mathbb{R}^N} \gamma_0(2y) (\pi t)^{-\frac{N}{2}} e^{-\frac{\|y-x/2\|^2}{t}} dy \\
&= \int_{\mathbb{R}^N} \gamma_0(y) (4\pi t)^{-\frac{N}{2}} e^{-\frac{\|y-x\|^2}{4t}} dy.
\end{aligned} \tag{3.103}$$

由于  $\gamma_0$  是  $\mathbb{R}^N$  上的连续概率测度, 它是这类有界的函数, 且满足

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^N \setminus \Omega_n} \gamma_0(x) dx = 0. \tag{3.104}$$

于是

$$\begin{aligned}
\int_{\mathbb{R}^N \setminus \Omega_n} \gamma(t, x) dx &= \pi^{-\frac{N}{2}} \int_{\mathbb{R}^N \setminus \Omega_n} dx \int_{\mathbb{R}^N} \gamma_0(x + 2\sqrt{t}\eta) e^{-\eta^2} d\eta \\
&= \pi^{-\frac{N}{2}} \int_{\mathbb{R}^N} e^{-\eta^2} d\eta \int_{\mathbb{R}^N \setminus \Omega_n} \gamma_0(x + 2\sqrt{t}\eta) dx \\
\overline{\lim}_{n \rightarrow \infty} \int_{\mathbb{R}^N \setminus \Omega_n} \gamma(t, x) dx &\leq \pi^{-\frac{N}{2}} \int_{\mathbb{R}^N} e^{-\eta^2} d\eta \overline{\lim}_{n \rightarrow \infty} \int_{\mathbb{R}^N \setminus \Omega_n} \gamma_0(x + 2\sqrt{t}\eta) dx = 0.
\end{aligned} \tag{3.105}$$

因此, 存在一个子序列  $\{\gamma_{n_k}\}$ , 对于任意  $t > 0$  几乎处处满足  $\gamma_{n_k} \rightarrow \gamma$  (当  $n_k \rightarrow \infty$  时), 因为  $L^1$  收敛的序列必然包含一个几乎处处收敛的子序列。

此外, 由比较原理, 我们得到:

$$\gamma_{n_k}(t, x) \leq \gamma_{n_{k+1}}(t, x) \leq \gamma(t, x), \text{ 在 } \mathbb{R}^N \times (0, \infty) \text{ 上,} \tag{3.106}$$

$$\lim_{n_k \rightarrow \infty} \gamma_{n_k}(t, x) = \gamma(t, x), \text{ 在 } \mathbb{R}^N \times (0, \infty) \text{ 上.} \tag{3.107}$$

由此我们证明了对于任意  $t > 0$ ,  $\gamma(t, \cdot)$  在  $\Omega$  上是对数凹的。至此, 定理 3.3 得证。

### 3.5.7 定理 3.9 的证明

我们计算 KL 散度的时间导数:

$$\begin{aligned}
\frac{d}{dt} \text{KL}(\mu_t \| \nu_t) &= \frac{d}{dt} \int_{\mathbb{R}} \mu(x, t) \log \frac{\mu(x, t)}{\nu(x, t)} dx \\
&= \int (\partial_t \mu_t) \log \frac{\mu_t}{\nu_t} dx + \int \mu_t \cdot \frac{1}{\mu_t} (\partial_t \mu_t) dx - \int \frac{\mu_t}{\nu_t} \cdot (\partial_t \nu_t) dx.
\end{aligned} \tag{3.108}$$

代入定理 3.6 中的演化方程, 并将积分分为漂移、扩散和跳跃三部分分别处理。

1. 漂移项部分:

$$\begin{aligned} \text{Drift} &= - \int \frac{\partial}{\partial x}(a\mu) \log \frac{\mu}{\nu} dx + \int a\mu \frac{\partial}{\partial x} \left( \frac{\mu}{\nu} \right) \frac{\nu}{\mu} dx \\ &= \int a\mu \frac{\partial}{\partial x} \log \frac{\mu}{\nu} dx - \int a\nu \frac{\mu}{\nu} \frac{\partial}{\partial x} \log \frac{\mu}{\nu} dx = 0. \end{aligned} \quad (3.109)$$

注意, 这里我们应用了分部积分, 并假设边界项为零。上式中第三项 (对应  $\partial_t \nu$  的贡献) 与第一项抵消。更详细地:

$$\begin{aligned} - \int a(x, t)\nu(x, t) \frac{\partial}{\partial x} \left( \frac{\mu_t}{\nu_t} \right) dx &= (-1) \int a(x, t)\nu(x, t) \frac{\mu(x, t)}{\nu(x, t)} \frac{\partial}{\partial x} \left( \log \frac{\mu_t}{\nu_t} \right) dx \\ &= - \int a(x, t)\mu(x, t) \frac{\partial}{\partial x} \left( \log \frac{\mu_t}{\nu_t} \right) dx. \end{aligned} \quad (3.110)$$

该项与  $\partial_t \mu_t$  展开式中的漂移部分抵消。

2. 扩散项部分: 我们重点关注扩散项的交叉项处理。

$$\begin{aligned} -\frac{1}{2} \int \frac{\mu_t}{\nu_t} \frac{\partial^2}{\partial x^2} (b(x, t)\nu(x, t)) dx &= \frac{1}{2} \int \frac{\partial}{\partial x} (b(x, t)\nu(x, t)) \frac{\partial}{\partial x} \left( \frac{\mu_t}{\nu_t} \right) dx \\ &= \frac{1}{2} \int \frac{\partial}{\partial x} (b\nu) \frac{\mu}{\nu} \frac{\partial}{\partial x} \log \frac{\mu_t}{\nu_t} dx \\ &= \frac{1}{2} \int \left[ \frac{\partial}{\partial x} (b\mu) - b\partial_x \mu + b\mu \frac{\partial_x \nu}{\nu} \right] \frac{\partial}{\partial x} \log \frac{\mu_t}{\nu_t} dx \\ &= \frac{1}{2} \int \frac{\partial}{\partial x} (b\mu) \frac{\partial}{\partial x} \log \frac{\mu_t}{\nu_t} dx - \frac{1}{2} \int b\mu \left\| \frac{\partial}{\partial x} \log \frac{\mu_t}{\nu_t} \right\|^2 dx. \end{aligned} \quad (3.111)$$

同样地, 上式第一项与  $\partial_t \mu_t$  展开式中的扩散部分抵消, 仅留下 Fisher 信息项:

$$-\frac{1}{2} \int b\mu \|\partial_x \log(\mu/\nu)\|^2 dx.$$

3. 跳跃项部分: 整理跳跃部分的所有积分项, 得到:

$$\begin{aligned} \text{Jump} &= \iint \mu(x - \xi) \lambda(x - \xi) w(\xi|x - \xi) \log \frac{\mu(x)}{\nu(x)} d\xi dx - \int \mu \lambda \log \frac{\mu}{\nu} dx \\ &\quad - \iint \frac{\mu(x)}{\nu(x)} \nu(x - \xi) \lambda(x - \xi) w(\xi|x - \xi) d\xi dx + \int \frac{\mu}{\nu} \nu \lambda dx. \end{aligned} \quad (3.112)$$

通过变量代换  $y = x - \xi$ , 上述积分可重组为:

$$\iint \mu(y) \lambda(y) w(\xi|y) \left[ \log \frac{\mu(y + \xi)}{\nu(y + \xi)} - \log \frac{\mu(y)}{\nu(y)} - \frac{\mu(y + \xi)\nu(y)}{\nu(y + \xi)\mu(y)} + 1 \right] d\xi dy. \quad (3.113)$$

利用对数恒等式，括号内的项即为：

$$\log \frac{\mu(y+\xi)\nu(y)}{\nu(y+\xi)\mu(y)} - \frac{\mu(y+\xi)\nu(y)}{\nu(y+\xi)\mu(y)} + 1.$$

令  $z = \frac{\mu(y+\xi)\nu(y)}{\nu(y+\xi)\mu(y)}$ ，则该项形式为  $\log z - z + 1$ 。由于对于任意  $z > 0$ ，均有  $\log z \leq z - 1$ ，因此该项始终非正。

将扩散部分与跳跃部分合并，即证得定理结论：

$$\frac{d}{dt} \text{KL}(\mu_t \| \nu_t) \leq -\frac{1}{2} \int_{\mathbb{R}} b(x, t) \mu(x, t) \left\| \frac{\partial}{\partial x} \log \frac{\mu_t}{\nu_t} \right\|^2 dx \leq 0. \quad (3.114)$$

### 3.6 本章小结

本章围绕连续时间马尔可夫过程中的信息演化性质展开了深入研究，主要涵盖了 Fokker-Planck 流的几何性质与跳-扩散信道的信息微分理论两个层面。

首先，我们建立了 Fokker-Planck 流经典解的适定性理论，并基于最优传输理论中的梯度流框架，推导了相对熵与互信息的高阶时间导数公式。结合 Ishige 等人关于抛物型方程解的空间对数凹性理论 [85]，我们证明了在势函数与初始分布满足特定的几何约束（如  $V/2$ -相对对数凹）时，互信息始终保持为时间的凸函数。这一结论成功将热流和 OU 流中的经典结果 [5-6] 推广至了一般非线性势函数的情形，且该证明方法不依赖于方程的显式解，具有更广泛的普适性。

其次，为了刻画复杂系统中常见的突发性干扰，本章将信道模型从连续扩散扩展至跳-扩散过程。通过引入 Kramers-Moyal 展开与 Kolmogorov 积分-微分方程，我们系统推导了跳-扩散信道下熵与互信息的演化规律。研究表明，互信息的时间导数可分解为扩散部分的 Fisher 型信息与跳跃部分的错配 KL 散度之和，从而将经典的 de Bruijn 恒等式与 I-MMSE 关系推广到了包含跳跃的更一般框架中。在此基础上，我们进一步推出了互信息与 KL 散度沿跳-扩散过程演化时的单调性结果，推广了数据处理不等式在连续时间马尔可夫过程中的适用性。

未来的研究方向主要集中在以下几个维度：

1. 弱化凸性条件：探究在初始状态不满足  $V/2$ -相对对数凹时，互信息是否具备“最终凸性”（Eventual Convexity），以及是否能在更弱的矩条件或 Fisher 信息约束下保持凸性。
2. 跳跃过程的几何分析：在跳-扩散框架下，进一步探索概率测度演化的几何性质，例如寻找使得 KL 散度指数衰减的 Log-Sobolev 不等式的充分条件，以及建立 Wasserstein 空间中适应非连续路径的度量收缩理论。

## 第 4 章 广义相对 Fisher 信息的强数据处理不等式

### 4.1 本章引言

Fisher 信息 (Fisher Information, FI) 最初由 R. A. Fisher [25] 在统计学中引入, 作为参数估计精度的度量。然而, 在过去的几十年里, 这一概念已被广泛推广为定义在概率测度上的非参数泛函, 并成为信息论 [71]、最优传输 [17, 72, 102]、马尔可夫扩散算子的分析与几何 [4, 103] 以及偏微分方程 (PDE) 研究中的核心对象。

(非参数) Fisher 信息, 有时被称为 Fisher 信息泛函, 量化了概率密度的内在结构和平滑度。与依赖于特定模型的经典参数 Fisher 信息不同, 它是概率测度空间上的一个泛函, 捕捉了信息、扩散与概率空间几何之间的深刻联系。对于  $\mathbb{R}^d$  上具有足够正则性的绝对连续概率测度  $\mu$ , 其 Fisher 信息定义为:

$$\text{FI}(\mu) := \int_{\mathbb{R}^d} \|\nabla \log \mu(x)\|^2 d\mu. \quad (4.1)$$

这一定义可以直观地理解为得分函数 (即对数密度的梯度) 模平方在概率测度  $\mu$  下的期望。因此,  $\text{FI}(\mu)$  反映了  $\mu$  的“尖锐程度”或集中程度: 对于平滑且分散的分布,  $\nabla \log \mu$  几乎处处有限甚至有界, 导致  $\text{FI}(\mu)$  有限; 反之, 如果  $\mu$  高度集中或存在尖峰区域, 梯度变大,  $\text{FI}(\mu)$  也会相应增加甚至爆破。事实上, 在所有具有固定方差的分布中, 高斯分布具有最小的 Fisher 信息。这一性质在 PDE 理论中也得到了广泛利用, 特别是在证明动力学方程解是否在有限时间内爆破的问题上 [26, 104-109]。

H. P. McKean 在 1966 年做出了里程碑式的贡献, 受 Kac 对 Boltzmann 方程一维简化模型的启发, 他将 Fisher 信息引入了动力学理论 [26]。McKean 证明了: (a) 对于 Kac 模型, Fisher 信息是非增的; (b) 一维 Boltzmann 方程的解随着  $t \rightarrow \infty$  在  $L^1$  意义下指数收敛于平衡态。这一研究路线随后通过泛函不等式、概率度量、熵方法、傅里叶分析和线性化等工具得到了进一步发展, Arkeryd [27]、Bobylev [28]、Carvalho [29] 和 Villani 等人 [30] 做出了重要贡献。熵沿 Boltzmann 流的单调性最初被 McKean 认为是一维现象; 然而, Toscani 在 1992 年将其扩展到二维 Maxwellian 碰撞 [110], Villani 在 1998 年的博士论文中进一步将其推广到所有维度 [105]。

另一个基本方程是 Landau 方程, 由 Landau 于 1936 年引入, 用于模拟等离子体中库仑碰撞下的粒子密度演化 [111], 并可作为 Boltzmann 方程的掠尾碰撞极限

---

本章内容已投递在 IEEE Transactions on Information Theory (TIT), 目前在审.

(Grazing-collision Limit) 导出 [112]。此后，它成为动力学理论的核心，特别是在分析解的爆破与全局正则性方面。最近，Guillen 和 Silvestre 证明了 Fisher 信息沿 Landau 流递减，确立了 Landau 方程不会爆破这一长期悬而未决的结论 [109]。

从信息论的角度来看，Fisher 信息与微分熵也密切相关。考虑  $X_0 \sim p(x)$  被独立的 Gaussian 噪声  $Z \sim \mathcal{N}(0, I)$  扰动，形成  $X_t = X_0 + \sqrt{t}Z$ 。  $X_t$  的密度  $p_t$  在热方程（或 Fokker-Planck 方程，FPE）下演化。著名的 de Bruijn 恒等式断言：

$$\left. \frac{d}{dt} h(p_t) \right|_{t=0} = \frac{1}{2} \text{FI}(p), \quad (4.2)$$

其中  $h$  表示微分熵。从历史上看，这一关系并非由 de Bruijn 本人首次发表，而是出现在 Stam 1959 年关于信息论不等式的论文中 [2]，Stam 明确指出是 de Bruijn 私下与其交流了这一结果。此后，该恒等式便以 de Bruijn 的名字命名，并成为连接熵、Fisher 信息和 Gaussian 扰动的基石。该恒等式强调 Fisher 信息是 Gaussian 扰动下熵产生的精确速率，这为证明熵幂不等式 (EPI) 提供了关键的分析步骤。EPI 最初由 Shannon [113]、Stam 和 Blachman [114] 提出，断言熵幂在独立随机变量的卷积下是超可加性的，在信息论和统计物理中具有深远意义 [115-116]。在现代文献中，de Bruijn 恒等式常与其等价形式——由 Guo, Shamai 和 Verdú 提出的 I-MMSE 关系 [31] 并列提出。I-MMSE 关系后来被推广到 Fokker-Planck 信道 [3]（我们也称之为 FP 流）、概率扩散模型 [117] 和跳-扩散信道 [76]，进一步凸显了其在连接信息量度与随机动力学方面的核心地位。

Fisher 信息的几何解释随着最优传输和 Otto 演算的出现而显现：具有有限二阶矩的概率测度空间  $\mathcal{P}_2(\mathbb{R}^d)$  被视为赋予 2-Wasserstein 度量的无限维（伪）黎曼流形 [17]。在此背景下，(Boltzmann) 微分熵  $H(p) = \int p \log p \, dx$  充当势泛函。热方程  $\partial_t p = \Delta p$  正是  $H(p)$  在 Wasserstein 空间中的梯度流，且沿该流有：

$$\frac{d}{dt} H(p_t) = -\text{FI}(p_t). \quad (4.3)$$

因此，Fisher 信息刻画了熵的时间导数的负值，反映了熵的耗散速率。更一般地，考虑 Fokker-Planck 方程  $\partial_t p = \nabla \cdot (p \nabla V) + \Delta p$ ，其平衡分布为  $\pi(x) \propto e^{-V(x)}$ 。FPE 仍然是相对于参考测度  $\pi$  的相对熵  $H_\pi(p) = \int p \log \frac{p}{\pi} \, dx$  的梯度流 [63]。

在信息论中，数据处理不等式 (DPI) 断言，通过互信息、KL、Rényi 或  $\Phi$ -散度（或 Csiszár 的  $f$ -散度 [118]）衡量的信息，在通过 Gaussian 噪声信道时不会增加。强数据处理不等式 (SDPI) 通过量化收缩率进一步加强了这一点，其中收缩率依赖于信道和

输入分布。目前 SDPI 已成为分析采样算法（如未调整 Langevin 算法 (ULA) 和近端采样器）的有力工具，这些算法可以视为短时 Gaussian 信道（以及近端采样器的反向 Gaussian 信道）的迭代组合。其核心思想是建立沿这些信道的散度的 SDPI。通过迭代应用 SDPI，可以获得这些算法的混合时间保证。近期工作 [8-9, 119-121] 已将此框架应用于各种散度。对于由热流诱导的 Gaussian 信道，Klartag 和 Ordentlich 获得了相应 SDPI 系数的上下界，以及 de Bruijn 恒等式的推广，并给出了基于庞加莱常数的互信息和 MMSE 的新下界 [122]。

相比之下，相对 Fisher 信息的 SDPI 发展相对较少。由于相对 Fisher 信息是相对熵沿 FPE 的时间导数，相对 Fisher 信息的 SDPI 将直接蕴含 KL 散度的 SDPI。相对 Fisher 信息测量两个分布得分函数之间的平方  $L^2$  距离，通常比 KL 或 TV 距离等散度更强。事实上，即使对于强对数凹测度，相对 Fisher 信息也可能比 KL 散度大任意倍（参见 [10] 附录 D 中的反例），这使其成为一个更具挑战性但也更有力的度量。最近，Wibisono 建立了热流和 OU 流下相对 Fisher 信息的 SDPI [10]，在  $\alpha$ -强对数凹和  $L$ -光滑势函数下得出了近端采样器的尖锐混合时间保证。具体而言，相对 Fisher 信息的混合时间按  $O(\log \epsilon^{-1})$  缩放，显著优于对采样分布与目标分布之间误差  $\epsilon^{-1}$  的多项式依赖。

这些进展与采样算法分析中的一个更广泛趋势相平行：从 KL 散度转向更强或更灵活的信息度量，如 Rényi 或  $\Phi$ -散度。事实上，Rényi 散度的分析特别具有吸引力，因为收敛保证可以通过比较不等式转移到 KL、总变差和 Wasserstein-2 距离，并且还高精度采样器（如 Metropolis-adjusted Langevin 算法 (MALA) [123] 或 zigzag 采样器 [124]）提供了“热启动”（Warm Start）。然而，使用 Rényi 散度引入了新的技术障碍，因为它排除了标准耦合离散化论证的使用。最近，Chewi 等人在对数索伯列夫不等式 (LSI) 下提供了 Langevin 蒙特卡洛 (LMC) 的真正 Rényi 收敛保证 [34]，从而通过更直接的证明得出了比 [9, 32-33] 更强的结果。

受这些进展的启发，我们期望知晓相对 Fisher 信息是否允许类似的推广，例如 Rényi 版本的相对 Fisher 信息（或相对 Rényi-Fisher 信息，RFI）。这实际上已被提议作为 Rényi 散度的 de Bruijn 恒等式的类比 [125]，用于建立几个尖锐的 Rényi 熵等周不等式，将经典的熵等周不等式 [126] 推广到 Rényi 设置。具体而言，我们将 RFI 视为 FI 的直接 Rényi 对应物：模仿 KL 散度的 de Bruijn 恒等式，RFI 被定义为 Rényi 散度沿热流的时间导数。除了 RFI，另一个自然的扩展是相对  $\Phi$ -Fisher 信息 ( $\Phi$ -FI)，它被定义为 Csiszár  $\Phi$ -散度在 Fisher 信息层面的自然类比。其构造动机与  $\Phi$ -散度一脉相

承：这类度量以统一的方式涵盖了 KL、 $\chi^2$ 、Hellinger 和总变差距离等经典情形作为特例，并与  $\Phi$ -Sobolev 不等式密切相关 [127]。因此，开发  $\Phi$ -FI 的 SDPI 将允许在比对数索伯列夫更弱的泛函不等式下获得收缩保证，从而拓宽 Fisher 型分析的适用性。此外， $\Phi$ -FI 继承了  $\Phi$ -散度的灵活性，使其特别适合研究现代采样和优化算法中的鲁棒性和隐私性。从结构上看，KL 散度与 Fisher 信息通过 de Bruijn 恒等式紧密关联；类似地，将  $\Phi$ -散度推广至  $\Phi$ -Fisher 信息，可在得分函数的意义下实现自然的微分类比，从而在统一框架下融合基于散度与基于 Fisher 信息的分析视角。

在本章中，我们关注这两个扩展：相对 Rényi-Fisher 信息 (RFI) 和相对  $\Phi$ -Fisher 信息 ( $\Phi$ -FI)。我们的第一个主要贡献是沿 FPE 的 RFI 时间导数的演算，进而推导出在所谓的“对数 Hessian 比较原理”下的 SDPI 型收缩，该原理断言对于适当的常数  $K_1, K_2$  和势函数  $V$ ，我们有：

$$\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V. \quad (4.4)$$

然后，在  $\Phi$  方面，我们推导了相对  $\Phi$ -FI 沿 FPE 的导数公式，建立了相应的 (S)DPI，并将其应用于获得近端采样器在  $\Phi$ -FI 中的混合时间保证。关键技术步骤之一是通过根据  $\Phi$ -导数的阶数对各项进行分类来简化表达式，从而导出  $\Phi$ -FI 时间导数的优雅的矩阵表示。为获得比标准庞加莱不等式更快的衰减速率（类似于 Wibisono 等人对 FI 的分析），我们引入一个更精细的测试函数，以有效控制演化方程中的高阶项。同时我们引入了  $\Phi$ -散度的  $\theta$ -容许生成函数的概念，该函数满足涉及包含前四阶导数的自然微分不等式。最后，我们提出了一套实用的验证策略来检验对数 Hessian 比较原理，其核心在于利用 Gozlan, Li 等人 [128] 提出的扩散半群的显式对数 Hessian 恒等式，该恒等式直接推出了逐点曲率下界。结合对数凹性的保持和势函数的曲率控制 [62]，这些恒等式导出了所需比较的可验证充分条件。

不过，上述保证的比较仅在预热期（burn-in period）之后成立，即它是一个最终比较原理。作为补充结果，我们建立了一个“全时保持”定理：如果  $\mu_0$  或  $\nu_0$  是高斯分布，且另一个是严格对数凹的，并且满足

$$\nabla^2 \log \nu_0 \preceq K \nabla^2 \log \mu_0 \quad (K \in [0, 1]), \quad (4.5)$$

则该不等式在热流和 OU 半群下对所有  $t > 0$  传播。该证明结合了 Brascamp-Lieb 和 Cramér-Rao 不等式 [129] 与显式半群表示，从而将一类自然初始数据的最终收缩性扩展为时间一致的保持性。对于非高斯初始测度，这个问题仍然是开放的，也是目前在我们的框架内推导 RFI 混合时间保证的主要障碍。

本章的结构安排如下：在第 4.2.1 节中，我们回顾了相对 FI 沿热流和 OU 流的 SDPI 的现有结果。作为前人工作的扩展，我们在定理 4.4 中证明了一个在没有对称假设下提升收敛率的变体定理。此外，基于 FP 流的对数凹性保持性质，我们在定理 4.5 中建立了沿一般 FP 流的相对 FI 的 SDPI。在第 4.2.2 节中，我们在定理 4.6 中推导了相对 RFI 沿一般 FP 流的时间导数，这作为 RFI 的 SDPI 的基础。在第 4.2.3 节中，我们在对数 Hessian 比较原理成立的前提下，分别在定理 4.7、4.8 和 4.9 中展示了关于 RFI 沿热流、OU 流和一般 FP 流的 SDPI 的主要结果。在第 4.3.1 节中，我们在定理 4.10 中推导了相对  $\Phi$ -FI 沿一般 FP 流的时间导数，随后在第 4.3.2 节中，我们分别在定理 4.11、4.12 和 4.13 中建立了  $\Phi$ -FI 沿热流、OU 流和一般 FP 流的 SDPI。作为应用，在第 4.3.3 节中，我们在定理 4.14 和推论 4.2 中提供了近端采样器的混合时间保证。在第 4.4 节中，我们提供了一条验证 RFI 的 SDPI 中使用的对数 Hessian 比较原理的实用路径。具体而言，在第 4.4.1 节中，我们在定理 4.15 和 4.16 中分别建立了热流的最终和永久对数 Hessian 比较原理，并在推论 4.3 中将其应用于相对 RFI 的 SDPI。在第 4.4.2 节中，我们在定理 4.17 和 4.18 中建立了 OU 流的最终和永久对数 Hessian 比较原理，并在推论 4.5 中导出了沿 OU 流的相对 RFI 的有效 SDPI。在第 4.4.3 节中，我们在定理 4.21 中建立了一般 FP 流的最终对数 Hessian 比较原理，随后在推论 4.7 中将其应用于沿一般 FP 流的相对 RFI 的 SDPI。在第 4.6 节中，我们讨论了一些进一步的问题并总结了全文。

## 4.2 相对 Rényi-Fisher 信息沿 Fokker-Planck 流的演化行为

### 4.2.1 相对 Fisher 信息的强数据处理不等式

Vempala 和 Wibisono [9] 表明, 庞加莱不等式或对数索伯列夫不等式蕴含了 Langevin 扩散在 Rényi 散度意义下的收敛性。

**定理 4.1** [定理 2, 3 [9]] 设  $(P_t)_{t \geq 0}$  为可逆扩散马尔可夫半群,  $(\pi_t)_{t \geq 0}$  表示与该半群关联的马尔可夫过程的分布,  $\pi$  为半群的平稳分布。假设  $\pi$  是  $\mathbb{R}^d$  上的概率测度, 相对于 Lebesgue 测度具有光滑密度  $\rho$ , 并且对于所有  $t \geq 0$ ,  $\pi_t$  相对于  $\pi$  绝对连续。则以下结论成立:

1. 假设常数为  $C_{\text{LSI}}$  的对数索伯列夫不等式成立。则对于所有  $q \geq 1$ ,

$$R_q(\pi_t \| \pi) \leq \exp\left(-\frac{2t}{qC_{\text{LSI}}}\right) R_q(\pi_0 \| \pi). \quad (4.6)$$

2. 假设常数为  $C_{PI}$  的庞加莱不等式成立。则对于所有  $q \geq 2$ ,

$$R_q(\pi_t|\pi) = \begin{cases} R_q(\pi_0|\pi) - \frac{2t}{qC_{PI}}, & \text{若 } R_q(\pi_t|\pi) \geq 1, \\ \exp\left(-\frac{2t}{qC_{PI}}\right) R_q(\pi_0|\pi), & \text{若 } R_q(\pi_t|\pi) \leq 1. \end{cases} \quad (4.7)$$

这些结果自然引发了一个问题：类似的收缩现象是否适用于更精细的信息论量，如 Fisher 信息 (FI) 及其推广 RFI。在下文中，我们回顾 Wibisono [10] 建立的相对 FI 沿 Gaussian 流和 Ornstein-Uhlenbeck (OU) 流的 SDPI。特别是，我们在不对测度施加对称假设的情况下，为 OU 流建立了一个新的 SDPI。此外，结合文献 [62] 的对数凹性结果，我们还在适当假设下推导了在共同 Langevin 动力学下演化的测度之间的相对 FI 的 SDPI。

**定理 4.2** [定理 3 [10]] 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度，且  $FI(\mu_0|\nu_0) < \infty$ 。设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 Gaussian 信道的输出：

$$\partial_t u_t = \frac{1}{2} \Delta u_t. \quad (4.8)$$

则我们有：

1. 假设  $\nu_0$  是对数凹的，则对于所有  $t \geq 0$ ：

$$FI(\mu_t|\nu_t) \leq FI(\mu_0|\nu_0). \quad (4.9)$$

2. 假设对于某些  $\alpha > 0$ ， $\nu_0$  满足  $\alpha$ -SLC（强对数凹），则对于所有  $t \geq 0$ ：

$$FI(\mu_t|\nu_t) \leq \frac{FI(\mu_0|\nu_0)}{(1 + \alpha t)^2}. \quad (4.10)$$

3. 假设对于某些  $\alpha, \beta > 0$ ， $\nu_0$  满足  $\alpha$ -SLC 且  $\mu_0$  满足  $\beta$ -PI（庞加莱不等式），并假设  $\mu_0, \nu_0$  是对称的，则对于所有  $t \geq 0$ ：

$$FI(\mu_t|\nu_t) \leq \frac{FI(\mu_0|\nu_0)}{(1 + \alpha t)^2(1 + \beta t)}. \quad (4.11)$$

4. 假设  $\nu_0 \propto e^{-g-\psi}$ ，其中  $g: \mathbb{R}^d \rightarrow \mathbb{R}$  是  $\alpha$ -强凸的， $\psi: \mathbb{R}^d \rightarrow \mathbb{R}$  是  $L$ -Lipschitz 连续的 ( $\alpha > 0, 0 \leq L < \infty$ )，则对于所有  $t \geq 0$ ：

$$FI(\mu_t|\nu_t) \leq \frac{FI(\mu_0|\nu_0)}{(1 + \alpha t)^2} \exp\left(\frac{2tL^2}{1 + \alpha t} + \frac{8L\sqrt{t}}{\sqrt{1 + \alpha t}}\right). \quad (4.12)$$

**定理 4.3** [定理 4 [10]] 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度, 且  $\text{FI}(\mu_0 \|\nu_0) < \infty$ . 设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 OU 流的输出, 目标分布为  $\pi = \mathcal{N}(0, \gamma^{-1}I)$ :

$$\partial_t u_t = \nabla \cdot (\gamma x u_t) + \Delta u_t. \quad (4.13)$$

则我们有:

1. 假设对于某些  $\alpha > 0$ ,  $\nu_0$  满足  $\alpha$ -SLC, 则对于所有  $t \geq 0$ :

$$\text{FI}(\mu_t \|\nu_t) \leq \frac{\gamma^2 e^{-2\gamma t}}{(\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} \text{FI}(\mu_0 \|\nu_0). \quad (4.14)$$

2. 假设对于某些  $\alpha, \beta > 0$ ,  $\nu_0$  满足  $\alpha$ -SLC 且  $\mu_0$  满足  $\beta$ -PI. 假设  $\mu_0, \nu_0$  是对称的, 则对于所有  $t \geq 0$ :

$$\text{FI}(\mu_t \|\nu_t) \leq \frac{\gamma^3 e^{-4\gamma t}}{(\beta + e^{-2\gamma t}(\gamma - \beta))(\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} \text{FI}(\mu_0 \|\nu_0). \quad (4.15)$$

特别地, 作为定理 4.3 的扩展, 我们证明了一个提升收敛率的变体定理, 其中我们假设初始化有界但不对称。

**定理 4.4** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度, 且  $\text{FI}(\mu_0 \|\nu_0) < \infty$ . 设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 OU 流的输出, 目标分布为  $\pi = \mathcal{N}(0, \gamma^{-1}I)$ :

$$\partial_t u_t = \nabla \cdot (\gamma x u_t) + \Delta u_t. \quad (4.16)$$

假设对于某些  $\alpha, \beta > 0$ ,  $\nu_0$  满足  $\alpha$ -SLC 且  $\mu_0$  满足  $\beta$ -PI. 假设  $\mu_0, \nu_0$  是中心化的 (即  $\int x \mu_0(x) dx = 0$  且  $\int x \nu_0(x) dx = 0$ ) 且具有紧支集, 满足  $\text{supp}(\mu_0) \subset \text{supp}(\nu_0)$ ,  $\mu_0 \preceq \nu_0$  (即存在  $M > 0$  使得  $\mu_0 \leq M\nu_0$ ). 则对于所有  $t \geq 0$ , 存在常数  $C' > 0$  使得:

$$\begin{aligned} \text{FI}(\mu_t \|\nu_t) &\leq \frac{\gamma^3 e^{-4\gamma t} \text{FI}(\mu_0 \|\nu_0)}{(\beta + e^{-2\gamma t}(\gamma - \beta))(\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} + \frac{2C' \beta \gamma e^{-4\gamma t}}{(\beta + e^{-2\gamma t}(\gamma - \beta))(\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} \\ &= O(e^{-4\gamma t}). \end{aligned} \quad (4.17)$$

证明见第 4.5.1 小节。

**定理 4.5** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度, 且  $\text{FI}(\mu_0 \|\nu_0) < \infty$ . 设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 FP 流 (3.3) 的输出:

$$\partial_t u_t = \nabla \cdot (u_t \nabla V) + \frac{c}{2} \Delta u_t. \quad (4.18)$$

如果假设 3.1 和 3.2 成立, 并且假设势函数  $V$  是  $\alpha$ -强凸的 (即  $\nabla^2 V \geq \alpha I > 0$ ), 则我们有:

1. 假设  $\nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 则对于所有  $t \geq 0$ :

$$\text{FI}(\mu_t \|\nu_t) \leq \text{FI}(\mu_0 \|\nu_0). \quad (4.19)$$

2. 假设  $\nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 且  $\mu_t$  满足  $\beta_t$ -PI ( $\beta_t \geq 0$ )。假设  $\mu_0, \nu_0, V$  是对称的, 则对于所有  $t \geq 0$ :

$$\text{FI}(\mu_t \|\nu_t) \leq e^{-c \int_0^t \beta_s ds} \text{FI}(\mu_0 \|\nu_0). \quad (4.20)$$

3. 假设  $\mu_0, \nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 并假设  $\mu_0, \nu_0, V$  是对称的, 则对于所有  $t \geq 0$ :

$$\text{FI}(\mu_t \|\nu_t) \leq e^{-\alpha t} \text{FI}(\mu_0 \|\nu_0). \quad (4.21)$$

4. 如果  $\nu_0 = \nu \propto e^{-\frac{2}{c}V}$ , 则对于所有  $t \geq 0$ :

$$\text{FI}(\mu_t \|\nu) \leq e^{-2\alpha t} \text{FI}(\mu_0 \|\nu). \quad (4.22)$$

5. 假设  $\mu_0$  是  $\frac{V}{c}$ -相对对数凹的,  $\nu_0 = \nu \propto e^{-\frac{2}{c}V}$ , 且  $\mu_0, \nu$  是对称的, 则对于所有  $t \geq 0$ :

$$\text{FI}(\mu_t \|\nu) \leq e^{-3\alpha t} \text{FI}(\mu_0 \|\nu_0). \quad (4.23)$$

证明见第 4.5.2 小节。

#### 4.2.2 相对 Rényi-Fisher 信息沿 Fokker-Planck 流的时间导数

我们的关键工具是相对 Rényi-Fisher 信息沿同步 Fokker-Planck 流的时间导数公式。此公式是经典相对 Fisher 信息恒等式在 Rényi 设置下的推广。

**定理 4.6** 假设  $\mu_t$  和  $\nu_t$  沿 FP 流演化:

$$\begin{aligned} \partial_t \mu_t &= -\nabla \cdot (\mu_t b_t) + \frac{c}{2} \Delta \mu_t, \\ \partial_t \nu_t &= -\nabla \cdot (\nu_t b_t) + \frac{c}{2} \Delta \nu_t. \end{aligned} \quad (4.24)$$

则对于任何  $t \geq 0, 1 \leq q < +\infty$ ,

$$\partial_t R_q(\mu_t \|\nu_t) = -\frac{cq}{2} \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t \right) \quad (4.25)$$

$$\begin{aligned}
\partial_t \text{RFI}_q(\mu_t \| \nu_t) &= \frac{cq^2(q-1)}{2} \left( \int \rho_t^{q-1} d\mu_t \right)^{-2} \left\{ \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t \right)^2 \right. \\
&\quad \left. - \left( \int \rho_t^{q-1} d\mu_t \right) \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^4 d\mu_t \right) \right\} \\
&\quad - cq \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\
&\quad - 2cq(q-1) \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \right\} \\
&\quad + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{A_t^{\nu, b}}^2 d\mu_t \right\} \\
&\leq -cq \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\
&\quad - 2cq(q-1) \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \right\} \\
&\quad + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{A_t^{\nu, b}}^2 d\mu_t \right\}
\end{aligned} \tag{4.26}$$

其中  $\rho_t = \frac{\mu_t}{\nu_t}$ ,  $A_t^{\nu, b} := c\nabla^2 \log \nu_t - (\nabla b_t)_{\text{sym}}$ .

证明见第 4.5.3 小节。

在上述证明中，我们假设积分号下的微分是有效的，并且分部积分产生的所有边界项均消失。这些假设在一般情况下（甚至对于热流；参见 [7] 中的猜想 4.2）仍是开放问题，但在某些标准设置下可以验证。例如，如果定理 4.6 证明中的每一项积分都是有限的，且  $\mu_t, \nu_t$  足够正则（如  $C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$ ），则对于任意紧时间区间  $J \subset \mathbb{R}_+$ ，相应的被积函数  $f(t, x)$  满足：

$$\int_J \int_{\mathbb{R}^d} |\partial_t f(t, x)| dx dt < \infty, \tag{4.27}$$

因此， $t \mapsto \int_{\mathbb{R}^d} f(t, x) dx$  在  $J$  上绝对连续，积分号下的微分是合理的。关于确保此交换的充分条件，请参见第 4.5.4.1 小节的补充说明。对于定理 4.6 证明中的分部积分步骤，我们需要确保当积分域扩张至  $\mathbb{R}^d$  时边界项消失。这可以在以下充分条件下得到保证：

- $\rho_t = \frac{\mu_t}{\nu_t}$  有上界；
- $\nabla \log \rho_t, b_t, \nabla \cdot b_t$  可被某些多项式函数界定；
- FPE 具有唯一的指数衰减光滑解，即存在常数  $C_1, C_2, p > 0$  使得对所有  $x \in \mathbb{R}^d$ ,  $\mu_t(x), \nu_t(x) \leq C_1 e^{-C_2 \|x\|^p}$ ,

因为在定理 4.6 证明的每一步中，被积函数都由  $\nabla \log \rho_t, b_t, \nabla \cdot b_t$  和  $\rho_t$  的多项式界定。因此，由于  $\mu_t, \nu_t$  的指数衰减，边界项在积分域扩张至  $\mathbb{R}^d$  时消失。更多细节参见第

4.5.4.2 小节。如果我们在定理 4.6 中取  $q = 1$ ，我们可以立即恢复关于 KL 散度和 FI 沿 Fokker-Planck 流时间导数的已知结果 [10]。

**推论 4.1** (引理 2 [10]) 假设  $\mu_t$  和  $\nu_t$  沿 FP 流演化：

$$\begin{aligned}\partial_t \mu_t &= -\nabla \cdot (\mu_t b_t) + \frac{c}{2} \Delta \mu_t, \\ \partial_t \nu_t &= -\nabla \cdot (\nu_t b_t) + \frac{c}{2} \Delta \nu_t.\end{aligned}\tag{4.28}$$

则对于任意  $t \geq 0$ ,

$$\partial_t \text{KL}(\mu_t \| \nu_t) = -\frac{c}{2} \int \|\nabla \log \rho_t\|^2 d\mu_t \tag{4.29}$$

$$\partial_t \text{FI}(\mu_t \| \nu_t) = -c \int \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t + 2 \int \|\nabla \log \rho_t\|_{A_t^{\nu, b}}^2 d\mu_t \tag{4.30}$$

其中  $\rho_t = \frac{\mu_t}{\nu_t}$ 。

### 4.2.3 相对 Rényi-Fisher 信息的强数据处理不等式

研究相对 Fisher 信息的强数据处理不等式 (SDPI)，有助于揭示那些无法仅通过散度类不等式 (如 KL 散度) 刻画的更精细的收缩行为。然而，经典的 Fisher 信息通过 de Bruijn 恒等式与 KL 散度紧密关联，因此其理论框架本质上受限于 KL 散度的视角。为了捕捉更广泛的信息度量，自然要考虑由 Rényi 散度衍生出的新的信息量，即相对 Rényi-Fisher 信息 (RFI)。将 SDPI 扩展到相对 RFI 的情形，不仅统一了散度和 Fisher 信息框架下的收缩理论，还揭示了在 KL/FI 层面不可见的新结构见解。在下文中，我们建立了相对 RFI 的相应 SDPI 并讨论了遇到的挑战。

注意，虽然我们提出了关于相对 RFI 的 SDPI 的若干理论结果，但它们取决于一个关键条件，我们称之为“对数 Hessian 比较原理”，即  $\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V$ ，或相关表述中的等价矩阵不等式。一般来说，验证这一比较原理仍然是一个主要挑战：目前尚无通用的检查方法，而且其难度因高维对数密度的高阶导数 (如梯度和 Hessian) 估计的内在困难而加剧。因此，在完全一般性下建立比较原理似乎是一个开放问题。令人鼓舞的是，本章在几种重要设置下提供了一条切实可行的路径。依靠我们早期的对数凹性估计 [62]，我们可以在 Ornstein-Uhlenbeck (OU) 流和平稳测度下的 Langevin 动力学中建立对数 Hessian 的上界 (受制于假设 3.1 和 3.2)。结合 Gozlan-Li 等人 [128] 的对数 Hessian 表示公式 (该公式产生扩散半群的显式下界)，这些要素提供了一种在实践中验证对数 Hessian 比较原理的具体方法。我们在第 4.4 节详细阐述这一策略并讨论其含义。

**定理 4.7** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度, 且  $\text{RFI}_q(\mu_0\|\nu_0) < \infty, q \geq 1$ . 设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 Gaussian 信道的输出:

$$\partial_t \nu_t = \frac{1}{2} \Delta \nu_t. \quad (4.31)$$

则我们有:

1. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq (1 - q^{-1}) \nabla^2 \log \mu_t$ . 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t\|\nu_t) \leq \text{RFI}_q(\mu_0\|\nu_0). \quad (4.32)$$

2. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq \nabla^2 \log \mu_t$ , 且  $\mu_0$  或  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ). 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t\|\nu_t) \leq \frac{\text{RFI}_q(\mu_0\|\nu_0)}{(1 + \alpha t)^2}. \quad (4.33)$$

3. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t$ , 其中  $K \geq 1 - q^{-1}$  且  $\mu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t\|\nu_t) \leq \frac{\text{RFI}_q(\mu_0\|\nu_0)}{(1 + \alpha t)^{2(1+(K-1)q)}}. \quad (4.34)$$

此处若  $K = 0, q = 1$ , 我们约定  $\frac{1 + (K-1)q}{K} = 1$ .

4. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq (1 - q^{-1}) \nabla^2 \log \mu_t$ , 且  $\mu_t$  满足  $\beta_t$ -PI ( $\beta_t \geq 0$ ), 并假设  $\mu_0, \nu_0$  是对称的. 此外, 假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ . 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t\|\nu_t) \leq \exp\left(-\left(\frac{C_1}{C_2}\right)^{2(q-1)} \int_0^t \beta_s ds\right) \text{RFI}_q(\mu_0\|\nu_0). \quad (4.35)$$

5. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq (1 - q^{-1}) \nabla^2 \log \mu_t$ , 且  $\mu_0$  满足  $\beta$ -PI ( $\beta > 0$ ), 并假设  $\mu_0, \nu_0$  是对称的. 此外, 假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ . 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t\|\nu_t) \leq \frac{\text{RFI}_q(\mu_0\|\nu_0)}{(1 + \beta t)^{(C_1/C_2)^{2(q-1)}}}. \quad (4.36)$$

6. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t$ , 其中  $K \geq 1 - q^{-1}$ ,  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ) 且  $\mu_0$  满足  $\beta$ -PI ( $\beta > 0$ ), 并假设  $\mu_0, \nu_0$  是对称的. 如果假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ , 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t\|\nu_t) \leq \frac{\text{RFI}_q(\mu_0\|\nu_0)}{(1 + \alpha t)^{2(1+(K-1)q)/K} (1 + \beta t)^{(C_1/C_2)^{2(q-1)}}}. \quad (4.37)$$

此处若  $K = 0, q = 1$ , 我们定义  $\frac{1 + (K-1)q}{K} = 1$ .

7. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t$ . 假设  $K \geq 1 - q^{-1}$ ,  $\mu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 并假设  $\mu_0, \nu_0$  是对称的. 如果假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ , 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \parallel \nu_t) \leq \frac{\text{RFI}_q(\mu_0 \parallel \nu_0)}{(1 + \alpha t)^{2(1+(K-1)q) + (C_1/C_2)^{2(q-1)}}}. \quad (4.38)$$

证明见第 4.5.5 小节。

在定理 4.15 和 4.16 中, 我们提供了验证对数 Hessian 比较的充分条件. 结合定理 4.7, 我们可以获得显式的 SDPI 速率, 这在推论 4.3 中给出。

**定理 4.8** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度, 且  $\text{RFI}_q(\mu_0 \parallel \nu_0) < \infty, q \geq 1$ . 设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 OU 流的输出, 目标分布为  $\pi = \mathcal{N}(0, \gamma^{-1}I)$ :

$$\partial_t u_t = \nabla \cdot (\gamma x u_t) + \Delta u_t. \quad (4.39)$$

则我们有:

1. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq (1 - q^{-1}) \nabla^2 \log \mu_t - \frac{\gamma q^{-1}}{2} I$ . 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \parallel \nu_t) \leq \text{RFI}_q(\mu_0 \parallel \nu_0). \quad (4.40)$$

2. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t$ , 其中  $K \geq 1 - q^{-1}$  且  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \parallel \nu_t) \leq \frac{\gamma^{2K^{-1}(1+q(K-1))} e^{-2\gamma K^{-1}[2(1+q(K-1))-K]t}}{(\alpha + (\gamma - \alpha)e^{-2\gamma t})^{2K^{-1}(1+q(K-1))}} \text{RFI}_q(\mu_0 \parallel \nu_0). \quad (4.41)$$

此处若  $K = 0, q = 1$ , 我们定义  $\frac{1 + (K-1)q}{K} = 1$ 。

3. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t$ . 如果  $K \geq 1 - q^{-1}$  且  $\mu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \parallel \nu_t) \leq \frac{\gamma^{2(1+q(K-1))} e^{-2\gamma[2q(K-1)+1]t}}{(\alpha + (\gamma - \alpha)e^{-2\gamma t})^{2(q(K-1)+1)}} \text{RFI}_q(\mu_0 \parallel \nu_0). \quad (4.42)$$

4. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 I$ , 其中  $K_1 \geq 1 - q^{-1}$  且  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \parallel \nu_t) \leq \frac{\gamma^{2K_1^{-1}(1+q(K_1-1))} e^{-2\gamma K_1^{-1}[2(1+q(K_1-1))-K_1-2(q-1)K_2\gamma^{-1}]t}}{(\alpha + (\gamma - \alpha)e^{-2\gamma t})^{2K_1^{-1}(1+q(K_1-1))}} \text{RFI}_q(\mu_0 \parallel \nu_0). \quad (4.43)$$

此处若  $K_1 = 0, q = 1$ , 我们定义  $\frac{1 + (K_1-1)q}{K_1} = 1$ 。

5. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 I$ 。如果  $K_1 \geq 1 - q^{-1}$  且  $\mu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq \frac{\gamma^{2(q(K_1-1)+1)} e^{-2\gamma[2q(K_1-1)+1-2qK_2\gamma^{-1}]t}}{(\alpha + (\gamma - \alpha)e^{-2\gamma t})^{2(q(K_1-1)+1)}} \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.44)$$

6. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq (1 - q^{-1})\nabla^2 \log \mu_t - \frac{\gamma q^{-1}}{2} I$ , 且  $\mu_t$  满足  $\beta_t$ -PI ( $\beta_t \geq 0$ ), 并假设  $\mu_0, \nu_0$  是对称的。此外, 假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ 。则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq \exp\left(-2 \left(\frac{C_1}{C_2}\right)^{2(q-1)} \int_0^t \beta_s ds\right) \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.45)$$

7. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 I$ , 其中  $K_1 \geq 1 - q^{-1}$ ,  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ),  $\mu_0$  满足  $\beta$ -PI ( $\beta > 0$ ), 且  $\mu_0, \nu_0$  是对称的。如果假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ , 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq \mathcal{K}_t \text{RFI}_q(\mu_0 \|\nu_0), \quad (4.46)$$

其中

$$\mathcal{K}_t := \frac{\gamma^{(C_1/C_2)^{2(q-1)} + 2K_1^{-1}(1+q(K_1-1))} e^{-2\gamma K_1^{-1}[2(1+q(K_1-1)) + K_1(C_1/C_2)^{2(q-1)} - K_1 - 2(q-1)K_2\gamma^{-1}]t}}{(\beta + (\gamma - \beta)e^{-2\gamma t})^{(C_1/C_2)^{2(q-1)}} (\alpha + (\gamma - \alpha)e^{-2\gamma t})^{2K_1^{-1}(1+q(K_1-1))}}. \quad (4.47)$$

此处若  $K_1 = 0, q = 1$ , 我们定义  $\frac{1 + (K_1 - 1)q}{K_1} = 1$ 。

8. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 I$ 。假设  $K_1 \geq 1 - q^{-1}$ ,  $\mu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 且  $\mu_0, \nu_0$  是对称的。如果假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ , 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq \mathcal{K}'_t \text{RFI}_q(\mu_0 \|\nu_0), \quad (4.48)$$

其中

$$\mathcal{K}'_t := \frac{\gamma^{(C_1/C_2)^{2(q-1)} + 2(q(K_1-1)+1)} e^{-2\gamma[2q(K_1-1)+1+(C_1/C_2)^{2(q-1)} - 2qK_2\gamma^{-1}]t}}{(\alpha + (\gamma - \alpha)e^{-2\gamma t})^{(C_1/C_2)^{2(q-1)} + 2(q(K_1-1)+1)}}. \quad (4.49)$$

证明见第 4.5.6 小节。

在定理 4.17 和 4.18 中, 我们提供了验证对数 Hessian 比较的充分条件。结合定理 4.8, 这些条件使我们能够获得显式的 SDPI 速率, 这在推论 4.5 中给出。

**定理 4.9** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度, 且  $\text{RFI}_q(\mu_0 \|\nu_0) < \infty, q \geq 1$ 。设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 FP 流 (3.3) 的输出:

$$\partial_t u_t = \nabla \cdot (u_t \nabla V) + \frac{c}{2} \Delta u_t. \quad (4.50)$$

如果假设 3.1 和 3.2 成立, 并且假设势函数  $V$  是  $\alpha$ -强凸的 (即  $\nabla^2 V \geq \alpha I > 0$ ), 则我们有:

1. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \leq (1 - q^{-1})\nabla^2 \log \mu_t - c^{-1}q^{-1}\nabla^2 V$ 。则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.51)$$

2. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \leq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V$ , 其中

$$\begin{aligned} K_1 &\geq 1 - q^{-1} \\ K_1 &\geq 1 + cK_2, \end{aligned} \quad (4.52)$$

且  $\nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq e^{-2(q-1)K_1^{-1}(K_1 - cK_2 - 1)\alpha t} \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.53)$$

此处若  $K_1 = 0, q = 1$ , 我们定义  $\frac{1-q}{K_1} = 0$ 。

3. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \leq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V$ 。如果

$$\begin{aligned} K_1 &\geq 1 - q^{-1} \\ K_1 &\geq cK_2 + 1, \end{aligned} \quad (4.54)$$

且  $\mu_0$  是  $\frac{V}{c}$ -相对对数凹的, 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq e^{-2(K_1 - cK_2 - 1)q\alpha t} \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.55)$$

4. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \leq (1 - q^{-1})\nabla^2 \log \mu_t - c^{-1}q^{-1}\nabla^2 V$ , 且  $\mu_t$  满足  $\beta_t$ -PI ( $\beta_t \geq 0$ ), 并假设  $\mu_0, \nu_0, V$  是对称的。此外, 假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ 。则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq \exp\left(-c\left(\frac{C_1}{C_2}\right)^{2(q-1)} \int_0^t \beta_s ds\right) \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.56)$$

5. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \leq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V$ , 其中

$$\begin{aligned} K_1 &\geq 1 - q^{-1} \\ K_1 &\geq 1 + cK_2, \end{aligned} \quad (4.57)$$

且  $\mu_0, \nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 并假设  $\mu_0, \nu_0, V$  是对称的。如果假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ , 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq e^{-((C_1/C_2)^{2(q-1)} + 2(q-1)K_1^{-1}(K_1 - cK_2 - 1))\alpha t} \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.58)$$

此处若  $K_1 = 0, q = 1$ , 我们定义  $\frac{q-1}{K_1} = 0$ 。

6. 假设对于每个  $t \geq 0$ ,  $\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V$ 。如果

$$\begin{aligned} K_1 &\geq 1 - q^{-1} \\ K_1 &\geq cK_2 + 1, \end{aligned} \quad (4.59)$$

且  $\mu_0$  是  $\frac{V}{c}$ -相对对数凹的, 并假设  $\mu_0, \nu_0, V$  是对称的。如果假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ , 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq e^{-((C_1/C_2)^{2(q-1)} + 2(K_1 - cK_2 - 1)q)\alpha t} \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.60)$$

7. 如果  $\nu_0 = \nu \propto e^{-\frac{2}{c}V}$ , 且  $\mu_t$  是  $KV$ -相对对数凸的, 即  $\nabla^2 \log \mu_t \geq K \nabla^2 V$ , 满足  $q \geq 1, q(2 + cK) \geq 1 + cK$ , 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq e^{-2q[q(2+cK) - (1+cK)]\alpha t} \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.61)$$

8. 假设  $\nu_0 = \nu \propto e^{-\frac{2}{c}V}$ ,  $\mu_0$  是  $\frac{V}{c}$ -相对对数凹的, 且  $\mu_t$  是  $KV$ -相对对数凸的, 即  $\nabla^2 \log \mu_t \geq K \nabla^2 V$ , 满足  $q \geq 1, q(2 + cK) \geq 1 + cK$ , 并假设  $\mu_0, \nu_0, V$  是对称的。此外, 假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty$ 。则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \|\nu_t) \leq e^{-[2q[q(2+cK) - (1+cK)] + (C_1/C_2)^{2(q-1)}]\alpha t} \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.62)$$

证明见第 4.5.7 小节。

在定理 4.21 中, 我们提供了验证对数 Hessian 比较的充分条件。结合定理 4.9, 我们可以推导显式的 SDPI 速率, 这在推论 4.7 中给出。

### 4.3 相对 $\Phi$ -Fisher 信息沿 Fokker-Planck 流的演化行为

在本节中, 我们从相对 Rényi-Fisher 信息转向相对  $\Phi$ -Fisher 信息, 这是 Csiszár  $\Phi$ -散度在得分函数层面的类比。

### 4.3.1 相对 $\Phi$ -Fisher 信息沿 Fokker-Planck 流的时间导数

我们的关键工具是相对  $\Phi$ -Fisher 信息沿同步 Fokker-Planck 流的时间导数公式。与 RFI 情况不同,  $\Phi$ -FI 的计算内在地涉及  $\Phi$  的高阶导数 (最高四阶), 并且根据  $\Phi$  导数的阶数分组后, 所得恒等式具有清晰的矩阵形式。

**定理 4.10** 假设  $\mu_t$  和  $\nu_t$  沿 FP 流演化:

$$\begin{aligned}\partial_t \mu_t &= -\nabla \cdot (\mu_t b_t) + \frac{c}{2} \Delta \mu_t, \\ \partial_t \nu_t &= -\nabla \cdot (\nu_t b_t) + \frac{c}{2} \Delta \nu_t.\end{aligned}\quad (4.63)$$

则对于任意  $t \geq 0$ ,

$$\begin{aligned}\partial_t D_\Phi(\mu_t \| \nu_t) &= -\frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi''(\rho_t) d\nu_t \\ &= -\frac{c}{2} \text{FI}_\Phi(\mu_t \| \nu_t).\end{aligned}\quad (4.64)$$

$$\begin{aligned}\partial_t \text{FI}_\Phi(\mu_t \| \nu_t) &= 2 \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|_{A_t^{\nu, b}}^2 d\mu_t - c \int \rho_t \Phi''(\rho_t) \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \\ &\quad - c \int \|\nabla \log \rho_t\|^4 \left( \rho_t \Phi''(\rho_t) + 2\rho_t^2 \Phi'''(\rho_t) + \frac{1}{2} \rho_t^3 \Phi^{(4)}(\rho_t) \right) d\mu_t \\ &\quad - 2c \int \|\nabla \log \rho_t\|_{\nabla^2 \log \rho_t}^2 \left( \rho_t \Phi''(\rho_t) + \rho_t^2 \Phi'''(\rho_t) \right) d\mu_t \\ &= 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{A_t^{\nu, b}}^2 d\nu_t - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t,\end{aligned}\quad (4.65)$$

其中

$$\rho_t = \frac{\mu_t}{\nu_t}, \mathbf{X} = \frac{c}{2} \begin{pmatrix} 2\Phi''(\rho_t) & 2\Phi'''(\rho_t) \\ 2\Phi'''(\rho_t) & \Phi^{(4)}(\rho_t) \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} \|\nabla^2 \rho_t\|_{\text{HS}}^2 & \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2 \\ \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2 & \|\nabla \rho_t\|_{\nabla^4 \rho_t}^2 \end{pmatrix}.\quad (4.66)$$

证明见第 4.5.8 小节。

**注** 矩阵  $\mathbf{Y}$  是半正定的, 因为  $\|\nabla^2 \rho_t\|_{\text{HS}}^2 \geq 0$  且

$$\begin{aligned}\det(\mathbf{Y}) &= \|\nabla^2 \rho_t\|_{\text{HS}}^2 \|\nabla \rho_t\|^4 - \|\nabla \rho_t\|_{\nabla^2 \rho_t}^4 \\ &= \left( \sum_{i,j=1}^n |\partial_{ij} \rho_t|^2 \right) \left( \sum_{k=1}^n |\partial_k \rho_t|^2 \right)^2 - \left( \sum_{i,j=1}^n \partial_i \rho_t (\partial_{ij} \rho_t) \partial_j \rho_t \right)^2 \\ &= \left( \sum_{i,j=1}^n |\partial_{ij} \rho_t|^2 \right) \left( \sum_{i,j=1}^n |\partial_i \rho_t|^2 |\partial_j \rho_t|^2 \right) - \left( \sum_{i,j=1}^n \partial_i \rho_t (\partial_{ij} \rho_t) \partial_j \rho_t \right)^2 \\ &\geq 0,\end{aligned}\quad (4.67)$$

由 Cauchy-Schwarz 不等式得证。因此, 如果我们假设  $\Phi$  是  $\theta$ -容许生成函数 ( $0 < \theta \leq 1$ ), 则矩阵  $\mathbf{X}$  也是半正定的, 从而  $\text{Tr}(\mathbf{X}\mathbf{Y}) \geq 0$ 。

**注** 如果  $\nu_t$  从  $t = 0$  开始就是平稳测度（如果 FPE 存在平稳测度），那么计算可以通过 [79] 引理 2.3 证明中的  $\Gamma$ -演算方法简化，其中作者研究了更一般的 FPE。

在上述论证中，我们假设在适当条件下积分号下的微分是有效的，并且分部积分产生的边界项消失。对于相对  $\Phi$ -Fisher 信息的情况，如果我们进一步假设  $\Phi \in C^4(\mathbb{R})$ ，则边界项消失成立。更多细节参见第 4.5.4 小节。

### 4.3.2 相对 $\Phi$ -Fisher 信息的强数据处理不等式

**定理 4.11** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度，且  $\text{FI}_\Phi(\mu_0 \parallel \nu_0) < \infty$ ，其中  $\Phi$  是  $\theta$ -容许生成函数 ( $0 < \theta \leq 1$ )。设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 Gaussian 信道的输出：

$$\partial_t u_t = \frac{1}{2} \Delta u_t. \quad (4.68)$$

则我们有：

1. 假设  $\nu_0$  是对数凹的，则对于所有  $t \geq 0$ ：

$$\text{FI}_\Phi(\mu_t \parallel \nu_t) \leq \text{FI}_\Phi(\mu_0 \parallel \nu_0). \quad (4.69)$$

2. 假设  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ )，则对于所有  $t \geq 0$ ：

$$\text{FI}_\Phi(\mu_t \parallel \nu_t) \leq \frac{\text{FI}_\Phi(\mu_0 \parallel \nu_0)}{(1 + \alpha t)^2}. \quad (4.70)$$

3. 假设  $\nu_0$  满足  $\alpha$ -SLC 并假设  $\mu_0, \nu_0$  是对称的，则对于所有  $t \geq 0$ ：

$$\text{FI}_\Phi(\mu_t \parallel \nu_t) \leq \frac{\text{FI}_\Phi(\mu_0 \parallel \nu_0)}{(1 + \alpha t)^{2 + \frac{4(1-\theta)}{4-3\theta}}}. \quad (4.71)$$

4. 假设  $\nu_0 \propto e^{-g-\psi}$ ，其中  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  是  $\alpha$ -强凸的， $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$  是  $L$ -Lipschitz 连续的 ( $\alpha > 0, 0 \leq L < \infty$ )，则对于所有  $t \geq 0$ ：

$$\text{FI}_\Phi(\mu_t \parallel \nu_t) \leq \frac{\text{FI}_\Phi(\mu_0 \parallel \nu_0)}{(1 + \alpha t)^2} \exp\left(\frac{2tL^2}{1 + \alpha t} + \frac{8L\sqrt{t}}{\sqrt{1 + \alpha t}}\right). \quad (4.72)$$

证明见第 4.5.9 小节。

**定理 4.12** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度，且  $\text{FI}_\Phi(\mu_0 \parallel \nu_0) < \infty$ ，其中  $\Phi$  是  $\theta$ -容许生成函数 ( $0 < \theta \leq 1$ )。设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 OU 流的输出，目标分布为  $\pi = \mathcal{N}(0, \gamma^{-1}I)$ ：

$$\partial_t u_t = \nabla \cdot (\gamma x u_t) + \Delta u_t. \quad (4.73)$$

则我们有：

1. 假设  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 则对于所有  $t \geq 0$ :

$$\text{FI}_\Phi(\mu_t \|\nu_t) \leq \frac{\gamma^2 e^{-2\gamma t}}{(\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} \text{FI}_\Phi(\mu_0 \|\nu_0). \quad (4.74)$$

2. 假设  $\nu_0$  满足  $\alpha$ -SLC 并假设  $\mu_0, \nu_0$  是对称的, 则对于所有  $t \geq 0$ :

$$\text{FI}_\Phi(\mu_t \|\nu_t) \leq \frac{\gamma^{2 + \frac{4(1-\theta)}{4-3\theta}} e^{-(2 + \frac{8(1-\theta)}{4-3\theta})\gamma t}}{(\alpha + e^{-2\gamma t}(\gamma - \alpha))^{2 + \frac{4(1-\theta)}{4-3\theta}}} \text{FI}_\Phi(\mu_0 \|\nu_0). \quad (4.75)$$

证明见第 4.5.10 小节。

**定理 4.13** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度, 且  $\text{FI}_\Phi(\mu_0 \|\nu_0) < \infty$ , 其中  $\Phi$  是  $\theta$ -容许生成函数 ( $0 < \theta \leq 1$ )。设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 FP 流 (3.3) 的输出:

$$\partial_t u_t = \nabla \cdot (u_t \nabla V) + \frac{c}{2} \Delta u_t. \quad (4.76)$$

如果假设 3.1 和 3.2 成立, 并且假设势函数  $V$  是  $\alpha$ -强凸的 (即  $\nabla^2 V \geq \alpha I > 0$ ), 则我们有:

1. 假设  $\nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 则对于所有  $t \geq 0$ :

$$\text{FI}_\Phi(\mu_t \|\nu_t) \leq \text{FI}_\Phi(\mu_0 \|\nu_0). \quad (4.77)$$

2. 假设  $\nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 并假设  $\mu_0, \nu_0, V$  是对称的, 则对于所有  $t \geq 0$ :

$$\text{FI}_\Phi(\mu_t \|\nu_t) \leq e^{-\frac{4(1-\theta)}{4-3\theta}\alpha t} \text{FI}_\Phi(\mu_0 \|\nu_0). \quad (4.78)$$

3. 如果  $\nu_0 = \nu \propto e^{-\frac{2}{c}V}$ , 则对于所有  $t \geq 0$ :

$$\text{FI}_\Phi(\mu_t \|\nu) \leq e^{-2\alpha t} \text{FI}_\Phi(\mu_0 \|\nu). \quad (4.79)$$

4. 假设  $\nu_0 = \nu \propto e^{-\frac{2}{c}V}$  并假设  $\mu_0, \nu$  是对称的, 则对于所有  $t \geq 0$ :

$$\text{FI}_\Phi(\mu_t \|\nu) \leq e^{-(2 + \frac{4(1-\theta)}{4-3\theta})\alpha t} \text{FI}_\Phi(\mu_0 \|\nu_0). \quad (4.80)$$

证明见第 4.5.11 小节。

### 4.3.3 相对 $\Phi$ -Fisher 信息下近端采样器的混合时间

在建立相对  $\Phi$ -Fisher 信息的收敛保证之前，我们首先回顾近端采样器 (Proximal Sampler) 的构造。

设  $\mathbb{R}^d$  上的目标分布  $\nu^X$  具有密度  $\nu^X(x) \propto e^{-g(x)}$ 。为了便于采样，引入辅助变量  $Y \in \mathbb{R}^d$  并定义扩展空间  $\mathbb{R}^{2d} = \mathbb{R}^d \times \mathbb{R}^d$  上的联合分布：

$$\nu^{XY}(x, y) \propto \exp\left(-g(x) - \frac{1}{2\eta}\|y - x\|^2\right), \quad \eta > 0. \quad (4.81)$$

根据构造，其  $x$ -边缘分布与  $\nu^X$  重合；因此，采样  $(X, Y) \sim \nu^{XY}$  并丢弃  $Y$  可获得目标分布的有效样本。

近端采样器 [130] 通过交替从该联合密度中进行条件采样来进行。在第  $k \geq 0$  次迭代中，给定当前状态  $x_k \sim \mu_k^X$ ，算法执行以下两个更新：

1. **前向步骤 (Forward step)**。在给定  $x_k$  的条件下抽取  $y_k$ ：  $y_k | x_k \sim \nu^{Y|X}(\cdot | x_k) = \mathcal{N}(x_k, \eta I)$ 。等价地，这相当于对  $x_k$  添加独立的 Gaussian 噪声。 $y_k$  的分布为  $\mu_k^Y = \mu_k^X * \mathcal{N}(0, \eta I)$ 。
2. **后向步骤 (Backward step)**。在给定  $y_k$  的条件下抽取  $x_{k+1}$ ：  $x_{k+1} | y_k \sim \nu^{X|Y}(\cdot | y_k)$ ，产生下一个迭代  $x_{k+1} \sim \mu_{k+1}^X$ 。

由于 Gibbs 采样是可逆的，联合分布  $\nu^{XY}$  是诱导马尔可夫链的平稳分布，确保方法是无偏的：如果  $\mu_k^X = \nu^X$ ，则  $\mu_{k+1}^X = \nu^X$ 。

前向更新计算简单，仅需添加 Gaussian 噪声，而具有挑战性的后向更新需要从以下条件分布中采样

$$\nu^{X|Y}(x | y_k) \propto \exp\left(-g(x) - \frac{1}{2\eta}\|x - y_k\|^2\right). \quad (4.82)$$

这种条件分布在 [130] 中被称为受限高斯预言机 (Restricted Gaussian Oracle, RGO)。当势函数  $g$  是对数平滑的 (Hessian 一致有界) 时，对于足够小的步长  $\eta$ ，可以通过拒绝采样实现近似 RGO，且期望评估  $g$  的次数为  $O(1)$ ；有关算法细节，请参见 [10, 130]。

通过 Fokker-Planck 动力学的视角，近端采样器的每次迭代都具有吸引人的信息论解释。具体而言，前向步骤对应于将当前分布  $\mu_k^X$  在 Gaussian (或热) 半群下演化  $\eta > 0$  时间，产生平滑分布

$$\mu_k^Y = \mu_k^X * \mathcal{N}(0, \eta I). \quad (4.83)$$

这一步可以视为将  $\mu_k^X$  传输通过 Gaussian 信道，或者等价地视为布朗噪声驱动的扩散过程的结果。

相比之下，后向步骤作为该扩散的逆过程运行。它可以理解为将  $\mu_k^Y$  通过时间反向 Gaussian 信道传输相同的时间  $\eta$ ，从而生成下一个迭代  $\mu_{k+1}^X$ 。从这个角度来看，后向步骤遵循反向 Fokker-Planck（或反向热）流，有效地消除了前向步骤引入的 Gaussian 扰动。当该反向步骤应用于平稳分布  $\nu^Y$  时，它精确地恢复了原始目标分布  $\nu^X$ ，证实了整体方案使  $\nu^X$  保持不变。

这种基于扩散的观点为采样算法和信息几何提供了自然的桥梁。通过调用前向和反向 Gaussian 信道的 SDPI，Chen 等人 [8] 在对数索伯列夫不等式 (LSI) 下建立了近端采样器在 KL 散度中的指数收敛。每个前向和后向更新都在 KL 散度中引入了严格的收缩，产生了迭代的几何衰减。随后，Wibisono [10] 将此分析扩展到相对 Fisher 信息，证明对于强对数凹目标测度，相同的信息论解释也保证了采样器在相对 Fisher 信息中的指数收敛。

**引理 4.1** [定理 3 [8]] 假设  $\nu^X$  满足  $\alpha$ -LSI ( $\alpha > 0$ )。沿着近端采样器的每次迭代（任意  $\eta > 0$ ），我们有：

$$\begin{aligned} \text{KL}(\mu_k^Y \parallel \nu^Y) &\leq \frac{\text{KL}(\mu_k^X \parallel \nu^X)}{1 + \alpha\eta}, \\ \text{KL}(\mu_{k+1}^X \parallel \nu^X) &\leq \frac{\text{KL}(\mu_k^Y \parallel \nu^Y)}{1 + \alpha\eta}. \end{aligned} \quad (4.84)$$

因此，对于任何  $k \geq 0$ ，KL 散度指数快速收敛：

$$\text{KL}(\mu_k^X \parallel \nu^X) \leq \frac{\text{KL}(\mu_0^X \parallel \nu^X)}{(1 + \alpha\eta)^{2k}}. \quad (4.85)$$

**引理 4.2** [定理 6 [10]] 假设  $\nu^X$  满足  $\alpha$ -SLC ( $\alpha > 0$ )。沿着近端采样器的每次迭代（任意  $\eta > 0$ ），我们有：

$$\begin{aligned} \text{FI}(\mu_k^Y \parallel \nu^Y) &\leq \frac{\text{FI}(\mu_k^X \parallel \nu^X)}{(1 + \alpha\eta)^2}, \\ \text{FI}(\mu_{k+1}^X \parallel \nu^X) &\leq \text{FI}(\mu_k^Y \parallel \nu^Y). \end{aligned} \quad (4.86)$$

因此，对于任何  $k \geq 0$ ，相对 Fisher 信息指数快速收敛：

$$\text{FI}(\mu_k^X \parallel \nu^X) \leq \frac{\text{FI}(\mu_0^X \parallel \nu^X)}{(1 + \alpha\eta)^{2k}}. \quad (4.87)$$

接下来，我们将上述收敛结果从经典的相对 Fisher 信息扩展到更广泛的相对  $\Phi$ -Fisher 信息族，其中  $\Phi$  是  $\theta$ -容许生成函数， $\theta \in (0, 1]$ 。

**定理 4.14** 假设  $\nu^X$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 且  $\Phi$  是  $\theta \in (0, 1]$  的  $\theta$ -容许生成函数。沿着近端采样器的每次迭代 (任意  $\eta > 0$ ), 我们有:

$$\begin{aligned} \text{FI}_\Phi(\mu_k^Y \|\nu^Y) &\leq \frac{\text{FI}_\Phi(\mu_k^X \|\nu^X)}{(1 + \alpha\eta)^2}, \\ \text{FI}_\Phi(\mu_{k+1}^X \|\nu^X) &\leq \text{FI}_\Phi(\mu_k^Y \|\nu^Y). \end{aligned} \quad (4.88)$$

因此, 对于任何  $k \geq 0$ , 相对 Fisher 信息指数快速收敛:

$$\text{FI}_\Phi(\mu_k^X \|\nu^X) \leq \frac{\text{FI}_\Phi(\mu_0^X \|\nu^X)}{(1 + \alpha\eta)^{2k}}. \quad (4.89)$$

如果进一步假设  $\mu_0^X, \nu^X$  是对称的, 则我们有:

$$\begin{aligned} \text{FI}_\Phi(\mu_k^Y \|\nu^Y) &\leq \frac{\text{FI}_\Phi(\mu_k^X \|\nu^X)}{(1 + \alpha\eta)^{2 + \frac{4(1-\theta)}{4-3\theta}}}, \\ \text{FI}_\Phi(\mu_{k+1}^X \|\nu^X) &\leq \text{FI}_\Phi(\mu_k^Y \|\nu^Y), \\ \text{FI}_\Phi(\mu_k^X \|\nu^X) &\leq \frac{\text{FI}_\Phi(\mu_0^X \|\nu^X)}{(1 + \alpha\eta)^{(2 + \frac{4(1-\theta)}{4-3\theta})k}}. \end{aligned} \quad (4.90)$$

证明见第 4.5.12 小节。

为 RGO 配备拒绝采样, 定理 4.14 直接转化为近端采样器在相对 Fisher 信息中的高精度迭代复杂度保证。

**推论 4.2** 假设  $\nu^X \propto e^{-g}$  满足  $\alpha$ -SLC 且  $L$ -对数平滑的 ( $0 < \alpha \leq L < \infty$ ),  $\Phi$  是  $\theta$  ( $0 < \theta \leq 1$ )-容许生成函数。设  $x^* = \operatorname{argmin}_{x \in \mathbb{R}^d} g(x)$  为  $g$  的极小值点。考虑具有步长  $\eta = \frac{1}{Ld}$  和拒绝采样实现 RGO 的近端采样器。假设算法从  $x_0 \sim \mu_0^X = \mathcal{N}(x^*, \frac{1}{L}I)$  开始。则对于任何  $\varepsilon > 0$ , 沿着近端采样器的迭代  $x_k \sim \mu_k^X$ , 只要迭代次数  $k$  满足下式, 就有  $\text{FI}_\Phi(\mu_k^X \|\nu^X) \leq \varepsilon$ :

$$k \geq \frac{dL}{\alpha} \log \frac{CdL}{\varepsilon} \geq \frac{dL}{\alpha} \log \frac{\text{FI}_\Phi(\mu_0^X \|\nu^X)}{\varepsilon}, \quad (4.91)$$

其中  $C > 0$  为常数, 且每次迭代中用于拒绝采样的  $g$  的期望查询次数为  $O(1)$ 。如果进一步假设  $x^* = 0$  且  $\nu^X$  是对称的, 则有

$$k \geq \frac{dL}{(1 + \frac{2(1-\theta)}{4-3\theta})\alpha} \log \frac{CdL}{\varepsilon} \geq \frac{dL}{(1 + \frac{2(1-\theta)}{4-3\theta})\alpha} \log \frac{\text{FI}_\Phi(\mu_0^X \|\nu^X)}{\varepsilon}. \quad (4.92)$$

证明见第 4.5.13 小节。

## 4.4 对数 Hessian 公式与对数 Hessian 比较原理

在本节中，我们推导了与一般扩散半群  $(P_t)_{t \geq 0}$  作用于初始数据  $f$  相关的  $\log P_t f$  空间 Hessian 的显式表示。将此表示专门用于 Fokker-Planck 演化，得出了演化密度的对数 Hessian 公式，从中可以得出对数 Hessian 的最终（大时间）下界。此外，通过抛物型方程的对数凹性保持，足够“温和”的初始数据（即相对于势函数  $V$  足够对数凹）意味着对数 Hessian 的上界。结合这两种估计产生了对数 Hessian 比较原理，这对于验证相对 Rényi-Fisher 信息 SDPI 所需的结构假设特别方便。

作为补充结果，假设在  $t = 0$  时比较成立，我们建立了高斯初始化下热流和 Ornstein-Uhlenbeck (OU) 流的永久（全时）比较原理。尽管我们使用对数 Hessian 公式的主要是为了分析 SDPI，但以下陈述在 PDE 和随机分析中具有独立的意义。作为热身，我们从 FP 流的最简单情况，即热流开始。

### 4.4.1 热流的对数 Hessian 比较原理

始终设  $X_t = X_0 + \sqrt{t}Z$ ,  $Z \sim \mathcal{N}(0, I)$  表示热流，其中  $Z$  独立于  $X_0$ 。设  $X_0 \sim \mu_0$ ,  $X_t \sim \mu_t$ ,  $(X_0, X_t) \sim \mu_{0t}$ ，我们将联合分布和条件分布写为

$$\mu_0(x)\mu_{t|0}(y|x) = \mu_{0t}(x, y) = \mu_t(y)\mu_{0|t}(x|y). \quad (4.93)$$

则对数 Hessian 具有以下表示。

**引理 4.3** [引理 8 [5]] 沿热流，对于所有  $y \in \mathbb{R}^d$ ，我们有

$$-\nabla_y^2 \log \mu_t(y) = \frac{1}{t} \left( I - \frac{1}{t} \text{Cov}(\mu_{0|t}(\cdot|y)) \right). \quad (4.94)$$

因此，

$$\nabla^2 \log \mu_t(y) = -\frac{1}{t} I + \frac{1}{t^2} \text{Cov}(\mu_{0|t}(\cdot|y)) \succeq -\frac{1}{t} I, \quad (4.95)$$

这给出了对数 Hessian 的下界。

对于上界，假设初始分布  $\mu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ )，即

$$\nabla^2 \log \mu_0 \preceq -\alpha I. \quad (4.96)$$

由于  $\mu_t = \mu_0 * \mathcal{N}(0, tI)$ ，强对数凹参数退化， $\mu_t$  是  $\frac{\alpha}{1+\alpha t}$ -SLC，因此

$$\nabla^2 \log \mu_t \preceq -\frac{\alpha}{1+\alpha t} I. \quad (4.97)$$

设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度，且  $\mu_t, \nu_t$  表示它们在 Gaussian 信道（热流）下的像。我们旨在建立如下形式的比较

$$\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t. \quad (4.98)$$

利用上述界限，只需

$$-\frac{\alpha}{1+\alpha t} I \preceq -\frac{K}{t} I. \quad (4.99)$$

这等价于

$$K + K\alpha t \leq \alpha t. \quad (4.100)$$

因此，对于任何  $K \in [0, 1)$  和  $t \geq \frac{K}{\alpha(1-K)}$ ，对数 Hessian 比较原理对热流最终成立，即当  $t \geq \frac{K}{\alpha(1-K)}$  时。

**定理 4.15** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度。设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 Gaussian 信道的输出：

$$\partial_t u_t = \frac{1}{2} \Delta u_t. \quad (4.101)$$

如果  $\mu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ )，则对于任何  $K \in [0, 1)$ ，对数 Hessian 比较原理最终成立：

$$\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t, \quad (4.102)$$

只要  $t \geq \frac{K}{\alpha(1-K)}$ 。

尽管如此，上述对数 Hessian 比较原理仅在足够大的时间成立，即最终比较原理。作为补充结果，我们接下来记录高斯初始化下的永久保持性声明：如果  $\mu_0$  或  $\nu_0$  是高斯的，且  $\nabla^2 \log \nu_0 \preceq K \nabla^2 \log \mu_0$  ( $K \in [0, 1]$ )，则该不等式在热流下对所有  $t > 0$  持续存在。根据引理 4.3 中的对数 Hessian 公式，我们有

$$-\nabla_y^2 \log \mu_t(y) = \frac{1}{t} \left( I - \frac{1}{t} \text{Cov}(\mu_{0|t}(\cdot|y)) \right), \quad -\nabla_y^2 \log \nu_t(y) = \frac{1}{t} \left( I - \frac{1}{t} \text{Cov}(\nu_{0|t}(\cdot|y)) \right). \quad (4.103)$$

因此，对数 Hessian 比较原理  $\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t$  等价于

$$\text{Cov}(\nu_{0|t}(\cdot|y)) \preceq K \text{Cov}(\mu_{0|t}(\cdot|y)) + (1-K)tI. \quad (4.104)$$

记  $\nu_0 \propto e^{-V_{\nu_0}}, \mu_0 \propto e^{-V_{\mu_0}}$ 。则初始比较  $\nabla^2 \log \nu_0 \preceq K \nabla^2 \log \mu_0$  等价于  $K \nabla^2 V_{\mu_0} \preceq \nabla^2 V_{\nu_0}$ 。此外，贝叶斯法则结合 Gaussian 似然意味着

$$\nu_{0|t}(x_0|x_t) = \frac{\nu_{t|0}(x_t|x_0)\nu_0(x_0)}{\int \nu_{t|0}(x_t|x_0)\nu_0(x_0)dx_0} = \frac{\exp(-\frac{\|x_0-x_t\|^2}{2t})\nu_0(x_0)}{\int \nu_{t|0}(x_t|x_0)\nu_0(x_0)dx_0}, \quad (4.105)$$

即,

$$\nu_{0|t}(dx_0|x_t) \propto \nu_0(x_0) \cdot \exp\left(-\frac{\|x_0 - x_t\|^2}{2t}\right) dx_0 \propto \exp\{-W_{\nu,t}(x_0)\} dx_0, \quad (4.106)$$

其中

$$W_{\mu,t}(x_0) = V_{\mu_0}(x_0) + \frac{\|x_0\|^2}{2t} - \frac{\langle x_0, x_t \rangle}{t}, W_{\nu,t}(x_0) = V_{\nu_0}(x_0) + \frac{\|x_0\|^2}{2t} - \frac{\langle x_0, x_t \rangle}{t}. \quad (4.107)$$

因此

$$\nabla_{x_0}^2 W_{\nu,t}(x_0) - K \nabla_{x_0}^2 W_{\mu,t}(x_0) = \nabla_{x_0}^2 V_{\nu_0}(x_0) - K \nabla_{x_0}^2 V_{\mu_0}(x_0) + \frac{1-K}{t} I \succeq \frac{1-K}{t} I. \quad (4.108)$$

为了建立  $\text{Cov}(\nu_{0|t}(\cdot|x_t)) \preceq K \text{Cov}(\mu_{0|t}(\cdot|x_t)) + (1-K)tI$ , 只需验证二次型的相应不等式。即, 对于每个方向  $u \in \mathbb{R}^d$  ( $\|u\| = 1$ ) 和每个  $x_t \in \mathbb{R}^d$ ,

$$u^\top \text{Cov}(\nu_{0|t}(\cdot|x_t))u \leq K u^\top \text{Cov}(\mu_{0|t}(\cdot|x_t))u + (1-K)t, \forall x_t \in \mathbb{R}^d. \quad (4.109)$$

由于对称矩阵之间的序等价于所有 Rayleigh 商的支配, 因此证明上述标量不等式对所有  $u$  成立等价于所需的矩阵不等式。回顾对于任意服从  $\mu$  分布的随机变量  $Y$  和任意向量  $u \in \mathbb{R}^d$ , 我们有:

$$\begin{aligned} \text{Var}_\mu(u^\top Y) &= \mathbb{E}[u^\top Y u^\top Y] - (\mathbb{E}[u^\top Y])^2 = u^\top \mathbb{E}[Y Y^\top] u - (u^\top \mathbb{E}[Y])^2 \\ &= u^\top (\mathbb{E}[Y Y^\top] - \mathbb{E}[Y] \mathbb{E}[Y]^\top) u = u^\top \text{Cov}(Y) u. \end{aligned} \quad (4.110)$$

将此恒等式应用于  $Y \sim \nu_{0|t}(\cdot|x_t)$  和  $Y \sim \mu_{0|t}(\cdot|x_t)$  正好得出协方差比较的标量形式。在我们的证明中, 我们利用以下关键不等式将曲率界转移为协方差界, 从而在所述假设下产生永久比较。

**引理 4.4** [引理 7 [129]] 设  $P = \exp(-V)$  为  $\mathbb{R}^d$  上的概率测度, 且假设  $V$  在其定义域内部二阶连续可微。则以下成立:

1. (Brascamp-Lieb 不等式) 如果进一步假设  $P$  是严格对数凹的, 则有

$$\text{Cov}_{X \sim P}(X) \preceq \mathbb{E}_{X \sim P}[(\nabla^2 V(X))^{-1}]. \quad (4.111)$$

2. (Cramér-Rao 不等式)

$$\text{Cov}_{X \sim P}(X) \succeq \left( \mathbb{E}_{X \sim P}[\nabla^2 V(X)] \right)^{-1}. \quad (4.112)$$

如果假设  $\mu_0$  是高斯分布且  $\nu_0$  是严格对数凹的, 则  $\mu_t$  对所有  $t \geq 0$  都是高斯分布。由引理 4.4 中的 Brascamp-Lieb 不等式 (BLI) 并取随机变量  $X \sim \mu_{0|t}, Y \sim \nu_{0|t}$ , 我们有

$$\begin{aligned} \text{Var}_{\nu_{0|t}}(u^\top Y) &\leq \int u^\top (\nabla^2 W_{\nu,t}(y))^{-1} u \, d\nu_{0|t} \\ &= \int u^\top (\nabla^2 W_{\nu,t}(T_{\mu \rightarrow \nu}(x)))^{-1} u \, d\mu_{0|t} \\ &\leq \int u^\top \left( K \nabla^2 W_{\mu,t}(T_{\mu \rightarrow \nu}(x)) + \frac{1-K}{t} I \right)^{-1} u \, d\mu_{0|t}, \end{aligned} \quad (4.113)$$

其中我们取  $T_{\mu \rightarrow \nu}$  为将  $\mu_{0|t}$  推送至  $\nu_{0|t}$  的传输映射, 满足

$$\nu_{0|t} = (T_{\mu \rightarrow \nu})\# \mu_{0|t}, \quad \mu_{0|t} - a.e. \quad (4.114)$$

**注** Brenier 定理 [17] 保证了这种映射的存在性。此外, 如果我们在初始比较  $\nabla^2 \log \nu_0 \preceq K \nabla^2 \log \mu_0$  中设  $K = 1$ , 则  $\nabla^2 W_{\nu,t}(y) \succeq \nabla^2 W_{\mu,t}(y)$ , 我们可以通过 Caffarelli 收缩定理 [131-132] 得到  $T_{\mu \rightarrow \nu}$  的收缩性, 即  $\|T_{\mu \rightarrow \nu}\|_{\text{op}} \leq 1$ , 其中  $\|\cdot\|_{\text{op}}$  是算子范数。

此外, 利用对于任意两个正定矩阵  $A, B$  和  $\lambda \in [0, 1]$ , 有  $(\lambda A + (1-\lambda)B)^{-1} \preceq \lambda A^{-1} + (1-\lambda)B^{-1}$  的事实。则有

$$\begin{aligned} \text{Var}_{\nu_{0|t}}(u^\top Y) &\leq \int u^\top \left( K \nabla^2 W_{\mu,t}(T_{\mu \rightarrow \nu}(x)) + \frac{1-K}{t} I \right)^{-1} u \, d\mu_{0|t} \\ &\leq \int u^\top \left( K (\nabla^2 W_{\mu,t}(T_{\mu \rightarrow \nu}(x)))^{-1} + (1-K)tI \right) u \, d\mu_{0|t} \\ &= K \int u^\top (\nabla^2 W_{\mu,t}(T_{\mu \rightarrow \nu}(x)))^{-1} u \, d\mu_{0|t} + (1-K)t \|u\|^2 \\ &= K \text{Var}_{\mu_{0|t}}(u^\top X) + (1-K)t, \quad \text{因为 } W_{\mu_t} \text{ 是二次的, 且 } \nabla^2 W_{\mu,t} \text{ 是常数,} \end{aligned} \quad (4.115)$$

这导出了所需的协方差比较的标量形式。

如果假设  $\nu_0$  是高斯分布且  $\mu_0$  是严格对数凹的, 则  $\nu_t$  对所有  $t \geq 0$  都是高斯分布。在此情形下,

$$K^{-1} \nabla^2 W_{\nu,t} - \frac{K^{-1} - 1}{t} I \succeq \nabla^2 W_{\mu,t} \succ 0. \quad (4.116)$$

应用 Cramér-Rao 不等式 (CRI) 于条件分布  $X \sim \mu_{0|t}, Y \sim \nu_{0|t}$ , 对于每个  $u \in \mathbb{R}^d$

( $\|u\| = 1$ ), 有

$$\begin{aligned}
\text{Var}_{\mu_{0|t}}(u^\top X) &\geq \left( \int u^\top \nabla^2 W_{\mu,t}(x) u \, d\mu_{0|t} \right)^{-1} \\
&\geq \left( \int u^\top \left( K^{-1} \nabla^2 W_{\nu,t}(x) - \frac{K^{-1} - 1}{t} I \right) u \, d\mu_{0|t} \right)^{-1} \\
&= \left( K^{-1} \int u^\top \nabla^2 W_{\nu,t}(x) u \, d\mu_{0|t} - \frac{K^{-1} - 1}{t} \right)^{-1} \quad (4.117) \\
&= K \left( \int u^\top \nabla^2 W_{\nu,t}(T_{\nu \rightarrow \mu}(y)) u \, d\nu_{0|t} - \frac{1 - K}{t} \right)^{-1} \\
&= K \left( \text{Var}_{\nu_{0|t}}(u^\top Y)^{-1} - \frac{1 - K}{t} \right)^{-1} > 0.
\end{aligned}$$

两边取倒数得:

$$K \text{Var}_{\mu_{0|t}}(u^\top X)^{-1} + \frac{1 - K}{t} \leq \text{Var}_{\nu_{0|t}}(u^\top Y)^{-1}. \quad (4.118)$$

因此,

$$\begin{aligned}
\text{Var}_{\nu_{0|t}}(u^\top Y) &\leq \left( K \text{Var}_{\mu_{0|t}}(u^\top X)^{-1} + \frac{1 - K}{t} \right)^{-1} \\
&\leq K \text{Var}_{\mu_{0|t}}(u^\top X) + (1 - K)t \quad (4.119) \\
&= K u^\top \text{Cov}(\mu_{0|t}(\cdot|x_t)) u + (1 - K)t,
\end{aligned}$$

其中最后一步利用了逆映射的算子凸性 (取  $A = \text{Var}_{\mu_{0|t}}(u^\top X)$  和  $B = t$ )。这就建立了所述假设下热流的永久比较原理。

**定理 4.16** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度。设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 Gaussian 信道的输出:

$$\partial_t u_t = \frac{1}{2} \Delta u_t. \quad (4.120)$$

如果  $\mu_0$  是 Gaussian (或  $\nu_0$  是 Gaussian) 且  $\nu_0$  (或  $\mu_0$ ) 是严格对数凹的, 并且如果对于某个  $K \in [0, 1]$ , 初始比较成立:

$$\nabla^2 \log \nu_0 \preceq K \nabla^2 \log \mu_0, \quad (4.121)$$

则对数 Hessian 比较原理全时成立:

$$\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t, \forall t \geq 0. \quad (4.122)$$

**注** 定理 4.16 对于  $K > 1$  不成立。例如, 设  $\mu_0 = \mathcal{N}(0, \sigma_u^2), \nu_0 = \mathcal{N}(0, \sigma_v^2)$ 。则初始比较  $\nabla^2 \log \nu_0 \preceq K \nabla^2 \log \mu_0$  等价于  $-\sigma_v^{-2} \leq -K \sigma_u^{-2}$ , 即

$$K \sigma_v^2 \leq \sigma_u^2. \quad (4.123)$$

由于  $\mu_t \sim \mathcal{N}(0, \sigma_u^2 + t), \nu_t \sim \mathcal{N}(0, \sigma_v^2 + t)$ , 对数 Hessian 比较  $\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t$  等价于

$$-(\sigma_v^2 + t)^{-1} \leq -K(\sigma_u^2 + t)^{-1}, \quad (4.124)$$

即  $(K - 1)t \leq \sigma_u^2 - K\sigma_v^2$ , 这在  $K > 1$  且  $t > \frac{\sigma_u^2 - K\sigma_v^2}{K-1}$  时违反了比较原理。

基于上述结果, 我们给出了一种在实践中验证对数 Hessian 比较原理的具体方法, 然后结合定理 4.7, 将其应用于建立沿热流的相对 Rényi-Fisher 信息的 SDPI, 如下所示。

**推论 4.3** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度, 且  $\text{RFI}_q(\mu_0 \| \nu_0) < \infty, q \geq 1$ 。设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 Gaussian 信道的输出:

$$\partial_t u_t = \frac{1}{2} \Delta u_t. \quad (4.125)$$

则我们有:

1. 假设  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 则对于任何  $q \in [1, +\infty)$  和  $\delta \in (0, 1)$ , 存在足够大的时间  $T > 0$ , 使得对于所有  $t \geq T$  成立:

$$\text{RFI}_q(\mu_t \| \nu_t) \leq O((1 + \alpha t)^{-(2-\delta)}). \quad (4.126)$$

2. 假设  $\nabla^2 \log \nu_0 \preceq \nabla^2 \log \mu_0$ ,  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 且  $\mu_0$  是 Gaussian, 则对于所有  $t \geq 0$ :

$$\text{RFI}_q(\mu_t \| \nu_t) \leq \frac{\text{RFI}_q(\mu_0 \| \nu_0)}{(1 + \alpha t)^2}, \quad \forall q \in [1, \infty). \quad (4.127)$$

**证明 第 (1) 部分:** 根据定理 4.15, 对于任意  $K \in [0, 1)$ , 存在足够大的时间  $T > 0$ , 使得对于所有  $t \geq T$ , 均有:

$$\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t. \quad (4.128)$$

进而依据定理 4.7, 对于给定的  $q \in [1, +\infty)$  以及满足  $\epsilon \leq q^{-1}$  的  $K = 1 - \epsilon \in [0, 1)$ , 存在常数  $C > 0$  使得对任意  $t \geq T$  成立:

$$\begin{aligned} \text{RFI}_q(\mu_t \| \nu_t) &\leq \frac{C}{(1 + \alpha t)^{2(1+(K-1)q)/K}} \\ &= \frac{C}{(1 + \alpha t)^{\frac{2(1-\epsilon q)}{1-\epsilon}}} \\ &\equiv \frac{C}{(1 + \alpha t)^{2-\delta}}, \end{aligned} \quad (4.129)$$

其中  $\delta \in (0, 1)$ ，且当  $\epsilon \rightarrow 0$  时  $\delta$  收敛至 0。

**第 (2) 部分：** 根据定理 4.16，对于所有  $t \geq 0$ ，均有：

$$\nabla^2 \log \nu_t \preceq \nabla^2 \log \mu_t. \quad (4.130)$$

类似地，依据定理 4.7，对于所有  $t \geq 0$  及任意  $q \in [1, \infty)$ ，成立：

$$\text{RFI}_q(\mu_t \| \nu_t) \leq \frac{\text{RFI}_q(\mu_0 \| \nu_0)}{(1 + \alpha t)^2}. \quad (4.131)$$

■

#### 4.4.2 OU 流的对数 Hessian 比较原理

我们考虑具有二次势函数  $V(x) = \frac{1}{2}x^\top Qx + q^\top x + c$  的 Ornstein–Uhlenbeck (OU) 动力学，其中  $Q \succ 0$ 。SDE 形式为：

$$dX_t = -QX_t dt + dW_t, \quad X_0 \sim \mu_0, \quad (4.132)$$

其温和解（或弱解）为

$$X_t = e^{-Qt} X_0 + \int_0^t e^{-Q(t-s)} dW_s = e^{-Qt} X_0 + Z_t. \quad (4.133)$$

这里  $e^{At} = \sum_{k \geq 0} (At)^k / k!$  是矩阵指数， $Z_t \sim \mathcal{N}(0, \Sigma_t)$  独立于  $X_0$ ，且

$$\Sigma_t = \int_0^t e^{-Q(t-s)} e^{-Q^\top(t-s)} ds, \quad Q\Sigma_t + \Sigma_t Q^\top = I - e^{-Qt} e^{-Q^\top t}. \quad (4.134)$$

第二个恒等式通过对  $F(s) = e^{-Q(t-s)} e^{-Q^\top(t-s)}$  求导并在  $s \in [0, t]$  上积分得到。注意，我们可以用  $\tau$  替换  $t - s$  作为新变量，写为

$$\Sigma_t = \int_0^t e^{-Q\tau} e^{-Q^\top \tau} d\tau, \quad (4.135)$$

因此  $\Sigma_t$  随  $t$  增加，并在  $t \rightarrow \infty$  时收敛到 Lyapunov 方程  $Q\Sigma_\infty + \Sigma_\infty Q^\top = I$  的解。根据 Lyapunov 稳定性定理 [133]，如果复矩阵  $Q$  是正稳定的（即每个特征值都有严格正的实部），则 Lyapunov 方程

$$QX + XQ^\top = I \quad (4.136)$$

具有唯一的正定解  $X \succ 0$ 。

记  $\mu_t = \text{Law}(X_t)$ 。如热流情况，将联合分布和条件分布写为

$$\mu_{0t}(x, y) = \mu_0(x) \mu_{t|0}(y | x) = \mu_t(y) \mu_{0|t}(x | y), \quad (4.137)$$

其中 OU 转移是高斯的：

$$\mu_{t|0}(y | x_0) = \mathcal{N}(y; e^{-Qt}x_0, \Sigma_t). \quad (4.138)$$

事实上，[6] 已经在  $Q = \alpha I, \alpha > 0$  的情况下推导了 OU 流的对数 Hessian 公式。这里我们将其扩展到一般的  $Q \succ 0$ 。

**引理 4.5** (OU 流的对数 Hessian 公式) 设  $Q \in \mathbb{R}^{d \times d}$  并考虑 OU 流

$$dX_t = -QX_t dt + dW_t, \quad X_0 \sim \mu_0. \quad (4.139)$$

设  $\Sigma_t := \int_0^t e^{-Q(t-s)} e^{-Q^\top(t-s)} ds$ ，并注意  $Z_t := \int_0^t e^{-Q(t-s)} dW_s \sim \mathcal{N}(0, \Sigma_t)$ 。则对于所有  $t > 0$ ，边缘分布  $\mu_t = \text{Law}(X_t)$  具有  $C^\infty$  密度且二阶矩有限，并且对于每个  $y \in \mathbb{R}^d$ ，

$$\nabla_y^2 \log \mu_t(y) = -\Sigma_t^{-1} + \Sigma_t^{-1} e^{-Qt} \text{Cov}(\mu_{0|t}(\cdot | y)) e^{-Q^\top t} \Sigma_t^{-1}. \quad (4.140)$$

此处  $\mu_{0|t}(\cdot | y)$  表示给定  $X_t = y$  时  $X_0$  的后验（条件）分布。

证明见第 4.5.14 小节。

为了确保引理 4.5 中微分与积分交换次序的合理性，我们需要以下关于 OU 核导数的一致控制以及积分号下求导的论证。

**引理 4.6** (OU 核导数的一致控制) 固定  $t > 0$  并设  $\Sigma_t = \int_0^t e^{-Q(t-s)} e^{-Q^\top(t-s)} ds$ 。定义 OU 转移核

$$k(y, x_0) := (2\pi)^{-d/2} (\det \Sigma_t)^{-1/2} \exp\left(-\frac{1}{2} \|\Sigma_t^{-1/2}(y - e^{-Qt}x_0)\|^2\right). \quad (4.141)$$

设  $K \subset \mathbb{R}^d$  为紧集。则存在常数  $c_t > 0$  和  $C_{t,K} < \infty$ ，使得对于所有  $y \in K, x_0 \in \mathbb{R}^d$ ，以及所有模长  $|\alpha| \leq 2$  的多重指标  $\alpha$ ，有

$$|\partial_y^\alpha k(y, x_0)| \leq C_{t,K} (1 + \|x_0\|^2) \exp\left(-c_t \|x_0\|^2\right). \quad (4.142)$$

**证明** 记  $A_t := \Sigma_t^{-1/2} e^{-Qt}$  和  $z := \Sigma_t^{-1/2} y$ 。则

$$k(y, x_0) = C_t \exp\left(-\frac{1}{2} \|z - A_t x_0\|^2\right), \quad (4.143)$$

其中  $C_t = (2\pi)^{-d/2} (\det \Sigma_t)^{-1/2}$ 。直接计算可得

$$\nabla_y \log k = -\Sigma_t^{-1}(y - e^{-Qt}x_0), \quad \nabla_y^2 \log k = -\Sigma_t^{-1}. \quad (4.144)$$

因此, 每一阶数  $|\alpha| \leq 2$  的  $y$  导数都具有形式  $\partial_y^\alpha k = k \cdot P_\alpha(y - e^{-Qt}x_0)$ , 其中  $P_\alpha$  是系数仅依赖于  $\Sigma_t^{-1}$  的次数不高于 2 的多项式。对于  $y \in K$ , 有  $\|P_\alpha(y - e^{-Qt}x_0)\| \leq C_{t,K}(1 + \|x_0\|^2)$ 。

对于高斯因子, 利用奇异值界  $\|A_t x_0\| \geq s_{\min}(A_t)\|x_0\|$  和  $\|z\| \leq \sup_{y \in K} \|\Sigma_t^{-1/2}y\| < \infty$ , 我们有

$$\|z - A_t x_0\|^2 \geq s_{\min}(A_t)^2 \|x_0\|^2 - 2\|A_t\| \|z\| \|x_0\| \geq \frac{1}{2}s_{\min}(A_t)^2 \|x_0\|^2 - C_{t,K}, \quad (4.145)$$

从而对于某些  $c_t > 0, C'_{t,K} < \infty$ , 有  $k(y, x_0) \leq C'_{t,K} \exp(-c_t \|x_0\|^2)$ 。这里  $s_{\min}(A_t) = \frac{1}{\|A_t^{-1}\|_2} > 0$  表示  $A_t$  的最小奇异值。结合这两个估计即可得出 (4.142)。 ■

**推论 4.4** (OU 边缘分布的积分号下求导) 设  $\mu_0$  为  $\mathbb{R}^d$  上具有有限二阶矩的概率密度:  $\int_{\mathbb{R}^d} \|x\|^2 \mu_0(x) dx < \infty$ 。对于每个固定的  $t > 0$ , OU 边缘分布  $\mu_t(y) = \int k(y, x_0) \mu_0(x_0) dx_0$  关于  $y$  是  $C^\infty$  的, 且对于  $|\alpha| \leq 2$ , 有

$$\partial_y^\alpha \mu_t(y) = \int \partial_y^\alpha k(y, x_0) \mu_0(x_0) dx_0, \quad \text{对所有 } y \in \mathbb{R}^d. \quad (4.146)$$

**证明** 固定紧集  $K \subset \mathbb{R}^d$ 。由引理 4.6,

$$\sup_{y \in K} \left| \partial_y^\alpha k(y, x_0) \right| \mu_0(x_0) \leq C_{t,K}(1 + \|x_0\|^2) e^{-c_t \|x_0\|^2} \mu_0(x_0), \quad (4.147)$$

其关于  $x_0$  的积分是有限的, 因为  $(1 + \|x_0\|^2) \mu_0(x_0) \in \mathbb{L}^1(\mathbb{R}^d)$  且  $e^{-c_t \|x_0\|^2} \leq 1$ 。由控制收敛定理可知, 微分运算可以在  $K$  上一致地移入积分号内; 由于  $K$  的任意性, 该恒等式对所有  $y$  成立。 ■

从 (4.140) 我们立即获得通用下界

$$\nabla_y^2 \log \mu_t(y) \succeq -\Sigma_t^{-1}, \quad (4.148)$$

因为 (4.140) 中的第二项是半正定的。

假设  $\mu_0$  满足  $\alpha$ -强对数凹 ( $\alpha$ -SLC), 即  $\nabla^2 \log \mu_0 \preceq -\alpha I$ , 则应用可逆线性变换  $T$ , Hessian 界相应缩放:

$$\nabla^2 \log(T_{\#} \mu_0) \preceq -\alpha T^{-\top} T^{-1}. \quad (4.149)$$

结合线性推送  $x \mapsto e^{-Qt}x$  和高斯卷积  $\mathcal{N}(0, \Sigma_t)$ , 得到矩阵并行和 (Parallel-sum) 上界

$$\nabla_y^2 \log \mu_t(y) \preceq -\left( \Sigma_t + \frac{1}{\alpha} e^{-Qt} e^{-Q^\top t} \right)^{-1}, \quad t > 0. \quad (4.150)$$

此估计依赖于两个标准要素: 可逆线性变换下 SLC 的保持性以及独立随机变量卷积下 SLC 的并行和规则。为完整起见, 我们在此记录这两个事实。

**引理 4.7** (可逆线性映射下的 SLC) 设  $\mu$  是  $A$ -SLC 的 (矩阵形式), 即其密度  $p = e^{-\varphi}$  满足

$$\nabla^2 \varphi(x) \succeq A \quad \text{对所有 } x, \text{ 其中 } A \text{ 正定.} \quad (4.151)$$

设  $T \in \mathbb{R}^{d \times d}$  可逆,  $\nu := T_{\#} \mu$  为推送测度. 则  $\nu$  是  $T^{-\top} A T^{-1}$ -SLC 的:

$$\nabla^2 (-\log \nu)(y) \succeq T^{-\top} A T^{-1}, \quad y = Tx. \quad (4.152)$$

**引理 4.8** (卷积下的 SLC 并行和规则) 设  $X, Y$  是独立的  $\mathbb{R}^d$  值随机向量, 密度分别为  $p_X = e^{-\varphi}, p_Y = e^{-\psi}$ . 假设对于某些正定矩阵  $A, B$ , 有  $\nabla^2 \varphi \succeq A$  和  $\nabla^2 \psi \succeq B$ . 则  $Z = X + Y$  是  $A:B$ -SLC 的, 其中

$$A:B := (A^{-1} + B^{-1})^{-1}. \quad (4.153)$$

等价地,

$$\nabla^2 (-\log p_Z)(z) \succeq (A^{-1} + B^{-1})^{-1}, \quad \forall z \in \mathbb{R}^d. \quad (4.154)$$

这两个陈述都是经典的; 有关全面处理和相关扩展, 请参见 [37] 中的定理 3.7. 对于  $Q = 0$  (热流),  $\Sigma_t = tI$ , (4.150) 简化为  $\nabla^2 \log \mu_t \preceq -\frac{\alpha}{1+\alpha t} I$ , 这恢复了与  $\mathcal{N}(0, tI)$  卷积下 SLC 曲率的经典并行和组合.

设  $\mu_0, \nu_0$  为概率测度,  $\mu_t, \nu_t$  为它们在相同  $Q$  下的 OU 演化. 我们寻求如下形式的矩阵比较不等式

$$\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t - K_2 \Sigma_{\infty}^{-1}. \quad (4.155)$$

通过 (4.148) 对  $\mu_t$  和 (4.150) 对  $\nu_t$  (假设  $\nu_0$  满足  $\alpha$ -SLC), 只需

$$-\left(\Sigma_t + \frac{1}{\alpha} e^{-Qt} e^{-Q^{\top}t}\right)^{-1} \preceq -K_1 \Sigma_t^{-1} - K_2 \Sigma_{\infty}^{-1}, \quad (4.156)$$

这等价于

$$\left(I + \frac{1}{\alpha} \Sigma_t^{-1/2} e^{-Qt} e^{-Q^{\top}t} \Sigma_t^{-1/2}\right)^{-1} \succeq K_1 I + K_2 \Sigma_t^{1/2} \Sigma_{\infty}^{-1} \Sigma_t^{1/2}. \quad (4.157)$$

由于当  $t \rightarrow \infty$  时  $e^{-Qt} e^{-Q^{\top}t} \rightarrow 0$ , 而  $\Sigma_t$  增加到唯一的  $\Sigma_{\infty} \succ 0$  (解  $Q \Sigma_{\infty} + \Sigma_{\infty} Q^{\top} = I$ ), 如果

$$1 - K_1 - K_2 > 0, \quad (4.158)$$

则条件 (4.157) 对所有足够大的  $t$  成立. 特别地, 在  $K_2 = 0$  的情况下, 当  $K_1 < 1$  时条件成立. 因此我们得到:

**定理 4.17** (OU 流的最终对数 Hessian 比较) 设  $Q \succ 0$ ,  $\mu_t, \nu_t$  为具有漂移  $Q$  的 OU 演化。如果  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ ), 则对于满足  $1 - K_1 - K_2 > 0$  的任意  $K_1, K_2$ , 其中  $Q\Sigma_\infty + \Sigma_\infty Q^\top = I$ , 存在  $T = T(Q, \alpha, K_1, K_2, \mu_0, \nu_0)$  使得

$$\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t - K_2 \Sigma_\infty^{-1}, \quad \text{对所有 } t \geq T. \quad (4.159)$$

定理 4.17 中的大时间比较仅利用了先验的上下界, 因此产生了由 Lyapunov 极限  $\Sigma_\infty$  驱动的最终陈述。当其中一个初始分布是高斯时, 可以证明更强的永久 (全时) 比较: 在这种设置下, 我们可以通过 Brascamp-Lieb 和 Cramér-Rao 不等式从曲率界过渡到后验协方差界, 然后将这些界代回 OU 对数 Hessian 公式 (4.140)。下一个定理记录了这一永久现象, 其基本形式类似于热流情况。

**定理 4.18** (OU 流的永久对数 Hessian 比较) 设  $Q$  正定并考虑 OU 流  $dX_t = -QX_t dt + dW_t$ 。设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  演化的分布。假设  $\mu_0, \nu_0$  中有一个是 Gaussian, 另一个是严格对数凹的。如果对于某个  $K \in [0, 1]$ ,

$$\nabla^2 \log \nu_0 \preceq K \nabla^2 \log \mu_0, \quad (4.160)$$

则对于所有  $t \geq 0$ ,

$$\nabla^2 \log \nu_t \preceq K \nabla^2 \log \mu_t. \quad (4.161)$$

证明见第 4.5.15 小节。

然而, 即使我们加强初始条件, 假设对于某些满足  $1 - K_1 - K_2 \geq 0$  的  $K_1, K_2 \in [0, 1]$ , 有

$$\nabla^2 \log \nu_0 \preceq K_1 \nabla^2 \log \mu_0 - K_2 \Sigma_\infty^{-1}, \quad (4.162)$$

也并不一定能通过定理 4.18 中使用的相同 Brascamp-Lieb 和 Cramér-Rao 论证推导出对于所有  $t > 0$  成立

$$\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t - K_2 \Sigma_\infty^{-1}. \quad (4.163)$$

事实上, 在  $\mu_0$  为高斯且  $\nu_0$  为严格对数凹, 以及  $K_1, K_2 \in [0, 1], 1 - K_1 - K_2 \geq 0$  的相同假设下, 可以严格获得的只有以下较弱的永久比较:

$$\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t - K_2 \Sigma_t^{-1} + K_2 \left( \Sigma_t + \Sigma_t e^{Q^\top t} \Sigma_\infty^{-1} e^{Qt} \Sigma_t \right)^{-1}, \quad \forall t > 0. \quad (4.164)$$

关键步骤是将各项拆分为三个凸组合

$$\begin{aligned} & K_1 \nabla^2 W_{\mu,t} + (1 - K_1) e^{-Q^\top t} \Sigma_t^{-1} e^{-Qt} + K_2 \Sigma_\infty^{-1} \\ & = K_1 \nabla^2 W_{\mu,t} + (1 - K_1 - K_2) e^{-Q^\top t} \Sigma_t^{-1} e^{-Qt} + K_2 \left( e^{-Q^\top t} \Sigma_t^{-1} e^{-Qt} + \Sigma_\infty^{-1} \right), \end{aligned} \quad (4.165)$$

然后对这三项应用 Brascamp-Lieb 和 Cramér-Rao 不等式。能否建立具有固定  $\Sigma_\infty^{-1}$  项的更强时间一致界，仍是未来研究的一个开放问题。

基于上述结果，我们给出了一种在实践中验证对数 Hessian 比较原理的具体方法，然后结合定理 4.8，将其应用于建立沿 OU 流的相对 Rényi-Fisher 信息的 SDPI，如下所示。证明与推论 4.3 类似，此处省略。

**推论 4.5** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度，且  $\text{RFI}_q(\mu_0 \| \nu_0) < \infty, q \geq 1$ 。设  $\mu_t, \nu_t$  分别为从  $\mu_0, \nu_0$  出发的 OU 流的输出，目标分布为  $\pi = \mathcal{N}(0, \gamma^{-1}I)$ ：

$$\partial_t u_t = \nabla \cdot (\gamma x u_t) + \Delta u_t. \quad (4.166)$$

则我们有：

1. 假设  $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ )，则对于任何  $q \in [1, +\infty)$  和  $\delta \in (0, 1)$ ，存在足够大的时间  $T > 0$ ，对于所有  $t \geq T$  成立：

$$\text{RFI}_q(\mu_t \| \nu_t) \leq O(e^{-(2-\delta)\gamma t}). \quad (4.167)$$

2. 假设  $\nabla^2 \log \nu_0 \leq \nabla^2 \log \mu_0$ ， $\nu_0$  满足  $\alpha$ -SLC ( $\alpha > 0$ )，且  $\mu_0$  是 Gaussian，则对于所有  $t \geq 0$ ：

$$\text{RFI}_q(\mu_t \| \nu_t) \leq \frac{\gamma^2 e^{-2\gamma t}}{(\alpha + (\gamma - \alpha)e^{-2\gamma t})^2} \text{RFI}_q(\mu_0 \| \nu_0). \quad (4.168)$$

#### 4.4.3 FP 流的对数 Hessian 比较原理

我们研究与强凸势函数  $V$  相关的 Fokker-Planck (FP) 方程：

$$\partial_t u_t = \frac{c}{2} \Delta u_t + \nabla \cdot (u_t \nabla V(x)). \quad (4.169)$$

与 Ornstein-Uhlenbeck (OU) 设置相比，(4.169) 没有关于初始数据  $u_0$  的闭式解。即使使用 Feynman-Kac 表示，获得类似于 (4.140) 的  $u_t$  对数 Hessian 的易处理表达式也并非易事。遵循 [128] 中开发的策略，我们通过  $h$ -变换将 (4.169) 与扰动 OU 半群联系起来。在本小节中，我们沿用 [128] 的符号。

对于  $a, \sigma > 0$ , 考虑  $\mathbb{R}^d$  上的解耦 OU 算子  $L_{\sigma,a}^{ou}$ :

$$L_{\sigma,a}^{ou} = \frac{1}{2}\sigma^2\Delta - a x \cdot \nabla, \quad (4.170)$$

及其关联的 OU 过程  $(X_s^x)_{s \geq 0}$  满足

$$X_t^x = x + \sigma W_t - a \int_0^t X_s^x ds. \quad (4.171)$$

其在时间  $t$  的分布记为  $\gamma_t^x$ , 由 Mehler 核给出

$$d\gamma_t^x(y) = \frac{1}{Z_t} M_t^{\sigma,a}(x,y) dy, \quad M_t^{\sigma,a}(x,y) = \exp\left(-\frac{a\|y - e^{-at}x\|^2}{\sigma^2(1 - e^{-2at})}\right), \quad (4.172)$$

不变测度为

$$\gamma(dy) = \frac{1}{Z} \exp\left(-\frac{a\|y\|^2}{\sigma^2}\right) dy, \quad Z = \left(\frac{\pi\sigma^2}{a}\right)^{n/2}. \quad (4.173)$$

当  $a = 1, \sigma = \sqrt{2}$  时, 这是标准 Gaussian 测度。

定义

$$L^h = L^{ou} - h, \quad (L^h f)(x) = L^{ou} f(x) - h(x)f(x), \quad (4.174)$$

并记其半群为  $(P_t^h)_{t \geq 0}$ 。我们将反复使用以下表示。

**命题 4.1** [命题 2.1 [128]] 假设  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  连续且有下界, 并定义  $t \geq 0$  时的算子  $P_t^h$  为

$$P_t^h f(x) = \mathbb{E} \left[ f(X_t^x) e^{-\int_0^t h(X_s^x) ds} \right], \quad \forall x \in \mathbb{R}^d, \forall f \in \mathbb{L}^2(\gamma). \quad (4.175)$$

则  $(P_t^h)_{t \geq 0}$  是  $\mathbb{L}^2(\gamma)$  上的半群, 其无穷小生成元为  $L^h$ 。

**定义 4.1** [定义 2.2 [128]] 设  $t > 0, x \in \mathbb{R}^d$ , 且  $f \in \mathbb{L}^2(\gamma) \setminus \{0\}$  为非负函数。定义概率测度

$$\mathbb{Q}_{f,x}(\Gamma) = \frac{1}{P_t^h f(x)} \int_{\Gamma} f(X_t^x) e^{-\int_0^t V(X_s^x) ds} d\mathbb{P}, \quad (4.176)$$

并记  $\mathbb{E}_{f,x}$  为  $\mathbb{Q}_{f,x}$  下的期望。

下一个结果对我们的方法至关重要, 它提供了  $P_t^h f$  对数 Hessian 的显式恒等式。

**定理 4.19** [定理 2.3 [128]] 假设  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  有下界且属于  $\mathcal{D}^{(2)}$ , 其中  $\mathcal{D}^{(n)}, n = 0, 1, \dots$  表示其导数及函数本身具有多项式增长的  $C^n$  实值函数集合。对于  $t > 0$  和  $x \in \mathbb{R}^d$ , 设

$$A_t^x := - \int_0^t \nabla h(X_s^x) \frac{\sinh(a(t-s))}{\sinh(at)} ds + \frac{2ae^{-at}}{\sigma^2(1 - e^{-2at})} (X_t^x - e^{-at}x). \quad (4.177)$$

设  $f \in \mathbb{L}^2(\gamma) \setminus \{0\}$  为非负函数；利用定义 4.1 的概念，有

$$\nabla \log P_t^h f(x) = \mathbb{E}_{f,x}(A_t^x), \quad (4.178)$$

以及

$$\begin{aligned} \nabla^2 \log P_t^h f(x) &= - \int_0^t \left( \frac{\sinh(a(t-s))}{\sinh(at)} \right)^2 \mathbb{E}_{f,x}(\nabla^2 h(X_s^x)) ds + \mathbb{E}_{f,x}(A_t^x \otimes A_t^x) - \mathbb{E}_{f,x}(A_t^x)^{\otimes 2} \\ &\quad - \frac{2ae^{-2at}}{\sigma^2(1-e^{-2at})} I. \end{aligned} \quad (4.179)$$

取  $\sigma = \sqrt{c}$ ,  $a = 1$ , 设  $L^h = \frac{c}{2}\Delta - x \cdot \nabla - h$  在  $\mathbb{L}^2(\gamma)$  上, 对于足够光滑的  $W : \mathbb{R}^d \rightarrow \mathbb{R}$ , 我们通过以下变换定义  $\mathbb{L}^2(e^{W/2}\gamma)$  上的算子  $\mathcal{L}^W$ :

$$\begin{aligned} \mathcal{L}^W f &:= e^{-W/2} L^h(e^{W/2} f) \\ &= e^{-W/2} \left( \frac{c}{2}(\nabla - x) \cdot \left( e^{W/2} \nabla f + \frac{1}{2} e^{W/2} f \nabla W \right) - h e^{W/2} f \right) \\ &= e^{-W/2} \left[ \frac{c}{2} \left( e^{W/2} \Delta f + e^{W/2} \langle \nabla f, \nabla W \rangle + \frac{1}{4} e^{W/2} f \|\nabla W\|^2 + \frac{1}{2} e^{W/2} f \Delta W \right) \right] \\ &\quad - \langle x, \nabla f \rangle - \frac{1}{2} f \langle x, \nabla W \rangle - h f \\ &= \frac{c}{2} \Delta f + \left( \frac{c}{2} \nabla W - x \right) \cdot \nabla f + \left( \frac{c}{8} \|\nabla W\|^2 + \frac{c}{4} \Delta W - \frac{1}{2} \langle x, \nabla W \rangle - h \right) f, \end{aligned} \quad (4.180)$$

其关联半群  $(\mathcal{P}_t^W)_{t \geq 0}$  与  $P_t^h$  通过下式交织

$$\mathcal{P}_t^W f = e^{-W/2} P_t^h(e^{W/2} f). \quad (4.181)$$

取

$$\nabla W(x) = \frac{2}{c}(x - \nabla V(x)), \quad h(x) = \frac{1}{2c} \|\nabla V\|^2 - \frac{1}{2c} \|x\|^2 - \frac{1}{2} \Delta V + \frac{d}{2}, \quad (4.182)$$

我们恢复了 Langevin 扩散算子

$$L_V = \frac{c}{2} \Delta - \nabla V \cdot \nabla \quad \text{在 } \mathbb{L}^2(\mu_V) \text{ 上}, \quad \mu_V(dx) \propto e^{-2V(x)/c} dx. \quad (4.183)$$

因此,

$$\nabla^2 \log \mathcal{P}_t^W f = -\frac{1}{2} \nabla^2 W + \nabla^2 \log P_t^h(e^{W/2} f). \quad (4.184)$$

这立即得出以下推论。

**推论 4.6** [ [128] 推论 2.7 的推广版本] 设  $V : \mathbb{R}^d \rightarrow \mathbb{R}$  属于  $\mathcal{D}^{(4)}$ 。设  $\mu_V(dx) \propto e^{-\frac{2}{c}V(x)}dx$  且

$$W(x) := \frac{\|x\|^2}{c} - V(x), \quad h(x) = \frac{1}{2c} \|\nabla V\|^2 - \frac{1}{2c} \|x\|^2 - \frac{1}{2} \Delta V + \frac{d}{2}. \quad (4.185)$$

假设  $h$  有下界。设  $f \in \mathbb{L}^2(\mu_V) \setminus \{0\}$  为非负函数，对于所有  $x \in \mathbb{R}^d$ ，记  $\mathbb{E}_{f,x}^W$  为关于定义 4.1 中概率测度  $\mathbb{Q}_{e^{W/2},x}$  的期望。则对于所有  $t > 0$  和  $x \in \mathbb{R}^d$ ，

$$\begin{aligned} \nabla^2 \log \mathcal{P}_t^W f(x) &= \frac{1}{c} (\nabla^2 V - I) - \int_0^t \left( \frac{\sinh(t-s)}{\sinh(t)} \right)^2 \mathbb{E}_{f,x}^W (\nabla^2 h(X_s^x)) ds + \mathbb{E}_{f,x}^W (A_t^x \otimes A_t^x) \\ &\quad - \mathbb{E}_{f,x}^W (A_t^x)^{\otimes 2} - \frac{2e^{-2t}}{c(1-e^{-2t})} I, \end{aligned} \quad (4.186)$$

其中

$$A_t^x := - \int_0^t \nabla h(X_s^x) \frac{\sinh(t-s)}{\sinh(t)} ds + \frac{2e^{-t}}{c(1-e^{-2t})} (X_t^x - e^{-t}x). \quad (4.187)$$

考虑具有漂移  $\nabla V$  的 FP 动力学，

$$\partial_t \mu_t = \frac{c}{2} \Delta \mu_t + \nabla \cdot (\mu_t \nabla V), \quad (4.188)$$

和平稳测度  $\mu \propto e^{-2V/c}$ 。记  $\rho_t = \mu_t/\mu$ ，则有  $\partial_t \rho_t = \mathcal{L}^W \rho_t$  和  $\rho_t = \mathcal{P}_t^W \rho_0$ 。确保  $\rho_0 \in \mathbb{L}^2(\mu)$  的一个充分条件是

$$\int \frac{\mu_0}{\mu} d\mu_0 = \int e^{\frac{2}{c}V(x)} \mu_0(x)^2 dx < \infty, \quad (4.189)$$

例如，如果  $\mu_0$  具有紧支集，或者对于某个  $\epsilon > 0$ ， $\mu_0(x) \lesssim e^{-(\epsilon+1/c)V(x)}$  且  $V$  是  $\alpha$ -强凸的，则该条件成立。我们随后得到下一个结果。

**定理 4.20 (FP 流的对数 Hessian 公式)** 设  $V : \mathbb{R}^d \rightarrow \mathbb{R}$  属于  $\mathcal{D}^{(4)}$ 。设  $\mu_V(dx) \propto e^{-\frac{2}{c}V(x)}dx$  且

$$W(x) := \frac{\|x\|^2}{c} - V(x), \quad h(x) = \frac{1}{2c} \|\nabla V\|^2 - \frac{1}{2c} \|x\|^2 - \frac{1}{2} \Delta V + \frac{d}{2}. \quad (4.190)$$

假设  $h$  有下界。设  $\mu_t$  为 FP 流 (3.3) 的解，初始数据  $\mu_0$  满足  $\rho_0 = \frac{\mu_0}{\mu_V} \in \mathbb{L}^2(\mu_V)$  且  $\rho_0$  非负。对于所有  $t > 0$  和  $x \in \mathbb{R}^d$ ，记  $\mathbb{E}_{\rho_0,x}^W$  为关于定义 4.1 中概率测度  $\mathbb{Q}_{e^{W/2},\rho_0,x}$  的期望。则对于所有  $t > 0$  和  $x \in \mathbb{R}^d$ ，

$$\begin{aligned} \nabla^2 \log \mu_t(x) &= -\frac{1}{c} (\nabla^2 V + I) - \int_0^t \left( \frac{\sinh(t-s)}{\sinh(t)} \right)^2 \mathbb{E}_{\rho_0,x}^W (\nabla^2 h(X_s^x)) ds \\ &\quad + \mathbb{E}_{\rho_0,x}^W (A_t^x \otimes A_t^x) - \mathbb{E}_{\rho_0,x}^W (A_t^x)^{\otimes 2} - \frac{2e^{-2t}}{c(1-e^{-2t})} I, \end{aligned} \quad (4.191)$$

其中

$$A_t^x := - \int_0^t \nabla h(X_s^x) \frac{\sinh(t-s)}{\sinh(t)} ds + \frac{2e^{-t}}{c(1-e^{-2t})} (X_t^x - e^{-t}x). \quad (4.192)$$

假设函数  $\tilde{h}(x) := \frac{1}{2c} \|\nabla V(x)\|^2 - \frac{1}{2} \Delta V(x)$  是  $M$ -平滑的。则我们有 FP 流解的对数 Hessian 的直接下界如下。

$$\begin{aligned} \nabla^2 \log \mu_t(x) &= -\frac{1}{c}(\nabla^2 V + I) - \int_0^t \left( \frac{\sinh(t-s)}{\sinh(t)} \right)^2 \mathbb{E}_{\rho_{0,x}}^W(\nabla^2 h(X_s^x)) ds \\ &\quad + \mathbb{E}_{\rho_{0,x}}^W(A_t^x \otimes A_t^x) - \mathbb{E}_{\rho_{0,x}}^W(A_t^x)^{\otimes 2} - \frac{2e^{-2t}}{c(1-e^{-2t})} I \\ &\geq -\frac{1}{c}(\nabla^2 V + I) - \frac{2e^{-2t}}{c(1-e^{-2t})} I - \left(M - \frac{1}{c}\right) \int_0^t \left( \frac{\sinh(t-s)}{\sinh(t)} \right)^2 ds \\ &= -\frac{1}{c}(\nabla^2 V + I) - \frac{2e^{-2t}}{c(1-e^{-2t})} I - \left(M - \frac{1}{c}\right) \cdot \frac{e^{2t} - e^{-2t} - 4t}{2(e^{2t} + e^{-2t} - 2)} I. \end{aligned} \quad (4.193)$$

当  $t \rightarrow \infty$  时，我们有渐近下界

$$\begin{aligned} \nabla^2 \log \mu_t(x) &\geq -\frac{1}{c}(\nabla^2 V + I) - \frac{2e^{-2t}}{c(1-e^{-2t})} I - \left(M - \frac{1}{c}\right) \cdot \frac{e^{2t} - e^{-2t} - 4t}{2(e^{2t} + e^{-2t} - 2)} I \\ &\rightarrow -\frac{1}{c}(\nabla^2 V + I) - \frac{1}{2} \left(M - \frac{1}{c}\right) I \\ &= -\frac{1}{c} \nabla^2 V - \frac{1}{2} \left(M + \frac{1}{c}\right) I. \end{aligned} \quad (4.194)$$

在下文中我们假设：

- $V$  非负且光滑，且  $\tilde{h}$  是凸的且有下界；
- $\nu_0(x) \lesssim e^{-V(x)/c}$ 。

在这些条件下，对数凹性沿 FP 流传播，类似于 [62] 中的定理 3：

$$\nabla^2 \log \nu_t \preceq -\frac{1}{c} \nabla^2 V, \quad t \geq 0. \quad (4.195)$$

结合上述下界，可得如下形式的定量比较原理

$$\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V, \quad (4.196)$$

只要

$$-\frac{1}{c} \nabla^2 V \preceq -\frac{K_1}{c} \nabla^2 V - \frac{K_1}{2} \left(M + \frac{1}{c}\right) I + K_2 \nabla^2 V, \quad (4.197)$$

即

$$\frac{K_1}{2} \left(M + \frac{1}{c}\right) I \preceq \left(\frac{1-K_1}{c} + K_2\right) \nabla^2 V. \quad (4.198)$$

如果  $1 - K_1 + cK_2 > 0$  且  $V$  是  $\alpha$ -强凸的, 则充分条件为

$$\frac{K_1}{2} \left( M + \frac{1}{c} \right) \leq \alpha \left( \frac{1 - K_1}{c} + K_2 \right), \quad \text{等价于} \quad 1 - \left( 1 + \frac{cM + 1}{2\alpha} \right) K_1 + cK_2 \geq 0. \quad (4.199)$$

**定理 4.21** (FP 流的对数 Hessian 比较原理) 设  $V : \mathbb{R}^d \rightarrow \mathbb{R}$  属于  $\mathcal{D}^{(4)}$ 。设  $\mu_V(dx) \propto e^{-\frac{2}{c}V(x)}dx$  且

$$W(x) := \frac{\|x\|^2}{c} - V(x), \quad h(x) = \frac{1}{2c} \|\nabla V\|^2 - \frac{1}{2c} \|x\|^2 - \frac{1}{2} \Delta V + \frac{d}{2}. \quad (4.200)$$

假设  $h$  有下界。设  $\mu_t, \nu_t$  为 FP 流 (3.3) 的解, 初始数据  $\mu_0, \nu_0$  满足  $\rho_0 = \frac{\mu_0}{\nu_0} \in \mathbb{L}^2(\mu_V)$ ,  $\nu_0(x) \lesssim e^{-\frac{V(x)}{c}}$  且  $\nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 即  $\nabla^2 \log \nu_0 \preceq -\frac{1}{c} \nabla^2 V$ 。假设  $\tilde{h}(x) = \frac{1}{2c} \|\nabla V(x)\|^2 - \frac{1}{2} \Delta V(x)$  是凸的且  $M$ -光滑 (即  $\nabla^2 \tilde{h} \preceq MI$ ), 并且  $V$  是  $\alpha$ -强凸的 ( $\alpha > 0$ )。如果常数  $K_1, K_2$  满足  $1 - \left( 1 + \frac{cM+1}{2\alpha} \right) K_1 + cK_2 > 0$ ,  $1 - K_1 + cK_2 > 0$ , 则我们有最终比较: 存在  $T = T(V, K_1, K_2, \alpha, M, c, \mu_0, \nu_0)$  使得

$$\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V, \quad \forall t \geq T. \quad (4.201)$$

作为一个实际例子, 可以考虑  $V(x) = \frac{x^2}{2} + (1 + x^2)^{p/2}$ , 其中  $0 < p \leq 2$ 。更多细节参见 4.5.17 小节。基于上述结果, 我们提出了一种在实践中验证对数 Hessian 比较原理的具体方法。结合定理 4.9, 此方法被应用于推导相对 Rényi-Fisher 信息沿 FP 流的 SDPI, 如下所示, 其中  $\nu_t$  取为平稳测度  $\nu_\infty \propto e^{-\frac{2}{c}V}$ 。

**推论 4.7** 设  $\mu_0, \nu_0$  为  $\mathbb{R}^d$  上的概率测度, 且  $\text{RFI}_q(\mu_0 \| \nu_0) < \infty, q \geq 1$ 。设  $\mu_t$  为 FP 流从  $\mu_0$  出发对于目标分布  $\pi \propto e^{-\frac{2}{c}V}$  的输出, 且  $\nu_0 = \nu_\infty = \pi$ :

$$\partial_t u_t = \frac{c}{2} \Delta u_t + \nabla \cdot (u_t \nabla V(x)), \quad (4.202)$$

其中势函数  $V$  是  $\alpha$ -强凸的 ( $\alpha > 0$ )。假设  $\tilde{h}(x) = \frac{1}{2c} \|\nabla V(x)\|^2 - \frac{1}{2} \Delta V(x)$  是凸的且  $M$ -光滑 (即  $\nabla^2 \tilde{h} \preceq MI$ ), 并且常数  $K$  满足:

- $q(2 + cK) \geq 1 + cK$ ,
- $1 + cK \leq 0$ ,
- $2\alpha(1 + cK) + cM + 1 < 0$ .

则我们有:

1. 如果  $1 + cM - 2\alpha \leq 0$ , 则对于任何  $\delta > 0$ , 存在足够大的时间  $T > 0$ , 对于所有  $t \geq T, q \in [1, +\infty)$  成立:

$$\text{RFI}_q(\mu_t \| \nu_\infty) \leq O \left( e^{-2q \left( \frac{2\alpha - 1 - cM}{2\alpha} q + \frac{1 + cM}{2\alpha} - \delta \right) \alpha t} \right). \quad (4.203)$$

2. 如果  $1 + cM - 2\alpha > 0$ , 则对于任何  $\delta > 0$ , 存在足够大的时间  $T > 0$ , 对于所有  $t \geq T, q \in [1, \frac{1+cM}{1+cM-2\alpha})$  成立:

$$\text{RFI}_q(\mu_t \|\nu_\infty) \leq O\left(e^{-2q\left(\frac{2\alpha-1-cM}{2\alpha}q + \frac{1+cM}{2\alpha} - \delta\right)\alpha t}\right). \quad (4.204)$$

证明见第 4.5.16 小节。

对于 SDPI 的应用, 我们希望将最终对数 Hessian 比较原理细化为对动态  $\nu_t$  (不一定等同于  $\nu_\infty$ ) 成立, 这有利于我们直接应用定理 4.9 来获得相对 Rényi-Fisher 信息沿一般 FP 流的 SDPI。此外, 沿一般 FP 流的永久对数 Hessian 比较原理仍然是开放的。我们将这些改进留待进一步研究。

## 4.5 本章定理证明与技术细节

### 4.5.1 定理 4.4 的证明

首先, 我们推导  $\mu_t$  和  $\nu_t$  的显式表达式:

$$\begin{aligned} \mu_t(x) &= \left(\frac{2\pi(1-e^{-2\gamma t})}{\gamma}\right)^{-d/2} \int_{\mathbb{R}_y^d} \mu_0(y) \exp\left\{-\frac{\gamma\|x-e^{-\gamma t}y\|^2}{2(1-e^{-2\gamma t})}\right\} dy \\ &:= K_t^{-d/2} \int_{\mathbb{R}_y^d} \mu_0(y) \exp\left\{-\frac{\gamma\|x-e^{-\gamma t}y\|^2}{2(1-e^{-2\gamma t})}\right\} dy, \end{aligned} \quad (4.205)$$

$$\nu_t(x) = K_t^{-d/2} \int_{\mathbb{R}_y^d} \nu_0(y) \exp\left\{-\frac{\gamma\|x-e^{-\gamma t}y\|^2}{2(1-e^{-2\gamma t})}\right\} dy. \quad (4.206)$$

随后, 计算  $\mu_t$  和  $\nu_t$  的梯度:

$$\nabla \mu_t(x) = \frac{\gamma K_t^{-d/2}}{1-e^{-2\gamma t}} \int_{\mathbb{R}_y^d} \mu_0(y) (e^{-\gamma t}y - x) \exp\left\{-\frac{\gamma\|x-e^{-\gamma t}y\|^2}{2(1-e^{-2\gamma t})}\right\} dy, \quad (4.207)$$

$$\nabla \nu_t(x) = \frac{\gamma K_t^{-d/2}}{1-e^{-2\gamma t}} \int_{\mathbb{R}_y^d} \nu_0(y) (e^{-\gamma t}y - x) \exp\left\{-\frac{\gamma\|x-e^{-\gamma t}y\|^2}{2(1-e^{-2\gamma t})}\right\} dy. \quad (4.208)$$

在推论 4.1 中取  $b_t(x) = \gamma x$  及  $c = 2$ , 可得相对 Fisher 信息沿 OU 流的时间导数:

$$\begin{aligned} \partial_t \text{FI}(\mu_t \|\nu_t) &= -2\mathbb{E}_{\mu_t} \left[ \left\| \nabla^2 \log \frac{\mu_t}{\nu_t} \right\|_{\text{HS}}^2 \right] \\ &\quad - 2\mathbb{E}_{\mu_t} \left[ \left\| \nabla \log \frac{\mu_t}{\nu_t} \right\|_{(-2\nabla^2 \log \nu_t - \gamma I)}^2 \right]. \end{aligned} \quad (4.209)$$

利用文献 [37, 127] 中关于强对数凹性和庞加莱不等式在卷积下的复合规则, 可

知  $\nu_t$  满足  $\alpha_t$ -SLC,  $\mu_t$  满足  $\beta_t$ -PI, 其中参数为:

$$\alpha_t = \left( \frac{e^{-2\gamma t}}{\alpha} + \frac{1 - e^{-2\gamma t}}{\gamma} \right)^{-1} = \frac{\gamma\alpha}{\alpha + e^{-2\gamma t}(\gamma - \alpha)}, \quad (4.210)$$

$$\beta_t = \left( \frac{e^{-2\gamma t}}{\beta} + \frac{1 - e^{-2\gamma t}}{\gamma} \right)^{-1} = \frac{\gamma\beta}{\beta + e^{-2\gamma t}(\gamma - \beta)}. \quad (4.211)$$

对测试函数  $\nabla \log \frac{\mu_t}{\nu_t}$  应用  $\mu_t$  的庞加莱不等式, 有:

$$\begin{aligned} \mathbb{E}_{\mu_t} \left[ \left\| \nabla^2 \log \frac{\mu_t}{\nu_t} \right\|_{\text{HS}}^2 \right] &\geq \beta_t \text{Var}_{\mu_t} \left( \nabla \log \frac{\mu_t}{\nu_t} \right) \\ &= \beta_t \text{FI}(\mu_t \| \nu_t) - \beta_t \left\| \mathbb{E}_{\mu_t} \left( \nabla \log \frac{\mu_t}{\nu_t} \right) \right\|^2. \end{aligned} \quad (4.212)$$

因此可得:

$$\begin{aligned} \partial_t \text{FI}(\mu_t \| \nu_t) &\leq -2(\beta_t + 2\alpha_t - \gamma) \text{FI}(\mu_t \| \nu_t) \\ &\quad + \beta_t \left\| \mathbb{E}_{\mu_t} \left( \nabla \log \frac{\mu_t}{\nu_t} \right) \right\|^2. \end{aligned} \quad (4.213)$$

现估计非对称项  $\left\| \mathbb{E}_{\mu_t} \left( \nabla \log \frac{\mu_t}{\nu_t} \right) \right\|^2$ :

$$\begin{aligned} \mathbb{E}_{\mu_t} \left( \nabla \log \frac{\mu_t}{\nu_t} \right) &= \int_{\mathbb{R}_x^d} \mu_t(x) \left( \frac{\nabla \mu_t(x)}{\mu_t(x)} - \frac{\nabla \nu_t(x)}{\nu_t(x)} \right) dx \\ &= \int_{\mathbb{R}_x^d} \nabla \mu_t(x) dx - \int_{\mathbb{R}_x^d} \mu_t(x) \frac{\nabla \nu_t(x)}{\nu_t(x)} dx \\ &= - \int_{\mathbb{R}_x^d} \mu_t(x) \frac{\nabla \nu_t(x)}{\nu_t(x)} dx, \end{aligned} \quad (4.214)$$

其中

$$\begin{aligned} \int_{\mathbb{R}_x^d} \nabla \mu_t(x) dx &= \frac{\gamma K_t^{-d/2}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_x^d} dx \int_{\mathbb{R}_y^d} \mu_0(y) (e^{-\gamma t} y - x) \\ &\quad \cdot \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy \\ &= \frac{\gamma K_t^{-d/2}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_y^d} \mu_0(y) dy \int_{\mathbb{R}_x^d} (e^{-\gamma t} y - x) \\ &\quad \cdot \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dx \\ &= 0, \\ \int_{\mathbb{R}_x^d} \frac{\mu_t \nabla \nu_t}{\nu_t} dx &= \frac{\gamma K_t^{-d/2}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_x^d} \left( \int_{\mathbb{R}_y^d} \mu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy \right) \end{aligned} \quad (4.215)$$

$$\begin{aligned}
& \frac{\int_{\mathbb{R}^d} \nu_0(y) (e^{-\gamma t} y - x) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy}{\int_{\mathbb{R}^d} \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy} dx \\
&= \frac{\gamma K_t^{-d/2}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}^d} \mu_0(y) \left( \int_{\mathbb{R}_x^d} (-x) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dx \right) dy \\
&+ \frac{\gamma e^{-\gamma t} K_t^{-d/2}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_y^d \times \mathbb{R}_x^d} y \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} \\
& \frac{\int_{\mathbb{R}_y^d} \mu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy}{\int_{\mathbb{R}_y^d} \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy} dy dx \\
&= -\frac{\gamma e^{-\gamma t}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_y^d} y \mu_0(y) dy \\
&+ \frac{\gamma e^{-\gamma t} K_t^{-d/2}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_y^d \times \mathbb{R}_x^d} y \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} \\
& \frac{\int_{\mathbb{R}_y^d} \mu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy}{\int_{\mathbb{R}_y^d} \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy} dy dx \\
&\stackrel{(*)}{=} \frac{\gamma e^{-\gamma t} K_t^{-d/2}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_y^d \times \mathbb{R}_x^d} y \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} \\
& \frac{\int_{\mathbb{R}_y^d} \mu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy}{\int_{\mathbb{R}_y^d} \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy} dy dx. \tag{4.216}
\end{aligned}$$

其中 (\*) 利用了  $\mu_0$  是中心化分布的假设。在给出  $\|\mathbb{E}_{\mu_t} (\nabla \log \frac{\mu_t}{\nu_t})\|^2$  的估计之前，我们先考察密度函数  $\mu_t$  沿 OU 流的收敛速率作为准备。

对于任意固定的  $x \in \mathbb{R}^d$ ，设  $K := K_\infty = \frac{2\pi}{\gamma}$ ：

$$\begin{aligned}
K_t^{d/2} \mu_t(x) - K^{d/2} \pi(x) &= \int_{\mathbb{R}_y^d} \mu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy \\
&\quad - \exp \left\{ -\frac{\gamma \|x\|^2}{2} \right\} \\
&= \int_{\mathbb{R}_y^d} \mu_0(y) \left[ \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} \right.
\end{aligned}$$

$$\begin{aligned}
& - \exp \left\{ -\frac{\gamma \|x\|^2}{2} \right\} \Big] dy \\
& = e^{-\frac{\gamma \|x\|^2}{2}} \int_{\mathbb{R}^d} \mu_0(y) \left[ \exp \left\{ \frac{\gamma \|x\|^2}{2} - \frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} - 1 \right] dy,
\end{aligned}$$

其中当  $t > 0$  较大时, 有:

$$\begin{aligned}
\frac{\gamma \|x\|^2}{2} - \frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} & = -\frac{\gamma e^{-2\gamma t} (\|x\|^2 + \|y\|^2)}{2(1 - e^{-2\gamma t})} + \frac{\gamma e^{-\gamma t} x^\top y}{1 - e^{-2\gamma t}} \quad (4.217) \\
\exp \left\{ \frac{\gamma \|x\|^2}{2} - \frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} - 1 & = -\frac{\gamma e^{-2\gamma t} (\|x\|^2 + \|y\|^2)}{2(1 - e^{-2\gamma t})} + \frac{\gamma e^{-\gamma t} x^\top y}{1 - e^{-2\gamma t}} \\
& \quad + \frac{1}{2} \cdot \frac{\gamma^2 e^{-2\gamma t} |x^\top y|^2}{(1 - e^{-2\gamma t})^2} + o(e^{-2\gamma t}) \\
& = \frac{\gamma e^{-\gamma t} x^\top y}{1 - e^{-2\gamma t}} - \frac{\gamma e^{-2\gamma t} (\|x\|^2 + \|y\|^2)}{2(1 - e^{-2\gamma t})} \\
& \quad + \frac{\gamma^2 e^{-2\gamma t} |x^\top y|^2}{2(1 - e^{-2\gamma t})^2} + o(e^{-2\gamma t}) \\
& \leq \frac{\gamma e^{-\gamma t} x^\top y}{1 - e^{-2\gamma t}} - \frac{\gamma e^{-2\gamma t} (\|x\|^2 + \|y\|^2)}{2(1 - e^{-2\gamma t})} \\
& \quad + \frac{\gamma^2 e^{-2\gamma t} \|x\| \|y\|}{2(1 - e^{-2\gamma t})^2} + o(e^{-2\gamma t}). \quad (4.218)
\end{aligned}$$

因此对于固定的  $x \in \mathbb{R}^d$ , 可得:

$$\begin{aligned}
K_t^{d/2} \mu_t(x) - K^{d/2} \pi(x) & = e^{-\frac{\gamma \|x\|^2}{2}} \int_{\mathbb{R}^d} \mu_0(y) \left[ \frac{\gamma e^{-\gamma t} x^\top y}{1 - e^{-2\gamma t}} \right. \\
& \quad \left. - \frac{\gamma e^{-2\gamma t} (\|x\|^2 + \|y\|^2)}{2(1 - e^{-2\gamma t})} \right. \\
& \quad \left. + \frac{\gamma^2 e^{-2\gamma t} \|x\| \|y\|}{2(1 - e^{-2\gamma t})^2} + o(e^{-2\gamma t}) \right] dy \\
& = e^{-\frac{\gamma \|x\|^2}{2}} \cdot \frac{\gamma e^{-\gamma t}}{1 - e^{-2\gamma t}} \left( x^\top \int_{\mathbb{R}^d} y \mu_0(y) dy \right) \\
& \quad - e^{-\frac{\gamma \|x\|^2}{2}} \frac{\gamma e^{-2\gamma t} \|x\|^2}{2(1 - e^{-2\gamma t})} \\
& \quad - e^{-\frac{\gamma \|x\|^2}{2}} \cdot \frac{\gamma e^{-2\gamma t}}{2(1 - e^{-2\gamma t})} \int_{\mathbb{R}^d} \|y\|^2 \mu_0(y) dy \\
& \quad + e^{-\frac{\gamma \|x\|^2}{2}} \cdot \frac{\gamma^2 e^{-2\gamma t} \|x\|}{2(1 - e^{-2\gamma t})^2} \int_{\mathbb{R}^d} \|y\| \mu_0(y) dy + o(e^{-2\gamma t}) \\
& \stackrel{(*)}{=} e^{-\frac{\gamma \|x\|^2}{2}} \cdot \frac{\gamma^2 e^{-2\gamma t} \|x\|}{2(1 - e^{-2\gamma t})^2} \int_{\mathbb{R}^d} \|y\| \mu_0(y) dy \\
& \quad - e^{-\frac{\gamma \|x\|^2}{2}} \frac{\gamma e^{-2\gamma t} \|x\|^2}{2(1 - e^{-2\gamma t})}
\end{aligned}$$

$$\begin{aligned}
& - e^{-\frac{\gamma\|x\|^2}{2}} \cdot \frac{\gamma e^{-2\gamma t}}{2(1 - e^{-2\gamma t})} \int_{\mathbb{R}_y^d} \|y\|^2 \mu_0(y) dy \\
& + o(e^{-2\gamma t}).
\end{aligned} \tag{4.219}$$

**注** 值得注意的是展开式中的  $o(e^{-2\gamma t})$  项。该项隐含了初始测度  $\mu_0$  的所有高阶矩。因此，如果  $\mu_0$  的矩随其阶数  $n$  增长过快，则不能保证该项为真正的可忽略量。为解决此问题，我们观察到  $n+1$  阶的积分项涉及形如  $\|x\|^p \|y\|^q$  ( $p+q=n+1$ ) 的因子，其前系数随时间呈指数衰减，速率至多为  $e^{-n\gamma t}$ 。利用  $\mu_0$  紧支集的性质，假设  $\text{supp}(\mu_0) \subset B_R(0) \subset \mathbb{R}^d$  ( $R \geq 1$ )。考虑足够大的  $t$  使得  $1 \leq R < \frac{\gamma t}{100}$  且  $\gamma t > 50 \log 2$ 。此外，在分析高阶项时，我们将关注区域限制在  $x \in \mathcal{S}_t := \{x \in \mathbb{R}^d : \|x\| \leq e^{\frac{\gamma t}{4}}\}$ 。在此条件下，高阶项可界定如下：

$$\begin{aligned}
o(e^{-2\gamma t}) & \leq C \sum_{n=3}^{\infty} e^{-n\gamma t} 2^n e^{\frac{n\gamma t}{4}} R^n \leq C \sum_{n=3}^{\infty} e^{(-\frac{37\gamma t}{50} + \log 2)n} \\
& \leq C \sum_{n=3}^{\infty} e^{-\frac{36n\gamma t}{50}} = O(e^{-2.16\gamma t}),
\end{aligned} \tag{4.220}$$

其中  $C > 0$  为一致常数。该计算表明所有高阶项之和衰减速率至少为  $e^{-2.16\gamma t}$ ，因此我们可以安全地使用  $o(e^{-2\gamma t})$  表示高阶贡献。

综上所述，对于任意  $x \in \mathcal{S}_t$ ，我们有：

$$\begin{aligned}
& \frac{\int_{\mathbb{R}_y^d} \mu_0(y) \exp\left\{-\frac{\gamma\|x - e^{-\gamma t}y\|^2}{2(1 - e^{-2\gamma t})}\right\} dy}{\int_{\mathbb{R}_y^d} \nu_0(y) \exp\left\{-\frac{\gamma\|x - e^{-\gamma t}y\|^2}{2(1 - e^{-2\gamma t})}\right\} dy} \\
& = \frac{1 - \frac{\gamma e^{-2\gamma t} \|x\|^2}{2(1 - e^{-2\gamma t})} + \frac{\gamma^2 e^{-2\gamma t} \|x\|}{2(1 - e^{-2\gamma t})^2} \mathbb{E}_{\mu_0}(\|y\|) - \frac{\gamma e^{-2\gamma t}}{2(1 - e^{-2\gamma t})} \mathbb{E}_{\mu_0}(\|y\|^2) + O(e^{-2.16\gamma t})}{1 - \frac{\gamma e^{-2\gamma t} \|x\|^2}{2(1 - e^{-2\gamma t})} + \frac{\gamma^2 e^{-2\gamma t} \|x\|}{2(1 - e^{-2\gamma t})^2} \mathbb{E}_{\nu_0}(\|y\|) - \frac{\gamma e^{-2\gamma t}}{2(1 - e^{-2\gamma t})} \mathbb{E}_{\nu_0}(\|y\|^2) + O(e^{-2.16\gamma t})} \\
& = 1 + \frac{\frac{\gamma^2 e^{-2\gamma t} \|x\|}{2(1 - e^{-2\gamma t})^2} \mathbb{E}_{(\mu_0 - \nu_0)}(\|y\|) - \frac{\gamma e^{-2\gamma t}}{2(1 - e^{-2\gamma t})} \mathbb{E}_{(\mu_0 - \nu_0)}(\|y\|^2) + O(e^{-2.16\gamma t})}{1 - \frac{\gamma e^{-2\gamma t} \|x\|^2}{2(1 - e^{-2\gamma t})} + \frac{\gamma^2 e^{-2\gamma t} \|x\|}{2(1 - e^{-2\gamma t})^2} \mathbb{E}_{\nu_0}(\|y\|) - \frac{\gamma e^{-2\gamma t}}{2(1 - e^{-2\gamma t})} \mathbb{E}_{\nu_0}(\|y\|^2) + O(e^{-2.16\gamma t})} \\
& = 1 + O(e^{-2\gamma t} \cdot e^{\gamma t/4}) = 1 + O(e^{-1.75\gamma t}).
\end{aligned} \tag{4.221}$$

进而有：

$$\int_{\mathbb{R}_x^d} \frac{\mu_t \nabla \nu_t}{\nu_t} dx$$

$$\begin{aligned}
&= \frac{\gamma e^{-\gamma t} K_t^{-d/2}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_y^d \times \mathbb{R}_x^d} y \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} \\
&\quad \cdot \frac{\int_{\mathbb{R}_y^d} \mu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy}{\int_{\mathbb{R}_y^d} \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy} dy dx \\
&= \frac{\gamma e^{-\gamma t} K_t^{-d/2}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_y^d \times \mathcal{S}_t} y \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} \\
&\quad \cdot \frac{\int_{\mathbb{R}_y^d} \mu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy}{\int_{\mathbb{R}_y^d} \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy} dy dx \\
&\quad + \frac{\gamma e^{-\gamma t} K_t^{-d/2}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_y^d \times \mathcal{S}_t^c} y \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} \\
&\quad \cdot \frac{\int_{\mathbb{R}_y^d} \mu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy}{\int_{\mathbb{R}_y^d} \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy} dy dx \\
&= \frac{\gamma e^{-\gamma t} K_t^{-d/2}}{1 - e^{-2\gamma t}} (J_t^{(1)} + J_t^{(2)}), \tag{4.222}
\end{aligned}$$

其中

$$\begin{aligned}
J_t^{(1)} &= \int_{\mathbb{R}_y^d \times \mathcal{S}_t} y \nu_0(y) e^{-\frac{\gamma \|x\|^2}{2}} \left( 1 + \frac{\gamma e^{-\gamma t} x^\top y}{1 - e^{-2\gamma t}} \right. \\
&\quad \left. - \frac{\gamma e^{-2\gamma t} (\|x\|^2 + \|y\|^2)}{2(1 - e^{-2\gamma t})} + \frac{\gamma^2 e^{-2\gamma t} |x^\top y|^2}{2(1 - e^{-2\gamma t})^2} + O(e^{-2.16\gamma t}) \right) \\
&\quad \cdot (1 + O(e^{-1.75\gamma t})) dx dy \\
&= \int_{\mathbb{R}_y^d \times \mathcal{S}_t} y \nu_0(y) e^{-\frac{\gamma \|x\|^2}{2}} dx dy \\
&\quad + \frac{\gamma e^{-\gamma t}}{1 - e^{-2\gamma t}} \int_{\mathbb{R}_y^d \times \mathcal{S}_t} y \nu_0(y) e^{-\frac{\gamma \|x\|^2}{2}} x^\top y dx dy \\
&\quad + O(e^{-2\gamma t} \cdot e^{\frac{\gamma t}{2}}) + O(e^{-1.75\gamma t}) \\
&= 0 + 0 + O(e^{-1.5\gamma t}) = O(e^{-1.5\gamma t}), \tag{4.223} \\
\|J_t^{(2)}\| &= \left\| \int_{\mathbb{R}_y^d \times \mathcal{S}_t^c} y \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} \right\|
\end{aligned}$$

$$\begin{aligned}
& \left\| \frac{\int_{\mathbb{R}^d} \mu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy}{\int_{\mathbb{R}^d} \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy} dy dx \right\| \\
& \leq \int_{\mathbb{R}^d \times \mathcal{S}_t^c} \|y\| \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} \\
& \quad \cdot \frac{\int_{\mathbb{R}^d} \mu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy}{\int_{\mathbb{R}^d} \nu_0(y) \exp \left\{ -\frac{\gamma \|x - e^{-\gamma t} y\|^2}{2(1 - e^{-2\gamma t})} \right\} dy} dy dx \\
& \leq M \exp \left\{ -\frac{\gamma}{2} e^{\frac{2\gamma t}{5}} \right\} \int_{\mathbb{R}^d} \|y\| \nu_0(y) dy = O(e^{-e^{\frac{2\gamma t}{5}}}). \tag{4.224}
\end{aligned}$$

因此我们得到  $\|\mathbb{E}_{\mu_t}(\nabla \log \frac{\mu_t}{\nu_t})\|^2$  的估计:

$$\begin{aligned}
\left\| \mathbb{E}_{\mu_t} \left( \nabla \log \frac{\mu_t}{\nu_t} \right) \right\|^2 &= \frac{2\gamma^2 e^{-2\gamma t} K_t^{-d}}{(1 - e^{-2\gamma t})^2} \left( \|J_t^{(1)}\|^2 + \|J_t^{(2)}\|^2 \right) \\
&= \frac{2\gamma^2 e^{-2\gamma t} K_t^{-d}}{(1 - e^{-2\gamma t})^2} \left( O(e^{-3\gamma t}) + O(e^{-2e^{\frac{2\gamma t}{5}}}) \right) \\
&= O(e^{-5\gamma t}), \tag{4.225}
\end{aligned}$$

对于足够大的  $t > 0$ 。回顾式 (4.213)，有:

$$\begin{aligned}
\partial_t \text{FI}(\mu_t \| \nu_t) &\leq -2(\beta_t + 2\alpha_t - \gamma) \text{FI}(\mu_t \| \nu_t) + \beta_t \left\| \mathbb{E}_{\mu_t} \left( \nabla \log \frac{\mu_t}{\nu_t} \right) \right\|^2 \\
&:= -\lambda_t \text{FI}(\mu_t \| \nu_t) + \beta_t g(t), \tag{4.226}
\end{aligned}$$

其中

$$\begin{aligned}
\lambda_t &= 2(\beta_t + 2\alpha_t - \gamma) \\
&= \frac{2\gamma\beta}{\beta + e^{-2\gamma t}(\gamma - \beta)} + 2\gamma - \frac{4\gamma(\gamma - \alpha)e^{-2\gamma t}}{\alpha + e^{-2\gamma t}(\gamma - \alpha)} \\
&= \frac{d}{dt} \left( 4\gamma t + \log \left[ (\beta + e^{-2\gamma t}(\gamma - \beta)) (\alpha + e^{-2\gamma t}(\gamma - \alpha))^2 \right] \right), \tag{4.227}
\end{aligned}$$

$$\begin{aligned}
\beta_t &= \frac{2\gamma\beta}{\beta + e^{-2\gamma t}(\gamma - \beta)} \\
&= \frac{d}{dt} \left( \gamma t + \frac{1}{2} \log (\beta + e^{-2\gamma t}(\gamma - \beta)) \right). \tag{4.228}
\end{aligned}$$

由 Grönwall 不等式, 可得:

$$\frac{d}{dt} \left( e^{\int_0^t \lambda_s ds} \text{FI}(\mu_t \| \nu_t) \right) \leq e^{\int_0^t \lambda_s ds} \beta_t g(t), \quad (4.229)$$

$$\begin{aligned} \text{FI}(\mu_t \| \nu_t) &\leq \frac{\gamma^3 e^{-4\gamma t}}{(\beta + e^{-2\gamma t}(\gamma - \beta)) (\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} \text{FI}(\mu_0 \| \nu_0) \\ &\quad + \frac{\gamma^3 e^{-4\gamma t}}{(\beta + e^{-2\gamma t}(\gamma - \beta)) (\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} \\ &\quad \cdot \int_0^t \frac{2\gamma\beta e^{4\gamma s} (\beta + e^{-2\gamma s}(\gamma - \beta)) (\alpha + e^{-2\gamma s}(\gamma - \alpha))^2}{\gamma^3 (\beta + e^{-2\gamma s}(\gamma - \beta))} g(s) ds \\ &= \frac{\gamma^3 e^{-4\gamma t}}{(\beta + e^{-2\gamma t}(\gamma - \beta)) (\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} \text{FI}(\mu_0 \| \nu_0) \\ &\quad + \frac{2\beta\gamma e^{-4\gamma t}}{(\beta + e^{-2\gamma t}(\gamma - \beta)) (\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} \\ &\quad \cdot \int_0^t e^{4\gamma s} (\alpha + e^{-2\gamma s}(\gamma - \alpha))^2 g(s) ds. \end{aligned} \quad (4.230)$$

由于对于较大的  $s \leq t$ , 根据 (4.225) 有  $g(s) = O(e^{-5\gamma s})$ , 则积分项

$$\int_0^t e^{4\gamma s} (\alpha + e^{-2\gamma s}(\gamma - \alpha))^2 g(s) ds \quad (4.231)$$

确实是有限的。因此我们可以得出结论:

$$\begin{aligned} \text{FI}(\mu_t \| \nu_t) &\leq \frac{\gamma^3 e^{-4\gamma t} \text{FI}(\mu_0 \| \nu_0)}{(\beta + e^{-2\gamma t}(\gamma - \beta)) (\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} \\ &\quad + \frac{2C'\beta\gamma e^{-4\gamma t}}{(\beta + e^{-2\gamma t}(\gamma - \beta)) (\alpha + e^{-2\gamma t}(\gamma - \alpha))^2} \\ &= O(e^{-4\gamma t}), \end{aligned} \quad (4.232)$$

其中  $C' > 0$  为常数。

**注** 初始测度  $\mu_0$  和  $\nu_0$  的紧支集条件似乎难以放松, 即使假设  $\mu_0, \nu_0$  的尾部呈超指数衰减。这是因为我们需要确保展开式中的  $o(e^{-2\gamma t})$  项确实代表一个可忽略的量。具体而言,  $\mu_0$  和  $\nu_0$  的  $n$  阶矩的增长率应按常数的  $n$  次方缩放, 即对于某些常数  $C > 0$ , 有  $\mathbb{E}_{\mu_0}(\|y\|^n) \leq C^n$ 。

事实上, 这一  $n$  阶矩性质等价于  $X$  具有紧支集。考虑  $X$  的  $L^n$  范数, 我们有:

$$\begin{aligned} \|X\|_n &:= (\mathbb{E} \|X\|^n)^{1/n} \\ &\leq C. \end{aligned} \quad (4.233)$$

取极限  $n \rightarrow \infty$ , 得:

$$\begin{aligned} \lim_{n \rightarrow \infty} \|X\|_n &= \|X\|_\infty \\ &= \text{ess sup } \|X\| \leq C, \end{aligned} \quad (4.234)$$

这证明了  $X$  具有紧支集。

#### 4.5.2 定理 4.5 的证明

在此情形下, 相对 Fisher 信息的时间导数由下式给出:

$$\partial_t \text{FI}(\mu_t \| \nu_t) = -c \int \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t + 2 \int \|\nabla \log \rho_t\|_{(c\nabla^2 \log \nu_t + \nabla^2 V)}^2 d\mu_t. \quad (4.235)$$

**第 (1) 部分:** 若  $\nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 则根据定理 3.4, 对于任意  $t > 0$ ,  $\nu_t$  也是  $\frac{V}{c}$ -相对对数凹的, 即

$$-\nabla^2 \log \nu_t \succeq \frac{1}{c} \nabla^2 V. \quad (4.236)$$

于是, 我们得到:

$$\begin{aligned} \partial_t \text{FI}(\mu_t \| \nu_t) &= -c \int \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t + 2 \int \|\nabla \log \rho_t\|_{(c\nabla^2 \log \nu_t + \nabla^2 V)}^2 d\mu_t \\ &\leq 2 \int \|\nabla \log \rho_t\|_{(c\nabla^2 \log \nu_t + \nabla^2 V)}^2 d\mu_t \\ &\leq 0. \end{aligned} \quad (4.237)$$

因此我们有:

$$\text{FI}(\mu_t \| \nu_t) \leq \text{FI}(\mu_0 \| \nu_0). \quad (4.238)$$

**第 (2) 部分:** 注意到, 如果  $\mu_0, \nu_0, V$  是对称的, 那么对于任意  $t \geq 0$ ,  $\mu_t, \nu_t$  也是对称的。这可以通过定理 3.1 中 FPE (3.3) 唯一经典解的表达式直接证明。这意味着  $\nabla \log \mu_t(x)$  和  $\nabla \log \nu_t(x)$  是奇函数, 且

$$\begin{aligned} \mathbb{E}_{\mu_t}[\nabla \log \nu_t] &= \int_{\mathbb{R}^d} \mu_t(x) \nabla \log \nu_t(x) dx \\ &= \frac{1}{2} \int_{\mathbb{R}^d} \mu_t(x) \nabla \log \nu_t(x) dx + \frac{1}{2} \int_{\mathbb{R}^d} \mu_t(-x) \nabla \log \nu_t(-x) dx \\ &= \frac{1}{2} \int_{\mathbb{R}^d} \mu_t(x) (\nabla \log \nu_t(x) + \nabla \log \nu_t(-x)) dx \\ &= 0. \end{aligned} \quad (4.239)$$

由于  $\mu_t$  在  $\mathbb{R}^d$  上具有全支集, 通过分部积分可知  $\mathbb{E}_{\mu_t}[\nabla \log \mu_t] = 0$ 。结合上述结果, 可得  $\mathbb{E}_{\mu_t}[\nabla \log \frac{\mu_t}{\nu_t}] = 0$ 。因此, 对函数  $\phi = \nabla \log \frac{\mu_t}{\nu_t}$  应用  $\mu_t$  的庞加莱不等式, 有:

$$\begin{aligned} \mathbb{E}_{\mu_t} \left[ \left\| \nabla^2 \log \frac{\mu_t}{\nu_t} \right\|_{\text{HS}}^2 \right] &\geq \beta_t \text{Var}_{\mu_t} \left[ \nabla \log \frac{\mu_t}{\nu_t} \right] \\ &= \beta_t \mathbb{E}_{\mu_t} \left[ \left\| \nabla \log \frac{\mu_t}{\nu_t} \right\|^2 \right] \\ &= \beta_t \text{FI}(\mu_t \| \nu_t). \end{aligned} \quad (4.240)$$

因此, 由公式 (4.30) 可知:

$$\partial_t \text{FI}(\mu_t \| \nu_t) \leq -c\beta_t \text{FI}(\mu_t \| \nu_t) + 0, \quad (4.241)$$

$$\text{FI}(\mu_t \| \nu_t) \leq e^{-c \int_0^t \beta_s ds} \text{FI}(\mu_0 \| \nu_0). \quad (4.242)$$

**第 (3) 部分:** 取  $\beta_t = \frac{\alpha}{c}$ 。第 (3) 部分的证明与第 (2) 部分的论证相同。

**第 (4) 部分:**

$$\begin{aligned} \partial_t \text{FI}(\mu_t \| \nu) &= -c \int \left\| \nabla^2 \log \rho_t \right\|_{\text{HS}}^2 d\mu_t + 2 \int \left\| \nabla \log \rho_t \right\|_{(c\nabla^2 \log \nu_t + \nabla^2 \nu)}^2 d\mu_t \\ &\leq 2 \int \left\| \nabla \log \rho_t \right\|_{(-2\nabla^2 \nu + \nabla^2 \nu)}^2 d\mu_t \\ &\leq -2 \int \left\| \nabla \log \rho_t \right\|_{\nabla^2 \nu}^2 d\mu_t \\ &\leq -2\alpha \int \left\| \nabla \log \rho_t \right\|^2 d\mu_t \\ &= -2\alpha \text{FI}(\mu_t \| \nu). \end{aligned} \quad (4.243)$$

因此对于所有  $t \geq 0$ :

$$\text{FI}(\mu_t \| \nu) \leq e^{-2\alpha t} \text{FI}(\mu_0 \| \nu). \quad (4.244)$$

**第 (5) 部分:** 此结论可通过结合第 (3) 部分和第 (4) 部分的论证得证。

### 4.5.3 定理 4.6 的证明

我们通过对被积函数求导并利用链式法则进行计算:

$$\begin{aligned} \partial_t R_q(\mu_t \| \nu_t) &= \frac{1}{q-1} \partial_t \log \int \rho_t^q d\nu_t \\ &= \frac{1}{q-1} \frac{1}{\int \rho_t^q d\nu_t} \partial_t \int \rho_t^q d\nu_t \\ &= \frac{1}{q-1} \frac{\int \rho_t^q \partial_t \nu_t + q \int \rho_t^{q-1} \cdot \frac{\nu_t \partial_t \mu_t - \mu_t \partial_t \nu_t}{\nu_t^2} d\nu_t}{\int \rho_t^q d\nu_t} \\ &= \frac{1}{q-1} \left( \int \rho_t^q d\nu_t \right)^{-1} (I_1 + I_2), \end{aligned} \quad (4.245)$$

其中

$$\begin{aligned}
I_1 &= \int \rho_t^q \partial_t \nu_t \\
&= \int \rho_t^q \left( -\nabla \cdot (\nu_t b_t) + \frac{c}{2} \Delta \nu_t \right) \\
&= \int \langle b_t, \nabla(\rho_t^q) \rangle d\nu_t - \frac{c}{2} \int \langle \nabla \log \nu_t, \nabla(\rho_t^q) \rangle d\nu_t \\
&= q \int \rho_t^{q-1} \langle b_t, \nabla \log \rho_t \rangle d\mu_t - \frac{cq}{2} \int \mu_t \rho_t^{q-1} \langle \nabla \log \mu_t, \nabla \log \rho_t \rangle d\mu_t \\
&\quad + \frac{cq}{2} \int \mu_t \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t.
\end{aligned} \tag{4.246}$$

注意：

$$\nabla(\rho_t^q) = q\rho_t^{q-1} \nabla \rho_t = q\rho_t^{q-1} \rho_t \nabla \log \rho_t = q\rho_t^q \nabla \log \rho_t, \tag{4.247}$$

$$\nu_t \rho_t^q = \mu_t \rho_t^{q-1}, \tag{4.248}$$

$$\nabla \log \nu_t = \nabla \log \mu_t - \nabla \log \rho_t. \tag{4.249}$$

我们可以计算：

$$\begin{aligned}
I_2 &= q \int \rho_t^{q-1} \frac{\nu_t \partial_t \mu_t - \mu_t \partial_t \nu_t}{\nu_t^2} d\nu_t \\
&= q \int \rho_t^{q-1} \partial_t \mu_t - q \int \rho_t^q \partial_t \nu_t \\
&= q \int \rho_t^{q-1} \left( -\nabla \cdot (\mu_t b_t) + \frac{c}{2} \Delta \mu_t \right) - qI_1 \\
&= q \int \langle \nabla(\rho_t^{q-1}), b_t \rangle d\mu_t - \frac{cq}{2} \int \langle \nabla \log \mu_t, \nabla(\rho_t^{q-1}) \rangle d\mu_t - qI_1 \\
&= q(q-1) \int \rho_t^{q-2} \langle \nabla \rho_t, b_t \rangle d\mu_t - \frac{cq(q-1)}{2} \int \rho_t^{q-2} \langle \nabla \rho_t, \nabla \log \mu_t \rangle d\mu_t - qI_1 \\
&= q(q-1) \int \rho_t^{q-1} \langle \nabla \log \rho_t, b_t \rangle d\mu_t \\
&\quad - \frac{cq(q-1)}{2} \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \log \mu_t \rangle d\mu_t - qI_1.
\end{aligned} \tag{4.250}$$

于是我们得到 Rényi 散度的时间导数：

$$\begin{aligned}
\partial_t R_q(\mu_t \| \nu_t) &= \frac{1}{q-1} \left( \int \rho_t^q d\nu_t \right)^{-1} \left\{ (1-q)I_1 + q(q-1) \int \rho_t^{q-1} \langle \nabla \log \rho_t, b_t \rangle d\mu_t \right. \\
&\quad \left. - \frac{cq(q-1)}{2} \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \log \mu_t \rangle d\mu_t \right\} \\
&= \left( \int \rho_t^q d\nu_t \right)^{-1} \left\{ -I_1 + q \int \rho_t^{q-1} \langle \nabla \log \rho_t, b_t \rangle d\mu_t \right. \\
&\quad \left. - \frac{cq}{2} \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \log \mu_t \rangle d\mu_t \right\} \\
&= -\frac{cq}{2} \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t \right).
\end{aligned} \tag{4.251}$$

现在我们可以计算相对 Rényi-Fisher 信息的时间导数。在此之前，我们需要计算一些量的时间导数作为准备：

$$\begin{aligned}
\partial_t \left( \int \rho_t^{q-1} d\mu_t \right) &= \int \partial_t \mu_t \rho_t^{q-1} + (q-1) \int \rho_t^{q-2} \partial_t \rho_t d\mu_t \\
&= \int \partial_t \mu_t \rho_t^{q-1} + (q-1) \int \rho_t^{q-2} \frac{\nu_t \partial_t \mu_t - \mu_t \partial_t \nu_t}{\nu_t^2} d\mu_t \\
&= q \int \rho_t^{q-1} \partial_t \mu_t - (q-1) \int \rho_t^q \partial_t \nu_t \\
&= q \int \langle b_t, \nabla(\rho_t^{q-1}) \rangle d\mu_t - \frac{cq}{2} \int \langle \nabla \rho_t^{q-1}, \nabla \log \mu_t \rangle d\mu_t \\
&\quad - (q-1) \int \rho_t^q \left( -\nabla \cdot (\nu_t b_t) + \frac{c}{2} \Delta \nu_t \right) \\
&= q(q-1) \int \rho_t^{q-1} \langle b_t, \nabla \log \rho_t \rangle d\mu_t \\
&\quad - \frac{cq(q-1)}{2} \int \rho_t^{q-1} \langle \nabla \log \mu_t, \nabla \log \rho_t \rangle d\mu_t \\
&\quad - q(q-1) \int \rho_t^{q-1} \langle b_t, \nabla \log \rho_t \rangle d\mu_t \\
&\quad + \frac{cq(q-1)}{2} \int \rho_t^{q-1} \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t \\
&= -\frac{cq(q-1)}{2} \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t, \tag{4.252}
\end{aligned}$$

$$\begin{aligned}
\partial_t \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t &= \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \partial_t \mu_t + (q-1) \int \rho_t^{q-2} \partial_t \rho_t \|\nabla \log \rho_t\|^2 d\mu_t \\
&\quad + 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \log \partial_t \rho_t \rangle d\mu_t \\
&= \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \partial_t \mu_t \\
&\quad + (q-1) \int \rho_t^{q-2} \|\nabla \log \rho_t\|^2 \frac{\nu_t \partial_t \mu_t - \mu_t \partial_t \nu_t}{\nu_t^2} d\mu_t \\
&\quad + 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \log \partial_t \rho_t \rangle d\mu_t \\
&= q \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \partial_t \mu_t - (q-1) \int \rho_t^q \|\nabla \log \rho_t\|^2 \partial_t \nu_t \\
&\quad + 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, \partial_t \nabla \log \rho_t \rangle d\mu_t \\
&:= qJ_1 - (q-1)J_2 + J_3, \tag{4.253}
\end{aligned}$$

其中

$$\begin{aligned}
J_1 &= \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \partial_t \mu_t \\
&= \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \left( -\nabla \cdot (\mu_t b_t) + \frac{c}{2} \Delta \mu_t \right)
\end{aligned}$$

$$= \int \left\langle b_t, \nabla \left( \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \right) \right\rangle d\mu_t - \frac{c}{2} \int \left\langle \nabla \log \mu_t, \nabla \left( \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \right) \right\rangle d\mu_t.$$

注意：

$$\begin{aligned} \nabla \left( \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \right) &= (q-1)\rho_t^{q-2} \|\nabla \log \rho_t\|^2 \nabla \rho_t + 2\rho_t^{q-1} (\nabla^2 \log \rho_t) (\nabla \log \rho_t) \\ &= (q-1)\rho_t^{q-1} \|\nabla \log \rho_t\|^2 \nabla \log \rho_t + 2\rho_t^{q-1} (\nabla^2 \log \rho_t) (\nabla \log \rho_t). \end{aligned}$$

我们有：

$$\begin{aligned} J_1 &= \int \left\langle b_t - \frac{c}{2} \nabla \log \mu_t, (q-1)\rho_t^{q-1} \|\nabla \log \rho_t\|^2 \nabla \log \rho_t + 2\rho_t^{q-1} (\nabla^2 \log \rho_t) (\nabla \log \rho_t) \right\rangle d\mu_t \\ &= (q-1) \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \left\langle b_t - \frac{c}{2} \nabla \log \mu_t, \nabla \log \rho_t \right\rangle d\mu_t \\ &\quad + 2 \int \rho_t^{q-1} \left\langle b_t - \frac{c}{2} \nabla \log \mu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \right\rangle d\mu_t. \end{aligned} \quad (4.254)$$

类似地，我们有：

$$\begin{aligned} J_2 &= q \int \rho_t^q \|\nabla \log \rho_t\|^2 \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, \nabla \log \rho_t \right\rangle d\nu_t \\ &\quad + 2 \int \rho_t^q \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \right\rangle d\nu_t \\ &= q \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, \nabla \log \rho_t \right\rangle d\mu_t \\ &\quad + 2 \int \rho_t^{q-1} \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \right\rangle d\mu_t. \end{aligned} \quad (4.255)$$

组合可得：

$$\begin{aligned} &qJ_1 - (q-1)J_2 \\ &= q(q-1) \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \left\langle b_t - \frac{c}{2} \nabla \log \mu_t, \nabla \log \rho_t \right\rangle d\mu_t \\ &\quad + 2q \int \rho_t^{q-1} \left\langle b_t - \frac{c}{2} \nabla \log \mu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \right\rangle d\mu_t \\ &\quad - q(q-1) \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, \nabla \log \rho_t \right\rangle d\mu_t \\ &\quad - 2(q-1) \int \rho_t^{q-1} \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \right\rangle d\mu_t \\ &= -\frac{cq(q-1)}{2} \int \rho_t^{q-1} \|\nabla \log \rho_t\|^4 d\mu_t + 2 \int \rho_t^{q-1} \left\langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \right\rangle d\mu_t \\ &\quad - c \int \rho_t^{q-1} \left\langle \nabla \log \mu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \right\rangle d\mu_t \\ &\quad - c(q-1) \int \rho_t^{q-1} \left\langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \right\rangle d\mu_t. \end{aligned} \quad (4.256)$$

接下来我们计算  $J_3$ 。定义  $u_t := \log \mu_t, w_t := \log \nu_t$ ，则可以推导出  $u_t, w_t$  满足以下 PDE:

$$\begin{aligned} \partial_t u_t &= \frac{\partial_t \mu_t}{\mu_t} = e^{-u_t} \left[ -\nabla \cdot (e^{u_t} b_t) + \frac{c}{2} \Delta(e^{u_t}) \right] \\ &= e^{-u_t} \left[ -e^{u_t} \langle \nabla u_t, b_t \rangle - e^{u_t} \nabla \cdot b_t + \frac{c}{2} e^{u_t} \|\nabla u_t\|^2 + \frac{c}{2} e^{u_t} \Delta u_t \right] \\ &= -\langle \nabla u_t, b_t \rangle - \nabla \cdot b_t + \frac{c}{2} \|\nabla u_t\|^2 + \frac{c}{2} \Delta u_t, \end{aligned} \quad (4.257)$$

$$\nabla \partial_t \mu_t = -(\nabla^2 u_t) b_t - (\nabla b_t) \nabla u_t - \nabla(\nabla \cdot b_t) + c(\nabla^2 u_t) \nabla u_t + \frac{c}{2} \nabla \Delta u_t. \quad (4.258)$$

于是我们得到:

$$\begin{aligned} J_3 &= 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, \partial_t \nabla \log \rho_t \rangle d\mu_t = 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \partial_t u_t - \nabla \partial_t w_t \rangle d\mu_t \\ &:= J_4 - J_5, \end{aligned} \quad (4.259)$$

$$\begin{aligned} J_4 &= 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \partial_t u_t \rangle d\mu_t \\ &= 2 \int \rho_t^{q-1} \left\langle \nabla \log \rho_t, -(\nabla^2 u_t) b_t - (\nabla b_t) \nabla u_t - \nabla(\nabla \cdot b_t) \right. \\ &\quad \left. + c(\nabla^2 u_t) \nabla u_t + \frac{c}{2} \nabla \Delta u_t \right\rangle d\mu_t \\ &= -2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 u_t) \nabla \log \rho_t \rangle d\mu_t - 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla b_t) \nabla u_t \rangle d\mu_t \\ &\quad - 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, \text{Tr}(\nabla^2 b_t) \rangle d\mu_t \\ &\quad + 2c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 u_t) \nabla u_t \rangle d\mu_t + c \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \Delta u_t \rangle d\mu_t, \end{aligned} \quad (4.260)$$

$$\begin{aligned} J_5 &= 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \partial_t w_t \rangle d\mu_t \\ &= 2 \int \rho_t^{q-1} \left\langle \nabla \log \rho_t, -(\nabla^2 w_t) b_t - (\nabla b_t) \nabla w_t - \nabla(\nabla \cdot b_t) \right. \\ &\quad \left. + c(\nabla^2 w_t) \nabla w_t + \frac{c}{2} \nabla \Delta w_t \right\rangle d\mu_t \\ &= -2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 w_t) \nabla \log \rho_t \rangle d\mu_t - 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla b_t) \nabla w_t \rangle d\mu_t \\ &\quad - 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, \text{Tr}(\nabla^2 b_t) \rangle d\mu_t \\ &\quad + 2c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 w_t) \nabla w_t \rangle d\mu_t + c \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \Delta w_t \rangle d\mu_t \end{aligned} \quad (4.261)$$

注意  $u_t - w_t = \log \frac{\mu_t}{\nu_t} = \log \rho_t$ ，我们有:

$$\begin{aligned} J_3 &= -2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t - 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla b_t) \nabla \log \rho_t \rangle d\mu_t \\ &\quad + 2c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 u_t) \nabla u_t - (\nabla^2 w_t) \nabla w_t \rangle d\mu_t \\ &\quad + c \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \Delta \log \rho_t \rangle d\mu_t. \end{aligned} \quad (4.262)$$

这里我们使用平坦黎曼流形设置下的 Bochner 公式来简化最后一项：

$$\frac{1}{2}\Delta(|\nabla u|^2) = |\nabla^2 u|^2 + \langle \nabla u, \nabla \Delta u \rangle. \quad (4.263)$$

因此我们可以得到：

$$\begin{aligned} J_3 &= -2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t - 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla b_t) \nabla \log \rho_t \rangle d\mu_t \\ &\quad + 2c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 u_t) \nabla u_t - (\nabla^2 w_t) \nabla w_t \rangle d\mu_t + c \int \rho_t^{q-1} \langle \nabla \log \rho_t, \nabla \Delta \log \rho_t \rangle d\mu_t \\ &= -2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t - 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla b_t) \nabla \log \rho_t \rangle d\mu_t \\ &\quad + 2c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 u_t) \nabla u_t - (\nabla^2 w_t) \nabla w_t \rangle d\mu_t \\ &\quad + c \int \rho_t^{q-1} \left( \frac{1}{2} \Delta \|\nabla \log \rho_t\|^2 - \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 \right) d\mu_t. \end{aligned} \quad (4.264)$$

注意：

$$\begin{aligned} (\nabla^2 u_t) \nabla u_t - (\nabla^2 w_t) \nabla w_t &= (\nabla^2 u_t) \nabla u_t - (\nabla^2 w_t) \nabla u_t + (\nabla^2 w_t) \nabla u_t - (\nabla^2 w_t) \nabla w_t \\ &= (\nabla^2 \log \rho_t) \nabla \log \mu_t + (\nabla^2 \log \nu_t) \nabla \log \rho_t. \end{aligned} \quad (4.265)$$

结合上述结果，我们有：

$$\begin{aligned} \partial_t \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t &= -\frac{cq(q-1)}{2} \int \rho_t^{q-1} \|\nabla \log \rho_t\|^4 d\mu_t \\ &\quad + 2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\ &\quad - c \int \rho_t^{q-1} \langle \nabla \log \mu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\ &\quad - c(q-1) \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\ &\quad - 2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\ &\quad - 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla b_t) \nabla \log \rho_t \rangle d\mu_t \\ &\quad + c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \mu_t \rangle d\mu_t \\ &\quad + 2c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \nu_t) \nabla \log \rho_t \rangle d\mu_t \\ &\quad + \frac{c}{2} \int \rho_t^{q-1} \Delta \|\nabla \log \rho_t\|^2 d\mu_t - c \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \\ &= -\frac{cq(q-1)}{2} \int \rho_t^{q-1} \|\nabla \log \rho_t\|^4 d\mu_t \\ &\quad + 2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\ &\quad - c(q-1) \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \end{aligned}$$

$$\begin{aligned}
& -2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla b_t) \nabla \log \rho_t \rangle d\mu_t \\
& + c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \mu_t \rangle d\mu_t \\
& + 2c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \nu_t) \nabla \log \rho_t \rangle d\mu_t \\
& + \frac{c}{2} \int \rho_t^{q-1} \Delta \|\nabla \log \rho_t\|^2 d\mu_t - c \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t,
\end{aligned}$$

其中

$$\begin{aligned}
\int \rho_t^{q-1} \Delta \|\nabla \log \rho_t\|^2 d\mu_t &= - \int \nabla(\mu_t \rho_t^{q-1}) \cdot \nabla \|\nabla \log \rho_t\|^2 \\
&= -2 \int (\rho_t^{q-1} \nabla \log \mu_t + (q-1) \rho_t^{q-1} \nabla \log \rho_t) \cdot (\nabla^2 \log \rho_t) \nabla \log \rho_t d\mu_t \\
&= -2 \int \rho_t^{q-1} \langle \nabla \log \mu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad - 2(q-1) \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t. \tag{4.266}
\end{aligned}$$

因此我们有：

$$\begin{aligned}
\partial_t \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t &= -\frac{cq(q-1)}{2} \int \rho_t^{q-1} \|\nabla \log \rho_t\|^4 d\mu_t \\
&\quad + 2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad - c(q-1) \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad - 2 \int \rho_t^{q-1} \langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad - 2 \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla b_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad + c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \mu_t \rangle d\mu_t \\
&\quad - c \int \rho_t^{q-1} \langle \nabla \log \mu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad + 2c \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \nu_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad - c \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \\
&\quad - c(q-1) \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
&= -\frac{cq(q-1)}{2} \int \rho_t^{q-1} \|\nabla \log \rho_t\|^4 d\mu_t - c \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \\
&\quad - 2c(q-1) \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad + 2 \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{A_t^{\nu, b}}^2 d\mu_t. \tag{4.268}
\end{aligned}$$

然后我们可以得到相对 Rényi-Fisher 信息的时间导数：

$$\begin{aligned}
\partial_t \text{RFI}_q(\mu_t \| \nu_t) &= q \left( \int \rho_t^{q-1} d\mu_t \right)^{-2} \left\{ \left( \partial_t \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t \right) \left( \int \rho_t^{q-1} d\mu_t \right) \right\} \\
&\quad - q \left( \int \rho_t^{q-1} d\mu_t \right)^{-2} \left\{ \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t \right) \left( \partial_t \int \rho_t^{q-1} d\mu_t \right) \right\} \\
&= -q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \frac{cq(q-1)}{2} \int \rho_t^{q-1} \|\nabla \log \rho_t\|^4 d\mu_t \right. \\
&\quad \left. + c \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right\} \\
&\quad - 2cq(q-1) \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \right\} \\
&\quad + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{A_t^{\nu, b}}^2 d\mu_t \right\} \\
&\quad + \frac{cq^2(q-1)}{2} \left( \int \rho_t^{q-1} d\mu_t \right)^{-2} \left\{ \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t \right)^2 \right\} \\
&= \frac{cq^2(q-1)}{2} \left( \int \rho_t^{q-1} d\mu_t \right)^{-2} \left\{ \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t \right)^2 \right. \\
&\quad \left. - \left( \int \rho_t^{q-1} d\mu_t \right) \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^4 d\mu_t \right) \right\} \\
&\quad - cq \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\
&\quad - 2cq(q-1) \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \right\} \\
&\quad + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{A_t^{\nu, b}}^2 d\mu_t \right\} \\
&\leq -cq \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\
&\quad - 2cq(q-1) \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \right\} \\
&\quad + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{A_t^{\nu, b}}^2 d\mu_t \right\}. \tag{4.269}
\end{aligned}$$

证毕。

#### 4.5.4 计算时间导数的技术性论证

在本节中，我们提供了计算各定理中时间导数的技术细节。

##### 4.5.4.1 积分与微分交换次序的合理性验证

本节记录了正文中使用的积分与微分交换次序的自洽性证明，并附带两个便捷的推论及关于适用范围的简要注记。

设  $I \subset \mathbb{R}$  为开区间,  $(\mathcal{X}, \mathcal{F}, \mu)$  为有限测度空间。以下所有函数均假设在  $I \times \mathcal{X}$  上可测。对于  $t \in I$ , 记

$$F(t) := \int_{\mathcal{X}} f(t, x) d\mu, \quad g(t, x) := \partial_t f(t, x) \quad (4.270)$$

只要 (关于  $t$  的) 逐点导数存在。除非另有说明, 积分均是关于测度  $\mu$  的。

**定理 4.22** 设  $f: I \times \mathcal{X} \rightarrow \mathbb{R}$  和  $g: I \times \mathcal{X} \rightarrow \mathbb{R}$  满足:

1. (逐点可微性) 对于  $\mu$ -几乎处处的  $x \in \mathcal{X}$ , 映射  $t \mapsto f(t, x)$  在  $I$  上可微, 且  $\partial_t f(t, x) =: g(t, x)$ 。

2. ( $(t, x)$  的局部可积性)  $g \in \mathbb{L}_{\text{loc}}^1(I; \mathbb{L}^1(\mu))$ ; 即, 对于任意紧子集  $J \subset I$ ,

$$\int_J \int_{\mathcal{X}} |g(t, x)| d\mu dt < \infty. \quad (4.271)$$

3. (单时刻切片可积性) 存在  $t_0 \in I$  使得  $f(t_0, \cdot) \in \mathbb{L}^1(\mu)$ 。

则  $F: I \rightarrow \mathbb{R}$  是绝对连续的, 且对于几乎处处的  $t \in I$ ,

$$F'(t) = \int_{\mathcal{X}} g(t, x) d\mu = \int_{\mathcal{X}} \partial_t f(t, x) d\mu. \quad (4.272)$$

**证明** 固定包含  $t_0$  的紧子集  $J \subset I$ 。由  $g(t, x)$  的局部可积性和 Fubini 定理, 对于  $\mu$ -几乎处处的  $x$ , 有  $g(\cdot, x) \in \mathbb{L}^1(J)$ 。结合逐点可微性, 一维微积分基本定理表明  $f(\cdot, x)$  在  $J$  上绝对连续且满足 Newton-Leibniz 公式, 即对于所有  $t \in J$ ,

$$f(t, x) = f(t_0, x) + \int_{t_0}^t g(s, x) ds \quad \text{对几乎处处的 } x. \quad (4.273)$$

对  $x$  积分, 并利用 Fubini 定理及  $f(t_0, \cdot) \in \mathbb{L}^1(\mu)$  交换积分次序, 可得

$$F(t) = F(t_0) + \int_{t_0}^t \left( \int_{\mathcal{X}} g(s, x) d\mu \right) ds. \quad (4.274)$$

因此  $F$  在  $J$  上绝对连续, 且对于  $J$  中几乎处处的  $t$ , 有  $F'(t) = \int_{\mathcal{X}} g(t, x) d\mu$ 。由于  $J \subset I$  的任意性, 结论在  $I$  上成立。 ■

**推论 4.8** 假设定理 4.22 的条件成立, 且额外满足:

4. (紧集上的一致控制) 对于任意紧子集  $J \subset I$ , 存在  $G_J \in \mathbb{L}^1(\mu)$  使得

$$\sup_{t \in J} |g(t, x)| \leq G_J(x) \quad \text{对 } \mu\text{-几乎处处的 } x. \quad (4.275)$$

则恒等式

$$\frac{d}{dt} \int_{\mathcal{X}} f(t, x) d\mu = \int_{\mathcal{X}} \partial_t f(t, x) d\mu \quad (4.276)$$

对所有  $t \in I$  成立 (而不仅仅是几乎处处)。

**证明** 由定理 4.22, 对于包含  $t_0$  的任意  $t \in J \subset I$ :

$$f(t, x) = f(t_0, x) + \int_{t_0}^t g(s, x) ds \quad \text{对几乎处处的 } x. \quad (4.277)$$

对于绝对值足够小的  $h$ , 使得  $t, t+h \in J$ , 我们有

$$\frac{f(t+h, x) - f(t, x)}{h} = \frac{1}{h} \int_t^{t+h} g(s, x) ds, \quad \text{对 } \mu\text{-几乎处处的 } x. \quad (4.278)$$

由于映射  $t \mapsto f(t, x)$  在  $t$  处对几乎处处的  $x$  可微, 两边取极限  $h \rightarrow 0$  可得:

$$\lim_{h \rightarrow 0} \frac{f(t+h, x) - f(t, x)}{h} = g(t, x) = \lim_{h \rightarrow 0} \frac{1}{h} \int_t^{t+h} g(s, x) ds, \quad \text{对 } \mu\text{-几乎处处的 } x. \quad (4.279)$$

根据一致控制假设, 对于上述固定的紧区间  $J$ , 有

$$\left| \frac{1}{h} \int_t^{t+h} g(s, x) dx \right| \leq \sup_{s \in J} |g(s, x)| \leq G_J(x) \in \mathbb{L}^1(\mu). \quad (4.280)$$

因此应用控制收敛定理可得

$$\lim_{h \rightarrow 0} \int_{\mathcal{X}} \frac{f(t+h, x) - f(t, x)}{h} d\mu \leq \int_{\mathcal{X}} g(t, x) d\mu, \quad (4.281)$$

即,

$$\lim_{h \rightarrow 0} \frac{F(t+h) - F(t)}{h} \leq \int_{\mathcal{X}} g(t, x) d\mu. \quad (4.282)$$

因此时间导数  $F'(t)$  对所有  $t \in J$  存在, 且

$$F'(t) = \int_{\mathcal{X}} g(t, x) d\mu = \int_{\mathcal{X}} \partial_t f(t, x) d\mu. \quad (4.283)$$

由于  $J \subset I$  的任意性, 结论在  $I$  上成立。 ■

**推论 4.9** 假设对于所有  $t \in I$  均有  $f(t, \cdot) \in \mathbb{L}^1(\mu)$ , 且  $t \mapsto f(t, \cdot)$  作为  $\mathbb{L}^1(\mu)$ -值映射是  $C^1$  的, 并且  $\partial_t f(t, \cdot) \in \mathbb{L}^1(\mu)$ 。则

$$\frac{d}{dt} \int_{\mathcal{X}} f(t, x) d\mu = \int_{\mathcal{X}} \partial_t f(t, x) d\mu \quad \text{对所有 } t \in I. \quad (4.284)$$

**注** 仅凭“关于  $t$  的逐点可微性”和“单个可积切片  $f(t_0, \cdot) \in \mathbb{L}^1(\mu)$ ”不足以保证微分与积分次序的交换。设  $I = (0, 1)$  且  $(\mathcal{X}, \mu) = ((0, 1), \mathcal{L}^1)$ , 其中  $\mathcal{L}^1$  为 Lebesgue 测度。设  $\psi \in C_c^\infty((-1, 1))$  为非负函数且  $\int_{\mathbb{R}} \psi = 1$ 。将 Cantor 集的补集写为不相交开区间的并集

$$(0, 1) \setminus \mathcal{C} = \bigcup_{n=1}^{\infty} I_n, \quad I_n = (a_n - \delta_n, a_n + \delta_n), \quad (4.285)$$

这在  $(0, 1)$  中是稠密的。对于每个  $n$ , 定义重缩放的函数

$$\psi_n(t) := \frac{1}{\delta_n} \psi\left(\frac{t - a_n}{\delta_n}\right), \quad \text{使得} \quad \text{supp}\psi_n \subset I_n, \quad \int_0^1 \psi_n(t) dt = 1. \quad (4.286)$$

由于区间  $\{I_n\}$  两两不相交, 任意  $t \in (0, 1)$  至多属于一个  $I_k$ , 因此

$$\sum_{n=1}^{\infty} \psi_n(t) \leq \frac{1}{\delta_k} \sup_{s \in I_k} \psi(s) < \infty. \quad (4.287)$$

在空间方面, 取  $(0, 1)$  中两两不相交的区间  $A_n := (2^{-n}, 2^{-(n-1)})$  并设

$$b_n(x) := \frac{\chi_{A_n}(x)}{\mu(A_n)} = \frac{\chi_{A_n}(x)}{2^{-n}}, \quad (4.288)$$

使得  $\|b_n\|_{\mathbb{L}^1(\mu)} = 1$  且几乎每个  $x \in (0, 1)$  至多属于一个  $A_n$ 。定义

$$g(t, x) := \sum_{n=1}^{\infty} \psi_n(t) b_n(x), \quad f(t, x) := \int_0^t g(s, x) ds. \quad (4.289)$$

逐点可微性与单个可积切片。对于几乎处处的  $x$ , 至多存在一个下标  $n(x)$  使得  $x \in A_{n(x)}$ 。因此

$$g(t, x) = \psi_{n(x)}(t) b_{n(x)}(x), \quad f(t, x) = b_{n(x)}(x) \int_0^t \psi_{n(x)}(s) ds, \quad (4.290)$$

所以  $t \mapsto f(t, x)$  对所有  $t \in I$  可微且  $\partial_t f(t, x) = g(t, x)$ 。此外,  $f(0, \cdot) \equiv 0 \in \mathbb{L}^1(\mu)$ 。

在  $(t, x)$  上局部  $\mathbb{L}^1$  可积性的失效。对于任意紧区间  $J \subset I$ ,  $\cup_n I_n$  的稠密性意味着存在无穷多个  $n$  使得  $I_n \subset J$ , 于是

$$\iint_{J \times \mathcal{X}} |g(t, x)| d\mu dt = \int_J \sum_n \psi_n(t) dt = \sum_n \int_J \psi_n(t) dt \geq \sum_{I_n \subset J} 1 = \infty. \quad (4.291)$$

因此  $g \notin \mathbb{L}_{\text{loc}}^1(I; \mathbb{L}^1(\mu))$ 。

有限瞬时积分但这导致原函数发散。对于每个  $t \in (0, 1)$ ,

$$\int_{\mathcal{X}} g(t, x) d\mu = \sum_n \psi_n(t) < \infty, \quad (4.292)$$

因为至多有一个  $\psi_n(t)$  非零。然而,

$$F(t) := \int_{\mathcal{X}} f(t, x) d\mu = \sum_{n=1}^{\infty} \int_0^t \psi_n(s) ds = \#\{n : I_n \subset (0, t)\} = \infty \quad \text{对所有 } t > 0. \quad (4.293)$$

因此  $F$  不是绝对连续的 (甚至在  $(0, 1)$  上都不是有限的), 这表明在缺乏局部  $\mathbb{L}^1$  条件  $g \in \mathbb{L}_{\text{loc}}^1(I; \mathbb{L}^1(\mu))$  的情况下, 仅凭关于  $t$  的逐点可微性和单个可积切片不足以保证积分号下的微分是合法的。

**注 ( $\sigma$ -有限测度)** 定理 4.22 中  $\mu$  的有限性并非本质要求。如果  $(\mathcal{X}, \mathcal{F}, \mu)$  仅是  $\sigma$ -有限的, 定理 4.22 和推论 4.8 的结论在相同假设下仍然成立 (将  $\mathbb{L}^1(\mu)$  理解为  $\sigma$ -有限意义下的)。

固定紧区间  $J \subset I$  并选取一个递增的穷竭序列  $\{E_n\}_{n \geq 1} \subset \mathcal{F}$ , 满足  $0 < \mu(E_n) < \infty$  且  $\bigcup_{n=1}^{\infty} E_n = \mathcal{X}$ 。在每个有限测度空间  $(E_n, \mathcal{F} \cap E_n, \mu|_{E_n})$  上, 定理 4.22 给出

$$F_n(t) := \int_{E_n} f(t, x) d\mu = F_n(t_0) + \int_{t_0}^t \left( \int_{E_n} g(s, x) d\mu \right) ds, \quad F'_n(t) = \int_{E_n} g(t, x) d\mu \quad (4.294)$$

对几乎处处的  $t \in J$ 。

由于  $g \in \mathbb{L}^1(J \times \mathcal{X})$ , 我们可以写为

$$\int_{E_n} g(s, x) d\mu = \int_{\mathcal{X}} g(s, x) \chi_{E_n}(x) d\mu. \quad (4.295)$$

当  $n \rightarrow \infty$  时, 对所有  $x \in \mathcal{X}$  有  $\chi_{E_n}(x) \nearrow 1$ , 且  $|g(s, x)|$  在  $J \times \mathcal{X}$  上可积。因此由单调收敛定理 (应用于  $|g|\chi_{E_n}$ ) 和控制收敛定理 (应用于  $g\chi_{E_n}$ ),

$$\lim_{n \rightarrow \infty} \int_{E_n} |g(s, x)| d\mu = \int_{\mathcal{X}} |g(s, x)| d\mu, \quad \lim_{n \rightarrow \infty} \int_{E_n} g(s, x) d\mu = \int_{\mathcal{X}} g(s, x) d\mu, \quad (4.296)$$

对几乎处处的  $s \in J$ 。类似地,

$$\lim_{n \rightarrow \infty} F_n(t) = \int_{\mathcal{X}} f(t, x) d\mu =: F(t). \quad (4.297)$$

在  $F_n$  的恒等式中取极限, 得到

$$F(t) = F(t_0) + \int_{t_0}^t \left( \int_{\mathcal{X}} g(s, x) d\mu \right) ds, \quad F'(t) = \int_{\mathcal{X}} g(t, x) d\mu \quad \text{对几乎处处的 } t \in J. \quad (4.298)$$

在推论 4.8 的控制假设下, 将“对几乎处处  $t$ ”改进为“对所有  $t$ ”可以通过对每个  $E_n$  应用相同的论证并令  $n \rightarrow \infty$  获得。

#### 4.5.4.2 分部积分的适用性论证

对于定理 4.6 证明中的分部积分步骤, 我们需要确保当积分域扩张至  $\mathbb{R}^d$  时边界项消失。这可以在以下充分条件下得到保证:

- $\rho_t = \frac{\mu_t}{\nu_t}$  有上界;
- $\nabla \log \rho_t, b_t, \nabla \cdot b_t$  可被某些多项式函数界定;
- FPE 具有唯一的指数衰减光滑解, 即存在常数  $C_1, C_2, p > 0$  使得对所有  $x \in \mathbb{R}^d$ ,  $\mu_t(x), \nu_t(x) \leq C_1 e^{-C_2 \|x\|^p}$ 。

这是因为在定理 4.6 证明的每一步中, 被积函数都由  $\nabla \log \rho_t, b_t, \nabla \cdot b_t$  和  $\rho_t$  的多项式界定。因此, 由于  $\mu_t, \nu_t$  的指数衰减, 边界项在积分域扩张至  $\mathbb{R}^d$  时消失。

对于第一个假设, 如果  $\rho_0$  有上界, 则通过抛物型 PDE 的比较原理可以满足。此外, 当  $b_t(x) = -\nabla V(x)$  且多项式势函数  $V$  满足  $e^{-V} \in \mathbb{L}^1(\mathbb{R}^d)$  时, 指数衰减性可由定理 3.1 中相应条件下的 Feynman-Kac 表示推导得出。

关于  $\nabla \log \rho_t$  的多项式有界性, 在 Langevin 情形  $b_t(x) = -\nabla V(x)$  下, 如第 4.4.3 节所述, 可以获得如下显式表示:

$$\nabla \log \rho_t(x) = \mathbb{E}_{f_{\mu,x}^W}(A_t^x) - \mathbb{E}_{f_{\nu,x}^W}(A_t^x), \quad (4.299)$$

其中

$$\begin{aligned} A_t^x &:= -\int_0^t \nabla h(X_s^x) \frac{\sinh(t-s)}{\sinh(t)} ds + \frac{2e^{-t}}{c(1-e^{-2t})}(X_t^x - e^{-t}x), \\ f_{\mu}(x) &:= \frac{d\mu_0}{d\mu_{\infty}}(x), \quad f_{\nu}(x) := \frac{d\nu_0}{d\nu_{\infty}}(x), \quad \mu_{\infty} = \nu_{\infty} \propto e^{-\frac{2}{c}V}, \\ W(x) &= \frac{2}{c} \left( \frac{\|x\|^2}{2} - V(x) \right), \\ h(x) &= \frac{1}{2c} (\|\nabla V(x)\|^2 - \|x\|^2) - \frac{\Delta V}{2} + \frac{d}{2}. \end{aligned}$$

这里  $(X_s^x)_{s \geq 0}$  表示以下方程的唯一强解:

$$X_t^x = x + \sqrt{c}W_t - \int_0^t X_s^x ds, \quad (4.300)$$

其中  $(W_t)_{t \geq 0}$  为标准布朗运动, 即

$$X_t^x = e^{-t} \left( x + \int_0^t e^s \sqrt{c} dW_s \right). \quad (4.301)$$

对于有界可测函数  $f$  和  $x \in \mathbb{R}^d$ , 通过下式定义 Wiener 空间上的概率测度  $\mathbb{Q}_{f,x}$ :

$$\mathbb{Q}_{f,x}(\Gamma) = \frac{1}{P_t^h f(x)} \int_{\Gamma} f(X_t^x) e^{-\int_0^t V(X_s^x) ds} d\mathbb{P}, \quad (4.302)$$

其中

$$P_t^h f(x) = \mathbb{E} \left[ f(X_t^x) e^{-\int_0^t V(X_s^x) ds} \right], \quad (4.303)$$

且  $\mathbb{E}_{f,x}^W$  表示关于  $\mathbb{Q}_{e^{W/2}, f, x}$  的期望。注意  $h$  具有多项式界 (因为我们假设  $V$  是多项式), 特别地, 我们假设  $h$  满足:

$$\max\{|h(x)|, \|\nabla h(x)\|\} \leq C_N \left( 1 + \sum_{k=1}^N \|x\|^k \right). \quad (4.304)$$

因此我们有

$$\|X_s^x\| \leq e^{-s} \left( \|x\| + \left\| \int_0^s e^r \sqrt{c} dW_r \right\| \right), \quad (4.305)$$

$$\begin{aligned} \|\nabla h(X_s^x)\| &\leq C \cdot \sum_{k=0}^N e^{-ks} \left( \|x\| + \left\| \int_0^s e^r \sqrt{c} dW_r \right\| \right)^k \\ &\leq C_N \sum_{k=0}^N e^{-ks} \left( \|x\|^k + \left\| \int_0^s e^r \sqrt{c} dW_r \right\|^k \right). \end{aligned} \quad (4.306)$$

此估计中还使用了核函数  $\frac{\sinh(t-s)}{\sinh(t)} \in [0, 1]$  (对于  $t \geq 0$ ) 的一致上下界。具体而言,

$$\begin{aligned} \|\nabla \log \rho_t(x)\| &\leq \mathbb{E}_{f_\mu, x}^W \left\| \int_0^t \nabla h(X_s^x) \frac{\sinh(t-s)}{\sinh(t)} ds \right\| \\ &\quad + \mathbb{E}_{f_\nu, x}^W \left\| \int_0^t \nabla h(X_s^x) \frac{\sinh(t-s)}{\sinh(t)} ds \right\| \\ &\quad + \frac{2e^{-t}}{c(1-e^{-2t})} \left( \mathbb{E}_{f_\mu, x}^W \|X_t^x\| + \mathbb{E}_{f_\nu, x}^W \|X_t^x\| + 2e^{-t} \|x\| \right) \\ &\leq \mathbb{E}_{f_\mu, x}^W \left[ \int_0^t \|\nabla h(X_s^x)\| ds \right] + \mathbb{E}_{f_\nu, x}^W \left[ \int_0^t \|\nabla h(X_s^x)\| ds \right] \\ &\quad + \frac{2e^{-t}}{c(1-e^{-2t})} \left( \mathbb{E}_{f_\mu, x}^W \|X_t^x\| + \mathbb{E}_{f_\nu, x}^W \|X_t^x\| + 2e^{-t} \|x\| \right). \end{aligned} \quad (4.307)$$

如果进一步假设  $\bar{\mu}_0 \propto e^{V/c} \mu_0, \bar{\nu}_0 \propto e^{V/c} \nu_0$  是  $\sigma^2$ -亚高斯的 (其中  $\sigma^2 < c$ ), 即  $\bar{\mu}_0, \bar{\nu}_0$  满足尾概率界:

$$\mathbb{P}(\|X\| \geq t) \leq 2 \exp(-t^2/\sigma^2), \text{ 当 } X \sim \bar{\mu}_0 \text{ 或 } X \sim \bar{\nu}_0 \text{ 时}, \quad (4.308)$$

则  $e^{W/2} f_\mu(x) = e^{\frac{V(x)+\|x\|^2}{c}} \mu_0(x)$  和  $e^{W/2} f_\nu(x) = e^{\frac{V(x)+\|x\|^2}{c}} \nu_0(x)$  也是亚高斯的, 这确保了  $\mathbb{Q}_{f_\mu, x}, \mathbb{Q}_{f_\nu, x}$  是合法的概率测度, 且由于亚高斯性,  $\mathbb{Q}_{f_\mu, x}, \mathbb{Q}_{f_\nu, x}$  的所有矩确实存在。因此我们可以获得关于  $\|x\|$  的  $\nabla \log \rho_t(x)$  的多项式界:

$$\begin{aligned} \|\nabla \log \rho_t(x)\| &\leq \mathbb{E}_{f_\mu, x}^W \left[ \int_0^t \|\nabla h(X_s^x)\| ds \right] + \mathbb{E}_{f_\nu, x}^W \left[ \int_0^t \|\nabla h(X_s^x)\| ds \right] \\ &\quad + \frac{2e^{-t}}{c(1-e^{-2t})} \left( \mathbb{E}_{f_\mu, x}^W \|X_t^x\| + \mathbb{E}_{f_\nu, x}^W \|X_t^x\| + 2e^{-t} \|x\| \right) \\ &\leq t \cdot \mathbb{E}_{f_\mu, x}^W \left[ \sup_{s \in [0, t]} \|\nabla h(X_s^x)\| \right] + t \cdot \mathbb{E}_{f_\nu, x}^W \left[ \sup_{s \in [0, t]} \|\nabla h(X_s^x)\| \right] \\ &\quad + \frac{2e^{-t}}{c(1-e^{-2t})} \left( \mathbb{E}_{f_\mu, x}^W \|X_t^x\| + \mathbb{E}_{f_\nu, x}^W \|X_t^x\| + 2e^{-t} \|x\| \right) \\ &\leq t \cdot C_N \sum_{k=0}^N \left\{ 2 \|x\|^k + \mathbb{E}_{f_\mu, x}^W \left[ \sup_{s \in [0, t]} \left\| \int_0^s e^r \sqrt{c} dW_r \right\|^k \right] \right\} \end{aligned}$$

$$\begin{aligned}
& + \mathbb{E}_{f_{\nu,x}}^W \left[ \sup_{s \in [0,t]} \left\| \int_0^s e^r \sqrt{c} dW_r \right\|^k \right] \Big\} + C_1(t) \|x\| + C_2(t) \\
& \leq C_N(t) \cdot \left\{ \sum_{k=0}^N \|x\|^k + C'_N(t) \cdot \sum_{k=0}^N \mathbb{E}_{f_{\mu,x}}^W \left[ \left( \int_0^t e^{2s} c ds \right)^{k/2} \right] \right. \\
& \quad \left. + C'_N(t) \cdot \sum_{k=0}^N \mathbb{E}_{f_{\nu,x}}^W \left[ \left( \int_0^t e^{2s} c ds \right)^{k/2} \right] \right\} + C_1(t) \|x\| + C_2(t) \\
& \leq C''_N(t) \cdot \left( 1 + \sum_{k=1}^N \|x\|^k \right), \tag{4.309}
\end{aligned}$$

其中我们在第四个不等号处使用了 Burkholder-Davis-Gundy 不等式 [134-135]。结合上述获得的估计，我们现在可以将主要结果总结为以下定理。

**定理 4.23** 假设  $\mu_t$  和  $\nu_t$  沿 FP 流演化：

$$\begin{aligned}
\partial_t \mu_t &= -\nabla \cdot (\mu_t b_t) + \frac{c}{2} \Delta \mu_t, \\
\partial_t \nu_t &= -\nabla \cdot (\nu_t b_t) + \frac{c}{2} \Delta \nu_t.
\end{aligned} \tag{4.310}$$

对于任意  $t \geq 0, 1 \leq q < +\infty$ ，假设

- 初始密度比  $\rho_0 = \frac{\mu_0}{\nu_0}$  有上界；
- $b_t(x) = -\nabla V(x)$ ，其中  $V$  为多项式势函数且满足  $e^{-V} \in \mathbb{L}^1(\mathbb{R}^d)$ ；
- 分布  $\bar{\mu}_0 \propto e^{V/c} \mu_0, \bar{\nu}_0 \propto e^{V/c} \nu_0$  对于某些  $\sigma^2 < c$  均为  $\sigma^2$ -亚高斯分布；
- 相对 Rényi-Fisher 信息有限，即  $\text{RFI}_q(\mu_t \| \nu_t) < +\infty$ 。

则定理 4.6 证明中的分部积分步骤是合理的。

关于相对  $\Phi$ -Fisher 信息，类似的条件可用于论证定理 4.10 证明中的分部积分步骤。具体而言，如果我们假设  $\Phi \in C^4(\mathbb{R})$ ，或者至少  $\Phi \in C^4(\mathbb{R}_+)$ ，则由于  $\rho_t$  的有界性， $|\Phi^{(i)}(\rho_t)|$  对于  $i = 1, 2, 3, 4$  均有界。因此，与定理 4.23 相同的条件可用于验证定理 4.10 证明中的分部积分步骤。

#### 4.5.5 定理 4.7 的证明

在 Gaussian 信道情形下，相对 Rényi-Fisher 信息的时间导数由下式给出：

$$\begin{aligned}
\text{RFI}_q(\mu_t \| \nu_t) &= q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t \right), \tag{4.311} \\
\partial_t \text{RFI}_q(\mu_t \| \nu_t) &\leq -q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\
&\quad - 2q(q-1) \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \right\}
\end{aligned}$$

$$\begin{aligned}
& + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{\nabla^2 \log \nu_t}^2 d\mu_t \right\} \\
= & -q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\
& + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \\
& \cdot \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{(q\nabla^2 \log \nu_t - (q-1)\nabla^2 \log \mu_t)}^2 d\mu_t \right\}. \tag{4.312}
\end{aligned}$$

**第 (1) 部分:** 如果  $\nabla^2 \log \nu_t \preceq (1 - q^{-1})\nabla^2 \log \mu_t$ , 则有

$$\partial_t \text{RFI}_q(\mu_t \| \nu_t) \leq 0. \tag{4.313}$$

因此:

$$\text{RFI}_q(\mu_t \| \nu_t) \leq \text{RFI}_q(\mu_0 \| \nu_0). \tag{4.314}$$

**第 (2) 部分:** 如果对于  $K > 0$ , 有  $\nabla^2 \log \nu_t \preceq K\nabla^2 \log \mu_t$ , 则:

$$\begin{aligned}
\partial_t \text{RFI}_q(\mu_t \| \nu_t) & \leq 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{(q\nabla^2 \log \nu_t - (q-1)\nabla^2 \log \mu_t)}^2 d\mu_t \right\} \\
& \leq 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{(q\nabla^2 \log \nu_t - \frac{q-1}{K}\nabla^2 \log \mu_t)}^2 d\mu_t \right\}.
\end{aligned}$$

若  $\nu_0$  满足  $\alpha$ -SLC, 则  $\nu_t$  满足  $\alpha_t = \frac{\alpha}{1 + \alpha t}$ -SLC, 从而

$$\partial_t \text{RFI}_q(\mu_t \| \nu_t) \leq 2 \cdot \frac{1 + (K-1)q}{K} \alpha_t \text{RFI}_q(\mu_t \| \nu_t). \tag{4.315}$$

因此, 我们得到

$$\text{RFI}_q(\mu_t \| \nu_t) \leq \frac{\text{RFI}_q(\mu_0 \| \nu_0)}{(1 + \alpha t)^{2(1+(K-1)q)/K}}. \tag{4.316}$$

此外, 如果  $K = 0, q = 1$ , 我们可以通过相同的论证获得相对 FI 的 SDPI, 此时我们约定  $\frac{1+(K-1)q}{K} = 1, \frac{q-1}{K} = 0$ .

**第 (3) 部分:**

$$\begin{aligned}
\partial_t \text{RFI}_q(\mu_t \| \nu_t) & \leq 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{(q\nabla^2 \log \nu_t - (q-1)\nabla^2 \log \mu_t)}^2 d\mu_t \right\} \\
& \leq 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{(1+(K-1)q)\nabla^2 \log \mu_t}^2 d\mu_t \right\} \\
& \leq 2(1 + (K-1)q) \alpha_t \text{RFI}_q(\mu_t \| \nu_t). \tag{4.317}
\end{aligned}$$

因此, 我们得到

$$\text{RFI}_q(\mu_t \| \nu_t) \leq \frac{\text{RFI}_q(\mu_0 \| \nu_0)}{(1 + \alpha t)^{2(1+(K-1)q)}}. \tag{4.318}$$

**第(4)部分:** 如果  $\nabla^2 \log \nu_t \preceq (1 - q^{-1})\nabla^2 \log \mu_t$ , 则有

$$\begin{aligned} \partial_t \text{RFI}_q(\mu_t \| \nu_t) &\leq -q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\ &:= -q \left( \int \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\tilde{\mu}_t \right), \end{aligned} \quad (4.319)$$

其中

$$\tilde{\mu}_t = \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \mu_t \rho_t^{q-1} \quad (4.320)$$

是相对于  $\mu_t$  和  $\rho_t$  的扰动概率测度。由于我们有控制条件

$$0 < C_1 \leq \frac{d\mu_0}{d\nu_0} = \rho_0 \leq C_2 < \infty, \quad (4.321)$$

由抛物型 PDE 的比较原理, 对于所有  $t \geq 0$ , 有

$$0 < C_1 \leq \frac{d\mu_t}{d\nu_t} = \rho_t \leq C_2 < \infty \quad (4.322)$$

成立。因此,  $\tilde{\mu}_t$  确实是  $\mu_t$  的有界扰动:

$$0 < \left( \frac{C_1}{C_2} \right)^{q-1} \leq \frac{d\tilde{\mu}_t}{d\mu_t} = \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \rho_t^{q-1} \leq \left( \frac{C_2}{C_1} \right)^{q-1} < \infty, \text{ 当 } q \geq 1. \quad (4.323)$$

**引理 4.9** [Holley-Stroock 扰动 [136]] 假设  $\pi$  满足庞加莱或对数索伯列夫不等式。如果  $\mu$  满足  $0 < c \leq \frac{d\mu}{d\pi} \leq C < \infty$ , 则  $\mu$  也满足相应的泛函不等式, 常数分别为

$$C_{\text{PI}}(\mu) \leq \frac{C}{c} C_{\text{PI}}(\pi), \text{ 或 } C_{\text{LSI}}(\mu) \leq \frac{C}{c} C_{\text{LSI}}(\pi). \quad (4.324)$$

这里常数  $C_{\text{PI}}(\mu)$  或  $C_{\text{LSI}}(\mu)$  定义为使得  $\mu$  满足相应泛函不等式的最小常数:

$$\begin{aligned} \text{Var}_\mu(f) &\leq C_{\text{PI}}(\mu) \mathbb{E}_\mu[\|\nabla f\|^2], \\ \text{Ent}_\mu(f^2) &\leq 2C_{\text{LSI}}(\mu) \mathbb{E}_\mu[\|\nabla f\|^2], \end{aligned} \quad (4.325)$$

其中  $\text{Ent}_\mu(f) := \mathbb{E}_\mu(f \log f) - \mathbb{E}_\mu(f) \log \mathbb{E}_\mu(f)$ 。

由上述 Holley-Stroock 扰动引理, 我们可以得出  $\tilde{\mu}_t$  满足庞加莱不等式:

$$\begin{aligned} \int \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\tilde{\mu}_t &\geq \left( \frac{C_1}{C_2} \right)^{2q-2} \beta_t \text{Var}_{\tilde{\mu}_t}(\nabla \log \rho_t) \\ &= \left( \frac{C_1}{C_2} \right)^{2q-2} \beta_t \left[ \frac{1}{q} \text{RFI}_q(\mu_t \| \nu_t) - \|\mathbb{E}_{\tilde{\mu}_t}(\nabla \log \rho_t)\|^2 \right] \\ &= \left( \frac{C_1}{C_2} \right)^{2q-2} \beta_t \frac{1}{q} \text{RFI}_q(\mu_t \| \nu_t), \end{aligned} \quad (4.326)$$

其中我们利用了由于  $\tilde{\mu}_t, \rho_t$  对称而成立的事实  $\|\mathbb{E}_{\tilde{\mu}_t}(\nabla \log \rho_t)\|^2 = 0$ 。

因此, 我们得到

$$\partial_t \text{RFI}_q(\mu_t \| \nu_t) \leq - \left( \frac{C_1}{C_2} \right)^{2q-2} \beta_t \text{RFI}_q(\mu_t \| \nu_t), \quad (4.327)$$

$$\text{RFI}_q(\mu_t \| \nu_t) \leq \exp \left( - \left( \frac{C_1}{C_2} \right)^{2q-2} \int_0^t \beta_s ds \right) \text{RFI}_q(\mu_0 \| \nu_0), \quad \text{当 } q \geq 1. \quad (4.328)$$

**第 (5) ~ (7) 部分:** 这些部分的证明与 **第 (1) ~ (4) 部分** 的论证相同, 唯一的区别在于我们提供了  $\mu_0$  的具体可计算的庞加莱不等式常数来代替  $\beta_t$ 。细节此处省略。

**注** 在 **第 (4) ~ (7) 部分** 中, 假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} \leq C_2 < \infty$  可以被其他确保  $\tilde{\mu}_t$  满足庞加莱不等式的条件所取代。例如, 设  $d\mu_t \propto e^{-U_t} dx, d\nu_t \propto e^{-V_t} dx$ , 其中  $\nabla^2 U_0 \succeq \beta I \succ 0, \nabla^2 V_0 \succeq \alpha I \succ 0$ , 即  $\nu_0$  满足  $\alpha$ -SLC 且  $\mu_0$  是  $\beta$ -SLC。我们进一步假设对于所有  $t \geq 0$ , 有  $K_1 \nabla^2 \log \mu_t \preceq \nabla^2 \log \nu_t \preceq K_2 \nabla^2 \log \mu_t$ , 其中  $0 < K_2 \leq 1 \leq K_1$ 。则对于  $q \geq 1$  成立:

$$\begin{aligned} -\nabla^2 \log(\rho_t^{q-1} \mu_t) &= \nabla^2 U_t + (q-1)(U_t - V_t) \\ &= [1 - (q-1)(K_1 - 1)] \nabla^2 U_t \\ &\succeq [1 - (q-1)(K_1 - 1)] \beta_t I. \end{aligned} \quad (4.329)$$

在这种情况下, 对于  $q \geq 1$ , 我们可以用  $[1 - (q-1)(K_1 - 1)]$  替换 **第 (4) ~ (7) 部分** 中的  $(C_1/C_2)^{2(q-1)}$ 。

#### 4.5.6 定理 4.8 的证明

在 OU 流情形下, 相对 Rényi-Fisher 信息的时间导数由下式给出:

$$\text{RFI}_q(\mu_t \| \nu_t) = q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t \right), \quad (4.330)$$

$$\begin{aligned} \partial_t \text{RFI}_q(\mu_t \| \nu_t) &\leq -2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\ &\quad - 4q(q-1) \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \right\} \\ &\quad + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{2\nabla^2 \log \nu_t + \gamma I}^2 d\mu_t \right\} \\ &= -2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\ &\quad + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \\ &\quad \cdot \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{(2q\nabla^2 \log \nu_t - 2(q-1)\nabla^2 \log \mu_t + \gamma I)}^2 d\mu_t \right\}. \end{aligned} \quad (4.331)$$

**第 (1) 部分:** 如果  $\nabla^2 \log \nu_t \preceq (1 - q^{-1})\nabla^2 \log \mu_t - \frac{\gamma q^{-1}}{2}I$ , 则有

$$2q\nabla^2 \log \nu_t - 2(q-1)\nabla^2 \log \mu_t + \gamma I \preceq 0, \quad (4.332)$$

这导出了  $\partial_t \text{RFI}_q(\mu_t \| \nu_t) \leq 0$ 。因此:

$$\text{RFI}_q(\mu_t \| \nu_t) \leq \text{RFI}_q(\mu_0 \| \nu_0). \quad (4.333)$$

**第 (2), (4) 部分:** 如果  $q > 1, \nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 I$ , 则有

$$\nabla^2 \log \mu_t \succeq K_1^{-1} \nabla^2 \log \nu_t - K_1^{-1} K_2 I, \quad (4.334)$$

这导出

$$2q\nabla^2 \log \nu_t - 2(q-1)\nabla^2 \log \mu_t + \gamma I \preceq \frac{2q(K_1 - 1) + 2}{K_1} \nabla^2 \log \nu_t + (\gamma + 2(q-1)K_1^{-1}K_2)I. \quad (4.335)$$

我们回顾 OU 流在时刻  $t \geq 0$  的显式解为  $X_t \stackrel{d}{=} e^{-\gamma t} X_0 + \sqrt{\frac{1 - e^{-2\gamma t}}{\gamma}} Z$ , 其中  $X_0 \sim \nu_0$ ,  $Z \sim \mathcal{N}(0, I)$  独立,  $X_t \sim \nu_t$ 。由于假设  $\nu_0$  满足  $\alpha$ -SLC, 则  $e^{-\gamma t} X_0$  的分布是  $(e^{2\gamma t} \alpha)$ -SLC。注意  $\sqrt{\frac{1 - e^{-2\gamma t}}{\gamma}} Z$  的分布是  $(\frac{\gamma}{1 - e^{-2\gamma t}})$ -SLC。根据卷积下强对数凹性的保持性 ([37] 中的定理 3.7), 可知  $\nu_t$  满足  $\alpha_t$ -SLC, 其中

$$\alpha_t = \left( \frac{e^{-2\gamma t}}{\alpha} + \frac{1 - e^{-2\gamma t}}{\gamma} \right)^{-1} = \frac{\gamma \alpha}{\alpha + e^{-2\gamma t}(\gamma - \alpha)}. \quad (4.336)$$

于是我们有

$$\begin{aligned} & 2q\nabla^2 \log \nu_t - 2(q-1)\nabla^2 \log \mu_t + \gamma I \\ & \leq \left[ \frac{2q(K_1 - 1) + 2}{K_1} \cdot \frac{-\gamma \alpha}{\alpha + e^{-2\gamma t}(\gamma - \alpha)} + (\gamma + 2(q-1)K_1^{-1}K_2) \right] I \\ & = \frac{(\gamma + 2(q-1)K_1^{-1}K_2)(\alpha + e^{-2\gamma t}(\gamma - \alpha)) - 2K_1^{-1}\gamma\alpha(q(K_1 - 1) + 1)}{\alpha + e^{-2\gamma t}(\gamma - \alpha)} I \\ & \equiv \frac{A + Be^{-2\gamma t}}{C + De^{-2\gamma t}} I, \end{aligned} \quad (4.337)$$

其中  $A = 2(q-1)K_1^{-1}K_2\alpha + [1 - 2(K_1^{-1} + q(1 - K_1^{-1}))]\gamma\alpha$ ,  $B = (\gamma - \alpha)[2(q-1)K_1^{-1}K_2 + \gamma]$ ,  $C = \alpha$ ,  $D = \gamma - \alpha$ 。由基本微积分可知, 对于  $C, D \neq 0$ :

$$\int_0^t \frac{A + Be^{-kt}}{C + De^{-kt}} dt = \frac{A}{C}t + \frac{1}{k} \left( \frac{A}{C} - \frac{B}{D} \right) \ln \frac{C + De^{-kt}}{C + D}. \quad (4.338)$$

由 Grönwall 不等式, 我们有

$$\begin{aligned} \partial_t \text{RFI}_q(\mu_t \| \nu_t) &\leq 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{\frac{A+Be^{-2\gamma t}}{C+De^{-2\gamma t}} I}^2 d\mu_t \right\} \\ &\leq 2 \frac{A + Be^{-2\gamma t}}{C + De^{-2\gamma t}} \text{RFI}_q(\mu_t \| \nu_t). \end{aligned} \quad (4.339)$$

$$\begin{aligned} \text{RFI}_q(\mu_t \| \nu_t) &\leq e^{\int_0^t 2 \frac{A+Be^{-2\gamma s}}{C+De^{-2\gamma s}} ds} \text{RFI}_q(\mu_0 \| \nu_0) \\ &= \frac{\gamma^{2K_1^{-1}(1+q(K_1-1))} e^{-2\gamma K_1^{-1}[2(1+q(K_1-1))-K_1-2(q-1)K_2\gamma^{-1}]t}}{(\alpha + (\gamma - \alpha)e^{-2\gamma t})^{2K_1^{-1}(1+q(K_1-1))}} \\ &\quad \cdot \text{RFI}_q(\mu_0 \| \nu_0). \end{aligned} \quad (4.340)$$

取  $K_1 = K, K_2 = 0$ , 我们可以恢复第 (2) 部分的结果。对于  $K_1 = 0, q = 1$  的情况, 我们可以通过相同的论证获得相对 FI 的 SDPI, 其中我们约定  $\frac{1+(K_1-1)q}{K_1} = 1, \frac{q-1}{K_1} = 0$ 。

**第 (3), (5) 部分:** 如果  $q > 1, \nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 I$ , 则有

$$2q \nabla^2 \log \nu_t - 2(q-1) \nabla^2 \log \mu_t + \gamma I \preceq 2(q(K_1-1)+1) \nabla^2 \log \mu_t + (\gamma + 2qK_2) I. \quad (4.341)$$

我们回顾 OU 流在时刻  $t \geq 0$  的显式解为  $X_t \stackrel{d}{=} e^{-\gamma t} X_0 + \sqrt{\frac{1-e^{-2\gamma t}}{\gamma}} Z$ , 其中  $X_0 \sim \mu_0$ ,  $Z \sim \mathcal{N}(0, I)$  独立,  $X_t \sim \mu_t$ 。由于假设  $\mu_0$  满足  $\alpha$ -SLC, 则  $e^{-\gamma t} X_0$  的分布是  $(e^{2\gamma t} \alpha)$ -SLC。注意  $\sqrt{\frac{1-e^{-2\gamma t}}{\gamma}} Z$  的分布是  $(\frac{\gamma}{1-e^{-2\gamma t}})$ -SLC。根据卷积下强对数凹性的保持性, 可知  $\mu_t$  满足  $\alpha_t$ -SLC, 其中

$$\alpha_t = \left( \frac{e^{-2\gamma t}}{\alpha} + \frac{1-e^{-2\gamma t}}{\gamma} \right)^{-1} = \frac{\gamma \alpha}{\alpha + e^{-2\gamma t}(\gamma - \alpha)}. \quad (4.342)$$

于是我们有

$$\begin{aligned} &2q \nabla^2 \log \nu_t - 2(q-1) \nabla^2 \log \mu_t + \gamma I \\ &\leq 2(q(K_1-1)+1) \frac{-\gamma \alpha}{\alpha + e^{-2\gamma t}(\gamma - \alpha)} I + (\gamma + 2qK_2) I \\ &= \frac{2qK_2 \alpha - (1+2q(K_1-1))\gamma \alpha + e^{-2\gamma t}(\gamma - \alpha)(2qK_2 + \gamma)}{\alpha + e^{-2\gamma t}(\gamma - \alpha)} I \\ &\equiv \frac{A + Be^{-2\gamma t}}{C + De^{-2\gamma t}} I, \end{aligned} \quad (4.343)$$

其中  $A = 2qK_2 \alpha - (1+2q(K_1-1))\gamma \alpha, B = (\gamma - \alpha)(2qK_2 + \gamma), C = \alpha, D = \gamma - \alpha$ 。由

Grönwall 不等式, 我们有

$$\begin{aligned} \partial_t \text{RFI}_q(\mu_t \| \nu_t) &\leq 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{\frac{A+Be^{-2\gamma t}}{C+De^{-2\gamma t}} I}^2 d\mu_t \right\} \\ &\leq 2 \frac{A + Be^{-2\gamma t}}{C + De^{-2\gamma t}} \text{RFI}_q(\mu_t \| \nu_t). \end{aligned} \quad (4.344)$$

$$\begin{aligned} \text{RFI}_q(\mu_t \| \nu_t) &\leq e^{\int_0^t 2 \frac{A+Be^{-2\gamma s}}{C+De^{-2\gamma s}} ds} \text{RFI}_q(\mu_0 \| \nu_0) \\ &= \frac{\gamma^{2(q(K_1-1)+1)} e^{-2\gamma[2q(K_1-1)+1-2qK_2\gamma^{-1}]t}}{(\alpha + (\gamma - \alpha)e^{-2\gamma t})^{2(q(K_1-1)+1)}} \text{RFI}_q(\mu_0 \| \nu_0). \end{aligned} \quad (4.345)$$

取  $K_1 = K, K_2 = 0$ , 我们可以恢复 **第 (3) 部分** 的结果。

**第 (6) ~ (8) 部分:** 这些部分的证明与定理 4.8 中的 **第 (1) ~ (5) 部分** 以及定理 4.7 中的 **第 (5) 部分** 的论证相同。细节此处省略。

**注** 在 **第 (6) ~ (8) 部分** 中, 假设  $0 < C_1 \leq \frac{d\mu_0}{d\nu_0} \leq C_2 < \infty$  可以被其他确保  $\tilde{\mu}_t$  满足庞加莱不等式的条件所取代。例如, 设  $d\mu_t \propto e^{-U_t} dx, d\nu_t \propto e^{-V_t} dx$ , 其中  $\nabla^2 U_0 \geq \beta I > 0, \nabla^2 V_0 \geq \alpha I > 0$ , 即  $\nu_0$  满足  $\alpha$ -SLC 且  $\mu_0$  是  $\beta$ -SLC。我们进一步假设对于所有  $t \geq 0$ , 有  $K_3 \nabla^2 \log \mu_t \preceq \nabla^2 \log \nu_t \preceq K_4 \nabla^2 \log \mu_t$ , 其中  $0 < K_4 \leq 1 \leq K_3$ 。则对于  $q \geq 1$  成立:

$$\begin{aligned} -\nabla^2 \log(\rho_t^{q-1} \mu_t) &= \nabla^2 U_t + (q-1)(U_t - V_t) \\ &= [1 - (q-1)(K_3 - 1)] \nabla^2 U_t \\ &\geq [1 - (q-1)(K_3 - 1)] \beta_t I. \end{aligned} \quad (4.346)$$

在这种情况下, 对于  $q \geq 1$ , 我们可以用  $[1 - (q-1)(K_3 - 1)]$  替换 **第 (6) ~ (8) 部分** 中的  $(C_1/C_2)^{2(q-1)}$ 。

#### 4.5.7 定理 4.9 的证明

在 FP 流情形下, 相对 Rényi-Fisher 信息的时间导数由下式给出:

$$\begin{aligned} \text{RFI}_q(\mu_t \| \nu_t) &= q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla \log \rho_t\|^2 d\mu_t \right), \quad (4.347) \\ \partial_t \text{RFI}_q(\mu_t \| \nu_t) &\leq -cq \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\ &\quad - 2cq(q-1) \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \right\} \\ &\quad + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{(e^{\nabla^2 \log \nu_t + \nabla^2 V})}^2 d\mu_t \right\} \\ &\leq -cq \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \left( \int \rho_t^{q-1} \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \right) \\ &\quad + 2q \left( \int \rho_t^{q-1} d\mu_t \right)^{-1} \end{aligned}$$

$$\cdot \left\{ \int \rho_t^{q-1} \|\nabla \log \rho_t\|_{(cq\nabla^2 \log \nu_t - c(q-1)\nabla^2 \log \mu_t + \nabla^2 V)}^2 d\mu_t \right\}. \quad (4.348)$$

**第(1)部分:** 如果  $\nabla^2 \log \nu_t \preceq (1 - q^{-1})\nabla^2 \log \mu_t - c^{-1}q^{-1}\nabla^2 V$ , 则有

$$cq\nabla^2 \log \nu_t - c(q-1)\nabla^2 \log \mu_t + \nabla^2 V \leq 0, \quad (4.349)$$

这导出  $\partial_t \text{RFI}_q(\mu_t \| \nu_t) \leq 0$ . 因此:

$$\text{RFI}_q(\mu_t \| \nu_t) \leq \text{RFI}_q(\mu_0 \| \nu_0). \quad (4.350)$$

**第(2)部分:** 如果  $\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V$ , 其中  $K_1 \geq 1 - q^{-1}, q \geq 1$ . 则有

$$\begin{aligned} cq\nabla^2 \log \nu_t - c(q-1)\nabla^2 \log \mu_t + \nabla^2 V &\leq (cq(K_1 - 1) + c)K_1^{-1}\nabla^2 \log \nu_t \\ &\quad + (1 + c(q-1)K_1^{-1}K_2)\nabla^2 V. \end{aligned} \quad (4.351)$$

由于  $\nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 即  $\nabla^2 \log \nu_0 \preceq -\frac{1}{c}\nabla^2 V$ , 我们有  $\nabla^2 \log \nu_t \preceq -\frac{1}{c}\nabla^2 V$  以及

$$\begin{aligned} &cq\nabla^2 \log \nu_t - c(q-1)\nabla^2 \log \mu_t + \nabla^2 V \\ &\leq (-q(K_1 - 1) - 1)K_1^{-1}\nabla^2 V + (1 + c(q-1)K_1^{-1}K_2)\nabla^2 V \\ &\leq -K_1^{-1}(q-1)(1 - K_1 + cK_2)\alpha I. \end{aligned} \quad (4.352)$$

由 Grönwall 不等式, 我们有

$$\text{RFI}_q(\mu_t \| \nu_t) \leq e^{-2K_1^{-1}(q-1)(1-K_1+cK_2)\alpha t} \text{RFI}_q(\mu_0 \| \nu_0). \quad (4.353)$$

**第(3)部分:** 如果  $\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V$ , 且  $K_1 \geq 1 - q^{-1}$ . 则有

$$\begin{aligned} cq\nabla^2 \log \nu_t - c(q-1)\nabla^2 \log \mu_t + \nabla^2 V &\leq c[1 + q(K_1 - 1)]\nabla^2 \log \mu_t \\ &\quad + (cqK_2 + 1)\nabla^2 V. \end{aligned} \quad (4.354)$$

由于  $\mu_0$  是  $\frac{V}{c}$ -相对对数凹的, 即  $\nabla^2 \log \mu_0 \leq -\frac{1}{c}\nabla^2 V$ , 我们有  $\nabla^2 \log \mu_t \leq -\frac{1}{c}\nabla^2 V$  以及

$$\begin{aligned} &cq\nabla^2 \log \nu_t - c(q-1)\nabla^2 \log \mu_t + \nabla^2 V \\ &\leq -[1 + q(K_1 - 1)]\nabla^2 V + (cqK_2 + 1)\nabla^2 V \\ &\leq (cK_2 - K_1 + 1)q\alpha I. \end{aligned} \quad (4.355)$$

由 Grönwall 不等式, 我们有

$$\text{RFI}_q(\mu_t \|\nu_t) \leq e^{-2(K_1 - cK_2 - 1)q\alpha t} \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.356)$$

**第 (4) ~ (6) 部分:** 这几部分的证明与 **第 (1) ~ (3) 部分** 相同, 区别仅在于引入了庞加莱不等式, 细节此处省略。

**第 (7) 部分:** 如果  $\nu_0 = \nu \propto e^{-\frac{2}{c}V}$ , 且  $\mu_t$  是  $KV$ -相对对数凸的, 即  $\nabla^2 \log \mu_t \geq K\nabla^2 V$ , 满足  $q \geq 1, q(2 + cK) \geq 1 + cK$ , 则对于所有  $t \geq 0$ :

$$\begin{aligned} & cq\nabla^2 \log \nu_t - c(q-1)\nabla^2 \log \mu_t + \nabla^2 V \\ &= -2q\nabla^2 V - c(q-1)\nabla^2 \log \mu_t + \nabla^2 V \\ &\leq -2q\nabla^2 V - cK(q-1)\nabla^2 V + \nabla^2 V \\ &= -(q(2+cK) - (1+cK))\nabla^2 V \\ &\leq -(q(2+cK) - (1+cK))\alpha I. \end{aligned} \quad (4.357)$$

由 Grönwall 不等式, 我们有

$$\text{RFI}_q(\mu_t \|\nu_t) \leq e^{-2q(q(2+cK) - (1+cK))\alpha t} \text{RFI}_q(\mu_0 \|\nu_0). \quad (4.358)$$

**第 (8) 部分:** 此部分的证明可通过与 **第 (6), (7) 部分** 相同的论证 (结合庞加莱不等式) 导出, 细节此处省略。

#### 4.5.8 定理 4.10 的证明

我们从相对  $\Phi$ -Fisher 信息的时间导数开始:

$$\begin{aligned} \partial_t \text{FI}_\Phi(\mu_t \|\nu_t) &= \int \|\nabla \rho_t\|^2 \Phi''(\rho_t) \partial_t \nu_t + 2 \int \Phi''(\rho_t) \langle \nabla \rho_t, \nabla \partial_t \rho_t \rangle d\nu_t \\ &\quad + \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \partial_t \rho_t d\nu_t \\ &:= I_1 + I_2 + I_3. \end{aligned} \quad (4.359)$$

定义  $u_t := \log \mu_t$ , 则我们有:

$$\begin{aligned} \partial_t u_t &= e^{-u_t} \left( -\nabla \cdot (e^{u_t} b_t) + \frac{c}{2} \Delta e^{u_t} \right) \\ &= -\nabla \cdot b_t - \langle b_t, \nabla u_t \rangle + \frac{c}{2} \Delta u_t + \frac{c}{2} \|\nabla u_t\|^2, \\ \partial_t \rho_t &= \partial_t \frac{\mu_t}{\nu_t} = \frac{\nu_t \partial_t \mu_t - \mu_t \partial_t \nu_t}{\nu_t^2} \end{aligned}$$

$$\begin{aligned}
&= \rho_t \partial_t \log \mu_t - \rho_t \partial_t \log \nu_t \\
&= \rho_t \left[ -\langle b_t, \nabla \log \rho_t \rangle + \frac{c}{2} \Delta \log \rho_t + \frac{c}{2} \|\nabla \log \mu_t\|^2 - \frac{c}{2} \|\nabla \log \nu_t\|^2 \right] \\
&= \rho_t \left[ -\langle b_t, \nabla \log \rho_t \rangle + \frac{c}{2} \Delta \log \rho_t + \frac{c}{2} \langle \nabla \log(\mu_t \nu_t), \nabla \log \rho_t \rangle \right], \\
\partial_t \log \rho_t &= \frac{\partial_t \rho_t}{\rho_t} = -\langle b_t, \nabla \log \rho_t \rangle + \frac{c}{2} \Delta \log \rho_t + \frac{c}{2} \langle \nabla \log(\mu_t \nu_t), \nabla \log \rho_t \rangle, \\
\nabla \partial_t \rho_t &= \nabla \rho_t \left[ -\langle b_t, \nabla \log \rho_t \rangle + \frac{c}{2} \Delta \log \rho_t + \frac{c}{2} \langle \nabla \log(\mu_t \nu_t), \nabla \log \rho_t \rangle \right] \\
&\quad + \rho_t \left[ -(\nabla b_t) \nabla \log \rho_t - (\nabla^2 \log \rho_t) b_t + \frac{c}{2} \nabla \Delta \log \rho_t \right. \\
&\quad \left. + \frac{c}{2} (\nabla^2 \log(\mu_t \nu_t)) \nabla \log \rho_t + \frac{c}{2} (\nabla^2 \log \rho_t) \nabla \log(\mu_t \nu_t) \right], \\
\nabla^2 \rho_t &= \nabla (e^{\log \rho_t} \nabla \log \rho_t) \\
&= \rho_t \nabla^2 \log \rho_t + \rho_t (\nabla \log \rho_t)^{\otimes 2} \equiv \rho_t \nabla^2 \log \rho_t + \rho_t (\nabla \log \rho_t) (\nabla \log \rho_t)^\top, \\
\nabla^2 \log \rho_t &= \frac{\nabla^2 \rho_t}{\rho_t} - \frac{(\nabla \rho_t)^{\otimes 2}}{\rho_t^2}.
\end{aligned}$$

然后我们可以将前两项重写如下：

$$\begin{aligned}
I_1 &= \int \|\nabla \rho_t\|^2 \Phi''(\rho_t) \partial_t \nu_t \\
&= \int \|\nabla \rho_t\|^2 \Phi''(\rho_t) \left[ -\nabla \cdot (\nu_t b_t) + \frac{c}{2} \Delta \nu_t \right] \\
&= \int \|\nabla \rho_t\|^2 \Phi''(\rho_t) \left[ -\nabla \cdot (\nu_t b_t - \frac{c}{2} \nabla \nu_t) \right] \\
&= \int (\nu_t b_t - \frac{c}{2} \nabla \nu_t) \cdot \nabla (\|\nabla \rho_t\|^2 \Phi''(\rho_t)) \\
&= \int (b_t - \frac{c}{2} \nabla \log \nu_t) \cdot \left( 2(\nabla^2 \rho_t) \nabla \rho_t \Phi''(\rho_t) + \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \nabla \rho_t \right) d\nu_t \\
&= \int \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, 2(\nabla^2 \rho_t) \nabla \rho_t \Phi''(\rho_t) + \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \nabla \rho_t \right\rangle d\nu_t \\
&= 2 \int \rho_t \Phi''(\rho_t) \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, (\nabla^2 \log \rho_t + (\nabla \log \rho_t)^{\otimes 2}) \nabla \log \rho_t \right\rangle d\mu_t \\
&\quad + \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, \nabla \log \rho_t \right\rangle d\mu_t \\
&= 2 \int \rho_t \Phi''(\rho_t) \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \right\rangle d\mu_t \\
&\quad + 2 \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|^2 \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, \nabla \log \rho_t \right\rangle d\mu_t \\
&\quad + \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \left\langle b_t - \frac{c}{2} \nabla \log \nu_t, \nabla \log \rho_t \right\rangle d\mu_t. \tag{4.360}
\end{aligned}$$

$$I_2 = 2 \int \Phi''(\rho_t) \langle \nabla \rho_t, \nabla \partial_t \rho_t \rangle d\nu_t$$

$$\begin{aligned}
&= \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 [c\Delta \log \rho_t + c \langle \nabla \log \mu_t, \nabla \log \rho_t \rangle \\
&\quad + c \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle - 2 \langle b_t, \nabla \log \rho_t \rangle] d\nu_t \\
&\quad + c \int \Phi''(\rho_t) \langle \nabla \rho_t, \nabla \Delta \log \rho_t \rangle d\mu_t \\
&\quad + c \int \Phi''(\rho_t) \langle \nabla \rho_t, (\nabla^2 \log \mu_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad + c \int \Phi''(\rho_t) \langle \nabla \rho_t, (\nabla^2 \log \nu_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad + c \int \Phi''(\rho_t) \langle \nabla \rho_t, (\nabla^2 \log \rho_t) \nabla \log \mu_t \rangle d\mu_t \\
&\quad + c \int \Phi''(\rho_t) \langle \nabla \rho_t, (\nabla^2 \log \rho_t) \nabla \log \nu_t \rangle d\mu_t \\
&\quad - 2 \int \Phi''(\rho_t) \langle \nabla \rho_t, (\nabla b_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad - 2 \int \Phi''(\rho_t) \langle \nabla \rho_t, (\nabla^2 \log \rho_t) b_t \rangle d\mu_t, \tag{4.361}
\end{aligned}$$

其中我们分别用  $J_1, J_2$  表示以下各项:

$$\begin{aligned}
J_1 &= c \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 \nabla \cdot \nabla \log \rho_t d\nu_t \\
&= -c \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \rho_t, \nabla \log \nu_t \rangle d\nu_t \\
&\quad - c \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \rho_t, \nabla \log \rho_t \rangle d\mu_t \\
&\quad - 2c \int \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
&= -c \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \rho_t, \nabla \log \nu_t \rangle d\nu_t \\
&\quad - c \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \rho_t, \nabla \log \rho_t \rangle d\mu_t \\
&\quad - 2c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
&\quad - 2c \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|^4 d\mu_t. \tag{4.362}
\end{aligned}$$

由 Bochner 公式, 我们有:

$$\begin{aligned}
\langle \nabla \log \rho_t, \nabla \Delta \log \rho_t \rangle &= \frac{1}{2} \Delta \|\nabla \log \rho_t\|^2 - \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 \\
&= \nabla \cdot ((\nabla^2 \log \rho_t) \nabla \log \rho_t) - \|\nabla^2 \log \rho_t\|_{\text{HS}}^2, \tag{4.363} \\
J_2 &= c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, \nabla \Delta \log \rho_t \rangle d\mu_t \\
&= -c \int \rho_t \Phi''(\rho_t) \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \\
&\quad - c \int \langle (\nabla^2 \log \rho_t) \nabla \log \rho_t, \nabla(\mu_t \rho_t \Phi''(\rho_t)) \rangle \\
&= -c \int \rho_t \Phi''(\rho_t) \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t
\end{aligned}$$

$$\begin{aligned}
& -c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \mu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t^2 \Phi'''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t. \quad (4.364)
\end{aligned}$$

结合  $I_1, I_2, J_1, J_2$  的结果, 我们得到:

$$\begin{aligned}
I_1 + I_2 = & 2 \int \rho_t \Phi''(\rho_t) \langle b_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \nu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& + 2 \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|^2 \langle b_t, \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|^2 \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t \\
& + \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \langle b_t, \nabla \log \rho_t \rangle d\mu_t \\
& -\frac{c}{2} \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \rho_t, \nabla \log \nu_t \rangle d\nu_t \\
& -c \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \rho_t, \nabla \log \rho_t \rangle d\mu_t \\
& -2c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -2c \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|^4 d\mu_t \\
& +c \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \rho_t, \nabla \log \mu_t \rangle d\nu_t \\
& +c \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \rho_t, \nabla \log \nu_t \rangle d\nu_t \\
& -2 \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 \langle b_t, \nabla \log \rho_t \rangle d\nu_t \\
& -c \int \rho_t \Phi''(\rho_t) \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \\
& -c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \mu_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t^2 \Phi'''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& +c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \mu_t) \nabla \log \rho_t \rangle d\mu_t \\
& +c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \nu_t) \nabla \log \rho_t \rangle d\mu_t \\
& +c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \mu_t \rangle d\mu_t \\
& +c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \nu_t \rangle d\mu_t
\end{aligned}$$

$$\begin{aligned}
& -2 \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla b_t) \nabla \log \rho_t \rangle d\mu_t \\
& -2 \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) b_t \rangle d\mu_t.
\end{aligned} \tag{4.365}$$

合并同类项，我们得到：

$$\begin{aligned}
I_1 + I_2 = & \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \langle b_t, \nabla \log \rho_t \rangle d\mu_t \\
& - \frac{c}{2} \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t \\
& - c \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \rho_t, \nabla \log \rho_t \rangle d\mu_t \\
& - c \int \rho_t \Phi''(\rho_t) \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \\
& - 3c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& - c \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|^4 d\mu_t \\
& - c \int \rho_t^2 \Phi'''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& + c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \mu_t) \nabla \log \rho_t \rangle d\mu_t \\
& + c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \nu_t) \nabla \log \rho_t \rangle d\mu_t \\
& - 2 \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla b_t) \nabla \log \rho_t \rangle d\mu_t.
\end{aligned} \tag{4.366}$$

$$\begin{aligned}
I_3 = & \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \rho_t \partial_t \log \rho_t d\nu_t \\
= & \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \partial_t \log \rho_t d\mu_t \\
= & - \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle b_t, \nabla \log \rho_t \rangle d\mu_t + \frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \Delta \log \rho_t d\mu_t \\
& + \frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle \nabla \log \mu_t, \nabla \log \rho_t \rangle d\mu_t \\
& + \frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t \\
= & - \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle b_t, \nabla \log \rho_t \rangle d\mu_t \\
& - \frac{c}{2} \int \langle \nabla \log \rho_t, \nabla (\mu_t \|\nabla \rho_t\|^2 \Phi'''(\rho_t)) \rangle \\
& + \frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle \nabla \log \mu_t, \nabla \log \rho_t \rangle d\mu_t \\
& + \frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t \\
= & - \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle b_t, \nabla \log \rho_t \rangle d\mu_t
\end{aligned}$$

$$\begin{aligned}
& -\frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle \nabla \log \mu_t, \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t \Phi'''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -\frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi^{(4)}(\rho_t) \rho_t \langle \nabla \log \rho_t, \nabla \log \rho_t \rangle d\mu_t \\
& +\frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle \nabla \log \mu_t, \nabla \log \rho_t \rangle d\mu_t \\
& +\frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t \\
= & -\int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle b_t, \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t \Phi'''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -\frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi^{(4)}(\rho_t) \rho_t \langle \nabla \log \rho_t, \nabla \log \rho_t \rangle d\mu_t \\
& +\frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t \\
= & -\int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle b_t, \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t^2 \Phi'''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t^2 \Phi'''(\rho_t) \|\nabla \log \rho_t\|^4 d\mu_t \\
& -\frac{c}{2} \int \rho_t \|\nabla \rho_t\|^2 \Phi^{(4)}(\rho_t) \langle \nabla \log \rho_t, \nabla \log \rho_t \rangle d\mu_t \\
& +\frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t. \tag{4.367}
\end{aligned}$$

$$\begin{aligned}
I_1 + I_2 + I_3 = & \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \langle b_t, \nabla \log \rho_t \rangle d\mu_t \\
& -\frac{c}{2} \int \Phi'''(\rho_t) \|\nabla \rho_t\|^2 \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t^2 \Phi'''(\rho_t) \|\nabla \log \rho_t\|^4 d\mu_t - c \int \rho_t \Phi''(\rho_t) \left\| \nabla^2 \log \rho_t \right\|_{\text{HS}}^2 d\mu_t \\
& -3c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|^4 d\mu_t \\
& -c \int \rho_t^2 \Phi'''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& +c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \mu_t) \nabla \log \rho_t \rangle d\mu_t \\
& +c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \nu_t) \nabla \log \rho_t \rangle d\mu_t \\
& -2 \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla b_t) \nabla \log \rho_t \rangle d\mu_t \\
& -\int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle b_t, \nabla \log \rho_t \rangle d\mu_t
\end{aligned}$$

$$\begin{aligned}
& -c \int \rho_t^2 \Phi'''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -c \int \rho_t^2 \Phi'''(\rho_t) \|\nabla \log \rho_t\|^4 d\mu_t \\
& -\frac{c}{2} \int \rho_t \|\nabla \rho_t\|^2 \Phi^{(4)}(\rho_t) \langle \nabla \log \rho_t, \nabla \log \rho_t \rangle d\mu_t \\
& +\frac{c}{2} \int \|\nabla \rho_t\|^2 \Phi'''(\rho_t) \langle \nabla \log \nu_t, \nabla \log \rho_t \rangle d\mu_t \\
= & -2c \int \rho_t \Phi'''(\rho_t) \|\nabla \log \rho_t\|^4 d\mu_t \\
& -2c \int \rho_t^2 \Phi'''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& -2c \int \rho_t \Phi''(\rho_t) \langle \nabla \log \rho_t, (\nabla^2 \log \rho_t) \nabla \log \rho_t \rangle d\mu_t \\
& +2 \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|_{(c\nabla^2 \log \nu_t - (\nabla b_t)_{\text{sym}})}^2 d\mu_t \\
& -c \int \rho_t \Phi''(\rho_t) \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t -c \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|^4 d\mu_t \\
& -\frac{c}{2} \int \rho_t \|\nabla \rho_t\|^2 \Phi^{(4)}(\rho_t) \langle \nabla \log \rho_t, \nabla \log \rho_t \rangle d\mu_t \\
= & -c \int \|\nabla \log \rho_t\|^4 \left( \rho_t \Phi''(\rho_t) + 2\rho_t^2 \Phi'''(\rho_t) + \frac{1}{2}\rho_t^3 \Phi^{(4)}(\rho_t) \right) d\mu_t \\
& +2 \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|_{(c\nabla^2 \log \nu_t - (\nabla b_t)_{\text{sym}})}^2 d\mu_t \\
& -2c \int \|\nabla \log \rho_t\|_{\nabla^2 \log \rho_t}^2 \left( \rho_t \Phi''(\rho_t) + \rho_t^2 \Phi'''(\rho_t) \right) d\mu_t \\
& -c \int \rho_t \Phi''(\rho_t) \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t. \tag{4.368}
\end{aligned}$$

至此我们得到了相对  $\Phi$ -Fisher 信息沿 FP 流时间导数的第一个表达式。此外，通过  $\Phi$  导数的阶数对各项进行分类，我们可以继续简化表达式：

$$\begin{aligned}
\partial_t \mathbf{FI}_\Phi(\mu_t \| \nu_t) & = -c \int \|\nabla \log \rho_t\|^4 \left( \rho_t \Phi''(\rho_t) + 2\rho_t^2 \Phi'''(\rho_t) + \frac{1}{2}\rho_t^3 \Phi^{(4)}(\rho_t) \right) d\mu_t \\
& +2 \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|_{(c\nabla^2 \log \nu_t - (\nabla b_t)_{\text{sym}})}^2 d\mu_t \\
& -2c \int \|\nabla \log \rho_t\|_{\nabla^2 \log \rho_t}^2 \left( \rho_t \Phi''(\rho_t) + \rho_t^2 \Phi'''(\rho_t) \right) d\mu_t \\
& -c \int \rho_t \Phi''(\rho_t) \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 d\mu_t \\
= & \int \left[ -c \|\nabla \log \rho_t\|^4 - 2c \|\nabla \log \rho_t\|_{\nabla^2 \log \rho_t}^2 - c \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 \right] \rho_t \Phi''(\rho_t) d\mu_t \\
& +2 \int \rho_t \Phi''(\rho_t) \|\nabla \log \rho_t\|_{(c\nabla^2 \log \nu_t - (\nabla b_t)_{\text{sym}})}^2 d\mu_t \\
& -2c \int \left[ \|\nabla \log \rho_t\|^4 + \|\nabla \log \rho_t\|_{\nabla^2 \log \rho_t}^2 \right] \rho_t^2 \Phi'''(\rho_t) d\mu_t \\
& -\frac{c}{2} \int \rho_t^3 \Phi^{(4)}(\rho_t) \|\nabla \log \rho_t\|^4 d\mu_t \\
= & -c \int \left[ \rho_t^2 \|\nabla \log \rho_t\|^4 + 2\rho_t^2 \|\nabla \log \rho_t\|_{\nabla^2 \log \rho_t}^2 + \rho_t^2 \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 \right] \Phi''(\rho_t) d\mu_t
\end{aligned}$$

$$\begin{aligned}
& + 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{(c\nabla^2 \log \nu_t - (\nabla b_t)_{\text{sym}})}^2 d\nu_t \\
& - 2c \int \left[ \rho_t^3 \|\nabla \log \rho_t\|^4 + \rho_t^3 \|\nabla \log \rho_t\|_{\nabla^2 \log \rho_t}^2 \right] \Phi'''(\rho_t) d\nu_t \\
& - \frac{c}{2} \int \Phi^{(4)}(\rho_t) \|\nabla \rho_t\|^4 d\nu_t.
\end{aligned} \tag{4.369}$$

注意到

$$\begin{aligned}
& \rho_t^2 \|\nabla \log \rho_t\|^4 + 2\rho_t^2 \|\nabla \log \rho_t\|_{\nabla^2 \log \rho_t}^2 + \rho_t^2 \|\nabla^2 \log \rho_t\|_{\text{HS}}^2 \\
& = \rho_t^{-2} \|\nabla \rho_t\|^4 + 2(\nabla \rho_t)^\top \left( \frac{\nabla^2 \rho_t}{\rho_t} - \frac{(\nabla \rho_t)^{\otimes 2}}{\rho_t^2} \right) \nabla \rho_t \\
& \quad + \rho_t^2 \left\| \frac{\nabla^2 \rho_t}{\rho_t} - \frac{(\nabla \rho_t)^{\otimes 2}}{\rho_t^2} \right\|_{\text{HS}}^2 \\
& = \rho_t^{-2} \|\nabla \rho_t\|^4 + 2\rho_t^{-1} (\nabla \rho_t)^\top (\nabla^2 \rho_t) \nabla \rho_t - 2\rho_t^{-2} \|\nabla \rho_t\|^4 \\
& \quad + \|\nabla^2 \rho_t\|_{\text{HS}}^2 + \rho_t^{-2} \|\nabla \rho_t\|^4 - 2\rho_t^{-1} (\nabla \rho_t)^\top (\nabla^2 \rho_t) \nabla \rho_t \\
& = \|\nabla^2 \rho_t\|_{\text{HS}}^2,
\end{aligned} \tag{4.370}$$

以及

$$\begin{aligned}
& \rho_t^3 \|\nabla \log \rho_t\|^4 + \rho_t^3 \|\nabla \log \rho_t\|_{\nabla^2 \log \rho_t}^2 \\
& = \rho_t^{-1} \|\nabla \rho_t\|^4 + \rho_t (\nabla \rho_t)^\top \left( \frac{\nabla^2 \rho_t}{\rho_t} - \frac{(\nabla \rho_t)^{\otimes 2}}{\rho_t^2} \right) \nabla \rho_t \\
& = \rho_t^{-1} \|\nabla \rho_t\|^4 + \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2 - \rho_t^{-1} \|\nabla \rho_t\|^4 \\
& = \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2.
\end{aligned} \tag{4.371}$$

因此我们有：

$$\begin{aligned}
\partial_t \text{FI}_\Phi(\mu_t \| \nu_t) & = -c \int \Phi''(\rho_t) \|\nabla^2 \rho_t\|_{\text{HS}}^2 d\nu_t - 2c \int \Phi'''(\rho_t) \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2 d\nu_t \\
& \quad - \frac{c}{2} \int \Phi^{(4)}(\rho_t) \|\nabla \rho_t\|^4 d\nu_t \\
& \quad + 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{(c\nabla^2 \log \nu_t - (\nabla b_t)_{\text{sym}})}^2 d\nu_t \\
& = 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{A_t^{\nu, b}}^2 d\nu_t - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t,
\end{aligned} \tag{4.372}$$

其中

$$\mathbf{X} = \frac{c}{2} \begin{pmatrix} 2\Phi''(\rho_t) & 2\Phi'''(\rho_t) \\ 2\Phi'''(\rho_t) & \Phi^{(4)}(\rho_t) \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} \|\nabla^2 \rho_t\|_{\text{HS}}^2 & \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2 \\ \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2 & \|\nabla \rho_t\|^4 \end{pmatrix}. \tag{4.373}$$

### 4.5.9 定理 4.11 的证明

在高斯信道情形下, 相对  $\Phi$ -Fisher 信息及其时间导数由下式给出:

$$\text{FI}_{\Phi}(\mu_t \|\nu_t) = \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 d\nu_t, \quad (4.374)$$

$$\partial_t \text{FI}_{\Phi}(\mu_t \|\nu_t) = 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{\nabla^2 \log \nu_t}^2 d\nu_t - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t, \quad (4.375)$$

其中

$$\mathbf{X} = \frac{1}{2} \begin{pmatrix} 2\Phi''(\rho_t) & 2\Phi'''(\rho_t) \\ 2\Phi'''(\rho_t) & \Phi^{(4)}(\rho_t) \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} \|\nabla^2 \rho_t\|_{\text{HS}}^2 & \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2 \\ \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2 & \|\nabla \rho_t\|_{\nabla^2 \rho_t}^4 \end{pmatrix}. \quad (4.376)$$

对于 **第 (1), (2), (4) 部分**, 其论证过程与文献 [10] 中 *Theorem 3* 的证明类似, 故此处略去不证。我们仅给出 **第 (3) 部分** 的证明。

由于  $\nu_0$  是  $\alpha$ -SLC 的, 因此  $\nu_t$  是  $\frac{\alpha}{1+\alpha t}$ -SLC 的, 这意味着其满足  $\frac{\alpha}{1+\alpha t}$ -庞加莱不等式 (PI)。因此, 对测试函数  $\phi = \sqrt{\Phi''(\rho_t)} \nabla \rho_t$  应用  $\nu_t$  的庞加莱不等式, 可得:

$$\int \|\phi - \mathbb{E}_{\nu_t} \phi\|^2 d\nu_t \leq \frac{1 + \alpha t}{\alpha} \int \|\nabla \phi\|_{\text{HS}}^2 d\nu_t. \quad (4.377)$$

由于  $\mu_0, \nu_0$  是对称的, 则对于所有  $t \geq 0$ ,  $\mu_t, \nu_t$  亦保持对称性。从而有  $\mathbb{E}_{\nu_t} \phi = 0$ 。注意到:

$$\begin{aligned} \nabla \phi &= \nabla(\sqrt{\Phi''(\rho_t)} \nabla \rho_t) \\ &= \frac{\Phi'''(\rho_t)}{2\sqrt{\Phi''(\rho_t)}} (\nabla \rho_t)^{\otimes 2} + \sqrt{\Phi''(\rho_t)} \nabla^2 \rho_t, \\ \|\nabla \phi\|_{\text{HS}}^2 &= \frac{(\Phi'''(\rho_t))^2}{4\Phi''(\rho_t)} \|\nabla \rho_t\|^4 + \Phi''(\rho_t) \|\nabla^2 \rho_t\|_{\text{HS}}^2 \\ &\quad + \Phi'''(\rho_t) \langle \nabla \rho_t, (\nabla^2 \rho_t) \nabla \rho_t \rangle \\ &\leq \frac{\theta}{8} \Phi^{(4)}(\rho_t) \|\nabla \rho_t\|^4 + \Phi''(\rho_t) \|\nabla^2 \rho_t\|_{\text{HS}}^2 \\ &\quad + \Phi'''(\rho_t) \langle \nabla \rho_t, (\nabla^2 \rho_t) \nabla \rho_t \rangle \\ &= \text{Tr}(\mathbf{X}_{\theta} \mathbf{Y}), \end{aligned} \quad (4.378)$$

其中

$$\mathbf{X}_{\theta} = \begin{pmatrix} \Phi''(\rho_t) & \frac{1}{2} \Phi'''(\rho_t) \\ \frac{1}{2} \Phi'''(\rho_t) & \frac{\theta}{8} \Phi^{(4)}(\rho_t) \end{pmatrix}. \quad (4.380)$$

设定常数  $k_\theta$  为使得  $\mathbf{X} - k_\theta \mathbf{X}_\theta$  半正定的最大数值, 即:

$$\begin{aligned}
k_\theta &= \sup \{k \in [0, 1] : \mathbf{X} - k \mathbf{X}_\theta \succeq 0\} \\
&= \sup \left\{ k \in [0, 1] : \begin{pmatrix} (1-k)\Phi''(\rho_t) & (1-\frac{k}{2})\Phi'''(\rho_t) \\ (1-\frac{k}{2})\Phi'''(\rho_t) & (\frac{1}{2}-\frac{k\theta}{8})\Phi^{(4)}(\rho_t) \end{pmatrix} \succeq 0 \right\} \\
&= \sup \left\{ k \in [0, 1] : (1-k) \left( \frac{1}{2} - \frac{k\theta}{8} \right) \Phi''(\rho_t) \Phi^{(4)}(\rho_t) \right. \\
&\quad \left. - \left( 1 - \frac{k}{2} \right)^2 (\Phi'''(\rho_t))^2 \geq 0 \right\} \\
&= \sup \left\{ k \in [0, 1] : (1-k) \left( \frac{1}{2} - \frac{k\theta}{8} \right) - \frac{\left( 1 - \frac{k}{2} \right)^2 (\Phi'''(\rho_t))^2}{\Phi''(\rho_t) \Phi^{(4)}(\rho_t)} \geq 0 \right\} \\
&\geq \sup \left\{ k \in [0, 1] : (1-k) \left( \frac{1}{2} - \frac{k\theta}{8} \right) - \left( 1 - \frac{k}{2} \right)^2 \frac{\theta}{2} \geq 0 \right\} \\
&= \frac{4(1-\theta)}{4-3\theta}. \tag{4.381}
\end{aligned}$$

进而, 我们有如下估计:

$$\begin{aligned}
\partial_t \mathbf{Fl}_\Phi(\mu_t \| \nu_t) &= 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{\nabla^2 \log \nu_t}^2 d\nu_t - \int \text{Tr}(\mathbf{X} \mathbf{Y}) d\nu_t \\
&= 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{\nabla^2 \log \nu_t}^2 d\nu_t \\
&\quad - \int \text{Tr}((\mathbf{X} - k_\theta \mathbf{X}_\theta) \mathbf{Y}) d\nu_t - \int \text{Tr}(k_\theta \mathbf{X}_\theta \mathbf{Y}) d\nu_t \\
&\leq 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{\nabla^2 \log \nu_t}^2 d\nu_t - k_\theta \int \text{Tr}(\mathbf{X}_\theta \mathbf{Y}) d\nu_t \\
&\leq 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{\nabla^2 \log \nu_t}^2 d\nu_t - k_\theta \int \|\nabla \phi\|_{\text{HS}}^2 d\nu_t \\
&\leq 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{\nabla^2 \log \nu_t}^2 d\nu_t - \frac{k_\theta \alpha}{1 + \alpha t} \int \|\phi\|^2 d\nu_t \\
&\leq -(2 + k_\theta) \frac{\alpha}{1 + \alpha t} \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 d\nu_t \\
&= - \left( 2 + \frac{4(1-\theta)}{4-3\theta} \right) \frac{\alpha}{1 + \alpha t} \mathbf{Fl}_\Phi(\mu_t \| \nu_t). \tag{4.382}
\end{aligned}$$

由 Grönwall 不等式, 可得:

$$\mathbf{Fl}_\Phi(\mu_t \| \nu_t) \leq \frac{\mathbf{Fl}_\Phi(\mu_0 \| \nu_0)}{(1 + \alpha t)^{2 + \frac{4(1-\theta)}{4-3\theta}}}. \tag{4.383}$$

## 4.5.10 定理 4.12 的证明

在 OU 流情形下, 相对  $\Phi$ -Fisher 信息及其时间导数由下式给出:

$$\text{FI}_\Phi(\mu_t \|\nu_t) = \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 d\nu_t, \quad (4.384)$$

$$\partial_t \text{FI}_\Phi(\mu_t \|\nu_t) = 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{(2\nabla^2 \log \nu_t + \gamma I)}^2 d\nu_t - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t, \quad (4.385)$$

其中

$$\mathbf{X} = \begin{pmatrix} 2\Phi''(\rho_t) & 2\Phi'''(\rho_t) \\ 2\Phi'''(\rho_t) & \Phi^{(4)}(\rho_t) \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} \|\nabla^2 \rho_t\|_{\text{HS}}^2 & \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2 \\ \|\nabla \rho_t\|_{\nabla^2 \rho_t}^2 & \|\nabla \rho_t\|_{\nabla^2 \rho_t}^4 \end{pmatrix}. \quad (4.386)$$

对于 **第(1)部分**, 其论证过程与文献 [10] 中 *Theorem 4* 的证明相同。此处我们仅给出 **第(2)部分** 的证明。

由于  $\nu_0$  是  $\alpha$ -SLC 的, 因此  $\nu_t$  是  $\frac{\gamma\alpha}{\alpha + e^{-2\gamma t}(\gamma - \alpha)}$ -SLC 的, 这意味着其满足  $\frac{\gamma\alpha}{\alpha + e^{-2\gamma t}(\gamma - \alpha)}$ -庞加莱不等式 (PI)。因此, 对测试函数  $\phi = \sqrt{\Phi''(\rho_t)} \nabla \rho_t$  应用  $\nu_t$  的庞加莱不等式, 可得:

$$\int \|\phi - \mathbb{E}_{\nu_t} \phi\|^2 d\nu_t \leq \frac{\alpha + e^{-2\gamma t}(\gamma - \alpha)}{\gamma\alpha} \int \|\nabla \phi\|_{\text{HS}}^2 d\nu_t. \quad (4.387)$$

由于  $\mu_0, \nu_0$  是对称的, 则对于所有  $t \geq 0$ ,  $\mu_t, \nu_t$  亦保持对称性。从而有  $\mathbb{E}_{\nu_t} \phi = 0$ 。注意到:

$$\begin{aligned} \nabla \phi &= \nabla(\sqrt{\Phi''(\rho_t)} \nabla \rho_t) \\ &= \frac{\Phi'''(\rho_t)}{2\sqrt{\Phi''(\rho_t)}} (\nabla \rho_t)^{\otimes 2} + \sqrt{\Phi''(\rho_t)} \nabla^2 \rho_t, \\ \|\nabla \phi\|_{\text{HS}}^2 &= \frac{(\Phi'''(\rho_t))^2}{4\Phi''(\rho_t)} \|\nabla \rho_t\|^4 + \Phi''(\rho_t) \|\nabla^2 \rho_t\|_{\text{HS}}^2 \\ &\quad + \Phi'''(\rho_t) \langle \nabla \rho_t, (\nabla^2 \rho_t) \nabla \rho_t \rangle \\ &\leq \frac{\theta}{8} \Phi^{(4)}(\rho_t) \|\nabla \rho_t\|^4 + \Phi''(\rho_t) \|\nabla^2 \rho_t\|_{\text{HS}}^2 \\ &\quad + \Phi'''(\rho_t) \langle \nabla \rho_t, (\nabla^2 \rho_t) \nabla \rho_t \rangle \\ &= \text{Tr}(\mathbf{X}_\theta \mathbf{Y}), \end{aligned} \quad (4.389)$$

其中

$$\mathbf{X}_\theta = \begin{pmatrix} \Phi''(\rho_t) & \frac{1}{2}\Phi'''(\rho_t) \\ \frac{1}{2}\Phi'''(\rho_t) & \frac{\theta}{8}\Phi^{(4)}(\rho_t) \end{pmatrix}. \quad (4.390)$$

设定常数  $k_\theta$  为使得  $\mathbf{X} - k_\theta \mathbf{X}_\theta$  半正定的最大数值, 即:

$$\begin{aligned}
k_\theta &= \sup \{k \in [0, 2] : \mathbf{X} - k\mathbf{X}_\theta \succeq 0\} \\
&= \sup \left\{ k \in [0, 2] : \begin{pmatrix} (2-k)\Phi''(\rho_t) & (2-\frac{k}{2})\Phi'''(\rho_t) \\ (2-\frac{k}{2})\Phi'''(\rho_t) & (1-\frac{k\theta}{8})\Phi^{(4)}(\rho_t) \end{pmatrix} \succeq 0 \right\} \\
&= \sup \left\{ k \in [0, 2] : (2-k) \left(1 - \frac{k\theta}{8}\right) \Phi''(\rho_t) \Phi^{(4)}(\rho_t) \right. \\
&\quad \left. - \left(2 - \frac{k}{2}\right)^2 (\Phi'''(\rho_t))^2 \geq 0 \right\} \\
&= \sup \left\{ k \in [0, 2] : (2-k) \left(1 - \frac{k\theta}{8}\right) - \frac{\left(2 - \frac{k}{2}\right)^2 (\Phi'''(\rho_t))^2}{\Phi''(\rho_t) \Phi^{(4)}(\rho_t)} \geq 0 \right\} \\
&\geq \sup \left\{ k \in [0, 2] : (2-k) \left(1 - \frac{k\theta}{8}\right) - \left(2 - \frac{k}{2}\right)^2 \frac{\theta}{2} \geq 0 \right\} \\
&= \frac{8(1-\theta)}{4-3\theta}. \tag{4.391}
\end{aligned}$$

进而, 我们有如下估计:

$$\begin{aligned}
\partial_t \text{Fl}_\Phi(\mu_t \| \nu_t) &= 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{(2\nabla^2 \log \nu_t + \gamma I)}^2 d\nu_t - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t \\
&= 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{(2\nabla^2 \log \nu_t + \gamma I)}^2 d\nu_t \\
&\quad - \int \text{Tr}((\mathbf{X} - k_\theta \mathbf{X}_\theta)\mathbf{Y}) d\nu_t - \int \text{Tr}(k_\theta \mathbf{X}_\theta \mathbf{Y}) d\nu_t \\
&\leq 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{(2\nabla^2 \log \nu_t + \gamma I)}^2 d\nu_t - k_\theta \int \text{Tr}(\mathbf{X}_\theta \mathbf{Y}) d\nu_t \\
&\leq 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{(2\nabla^2 \log \nu_t + \gamma I)}^2 d\nu_t - k_\theta \int \|\nabla \phi\|_{\text{HS}}^2 d\nu_t \\
&\leq 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{(2\nabla^2 \log \nu_t + \gamma I)}^2 d\nu_t - \frac{k_\theta \gamma \alpha}{\alpha + e^{-2\gamma t}(\gamma - \alpha)} \int \|\phi\|^2 d\nu_t \\
&\leq - \left( \frac{(4+k_\theta)\gamma\alpha}{\alpha + e^{-2\gamma t}(\gamma - \alpha)} - 2\gamma \right) \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 d\nu_t \\
&= - \left( \frac{(4+k_\theta)\gamma\alpha}{\alpha + e^{-2\gamma t}(\gamma - \alpha)} - 2\gamma \right) \text{Fl}_\Phi(\mu_t \| \nu_t). \tag{4.392}
\end{aligned}$$

由 Grönwall 不等式, 可得:

$$\text{Fl}_\Phi(\mu_t \| \nu_t) \leq \frac{\gamma^{2+\frac{4(1-\theta)}{4-3\theta}} e^{-(2+\frac{8(1-\theta)}{4-3\theta})\gamma t}}{(\alpha + e^{-2\gamma t}(\gamma - \alpha))^{2+\frac{4(1-\theta)}{4-3\theta}}} \text{Fl}_\Phi(\mu_0 \| \nu_0). \tag{4.393}$$

## 4.5.11 定理 4.13 的证明

在 FP 流情形下, 相对  $\Phi$ -Fisher 信息及其时间导数由下式给出:

$$\text{FI}_\Phi(\mu_t|\nu_t) = \int \Phi''(\rho_t) \|\nabla\rho_t\|^2 d\nu_t, \quad (4.394)$$

$$\partial_t \text{FI}_\Phi(\mu_t|\nu_t) = 2 \int \Phi''(\rho_t) \|\nabla\rho_t\|_{(c\nabla^2 \log \nu_t + \nabla^2 V)}^2 d\nu_t - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t, \quad (4.395)$$

其中

$$\mathbf{X} = \frac{c}{2} \begin{pmatrix} 2\Phi''(\rho_t) & 2\Phi'''(\rho_t) \\ 2\Phi'''(\rho_t) & \Phi^{(4)}(\rho_t) \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} \|\nabla^2\rho_t\|_{\text{HS}}^2 & \|\nabla\rho_t\|_{\nabla^2\rho_t}^2 \\ \|\nabla\rho_t\|_{\nabla^2\rho_t}^2 & \|\nabla\rho_t\|_{\nabla^2\rho_t}^4 \end{pmatrix}. \quad (4.396)$$

对于 **第 (1)、(3) 部分**, 其论证过程与定理 4.5 的证明相同。我们首先给出 **第 (2) 部分** 的证明。

由于  $\nu_0$  是  $\frac{V}{c}$ -相对对数凹的, 因此  $\nu_t$  也是  $\frac{V}{c}$ -相对对数凹的, 这意味着其满足  $\frac{\alpha}{c}$ -庞加莱不等式 (PI)。因此, 对测试函数  $\phi = \sqrt{\Phi''(\rho_t)}\nabla\rho_t$  应用  $\nu_t$  的庞加莱不等式, 可得:

$$\int \|\phi - \mathbb{E}_{\nu_t}\phi\|^2 d\nu_t \leq \frac{c}{\alpha} \int \|\nabla\phi\|_{\text{HS}}^2 d\nu_t. \quad (4.397)$$

由于  $\mu_0, \nu_0, V$  是对称的, 则对于所有  $t \geq 0$ ,  $\mu_t, \nu_t$  亦保持对称性。从而有  $\mathbb{E}_{\nu_t}\phi = 0$ 。注意到:

$$\begin{aligned} \nabla\phi &= \nabla(\sqrt{\Phi''(\rho_t)}\nabla\rho_t) \\ &= \frac{\Phi'''(\rho_t)}{2\sqrt{\Phi''(\rho_t)}}(\nabla\rho_t)^{\otimes 2} + \sqrt{\Phi''(\rho_t)}\nabla^2\rho_t, \\ \|\nabla\phi\|_{\text{HS}}^2 &= \frac{(\Phi'''(\rho_t))^2}{4\Phi''(\rho_t)}\|\nabla\rho_t\|^4 + \Phi''(\rho_t)\|\nabla^2\rho_t\|_{\text{HS}}^2 \\ &\quad + \Phi'''(\rho_t)\langle\nabla\rho_t, (\nabla^2\rho_t)\nabla\rho_t\rangle \\ &\leq \frac{\theta}{8}\Phi^{(4)}(\rho_t)\|\nabla\rho_t\|^4 + \Phi''(\rho_t)\|\nabla^2\rho_t\|_{\text{HS}}^2 \\ &\quad + \Phi'''(\rho_t)\langle\nabla\rho_t, (\nabla^2\rho_t)\nabla\rho_t\rangle \\ &= \text{Tr}(\mathbf{X}_\theta\mathbf{Y}), \end{aligned} \quad (4.399)$$

其中

$$\mathbf{X}_\theta = \begin{pmatrix} \Phi''(\rho_t) & \frac{1}{2}\Phi'''(\rho_t) \\ \frac{1}{2}\Phi'''(\rho_t) & \frac{\theta}{8}\Phi^{(4)}(\rho_t) \end{pmatrix}. \quad (4.400)$$

设定常数  $k_\theta$  为使得  $\mathbf{X} - k_\theta \mathbf{X}_\theta$  半正定的最大数值, 即:

$$\begin{aligned}
k_\theta &= \sup \{k \in [0, c] : \mathbf{X} - k \mathbf{X}_\theta \succeq 0\} \\
&= \sup \left\{ k \in [0, c] : \begin{pmatrix} (c-k)\Phi''(\rho_t) & (c-\frac{k}{2})\Phi'''(\rho_t) \\ (c-\frac{k}{2})\Phi'''(\rho_t) & (\frac{c}{2}-\frac{k\theta}{8})\Phi^{(4)}(\rho_t) \end{pmatrix} \succeq 0 \right\} \\
&= \sup \left\{ k \in [0, c] : (c-k) \left( \frac{c}{2} - \frac{k\theta}{8} \right) \Phi''(\rho_t) \Phi^{(4)}(\rho_t) \right. \\
&\quad \left. - \left( c - \frac{k}{2} \right)^2 (\Phi'''(\rho_t))^2 \geq 0 \right\} \\
&= \sup \left\{ k \in [0, c] : (c-k) \left( \frac{c}{2} - \frac{k\theta}{8} \right) - \frac{\left( c - \frac{k}{2} \right)^2 (\Phi'''(\rho_t))^2}{\Phi''(\rho_t) \Phi^{(4)}(\rho_t)} \geq 0 \right\} \\
&\geq \sup \left\{ k \in [0, c] : (c-k) \left( \frac{c}{2} - \frac{k\theta}{8} \right) - \left( c - \frac{k}{2} \right)^2 \frac{\theta}{2} \geq 0 \right\} \\
&= \frac{4(1-\theta)}{4-3\theta} c.
\end{aligned} \tag{4.401}$$

进而, 我们有如下估计:

$$\begin{aligned}
\partial_t \text{Fl}_\Phi(\mu_t \| \nu_t) &= 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{(c\nabla^2 \log \nu_t + \nabla^2 V)}^2 d\nu_t - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t \\
&\leq - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t \\
&= - \int \text{Tr}((\mathbf{X} - k_\theta \mathbf{X}_\theta)\mathbf{Y}) d\nu_t - \int \text{Tr}(k_\theta \mathbf{X}_\theta \mathbf{Y}) d\nu_t \\
&\leq -k_\theta \int \text{Tr}(\mathbf{X}_\theta \mathbf{Y}) d\nu_t \\
&\leq -k_\theta \int \|\nabla \phi\|_{\text{HS}}^2 d\nu_t \\
&\leq -\frac{k_\theta \alpha}{c} \int \|\phi\|^2 d\nu_t \\
&= -\frac{4(1-\theta)}{4-3\theta} \alpha \text{Fl}_\Phi(\mu_t \| \nu_t).
\end{aligned} \tag{4.402}$$

由 Grönwall 不等式, 可得:

$$\text{Fl}_\Phi(\mu_t \| \nu_t) \leq e^{-\frac{4(1-\theta)}{4-3\theta} \alpha t} \text{Fl}_\Phi(\mu_0 \| \nu_0). \tag{4.403}$$

**第 (4) 部分:** 除以下估计外, 其论证过程与 **第 (2) 部分** 相同:

$$\begin{aligned}
\partial_t \text{Fl}_\Phi(\mu_t \| \nu_t) &= 2 \int \Phi''(\rho_t) \|\nabla \rho_t\|_{(c\nabla^2 \log \nu_t + \nabla^2 V)}^2 d\nu_t - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t \\
&\leq -2\alpha \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 d\nu_t - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t
\end{aligned}$$

$$\begin{aligned}
&= -2\alpha \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 d\nu_t \\
&\quad - \int \text{Tr}((\mathbf{X} - k_\theta \mathbf{X}_\theta) \mathbf{Y}) d\nu_t - \int \text{Tr}(k_\theta \mathbf{X}_\theta \mathbf{Y}) d\nu_t \\
&\leq -2\alpha \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 d\nu_t - k_\theta \int \text{Tr}(\mathbf{X}_\theta \mathbf{Y}) d\nu_t \\
&\leq -2\alpha \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 d\nu_t - k_\theta \int \|\nabla \phi\|_{\text{HS}}^2 d\nu_t \\
&\leq -2\alpha \int \Phi''(\rho_t) \|\nabla \rho_t\|^2 d\nu_t - \frac{k_\theta \alpha}{c} \int \|\phi\|^2 d\nu_t \\
&= - \left( 2 + \frac{4(1-\theta)}{4-3\theta} \right) \alpha \text{Fl}_\Phi(\mu_t \| \nu_t). \tag{4.404}
\end{aligned}$$

由 Grönwall 不等式, 可得:

$$\text{Fl}_\Phi(\mu_t \| \nu_t) \leq e^{-(2 + \frac{4(1-\theta)}{4-3\theta})\alpha t} \text{Fl}_\Phi(\mu_0 \| \nu_0). \tag{4.405}$$

#### 4.5.12 定理 4.14 的证明

Proximal Sampler 的第一步对应于将当前分布与高斯核进行卷积, 即  $\mu_k^Y = \mu_k^X * \mathcal{N}(0, \eta I)$ 。由于假设  $\nu^X$  是  $\alpha$ -SLC 的, 应用高斯信道的强数据处理不等式 (SDPI, 参见定理 4.11) 可直接得到:

$$\text{Fl}_\Phi(\mu_k^Y | \nu^Y) \leq \frac{\text{Fl}_\Phi(\mu_k^X | \nu^X)}{(1 + \alpha\eta)^2}. \tag{4.406}$$

第二步可以解释为反向高斯信道的输出。为了精确建立这种联系, 定义  $\pi_0 = \nu^X$ , 并令  $\pi_t = \pi_0 * \mathcal{N}(0, tI)$  (其中  $0 \leq t \leq \eta$ ), 使得  $\pi_\eta = \nu^Y$ 。分布族  $(\pi_t)_{t \in [0, \eta]}$  的演化遵循热方程  $\partial_t \pi_t = \frac{1}{2} \Delta \pi_t$ 。现在定义其时间反转  $\pi_t^\leftarrow := \pi_{\eta-t}$ , 使得  $\pi_0^\leftarrow = \nu^Y$  且  $\pi_\eta^\leftarrow = \nu^X$ 。则  $\pi_t^\leftarrow$  满足反向热流方程:

$$\partial_t \pi_t^\leftarrow = -\frac{1}{2} \Delta \pi_t^\leftarrow = -\nabla \cdot (\pi_t^\leftarrow \nabla \log \pi_t^\leftarrow) + \frac{1}{2} \Delta \pi_t^\leftarrow. \tag{4.407}$$

我们现在通过如下随机微分方程定义反向高斯信道:

$$dX_t = \nabla \log \pi_t^\leftarrow(X_t) dt + dW_t, \tag{4.408}$$

其中  $W_t$  是标准布朗运动。如果  $X_t$  遵循此动力学演化, 则其分布  $\xi_t = \text{Law}(X_t)$  满足 Fokker-Planck 方程:

$$\partial_t \xi_t = -\nabla \cdot (\xi_t \nabla \log \pi_t^\leftarrow) + \frac{1}{2} \Delta \xi_t. \tag{4.409}$$

该反向信道具有一个有用的性质: 如果它初始化为某个  $y \in \mathbb{R}^d$  的 Dirac 测度  $\xi_0 = \delta_y$ , 则在时刻  $t = \eta$  产生的分布等于条件分布  $\xi_\eta = \nu^{X|Y}(\cdot | y)$ , 这正是受限高斯预

言机 (RGO) 所采样的分布。特别地, 如果过程从  $\xi_0 = \mu_k^Y$  开始, 则  $t = \eta$  时的分布与下一次迭代的分布一致:

$$\xi_\eta = \int_{\mathbb{R}^d} \mu_k^Y(y) \nu^{X|Y}(\cdot|y) dy = \mu_{k+1}^X. \quad (4.410)$$

因此, 我们沿着反向热流 (反向高斯流) 同时实现了从  $\pi_0^\leftarrow = \nu^Y$  到  $\pi_\eta^\leftarrow = \nu^X$ , 以及从  $\xi_0 = \mu_k^Y$  到  $\xi_\eta = \mu_{k+1}^X$  的插值。这种演化是漂移项为  $b_t = \nabla \log \pi_t^\leftarrow$  且扩散系数为  $c = 1$  的 Fokker-Planck 流的一个特例。为了符号的一致性, 我们将  $\pi_t^\leftarrow$  重标记为  $\nu_t$ , 将  $\xi_t$  重标记为  $\mu_t$ 。在此情形下, 曲率项消失, 因为:

$$-c \nabla^2 \log \nu_t + (\nabla b_t)_{\text{sym}} = 0. \quad (4.411)$$

因此, 由时间导数恒等式 (4.65), 我们得到:

$$\frac{d}{dt} \text{Fl}_\Phi(\mu_t | \nu_t) = - \int \text{Tr}(\mathbf{X}\mathbf{Y}) d\nu_t \leq 0. \quad (4.412)$$

对  $t \in [0, \eta]$  进行积分可得:

$$\text{Fl}_\Phi(\mu_{k+1}^X \| \nu^X) \leq \text{Fl}_\Phi(\mu_k^Y \| \nu^Y). \quad (4.413)$$

结合前向步骤的收缩性并对  $k$  进行迭代, 即可得到所需的指数收敛率。

此外, 如果  $\nu^X$  和  $\mu_0^X$  均为对称的 (即  $\nu^X(x) = \nu^X(-x)$  且  $\mu_0^X(x) = \mu_0^X(-x)$ ), 那么对称性在 Proximal Sampler 的所有前向和后向步骤中都得以保持。实际上, 前向步骤  $\mu_k^Y = \mu_k^X * \mathcal{N}(0, \eta I)$  显然保持偶对称性, 因为与各向同性高斯核 (偶函数) 的卷积将把任意对称密度映射为对称密度。对于后向步骤, 对称性的保持由以下引理保证。

**引理 4.10 (反向热流下的对称性保持)** 设  $(\nu_t)_{t \in [0, \eta]}$  为一族光滑的对称密度函数, 即对所有  $t$  均有  $\nu_t(x) = \nu_t(-x)$ 。考虑  $\mu_t$  满足如下反向 Fokker-Planck 方程:

$$\partial_t \mu_t = -\nabla \cdot (\mu_t \nabla \log \nu_t) + \frac{1}{2} \Delta \mu_t, \quad \mu_0 \text{ 为对称的。} \quad (4.414)$$

则对于所有  $t \in [0, \eta]$ ,  $\mu_t$  始终保持对称。

**证明** 该引理的证明是直接的。定义反射密度  $\mu_t^\sharp(x) := \mu_t(-x)$ 。由于  $\log \nu_t$  是偶函数, 其梯度是奇函数:  $\nabla \log \nu_t(-x) = -\nabla \log \nu_t(x)$ 。利用链式法则和  $\nabla \log \nu_t$  的奇偶性直接计算表明,  $\mu_t^\sharp$  满足:

$$\partial_t \mu_t^\sharp = -\nabla \cdot (\mu_t^\sharp \nabla \log \nu_t) + \frac{1}{2} \Delta \mu_t^\sharp. \quad (4.415)$$

即  $\mu_t^\sharp$  满足与  $\mu_t$  相同的偏微分方程 (PDE), 且具有相同的初始条件  $\mu_0^\sharp = \mu_0$ 。

注意到对于任意标量场  $f$  和向量场  $F$ , 我们定义其反射为  $f^\sharp(x) := f(-x)$ ,  $F^\sharp(x) := F(-x)$ 。则有:

$$\begin{aligned}\partial_t f^\sharp(x) &= (\partial_t f)(-x), \\ \nabla f^\sharp(x) &= -(\nabla f)(-x), \\ \Delta f^\sharp(x) &= (\Delta f)(-x), \\ \nabla \cdot F^\sharp(x) &= -(\nabla \cdot F)(-x).\end{aligned}$$

由线性 FPE 解的唯一性可知, 对所有  $t$  均有  $\mu_t^\sharp = \mu_t$ , 因此  $\mu_t$  是对称的。 ■

由此可见, 从对称的  $\nu^X$  和  $\mu_0^X$  出发, 我们得到  $\nu^Y$  以及每一对  $(\mu_k^X, \mu_k^Y)$  对所有  $k \in \mathbb{N}$  均保持对称。在此限制下, 高斯信道沿对称子空间具有更强的收缩性, 从而产生定理 4.14 中所述的增强收敛率。

#### 4.5.13 推论 4.2 的证明

该论证过程沿用了文献 [10] 中 *Corollary 7* 的证明思路。为保证内容的完整性, 我们在此给出具体过程。

由于假设  $\nu^X \propto e^{-g}$  是  $\alpha$ -SLC 且  $L$ -对数光滑的, 因此对于所有  $x \in \mathbb{R}^d$ , 均有  $\alpha I \leq \nabla^2 g(x) \leq LI$ 。令  $x^* = \arg \min_x g(x)$ , 从而有  $\nabla g(x^*) = 0$ 。

从  $\mu_0^X = \mathcal{N}(x^*, \frac{1}{L}I)$  出发, 我们可以如下估计初始相对  $\Phi$ -Fisher 信息:

$$\begin{aligned}\text{FI}_\Phi(\mu_0^X \parallel \nu) &= \int \Phi''\left(\frac{\mu_0^X}{\nu}\right) \left\| \nabla \frac{\mu_0^X}{\nu} \right\|^2 d\nu \\ &= \int \frac{\mu_0^X}{\nu} \Phi''\left(\frac{\mu_0^X}{\nu}\right) \left\| \nabla \log \frac{\mu_0^X}{\nu} \right\|^2 d\mu_0^X \\ &\leq \Phi''(0) \int \frac{\mu_0^X}{\nu} \left\| \nabla \log \frac{\mu_0^X}{\nu} \right\|^2 d\mu_0^X \\ &\leq \Phi''(0) \int \frac{\mu_0^X}{\nu} \|(Lx - \nabla g(x)) - (Lx^* - \nabla g(x^*))\|^2 d\mu_0^X \\ &\leq (L - \alpha)^2 \Phi''(0) C \int \exp(g(x) - L\|x - x^*\|^2) \|x - x^*\|^2 dx \\ &\leq C(L - \alpha)^2 \int \exp\left(-\frac{L}{2}\|x - x^*\|^2\right) \|x - x^*\|^2 dx \\ &\leq C(L - \alpha)^2 \frac{d}{L} \\ &\leq C \cdot dL.\end{aligned}\tag{4.416}$$

在上述估计中, 为简便起见, 所有常数乘法因子均已被吸收到  $C > 0$  中。该界利用了  $g$  的  $L$ -对数光滑性以及由  $\theta$ -容许性所蕴含的  $\Phi''$  的单调性。

事实上, 令  $h(x) = \Phi''(x)$ , 容许性条件

$$(h')^2 \leq \frac{\theta}{2} h h'' \quad (x > 0) \quad (4.417)$$

给出了不等式  $u' \geq (\frac{2}{\theta} - 1)u^2$ , 其中我们定义  $u = h'/h$ 。如果存在某个  $x_0 > 0$  使得  $u(x_0) > 0$ , 则  $u$  将在有限时间内爆破, 这与  $\Phi''$  的全局  $C^2$  正则性相矛盾。因此, 对所有  $x > 0$  均有  $u(x) \leq 0$ , 即:

$$u(x) = \frac{h'(x)}{h(x)} \leq 0 \implies h'(x) = \Phi'''(x) \leq 0. \quad (4.418)$$

故  $h(x) = \Phi''(x) \leq \Phi''(0)$ , 进而有  $x\Phi''(x) \leq \Phi''(0)x$ 。

取步长  $\eta = \frac{1}{dL}$ , 受限高斯预言机可以通过拒绝采样来实现, 且如文献 [8] 所示, 其对  $g$  的期望评估次数为  $O(1)$ 。此时, 根据定理 4.14 中的指数收敛率, 我们有:

$$\text{Fl}_{\Phi}(\mu_k^X | \nu^X) \leq \frac{\text{Fl}_{\Phi}(\mu_0^X | \nu^X)}{(1 + \alpha\eta)^{2k}} \leq \frac{CdL}{(1 + \frac{\alpha}{dL})^{2k}}. \quad (4.419)$$

为了保证  $\text{Fl}_{\Phi}(\mu_k^X | \nu^X) \leq \varepsilon$ , 只需  $\frac{CdL}{(1 + \frac{\alpha}{dL})^{2k}} \leq \varepsilon$ , 或者等价地:

$$\begin{aligned} k &\geq \frac{dL}{\alpha} \log \frac{CdL}{\varepsilon} \\ &\geq \frac{\log \frac{CdL}{\varepsilon}}{2 \log(1 + \frac{\alpha}{dL})}, \end{aligned}$$

其中最后一个不等式使用了当  $c = \frac{\alpha}{dL} \in (0, 1]$  时  $1 + c \geq e^{c/2}$  这一事实。

最后, 如果我们进一步假设  $x^* = 0$  且  $\nu^X$  是对称的, 则定理 4.14 中改进的收缩系数适用。在此情形下, 我们将  $2 \log(1 + \frac{\alpha}{dL})$  替换为

$$\left(2 + \frac{4(1 - \theta)}{4 - 3\theta}\right) \log \left(1 + \frac{\alpha}{dL}\right), \quad (4.420)$$

从而得到推论中所述的精细复杂度界。

#### 4.5.14 引理 4.5 的证明

根据 OU 方程解的温和形式 (Mild Form), 我们有:

$$X_t = e^{-Qt} X_0 + Z_t, \quad (4.421)$$

其中  $Z_t \sim \mathcal{N}(0, \Sigma_t)$  且独立于  $X_0$ 。因此，OU 过程的转移密度为：

$$k(y, x_0) := \mu_{t|0}(y | x_0) = (2\pi)^{-d/2} (\det \Sigma_t)^{-1/2} \exp\left(-\frac{1}{2} \|\Sigma_t^{-1/2} (y - e^{-Qt} x_0)\|^2\right). \quad (4.422)$$

从而有：

$$\mu_t(y) = \int_{\mathbb{R}^d} k(y, x_0) \mu_0(x_0) dx_0. \quad (4.423)$$

直接计算可得：

$$\nabla_y \log k(y, x_0) = -\Sigma_t^{-1} (y - e^{-Qt} x_0), \quad \nabla_y^2 \log k(y, x_0) = -\Sigma_t^{-1}. \quad (4.424)$$

利用恒等式  $\nabla_y^2 k = k(\nabla_y^2 \log k + (\nabla_y \log k)(\nabla_y \log k)^\top)$ ，我们得到：

$$\begin{aligned} \nabla_y k(y, x_0) &= -\Sigma_t^{-1} (y - e^{-Qt} x_0) k(y, x_0), \\ \nabla_y^2 k(y, x_0) &= \left[ -\Sigma_t^{-1} + \Sigma_t^{-1} (y - e^{-Qt} x_0) (y - e^{-Qt} x_0)^\top \Sigma_t^{-1} \right] k(y, x_0). \end{aligned} \quad (4.425)$$

通过在积分号下求导，可得：

$$\nabla_y \mu_t(y) = \int \nabla_y k(y, x_0) \mu_0(x_0) dx_0, \quad \nabla_y^2 \mu_t(y) = \int \nabla_y^2 k(y, x_0) \mu_0(x_0) dx_0. \quad (4.426)$$

事实上，对于每个  $t > 0$ ， $k(\cdot, x_0)$  及其关于  $y$  的一阶和二阶导数均被一个关于  $x_0$  的可积函数所控制，且该控制在  $y$  的紧集上是一致的；因此，由控制收敛定理（参见正文中的引理 4.6 和推论 4.4），我们可以在积分号下求导。

引入后验测度

$$\mu_{0|t}(dx_0 | y) = \frac{k(y, x_0) \mu_0(x_0)}{\mu_t(y)} dx_0, \quad (4.427)$$

并定义后验期望  $\mathbb{E}_y[\cdot] := \int (\cdot) \mu_{0|t}(dx_0 | y)$ 。记  $r = r(y, x_0) := y - e^{-Qt} x_0$ ，方程 (4.425) 给出：

$$\frac{\nabla_y \mu_t(y)}{\mu_t(y)} = -\Sigma_t^{-1} \mathbb{E}_y[r], \quad (4.428)$$

$$\frac{\nabla_y^2 \mu_t(y)}{\mu_t(y)} = -\Sigma_t^{-1} + \Sigma_t^{-1} \mathbb{E}_y[rr^\top] \Sigma_t^{-1}. \quad (4.429)$$

回顾如下矩阵恒等式：

$$\nabla_y^2 \log \mu_t = \frac{\nabla_y^2 \mu_t}{\mu_t} - \frac{\nabla_y \mu_t \nabla_y \mu_t^\top}{\mu_t^2}. \quad (4.430)$$

代入 (4.428) 和 (4.429)，并利用关系式  $\mathbb{E}_y[rr^\top] - \mathbb{E}_y[r] \mathbb{E}_y[r]^\top = \text{Cov}_y(r)$ ，我们得到：

$$\nabla_y^2 \log \mu_t(y) = -\Sigma_t^{-1} + \Sigma_t^{-1} \text{Cov}_y(r) \Sigma_t^{-1}. \quad (4.431)$$

在条件  $X_t = y$  下（等价地，在后验分布  $\mu_{0|t}(\cdot | y)$  下）， $y$  是确定性的，因此：

$$\begin{aligned}\text{Cov}_y(r) &= \text{Cov}_y(y - e^{-Qt}X_0) \\ &= e^{-Qt} \text{Cov}(X_0 | X_t = y) e^{-Q^\top t} \\ &= e^{-Qt} \text{Cov}(\mu_{0|t}(\cdot | y)) e^{-Q^\top t}.\end{aligned}$$

将其代入 (4.431) 即得所需公式：

$$\nabla_y^2 \log \mu_t(y) = -\Sigma_t^{-1} + \Sigma_t^{-1} e^{-Qt} \text{Cov}(\mu_{0|t}(\cdot | y)) e^{-Q^\top t} \Sigma_t^{-1}, \quad (4.432)$$

这也完成了对数 Hessian 公式 (4.140) 的证明。

#### 4.5.15 定理 4.18 的证明

OU 流的解给出：

$$X_t = e^{-Qt}X_0 + Z_t, \quad Z_t := \int_0^t e^{-Q(t-s)} dW_s \sim \mathcal{N}(0, \Sigma_t), \quad (4.433)$$

其中  $\Sigma_t = \int_0^t e^{-Q\tau} e^{-Q^\top \tau} d\tau > 0$ 。因此，前向 OU 转移核为：

$$k_t(y | x_0) = (2\pi)^{-d/2} (\det \Sigma_t)^{-1/2} \exp\left[-\frac{1}{2} \left\| \Sigma_t^{-1/2} (y - e^{-Qt}x_0) \right\|^2\right]. \quad (4.434)$$

设  $\mu_0(dx_0) = e^{-V_{\mu_0}(x_0)} dx_0 / Z_\mu$  且  $\nu_0(dx_0) = e^{-V_{\nu_0}(x_0)} dx_0 / Z_\nu$ 。由贝叶斯法则可得：

$$\begin{aligned}\mu_{0|t}(dx_0 | y) &= \frac{k_t(y | x_0) e^{-V_{\mu_0}(x_0)}}{\int k_t(y | z) e^{-V_{\mu_0}(z)} dz} dx_0 \\ &\propto \exp\left\{-W_{\mu,t}(x_0; y)\right\} dx_0,\end{aligned}$$

其中后验势函数定义为：

$$W_{\mu,t}(x_0; y) := V_{\mu_0}(x_0) + \frac{1}{2} \left\| \Sigma_t^{-1/2} (y - e^{-Qt}x_0) \right\|^2. \quad (4.435)$$

展开并略去仅依赖于  $y$  的常数项，可得：

$$\frac{1}{2} \left\| \Sigma_t^{-1/2} (y - e^{-Qt}x_0) \right\|^2 = \frac{1}{2} x_0^\top A_t x_0 - b_t(y)^\top x_0 + c_t(y), \quad (4.436)$$

其中

$$A_t := e^{-Q^\top t} \Sigma_t^{-1} e^{-Qt} \succ 0, \quad b_t(y) := e^{-Q^\top t} \Sigma_t^{-1} y. \quad (4.437)$$

因此有：

$$\nabla_{x_0}^2 W_{\mu,t}(x_0; y) = \nabla^2 V_{\mu_0}(x_0) + A_t, \quad \nabla_{x_0}^2 W_{\nu,t}(x_0; y) = \nabla^2 V_{\nu_0}(x_0) + A_t. \quad (4.438)$$

对于所有  $y \in \mathbb{R}^d$  (参见引理 4.5), 我们有:

$$\nabla_y^2 \log \mu_t(y) = -\Sigma_t^{-1} + \Sigma_t^{-1} e^{-Q_t} \text{Cov}(\mu_{0|t}(\cdot | y)) e^{-Q_t^\top} \Sigma_t^{-1}, \quad (4.439)$$

对  $\nu$  亦有类似结果。因此,  $\nabla_y^2 \log \nu_t \preceq K \nabla_y^2 \log \mu_t$  等价于:

$$\text{Cov}(\nu_{0|t}(\cdot | y)) \preceq K \text{Cov}(\mu_{0|t}(\cdot | y)) + (1 - K) A_t^{-1}, \quad \forall y \in \mathbb{R}^d, \quad (4.440)$$

或者, 对于任意单位向量  $u \in \mathbb{R}^d$ , 满足:

$$\text{Var}_{\nu_{0|t}}(u^\top X_0) \leq K \text{Var}_{\mu_{0|t}}(u^\top X_0) + (1 - K) u^\top A_t^{-1} u. \quad (4.441)$$

**情形 1:**  $\mu_0$  为高斯分布,  $\nu_0$  为严格对数凹分布。

由引理 4.4 中的 Brascamp-Lieb 不等式可知:

$$\text{Var}_{\nu_{0|t}}(u^\top X_0) \leq \int u^\top \left( \nabla^2 W_{\nu,t}(x_0; y) \right)^{-1} u d\nu_{0|t}. \quad (4.442)$$

初始曲率比较关系  $\nabla^2 \log \nu_0 \preceq K \nabla^2 \log \mu_0$  等价于  $\nabla^2 V_{\nu_0} \succeq K \nabla^2 V_{\mu_0}$ 。从而有:

$$\begin{aligned} \nabla^2 W_{\nu,t} &= \nabla^2 V_{\nu_0} + A_t \\ &\succeq K(\nabla^2 V_{\mu_0} + A_t) + (1 - K)A_t \\ &= K \nabla^2 W_{\mu,t} + (1 - K)A_t. \end{aligned}$$

利用映射  $X \mapsto X^{-1}$  在  $\mathbb{S}_{++}^d := \{X \in \mathbb{R}^{d \times d}; X \text{ 为正定矩阵}\}$  上的算子凸性, 可得:

$$\left( K \nabla^2 W_{\mu,t} + (1 - K)A_t \right)^{-1} \preceq K (\nabla^2 W_{\mu,t})^{-1} + (1 - K)A_t^{-1}. \quad (4.443)$$

由于  $\mu_0$  是高斯分布,  $\nabla^2 W_{\mu,t}$  为常数, 且  $\text{Var}_{\mu_{0|t}}(u^\top X_0) = u^\top (\nabla^2 W_{\mu,t})^{-1} u$ 。积分即得式 (4.441)。

**情形 2:**  $\nu_0$  为高斯分布,  $\mu_0$  为严格对数凹分布。

此时  $\nabla^2 W_{\nu,t}$  为常数, 且初始条件蕴含:

$$\begin{aligned} \nabla^2 W_{\mu,t} &= \nabla^2 V_{\mu_0} + A_t \\ &\preceq K^{-1} \nabla^2 V_{\nu_0} + A_t \\ &= K^{-1} \nabla^2 W_{\nu,t} - (K^{-1} - 1)A_t \succ 0. \end{aligned}$$

由 Cramér-Rao 不等式 (引理 4.4), 有:

$$\text{Var}_{\mu_{0|t}}(u^\top X_0) \geq \left( \mathbb{E}_y [u^\top \nabla^2 W_{\mu,t}(X_0; y) u] \right)^{-1}. \quad (4.444)$$

结合并求逆可得：

$$K \operatorname{Var}_{\mu_{0|t}}(u^\top X_0)^{-1} + (1-K) u^\top A_t u \leq u^\top \nabla^2 W_{\nu,t} u. \quad (4.445)$$

利用函数  $x \mapsto x^{-1}$  的标量凸性，我们有：

$$\begin{aligned} \operatorname{Var}_{\nu_{0|t}}(u^\top X_0) &\leq \left( K \operatorname{Var}_{\mu_{0|t}}(u^\top X_0)^{-1} + (1-K) u^\top A_t u \right)^{-1} \\ &\leq K \operatorname{Var}_{\mu_{0|t}}(u^\top X_0) + (1-K) u^\top A_t^{-1} u, \end{aligned}$$

这正是式 (4.441)，它等价于对所有  $t \geq 0$  成立的 OU 对数 Hessian 公式  $\nabla_y^2 \log \nu_t \preceq K \nabla_y^2 \log \mu_t$ 。此处我们使用了如下柯西-施瓦茨不等式：

$$(u^\top A_t u)(u^\top A_t^{-1} u) \geq \|u\|^4 = 1 \implies (u^\top A_t u)^{-1} \leq u^\top A_t^{-1} u. \quad (4.446)$$

#### 4.5.16 推论 4.7 的证明

根据 4.4.3 节中的结论，沿着 FP 流，我们有解的对数 Hessian 矩阵的如下渐近下界：

$$\begin{aligned} \nabla^2 \log \mu_t(x) &\succeq -\frac{1}{c}(\nabla^2 V + I) - \frac{2e^{-2t}}{c(1-e^{-2t})} I \\ &\quad - \left( M - \frac{1}{c} \right) \cdot \frac{e^{2t} - e^{-2t} - 4t}{2(e^{2t} + e^{-2t} - 2)} I \\ &\rightarrow -\frac{1}{c}(\nabla^2 V + I) - \frac{1}{2} \left( M - \frac{1}{c} \right) I \\ &= -\frac{1}{c} \nabla^2 V - \frac{1}{2} \left( M + \frac{1}{c} \right) I. \end{aligned}$$

因此，欲证条件  $\nabla^2 \log \mu_t \succeq K \nabla^2 V$ （其中  $K < 0$ ）最终成立（即当  $t$  充分大时成立），只需证明：

$$-\frac{1}{c} \nabla^2 V - \frac{1}{2} \left( M + \frac{1}{c} \right) I \succ K \nabla^2 V, \quad (4.447)$$

这等价于：

$$\left( -\frac{1}{c} - K \right) \nabla^2 V \succ \frac{1}{2} \left( M + \frac{1}{c} \right) I. \quad (4.448)$$

由于  $\nabla^2 V \succ \alpha I$  且  $1 + cK \leq 0$ ，若下式成立，则上述不等式成立：

$$\left( -\frac{1}{c} - K \right) \alpha > \frac{1}{2} \left( M + \frac{1}{c} \right). \quad (4.449)$$

结合  $q(2 + cK) \geq 1 + cK$ ，即  $K \geq \frac{2q-1}{c(1-q)}$ ，我们得到如下充分条件：

$$\frac{2q-1}{c(1-q)} \leq K < -\frac{M}{2\alpha} - \frac{1}{c} \left( \frac{1}{2\alpha} + 1 \right), \quad (4.450)$$

即：

$$(1 + cM - 2\alpha)q < 1 + cM. \quad (4.451)$$

**第(1)部分：**若  $1 + cM - 2\alpha \leq 0$ ，则上述条件对所有  $q \in [1, +\infty)$  均成立。因此，对于任意

$$K \in \left[ \frac{2q-1}{c(1-q)}, -\frac{M}{2\alpha} - \frac{1}{c} \left( \frac{1}{2\alpha} + 1 \right) \right), \quad (4.452)$$

存在  $T > 0$ ，使得对所有  $t \geq T$ ，均有  $\nabla^2 \log \mu_t \succeq K \nabla^2 V$ 。此时由定理 4.9，我们有：

$$\text{RFI}_q(\mu_t \|\nu) \leq e^{-2q[q(2+cK)-(1+cK)]\alpha t} \text{RFI}_q(\mu_T \|\nu), \quad \forall t \geq T. \quad (4.453)$$

对于某个微小的  $\delta > 0$ ，选取  $K = -\frac{M}{2\alpha} - \frac{1}{c} \left( \frac{1}{2\alpha} + 1 \right) - \delta$ ，即可得到所述收缩率。

**第(2)部分：**若  $1 + cM - 2\alpha > 0$ ，则上述条件对所有

$$1 \leq q < \frac{1 + cM}{1 + cM - 2\alpha} \quad (4.454)$$

均成立。因此，对于任意

$$K \in \left[ \frac{2q-1}{c(1-q)}, -\frac{M}{2\alpha} - \frac{1}{c} \left( \frac{1}{2\alpha} + 1 \right) \right), \quad (4.455)$$

存在  $T > 0$ ，使得对所有  $t \geq T$ ，均有  $\nabla^2 \log \mu_t \succeq K \nabla^2 V$ 。此时由定理 4.9，我们有：

$$\text{RFI}_q(\mu_t \|\nu) \leq e^{-2q[q(2+cK)-(1+cK)]\alpha t} \text{RFI}_q(\mu_T \|\nu), \quad \forall t \geq T. \quad (4.456)$$

同样，对于某个微小的  $\delta > 0$ ，选取  $K = -\frac{M}{2\alpha} - \frac{1}{c} \left( \frac{1}{2\alpha} + 1 \right) - \delta$ ，即可得到所述收缩率。

#### 4.5.17 定理 4.21 中势函数的一个例子

作为说明性示例，我们考察  $V(x) = \frac{x^2}{2} + (1+x^2)^{p/2}$ ，其中  $0 < p \leq 2$ 。此时有：

$$V'(x) = x + px(1+x^2)^{\frac{p}{2}-1},$$

$$V''(x) = 1 + p(1+x^2)^{\frac{p}{2}-1} + p(p-2)x^2(1+x^2)^{\frac{p}{2}-2},$$

$$V'''(x) = p(p-2) \left[ 3x(1+x^2)^{\frac{p}{2}-2} + (p-4)x^3(1+x^2)^{\frac{p}{2}-3} \right],$$

$$V^{(4)}(x) = p(p-2) \left[ 3(1+x^2)^{\frac{p}{2}-2} + 6(p-4)x^2(1+x^2)^{\frac{p}{2}-3} + (p-4)(p-6)x^4(1+x^2)^{\frac{p}{2}-4} \right].$$

因此， $V(x) \in \mathcal{D}^{(4)}$  是参数  $\alpha = 1$  的强凸函数，且  $V''$ ,  $V'''$ ,  $V^{(4)}$  均有界。故存在以下常数：

$$K_3 := \sup_x (V''(x))^2 < \infty, \quad K_4 := \sup_x |V'(x)V'''(x)| < \infty, \quad K_5 := \sup_x |V^{(4)}(x)| < \infty. \quad (4.457)$$

注意到:

$$\begin{aligned} h(x) &= \frac{1}{2c}x^2 + \frac{p}{c}x^2(1+x^2)^{\frac{p}{2}-1} + \frac{p^2}{2c}x^2(1+x^2)^{p-2} - \frac{1}{2c}x^2 - \frac{1}{2}V''(x) + \frac{1}{2} \\ &= \frac{p}{c}x^2(1+x^2)^{\frac{p}{2}-1} + \frac{p^2}{2c}x^2(1+x^2)^{p-2} - \frac{1}{2}V''(x) + \frac{1}{2} \\ &\geq -\frac{\sqrt{K_3}}{2} + \frac{1}{2} > -\infty, \end{aligned}$$

因此  $h$  有下界。此外,

$$\tilde{h}''(x) = \frac{1}{c} \left[ (V''(x))^2 + V'(x)V'''(x) \right] - \frac{1}{2}V^{(4)}(x) \leq \frac{K_3 + K_4}{c} + \frac{1}{2}K_5 =: M. \quad (4.458)$$

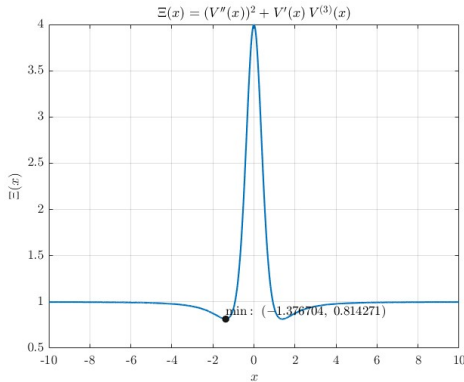
从而可知  $\tilde{h}$  是  $M$ -光滑的, 其中我们可以取  $M = \frac{K_3+K_4}{c} + \frac{1}{2}K_5$ 。为保证  $\tilde{h}$  的凸性, 定义:

$$\Xi(x) := (V''(x))^2 + V'(x)V'''(x), \quad \Upsilon(x) := V_+^{(4)}(x) = \max\{V^{(4)}(x), 0\}. \quad (4.459)$$

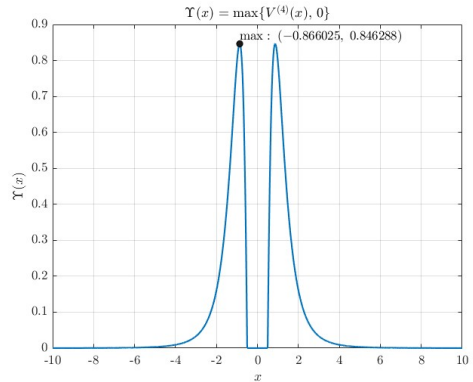
若  $\inf_x \Xi(x) > 0$ , 则任意

$$0 < c \leq c_{\max} := \frac{2 \inf_x \Xi(x)}{\sup_x \Upsilon(x)} \quad (4.460)$$

均能保证  $\tilde{h}$  在  $\mathbb{R}$  上是凸的。对于  $p = 1$ , 数值计算表明  $\inf_x \Xi(x) > 0.8$ ,  $\sup_x \Upsilon(x) < 0.85$ , 故  $c_{\max} > 1.88235$ 。因此, 任意  $c \in (0, 1.88235]$  均可保证凸性; 参见图 4.1 和图 4.2。



(a)  $\Xi(x) = (V''(x))^2 + V'(x)V^{(3)}(x)$



(b)  $\Upsilon(x) = \max\{V^{(4)}(x), 0\}$

图 4.1 对于  $p = 1$  的  $V(x) = \frac{x^2}{2} + (1+x^2)^{p/2}$ ,  $\Xi$  和  $\Upsilon$  的示意图。

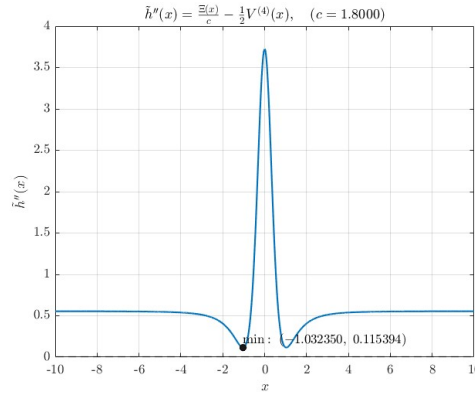


图 4.2 对于  $p = 1$  且  $c = 1.8$  的  $V(x) = \frac{x^2}{2} + (1 + x^2)^{p/2}$ ,  $\tilde{h}$  的二阶导数。

## 4.6 本章小结

本章推导了相对 Rényi-Fisher 信息和相对  $\Phi$ -Fisher 信息沿 Fokker-Planck 流的微分恒等式，并利用它们在曲率控制下建立了强数据处理不等式。对于 RFI，我们推导了时间导数公式，并在满足下述条件时获得了收缩性：

$$\nabla^2 \log \nu_t \preceq K_1 \nabla^2 \log \mu_t + K_2 \nabla^2 V. \quad (4.461)$$

对于  $\Phi$ -FI，我们提出了矩阵结构的导数恒等式，证明了  $\theta$ -容许生成函数的相应 (S)DPI，并推导了近端采样器在  $\Phi$ -FI 度量下的指数收敛保证。在方法上，该分析结合了 Gozlan-Li 的扩散半群对数 Hessian 公式、曲率界和对数凹性保持；特别地，我们展示了高斯初始化下热流和 Ornstein-Uhlenbeck 流的全时保持性。

主要的挑战在于在一般设置中验证和加强对数 Hessian 比较。控制高维对数密度的高阶导数仍然很微妙。目前，将比较原理细化为对动态  $\nu_t$ （不一定等同于  $\nu_\infty$ ）成立仍在研究中，这将允许直接应用定理 4.9 来获得相对 Rényi-Fisher 信息沿一般 FP 流的 SDPI。我们的永久比较结果也假设了高斯初始化；将全时控制扩展到此类之外是开放的。FP 模型的结构假设也很严格：在高维中，构造使得辅助函数  $\tilde{h}$  全局凸的非二次势函数  $V$  尚未解决。此外，一些  $\Phi$ -FI 速率依赖于  $\theta$ -容许性；消除这一技术条件将使算法结果更加清晰。此外，关于相对 Fisher 信息混合时间保证的相关进展包括 [11, 137]，其重点是非对数凹采样。鉴于这些结果，我们期望在目前的框架下自然涌现出非对数凹采样在  $\Phi$ -FI 中的类似收敛保证。

## 第 5 章 基于现代分析方法的率失真理论

### 5.1 本章引言

率失真理论作为信息论的一个基础分支，其核心任务是在满足给定失真约束的前提下，探究数据压缩的最小编码速率。这一理论最早由 Claude E. Shannon 在其 1948 年的开创性工作 [1] 以及 1959 年的经典论文 [12] 中提出，自此便成为现代数据压缩与通信系统设计的基石。率失真问题的本质在于设计一种编码方案，在特定的信源分布与失真度量下，寻求编码速率与重构质量之间的最优权衡。

经典的率失真问题主要关注有限字母表信源或定义在欧几里得空间上的连续信源，相关研究已在 Berger [139]、Blahut [14, 140] 以及 Cover [71] 等人的著作中得到了详尽的阐述。随着研究的深入，学者们开始探索更为复杂的问题，即抽象字母表上的率失真问题。其中，I. Csiszár 做出了里程碑式的贡献，他将 RD 理论严格推广至抽象字母表 [13]，并导出了 RD 函数的参数化表示，为 Blahut-Arimoto 算法 [14-15] 等数值计算方法奠定了坚实的理论基础。随后，F. Rezaei 等人 [16] 利用 Stone-Čech 紧化和局部凸拓扑向量空间中的不动点方法，显著推进了 Csiszár 的结果。他们不仅放宽了目标空间（重构空间）必须为紧致空间的假设，还在这些广义条件下对最优重构分布进行了全面的刻画。此外，E. Riegler 等人 [141-142] 聚焦于定义在紧流形和分形集合上的有损压缩理论，利用率失真维数（Rate-distortion dimension）的相关结果 [143]，在更一般的设定下导出了香农下界（Shannon Lower Bound, SLB）。在 RD 函数的参数化表示方面，V. Kostina 和 E. Tuncel [144] 避免了对 Karush-Kuhn-Tucker (KKT) 最优性条件的依赖，利用 Donsker-Varadhan 的最小相对熵变分公式 [145] 简化了 Csiszár 的论证，极大地推进了抽象信源的逐次细化（Successive Refinement）理论，推广了 Equitz 和 Cover 的早期成果 [146]。

尽管率失真理论已取得显著进展，但随着几何与分析等现代数学工具的渗透，这一经典理论在数学基础层面的局限性日益凸显，亟需利用新的数学框架进行重构。

首先，从几何视角来看，率失真理论与最优传输（Optimal Transport, OT）理论存在着天然的联系。2013 年，Robert M. Gray 深刻阐述了最优传输与信息论之间的联系 [18]，提供了信源编码的几何视角，并从量化的角度简要综述了最优传输理论。然而，传统的 OT 理论难以直接应用于 RD 问题，因为两者的基本设定存在本质差异：传统

---

本章内容已发布在预印本 Arxiv:2501.09362 [138]，并包含了 IEEE ISIT 2026 在投论文的部分内容。

OT 旨在最小化两个固定边缘分布之间的传输成本，而 RD 理论则是在固定源分布的前提下，通过优化目标测度（重构分布）的正则分解（Regular Disintegration）来平衡编码速率与失真。为了弥合这一鸿沟，Yang 等人 [147] 在熵正则化最优传输（Entropic Optimal Transport, EOT）的框架下，通过引入熵正则项建立了一定的等价性。RD 对偶问题的形式与 EOT 问题非常相似，仅差一个乘法常数，这使得利用 EOT 最优传输计划的闭式解来推导 RD 对偶问题的解成为可能。然而，这种方法在很大程度上依赖于单一失真约束的设定。当考虑广义的多约束 RD 问题时，基于 EOT 的框架往往失效，无法给出有效的参数化表示。

作为 OT 和 EOT 的推广，Gozlan 等人 [19] 近期提出的最优弱传输（Optimal Weak Transport, OWT）理论为 RD 问题提供了一个更自然的数学框架。OWT 天然地关注于优化目标测度相对于源测度的分解，这与 RD 问题的数学结构完美契合。在 OWT 框架下，我们不仅能够更简洁地重现 RD 理论中的若干关键结果，还能导出连接 RD 函数与薛定谔桥问题（Schrödinger Bridge Problem）[148] 的参数化表示。更重要的是，基于 OWT 问题解的存在性定理 [149-150]，我们可以极大地简化 RD 函数最优性必要条件的证明。这种方法提供了一种严谨而简洁的途径，能够在不依赖传统变分法的情况下，重现 K. Rose 关于香农下界可达性的经典结论 [151]，从而规避了变分方法中可能存在的缺乏严密性的风险。

其次，从分析视角来看，尽管 Rezaei 等人的工作放宽了紧性假设，但在处理一般的非紧波兰空间（如  $\mathbb{R}^n$  或一般的黎曼流形）时，最优重构分布的存在性仍然是一个棘手的分析难题。经典结果往往依赖于重构字母表的紧性或边缘分布的绝对连续性，这允许使用变分法或凸对偶理论来确立最优重构策略的存在性与特征。然而，在更一般的设定下——例如重构空间非紧且边缘分布缺乏绝对连续性时——这些结构性假设不再成立。在这种情况下，率失真泛函的极小化序列可能会表现出病态行为，如质量“逃逸”到无穷远处（Vanishing）或分裂成相距甚远的碎片（Dichotomy），导致“紧性缺失（Loss of Compactness）”。这促使我们寻找新的数学工具来分析抽象或非经典设定下的率失真问题。

为了解决这一存在性难题，本章引入了 P.-L. Lions 在变分法和非线性分析领域发展的集中紧性原理（Concentration-Compactness Principle）[21-22, 24]。这一原理提供了一个强大的框架，用于克服因平移不变性或质量逃逸现象导致的紧性丧失的困难，特别是在标准紧性论证失效的非紧空间设定中。

综上所述，本章旨在通过现代分析方法的双重视角——最优弱传输与集中紧性

原理——对率失真理论进行重构。我们首先在 OWT 框架下建立 RD 函数的几何结构与参数化表示，及其与薛定谔桥问题的联系；随后，利用集中紧性原理，在无需重构空间紧性或边缘分布绝对连续性的条件下，严格分析极小化序列的渐近行为（紧性、消散或二分），从而确立最优重构分布的存在性。这一融合不仅通过引入非线性泛函分析的深刻工具，为率失真理论构建了全新的概念框架，更为处理标准假设失效下的复杂信源编码问题（如函数空间信源、连续时间信号或高分辨率几何数据）开辟了新的研究路径。

## 5.2 最优弱传输视角下的率失真函数

在经典信息论中，率失真函数  $R(D)$  通常被定义为互信息  $I(X; Y)$  在给定失真约束  $\mathbb{E}[\rho(X, Y)] \leq D$  下的最小值。这一优化问题本质上是寻找最优的条件概率分布（即从源空间  $\mathcal{X}$  到重构空间  $\mathcal{Y}$  的转移概率核，或称为正则分解）。然而，正如引言所述，这种视角在处理一般度量空间时，难以直接利用最优传输理论中关于联合分布几何性质的深刻结果。

本节将利用 Gozlan 等人提出的最优弱传输理论框架[19]，对率失真问题进行重构。我们将证明，率失真问题可以自然地转化为一个以重构分布  $\nu$  为优化变量的变分问题，并利用 OWT 的对偶理论导出  $R(D)$  的显式参数化表示。这一结果不仅涵盖了 Csiszár 的经典结论，还揭示了率失真函数与随机控制中的薛定谔桥问题（Schrödinger Bridge Problem）[148] 之间的内在联系。

### 5.2.1 最优弱传输理论简介

最优传输理论旨在研究两个概率测度之间质量传输的最优方案。给定波兰空间（完备可分度量空间） $(\mathcal{X}, d_X)$  和  $(\mathcal{Y}, d_Y)$  上的两个概率测度  $\mu$  和  $\nu$ ，最优传输问题旨在寻找一个传输计划  $\pi$ ，以最小化传输成本：

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y) \quad (5.1)$$

其中， $\Pi(\mu, \nu)$  表示以  $\mu$  和  $\nu$  为边缘分布的所有传输计划的集合， $c(x, y)$  是将质量从  $x$  传输到  $y$  的成本函数。最优传输问题在数学领域得到了广泛研究，并已应用于机器学习 [152-155]、统计学 [156-158] 和信息论 [5, 159-161] 等多个领域。关于最优传输的背景知识，读者可参考 [17]。

弱传输理论最初是受几何不等式应用的启发而发展起来的，它引发了各研究小组的广泛研究兴趣，并已在多个领域得到应用 [149-150, 162-166]。在本章中， $\mathcal{X}$  和

$\mathcal{Y}$  始终表示波兰空间。给定概率测度  $\mu \in \mathcal{P}(\mathcal{X})$  和  $\nu \in \mathcal{P}(\mathcal{Y})$ ，我们用  $\Pi(\mu, \nu)$  表示  $\mathcal{X} \times \mathcal{Y}$  上所有以  $\mu$  和  $\nu$  为边缘分布的耦合 (couplings) 的集合。给定  $\mathcal{X} \times \mathcal{Y}$  上的一个耦合  $\pi$ ，我们用  $(\pi_x)_{x \in \mathcal{X}}$  表示其关于第一个边缘分布的正则分解 (即条件分布)。我们考虑形式如下的成本泛函：

$$C : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \rightarrow \mathbb{R} \cup \{+\infty\} \quad (5.2)$$

其中  $C$  是有下界且下半连续的，并且对于所有的  $x \in \mathcal{X}$ ， $C(x, \cdot)$  在  $\mathcal{P}(\mathcal{Y})$  上是凸的。弱传输问题旨在寻找一个耦合  $\pi$ ，以最小化如下成本泛函：

$$V_C(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X}} C(x, \pi_x) d\mu(x). \quad (5.3)$$

经典的最优传输问题可以通过取  $C(x, p) = \int c(x, y) dp(y)$  (其中  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$  为某种成本函数) 而被包含在内。对于  $t \geq 1$ ， $\mathcal{P}_{d_Y}^t(\mathcal{Y})$  (简记为  $\mathcal{P}_t(\mathcal{Y})$ ) 表示关于某个固定度量  $d_Y$  (与  $\mathcal{Y}$  上的拓扑相容) 具有有限  $t$  阶矩的 Borel 概率测度的集合。也就是说，一个 Borel 概率测度  $\nu$  属于  $\mathcal{P}_t(\mathcal{Y})$  当且仅当对于某个  $y_0 \in \mathcal{Y}$ ，有

$$\int_{\mathcal{Y}} d_Y(y, y_0)^t d\nu(y) < \infty. \quad (5.4)$$

$\Phi_t(\mathcal{Y})$  表示  $\mathcal{Y}$  上被  $1 + d_Y(y, y_0)^t$  的倍数所控制的连续函数的集合。我们在概率测度集  $\mathcal{P}_t(\mathcal{Y})$  上赋予  $t$  阶 Wasserstein 拓扑。具体而言，一个序列  $(\nu_k)_{k \in \mathbb{N}}$  收敛于  $\nu \in \mathcal{P}_t(\mathcal{Y})$ ，如果对于所有的  $f \in \Phi_t(\mathcal{Y})$ ， $\nu_k(f) := \int f d\nu_k$  收敛于  $\nu(f)$ 。空间  $\mathcal{P}(\mathcal{Y})$  本身赋予通常的弱拓扑。同样的约定也适用于  $\mathcal{X}$ 。

**定义 5.1** (A) 成本泛函  $C : \mathcal{X} \times \mathcal{P}_t(\mathcal{Y}) \rightarrow \mathbb{R} \cup \{+\infty\}$  满足性质 (A) 当且仅当：

1.  $C$  关于  $\mathcal{X} \times \mathcal{P}_t(\mathcal{Y})$  上的乘积拓扑是下半连续的；
2.  $C$  是有下界的；
3. 映射  $p \mapsto C(x, p)$  是凸的，即对于所有的  $x \in \mathcal{X}$ ，所有的  $p, q \in \mathcal{P}_t(\mathcal{Y})$  以及  $\lambda \in [0, 1]$ ，有

$$C(x, \lambda p + (1 - \lambda)q) \leq \lambda C(x, p) + (1 - \lambda)C(x, q). \quad (5.5)$$

**定理 5.1** (存在性与半连续性)[Theorem 3.2 in [150]] 如果成本泛函  $C$  满足性质 (A)，则 (5.3) 中的下确界是可以达到的，并且值  $V_C(\mu, \nu)$  依赖于边缘分布  $(\mu, \nu) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}_t(\mathcal{Y})$  的方式是下半连续的。

### 5.2.2 薛定谔桥问题简介

(静态) 薛定谔问题 (Schödinger Problem, SP), 也称为薛定谔桥问题, 最早在 1931 年的文献 [20] 中被提出。虽然这个问题历史悠久, 但在过去的几十年里, 由于其与各个领域 (特别是熵正则化最优传输理论) 的深刻联系, 引起了广泛关注。用现代术语来说, SP 可以表述如下: 对于给定的两个波兰空间  $\mathcal{X}, \mathcal{Y}$  及其上的概率测度  $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ , 以及给定的参考概率测度  $\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ , 我们用  $\Pi(\mu, \nu) \subset \mathcal{P}(\mathcal{X} \times \mathcal{Y})$  表示耦合的集合, 即满足以下条件的所有  $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$  的集合:

$$\begin{aligned} \int_{\mathcal{X} \times \mathcal{Y}} f(x) d\pi(x, y) &= \int_{\mathcal{X}} f(x) d\mu(x) \\ \int_{\mathcal{X} \times \mathcal{Y}} g(y) d\pi(x, y) &= \int_{\mathcal{Y}} g(y) d\nu(y) \end{aligned} \quad (5.6)$$

其中  $f, g$  为任意有界可测函数。薛定谔问题旨在寻找一个耦合  $\pi \in \Pi(\mu, \nu)$ , 以最小化相对于参考测度  $\gamma$  的相对熵:

$$\text{KL}(\pi \parallel \gamma) \equiv D_{KL}(\pi \parallel \gamma) = \begin{cases} \int_{\mathcal{X} \times \mathcal{Y}} \log \frac{d\pi}{d\gamma} d\pi, & \text{若 } \pi \ll \gamma \\ +\infty, & \text{否则} \end{cases} \quad (5.7)$$

这个问题可以被视为最优弱传输 (OWT) 问题的一个特例。假设值  $\inf_{\pi \in \Pi(\mu, \nu)} D_{KL}(\pi \parallel \gamma)$  是有限的, 那么 OWT 理论保证了 SP 中极小值的存在性, 并提供了优化器的结构。

**引理 5.1** [Corollary 4.3 in [150]] 设  $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ , 且  $\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$  是一个与  $\mu \times \nu$  等价的概率测度。假设相应的熵正则化最优传输问题的值是有限的, 即

$$\inf_{\pi \in \Pi(\mu, \nu)} D_{KL}(\pi \parallel \gamma) < +\infty. \quad (5.8)$$

一个耦合  $\pi^*$  是上述问题的极小值当且仅当  $D_{KL}(\pi^* \parallel \gamma) < \infty$  且存在可测函数  $f, g$  使得

$$\frac{d\pi^*}{d\gamma}(x, y) = f(x)g(y), \quad \gamma\text{-几乎处处}. \quad (5.9)$$

事实上, 上述引理中的  $f(x)$  和  $g(y)$  是非负的, 且在相差一个乘法常数的意义下是唯一的 [167]。在  $d\gamma = Ke^{-\beta\rho(x, y)}d\mu \times d\nu$  (其中  $\rho(x, y)$  为非负可测函数) 的情况下, 由于  $\pi^* \in \Pi(\mu, \nu)$ , 我们有  $\int_{\mathcal{Y}} d\pi^* = d\mu, \int_{\mathcal{X}} d\pi^* = d\nu$ , 这可以写成如下的薛定谔方程组 [167]:

$$\begin{aligned} 1 &= Kf(x) \int_{\mathcal{Y}} g(y)e^{-\beta\rho(x, y)} d\nu \\ &= Kg(y) \int_{\mathcal{X}} f(x)e^{-\beta\rho(x, y)} d\mu \\ &= \int_{\mathcal{X}} \frac{g(y')e^{-\beta\rho(x, y')}}{\int_{\mathcal{Y}} g(y)e^{-\beta\rho(x, y)} d\nu} d\mu, \quad \forall y' \in \mathcal{Y}. \end{aligned} \quad (5.10)$$

其中

$$\begin{aligned} K &= \left( \int_{\mathcal{X} \times \mathcal{Y}} e^{-\beta \rho(x,y)} d\mu d\nu \right)^{-1} \\ &= \left( \int_{\mathcal{X} \times \mathcal{Y}} f(x)g(y)e^{-\beta \rho(x,y)} d\mu d\nu \right)^{-1} \end{aligned} \quad (5.11)$$

在接下来的章节中，我们将回顾率失真理论，并在弱传输理论的框架下探讨其与薛定谔问题的联系。

### 5.2.3 最优弱传输视角下的率失真理论

在本节中，我们将利用 OWT 的框架重新审视率失真理论。我们首先将原本关注于优化分解（disintegration）的 RD 问题转化为一个在给定目标测度（或重构分布）条件下的 OWT 问题，即优化联合分布。随后，我们对目标测度求下确界。我们将利用弱传输问题的存在性定理及其在薛定谔问题中的应用，给出 RD 函数的参数化表示。此外，在最优重构分布存在的条件下，我们将证明 RD 函数的最优性必要条件，从而提供 K. Rose [151] 提出的 SLB 可达性的简洁证明。我们的主要贡献总结在定理 5.2 至定理 5.4 中。

**引理 5.2** 在假设 2.1 和 2.2 下，率失真函数  $R(D)$  可以表示为：

$$\inf_{\pi_x: \mathbb{E}\pi \rho \leq D} I(X;Y) = \inf_{\nu \in \mathcal{P}(\mathcal{Y})} \inf_{\pi \in \Pi(\mu, \nu): \mathbb{E}\pi \rho \leq D} I(X;Y) \quad (5.12)$$

证明见第 5.4.1 小节。

这一引理在 I. Csiszár 的论文 [13] 第一部分中已有隐含证明。它也是 Eric Lei 等人 [168] 工作中的主要结果。然而，他们的证明存在漏洞：直接依赖了 M. Cuturi 的书 [169] 中关于离散设置下 EOT 问题的结果，而该结果并不能直接推广到连续或抽象字母表设置。因此，我们在本文中将该结果作为一个独立的引理重新陈述，并提供了另一种证明。

**注** 注意  $I(X;Y) = \int_{\mathcal{X} \times \mathcal{Y}} \log \frac{d\pi}{d\mu d\nu} d\pi = \int_{\mathcal{X}} d\mu(x) \int_{\mathcal{Y}} \log \frac{d\pi_x}{d\nu} d\pi_x$ ，且  $\pi \ll \mu \times \nu \iff \pi_x \ll \nu$ 。因此，我们始终假设  $\pi_x \ll \nu$ ，以确保下确界是有意义的。

考虑率失真问题的（弱）对偶问题：

$$\begin{aligned} R(D) &= \inf_{\nu \in \mathcal{P}(\mathcal{Y})} \inf_{\pi \in \Pi(\mu, \nu): \mathbb{E}\pi \rho \leq D} I(X;Y) \\ &\geq \inf_{\nu \in \mathcal{P}(\mathcal{Y})} \sup_{\beta \geq 0} \inf_{\pi \in \Pi(\mu, \nu)} \{I(X;Y) + \beta(\mathbb{E}\pi \rho - D)\}. \end{aligned} \quad (5.13)$$

定义

$$\begin{aligned}
 J(\nu, \beta) &:= \inf_{\pi \in \Pi(\mu, \nu)} \{I(X; Y) + \beta(\mathbb{E}_\pi \rho - D)\} \\
 &= \int_{\mathcal{X}} d\mu(x) \int_{\mathcal{Y}} \left[ \log \frac{d\pi_x}{d\nu} + \beta\rho(x, y) - \beta D \right] d\pi_x \\
 d\gamma(x, y) &:= K e^{-\beta\rho(x, y)} d\mu(x) d\nu(y)
 \end{aligned} \tag{5.14}$$

其中  $K = \left( \int_{\mathcal{X} \times \mathcal{Y}} e^{-\beta\rho(x, y)} d\mu d\nu \right)^{-1}$ , 使得  $\gamma \ll \mu \times \nu$  是  $\mathcal{X} \times \mathcal{Y}$  上的一个概率测度 (注意  $\gamma$  实际上与  $\mu \times \nu$  等价)。我们有

$$\begin{aligned}
 &I(X; Y) + \beta(\mathbb{E}_\pi \rho - D) \\
 &= \int_{\mathcal{X}} D_{KL}(\pi_x \| \gamma_x) d\mu + \int_{\mathcal{X}} \log K_x d\mu(x) - \beta D \\
 &= \int_{\mathcal{X} \times \mathcal{Y}} \log \frac{d\pi_x}{d\gamma_x} d\pi - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta\rho(x, y)} d\nu \right) d\mu(x) - \beta D \\
 &= \int_{\mathcal{X} \times \mathcal{Y}} \log \frac{d\pi}{d\gamma} d\pi - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta\rho(x, y)} d\nu \right) d\mu(x) - \beta D + \int_{\mathcal{X}} \log \frac{d\gamma_0}{d\mu} d\mu
 \end{aligned} \tag{5.15}$$

其中  $\gamma_0$  是  $\gamma$  的  $X$ -边缘分布, 且

$$d\gamma_x = K_x e^{-\beta\rho(x, y)} d\nu, \quad K_x = \left( \int_{\mathcal{Y}} e^{-\beta\rho(x, y)} d\nu \right)^{-1}. \tag{5.16}$$

因此, 我们只需要考虑  $D_{KL}(\pi \| \gamma)$  在  $\pi \in \Pi(\mu, \nu)$  上的下确界。

**注** 或许有人会担心  $K, K_x$  的有限性以及  $\log K_x$  的可积性, 但实际上它们总是成立的。值得注意的是,  $K = +\infty$  当且仅当  $\rho(x, y) = +\infty$  在  $\mu \times \nu$ -几乎处处成立, 这意味着对于每个满足  $\pi_x \ll \nu$  的  $\pi \in \Pi(\mu, \nu)$ ,  $\rho = +\infty$  在  $\pi$ -几乎处处成立。在这种情况下,  $\mathbb{E}_\pi \rho = +\infty$  且  $J(\nu, \beta) = +\infty$ 。因此我们只需要考虑  $K < +\infty$  的情况。类似地,  $K_x = +\infty$  当且仅当对于给定的  $x \in \mathcal{X}$ ,  $\rho(x, y) = +\infty$  在  $\nu$ -几乎处处成立, 这意味着对于  $\pi_x \ll \nu$ ,  $\rho = +\infty$  在  $\pi_x$ -几乎处处成立。于是我们必须有  $K_x < +\infty$  在  $\mu$ -几乎处处成立, 否则  $J(\nu, \beta)$  将变为无穷大。因此  $K_x = +\infty$  的情况可以忽略。此外, 如果  $\int_{\mathcal{X}} \log K_x d\mu(x) = +\infty$ , 我们同样有  $J(\nu, \beta) = +\infty$ , 这种情况也可以排除。

我们考虑  $J(\nu, \beta)$  极小值的存在性。定义

$$C(x, p) := \int_{\mathcal{Y}} \left[ \log \frac{dp}{d\nu} + \beta\rho(x, y) - \beta D \right] dp, \quad p \in \mathcal{P}(\mathcal{Y}). \tag{5.17}$$

**假设 5.1**  $\int_{\mathcal{Y}} \rho(x, y) dp$  关于  $(x, p) \in \mathcal{X} \times \mathcal{P}(\mathcal{Y})$  是联合下半连续的。

如果  $\mathcal{X}, \mathcal{Y}$  是波兰空间且  $\rho \in C_b(\mathcal{X} \times \mathcal{Y})$ , 则此假设成立。可以验证  $(x, p) \rightarrow \int_{\mathcal{Y}} \rho(x, y) dp$  确实是有界且联合连续的。此外, 如果  $\rho(x, y)$  是联合下半连续且有上界的 (下界为 0), 那么它可以被一系列有界  $k$ -Lipschitz 函数递增逼近, 这意味着  $C(x, p)$  关于  $\mathcal{X} \times \mathcal{P}(\mathcal{Y})$  上的乘积拓扑是联合下半连续的。验证这些断言的一种方法是使用 [170] 中的 *Theorem A.3.12*。

**引理 5.3** 在假设 5.1 下, 成本泛函  $C(x, p)$  是联合下半连续的, 有下界的, 并且关于第二个变量是凸的。

证明见第 5.4.2 小节。

**推论 5.1** 在假设 5.1 下,  $J(\nu, \beta)$  存在极小值, 即存在  $\pi^* \in \Pi(\mu, \nu)$  使得

$$\begin{aligned} J(\nu, \beta) &= \int_{\mathcal{X}} d\mu(x) \int_{\mathcal{Y}} \left[ \log \frac{d\pi_x^*}{d\nu} + \beta\rho(x, y) - \beta D \right] d\pi_x^* \\ &= \int_{\mathcal{X} \times \mathcal{Y}} \log \frac{d\pi^*}{d\gamma} d\pi^* - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta\rho(x, y)} d\nu \right) d\mu(x) - \beta D + \int_{\mathcal{X}} \log \frac{d\gamma_0}{d\mu} d\mu. \end{aligned} \quad (5.18)$$

这一推论可以直接由 [149] 中的 *Theorem 1.1* 导出。然而, 假设 5.1 在许多情况下可能不成立, 例如  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$  且  $\rho(x, y) = \|x - y\|^2$  的情况。因此, 我们需要利用改进的存在性定理 5.1 来处理一般情况。在此之前, 我们需要对  $\rho(x, y)$  和信源  $\mu$  施加一些额外条件。

**假设 5.2** 对于空间  $\mathcal{Y}$  上给定的度量  $d_{\mathcal{Y}}$ ,  $\int_{\mathcal{Y}} \rho(x, y) dp$  关于  $(x, p) \in \mathcal{X} \times \mathcal{P}_t(\mathcal{Y})$  是联合下半连续的。

这一假设比假设 5.1 更弱, 因为  $\mathcal{P}_t(\mathcal{Y})$  中的收敛 (或拓扑) 严格强于  $\mathcal{P}(\mathcal{Y})$  中的收敛。这一扩展拓宽了适用场景, 例如当成本函数形式为  $\rho(x, y) = \|x - y\|^t$  (其中  $t \geq 1$ ) 且  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$  赋予欧几里得度量时。我们也可以证明成本泛函  $C(x, p)$  满足性质 (A)。此外, 我们还提供了一个反例说明该假设并不总是成立; 详情见第 5.4.3 小节。

**定理 5.2** 在假设 5.2 下, 成本泛函  $C(x, p)$  满足性质 (A)。假设  $\nu \in \mathcal{P}_t(\mathcal{Y})$ , 则  $J(\nu, \beta)$  存在极小值  $\pi^*$ 。

证明见第 5.4.4 小节。

为了将重构分布限制在空间  $\mathcal{P}_t(\mathcal{Y})$  中, 我们引入以下假设, 这允许我们应用 OWT 的改进存在性定理 5.1 来保证  $\pi^*$  的存在性。虽然这一假设比 I. Csiszár [13] 和 F. Rezaei [16] 等人工作中的假设更强, 但在实际场景中仍然是合理且广泛适用的。

**假设 5.3**  $\mathcal{X} = \mathcal{Y}$  为波兰空间, 且信源  $\mu \in \mathcal{P}_t(\mathcal{X})$ 。失真函数  $\rho(x, y)$  满足

$$\rho(x, y) \geq c \cdot d_{\mathcal{X}}(x, y)^t \text{ 对于某个常数 } c > 0. \quad (5.19)$$

**命题 5.1** 在假设 2.1、2.2 和 5.3 下, 对于任意  $D > D_{\min}$ , 重构分布  $\nu$  具有有限的  $t$  阶矩, 即  $\nu \in \mathcal{P}_t(\mathcal{Y})$ 。

证明见第 5.4.5 小节。

这一重要命题使得我们可以将  $\nu$  的可行集从  $\mathcal{P}(\mathcal{Y})$  限制到  $\mathcal{P}_t(\mathcal{Y})$ 。

事实上, 通过应用 OWT 的存在性定理以及薛定谔问题的相关结果 [148, 167], 我们可以直接获得  $J(\nu, \beta)$  的表达式, 进而为  $R(D)$  提供新的参数化表示。

由引理 5.1, 我们可以得到  $R(D)$  的参数化表示以及 RD 函数的一系列等价形式。

**定理 5.3** 在假设 2.1–5.3 下, RD 函数  $R(D)$  可以表示为:

$$\begin{aligned} R(D) &= \inf_{\nu} \max_{\beta \in \partial R(D)} \left\{ - \int \log \left( \int e^{-\beta \rho(x, y)} d\nu \right) d\mu - \beta D + L(\nu, \beta) \right\} \\ &= \max_{\beta \in \partial R(D)} \inf_{\nu} \left\{ - \int \log \left( \int e^{-\beta \rho(x, y)} d\nu \right) d\mu - \beta D + L(\nu, \beta) \right\} \\ &= \inf_{\nu} \left\{ - \int \log \left( \int e^{-\beta \rho(x, y)} d\nu \right) d\mu - \beta D + L(\nu, \beta) \right\}, \forall \beta \in \partial R(D) \\ &\geq \inf_{\nu} \left\{ - \int \log \left( \int e^{-\beta \rho(x, y)} d\nu \right) d\mu - \beta D \right\}, \forall \beta \in \partial R(D). \end{aligned} \quad (5.20)$$

其中

$$L(\nu, \beta) = \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu}{\int_{\mathcal{Y}} g(y) e^{-\beta \rho(x, y)} d\nu} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu. \quad (5.21)$$

且  $g(y)$  满足薛定谔方程 (5.10)。

证明见第 5.4.6 小节。

如果我们进一步假设最优重构分布  $\nu^*$  存在, 则  $L(\nu, \beta)$  等于 0, 从而可以复现参数化表示 (5.23)。F. Rezaei 等人 [16] 已经深入研究了最优重构分布的存在性; 参见 [16] 中的 *Lemma 2.4* 和 *Theorem 3.1*。

**定理 5.4** 在定理 5.3 的设定下, 如果额外假设最优重构分布  $\nu^*$  的存在性, 则对应的最优联合分布满足

$$d\pi^* = \frac{e^{-\beta \rho(x, y)}}{\int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu^*} d\mu \times d\nu^* \quad (5.22)$$

这等价于引理 5.1 中的  $g(y)$  为常数，即定理 5.3 中的  $L(\nu^*, \beta) = 0$ 。此时，RD 函数  $R(D)$  可以表示为

$$\begin{aligned} R(D) &= \inf_{\nu} \left\{ - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta \rho(x,y)} d\nu \right) d\mu - \beta D \right\} \\ &= - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta \rho(x,y)} d\nu^* \right) d\mu - \beta D \end{aligned} \quad (5.23)$$

该式对任意  $\beta \in \partial R(D)$  成立，并且即使  $\beta$  在  $\partial R(D)$  中变化， $\nu^*$  保持不变。

证明见第 5.4.7 小节。

利用 (5.22) 以及 K. Rose 利用 Hermite 多项式完备性的方法，我们可以重现以下结论。

**推论 5.2** [Theorem 2 and Corollary 2 in [151]] 对于  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$  和  $\rho(x, y) = \|x - y\|^2$ ，如果最优重构随机变量的支集具有聚点，则 RD 函数与 SLB 重合。如果 SLB 不取等号，则最优重构随机变量的支集由孤立的奇点组成。进一步地，如果该支集是有界的，则  $Y$  是离散且有限的。

### 5.3 非紧空间下最优重构分布的存在性分析

在上一节中，我们利用最优弱传输理论建立了率失真函数的几何结构与参数化表示。然而，这些结论的严格性依赖于一个核心前提：最优重构分布  $\nu^*$  的存在性。在经典的率失真理论中，离散字母表情形下的存在性可由 Blahut-Arimoto 算法直接给出；而在抽象字母表情形下，I. Csiszár 在重构空间  $\mathcal{Y}$  满足紧性的假设下证明了最优重构分布的存在性。然而，在更一般的非紧波兰空间（如  $\mathcal{Y} = \mathbb{R}^n$  或一般的黎曼流形）中，传统的紧性论证不再适用。

本节旨在推进这一长期存在的分析难题。我们将引入非线性泛函分析中的集中紧性原理（Concentration-Compactness Principle），通过对率失真泛函极小化序列的精细渐近分析，在极宽泛的条件下严格证明最优重构分布的存在性。

为了克服紧性假设的限制，F. Rezaei 等人在 2006 年的工作中 [16] 取得了显著进展。他们通过引入 Stone-Čech 紧化以及局部凸拓扑向量空间中的不动点方法，成功移除了目标空间的紧性假设，从而推广了 Csiszár 的结果。

**定理 5.5** (Theorem 3.1 in [16]) 设  $\mathcal{X}, \mathcal{Y}$  为两个波兰空间，失真函数  $\rho: \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty]$  为非负广义实值可测函数。设  $\mu$  为源空间  $\mathcal{X}$  上的一个概率测度。若对于  $\mu$ -几乎所有

的  $x \in \mathcal{X}$ , 映射  $y \mapsto \rho(x, y)$  在  $\mathcal{Y}$  上是连续的。且对于任意  $D \in [0, \infty)$ , 容许集非空。则率失真问题存在最小值, 且最优重构分布存在。

尽管 Rezaei 等人的工作通过移除紧性假设推进了理论发展, 但其定理 5.5 引入了另一个较强的限制条件: 要求失真函数  $y \mapsto \rho(x, y)$  处处连续 (对于  $\mu$ -a.e. 的  $x$ )。而本文的研究表明, Rezaei 工作中这一关于  $\rho$  的连续性假设是可以被彻底移除的。相应地, 为了在非紧空间上保证变分问题的适定性, 我们引入了失真函数的强制性 (Coercivity) 条件作为替代。

具体而言, 我们仅需假设失真函数是适当的、可测的且强制 (Coercive) 的。这一框架不仅涵盖了经典的连续情形, 更能够严格处理以下带有跳变、截断乃至无穷势垒的广义失真问题, 而这些问题在 Rezaei 的连续性框架下是无法定义的:

1. **带硬约束的二次失真 (Quadratic Distortion with Hard Constraint):** 此类函数常用于安全控制领域, 要求误差严格限制在阈值  $\Delta$  内, 否则代价无穷大。

$$\rho(x, y) = \begin{cases} |x - y|^2, & \text{若 } |x - y| \leq \Delta \\ +\infty, & \text{若 } |x - y| > \Delta \end{cases} \quad (5.24)$$

显然, 该函数在边界处不连续, Rezaei 的理论无法直接适用, 而本文框架可轻松处理。

2. **阶梯型增长失真 (Staircase Growth Distortion):** 模拟通信计费或量化误差中的分级代价, 例如  $\rho(x, y) = \lceil |x - y| \rceil$ 。这类函数具有密集的间断点, 属于典型的间断函数。
3. **带惩罚跳变的鲁棒失真 (Robust Distortion with Penalty Jump):** 在异常检测中, 当误差超过阈值  $\tau$  时, 施加一个巨大的惩罚常数  $J$ :

$$\rho(x, y) = \begin{cases} |x - y|^2, & \text{若 } |x - y| \leq \tau \\ |x - y|^2 + J, & \text{若 } |x - y| > \tau \end{cases} \quad (5.25)$$

得益于集中紧性原理, 我们将证明只要失真函数满足前述宽泛条件, 最优重构分布的存在性即可得到严格保证。在给出主要存在性定理之前, 我们需要引入非线性泛函分析中处理非紧性问题的核心工具——集中紧性引理。

**定理 5.6 (集中紧性引理 I [171])** 设  $\{\mu_m\}$  是  $\mathbb{R}^n$  上的概率测度序列。存在一个子序列  $\{\mu_{m_k}\}$  (为记号简便, 仍记为原序列) 满足以下三种情况之一:

1. **紧性 (Compactness):** 存在序列  $\{x_m\} \subset \mathbb{R}^n$ , 使得  $\forall \epsilon > 0, \exists R > 0$ , 满足

$$\int_{B_R(x_m)} d\mu_m \geq 1 - \epsilon, \quad \forall m \in \mathbb{N}. \quad (5.26)$$

2. **消散 (Vanishing):** 对于每个固定的  $R > 0$ , 有

$$\lim_{m \rightarrow \infty} \left( \sup_{x \in \mathbb{R}^n} \int_{B_R(x)} d\mu_m \right) = 0. \quad (5.27)$$

3. **二分 (Dichotomy):** 存在  $\lambda \in (0, 1)$ , 使得  $\forall \epsilon > 0, \exists R > 0$  以及序列  $\{x_m\} \subset \mathbb{R}^n$  满足以下性质: 给定  $R' > R$ , 存在非负测度  $\mu_m^1, \mu_m^2$  使得

$$\begin{aligned} 0 &\leq \mu_m^1 + \mu_m^2 \leq \mu_m, \\ \text{supp}(\mu_m^1) &\subset B_R(x_m), \quad \text{supp}(\mu_m^2) \subset B_{R'}^c(x_m), \\ \limsup_{m \rightarrow \infty} \left( \left| \lambda - \int_{\mathbb{R}^n} d\mu_m^1 \right| + \left| (1 - \lambda) - \int_{\mathbb{R}^n} d\mu_m^2 \right| \right) &\leq \epsilon, \\ \text{dist}(\text{supp}(\mu_m^1), \text{supp}(\mu_m^2)) &\rightarrow \infty, \quad m \rightarrow \infty. \end{aligned} \quad (5.28)$$

事实上, 针对本文研究的抽象空间问题, 我们可以建立集中紧性引理 I 的一个广义版本。

**定理 5.7 (集中紧性引理 I 的广义版本)** 设  $\{\mu_m\}$  是无界波兰空间  $(\mathcal{X}, d)$  上的概率测度序列。存在一个子序列  $\{\mu_m\}$  (忽略符号的滥用) 满足以下三种情况之一:

1. **紧性 (Compactness):** 存在序列  $\{x_m\} \subset \mathcal{X}$ , 使得  $\forall \epsilon > 0, \exists R > 0$ , 满足

$$\int_{B_R(x_m)} d\mu_m \geq 1 - \epsilon, \quad \forall m \in \mathbb{N}. \quad (5.29)$$

2. **消散 (Vanishing):** 对于每个固定的  $R > 0$ , 有

$$\lim_{m \rightarrow \infty} \left( \sup_{x \in \mathcal{X}} \int_{B_R(x)} d\mu_m \right) = 0. \quad (5.30)$$

3. **二分 (Dichotomy):** 存在  $\lambda \in (0, 1)$ , 使得  $\forall \epsilon > 0, \exists R > 0$  以及序列  $\{x_m\} \subset \mathcal{X}$  满足以下性质: 给定  $R' > R$ , 存在非负测度  $\mu_m^1, \mu_m^2$  使得

$$\begin{aligned} 0 &\leq \mu_m^1 + \mu_m^2 \leq \mu_m, \\ \text{supp}(\mu_m^1) &\subset B_R(x_m), \quad \text{supp}(\mu_m^2) \subset B_{R'}^c(x_m), \\ \limsup_{m \rightarrow \infty} \left( \left| \lambda - \int_{\mathcal{X}} d\mu_m^1 \right| + \left| (1 - \lambda) - \int_{\mathcal{X}} d\mu_m^2 \right| \right) &\leq \epsilon, \\ \text{dist}(\text{supp}(\mu_m^1), \text{supp}(\mu_m^2)) &\rightarrow \infty, \quad m \rightarrow \infty. \end{aligned} \quad (5.31)$$

证明见第 5.4.8 小节。

此外, 我们还需要用到实分析中的经典引理:

**引理 5.4 (Helly 引理)** 设  $\{f_n\}$  是一个将实区间  $I$  映射到实直线  $\mathbb{R}$  的递增函数序列, 并且假设它是一致有界的: 即存在  $a, b \in \mathbb{R}$  使得对于每个  $n \in \mathbb{N}$  都有  $a \leq f_n \leq b$ 。则序列  $\{f_n\}$  存在一个逐点收敛的子序列。

**注** 这里我们强调, 定理 5.7 中所考虑的波兰空间  $(\mathcal{X}, d)$  是无界的。这是因为对于任意有界的波兰空间, 其空间本身即为紧空间。在这种情形下, 概率测度的紧性可以通过 Prokhorov 定理更加直接地建立, 这一点已由 Csiszár 在文献 [13] 中给出示例说明。

**注** 我们在此澄清定理 5.7 中“紧性 (compactness)”一词的使用语境。对于一般的无界波兰空间而言, 这一表述在严格意义上存在一定的记号滥用, 因为在无限维空间中, 闭球  $B_R(x_m)$  并不必然是紧集。在我们的框架下, “紧性”应被理解为测度在某个有界集合内的集中性 (concentration of measure)。关键在于, 要使这种集中性能够推出测度序列的紧性 (tightness), 底层度量空间必须满足适当性 (properness) 条件。

**定义 5.2 (适当度量空间)** 若一个度量空间中任意闭且有界的子集都是紧的, 则称该度量空间为适当的 (proper)。

该性质保证了: 若一系列测度集中于一列有界集合之上, 则该测度序列确实是紧的, 从而可以应用 Prokhorov 定理。为便于理解, 我们列举若干典型的适当波兰空间的例子:

- 配备欧氏度量的  $\mathbb{R}^n$  的任意闭子集。
- 任意紧的波兰空间, 例如  $[0, 1]^{\mathbb{N}}$ 。该空间虽为无限维, 但由于 Tychonoff 定理, 其紧性得以保证, 因此是适当的。
- 任意完备的次黎曼 (sub-Riemannian) 或次 Finsler 流形在满足第二可数性的条件下, 均为适当的波兰空间。这一结论可由 Chow–Rashevskii 定理 [172-173] 与 Ball–Box 定理 [174-175] 保证。对于完备的黎曼或 Finsler 流形, 该性质则直接来源于 Hopf–Rinow 定理 [176]。

利用上述广义集中紧性引理, 我们可以证明率失真泛函  $R(D)$  的任意极小化序列都具有紧子序列, 这是确立最优重构分布存在性的关键步骤。根据定理 5.4, 我们可以将率失真泛函重写如下:

$$\begin{aligned} R(D) &= \inf_{\nu \in \mathcal{P}(\mathcal{Y})} \left\{ - \int_{\mathcal{X}} \log \left( \int e^{-\beta^* \rho(x,y)} d\nu \right) d\mu - \beta^* D \right\} \\ &= - \sup_{\nu \in \mathcal{P}(\mathcal{Y})} \left\{ \int_{\mathcal{X}} \log \left( \int e^{-\beta^* \rho(x,y)} d\nu \right) d\mu \right\} + \beta^* D, \end{aligned} \quad (5.32)$$

其中  $\beta^*$  是与  $D$  关联的最优拉格朗日乘子。

为保证理论的适用性，我们需要对失真函数施加如下“强制性”假设，这在非紧空间分析中是标准的且必要的条件。

**假设 5.4** 设失真函数  $\rho: \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty]$  是下半连续的、适当的 (proper) 且关于  $y$  强制的 (coercive)，即满足：

1. 对于  $\mu$ -几乎所有的  $x \in \mathcal{X}$ ，映射  $y \mapsto \rho(x, y)$  在  $\mathcal{Y}$  上是下半连续的。
2. 对每个固定的  $x \in \mathcal{X}$ ，存在某个  $z \in \mathcal{Y}$ ，使得  $\rho(x, z) < \infty$ 。
3. 对任意的  $x \in \mathcal{X}$  以及某个固定的  $y' \in \mathcal{Y}$ ，有

$$\lim_{d_{\mathcal{Y}}(y', y) \rightarrow \infty} \rho(x, y) = +\infty, \quad \forall x \in \mathcal{X}. \quad (5.33)$$

其中  $(\mathcal{X}, d_{\mathcal{X}}), (\mathcal{Y}, d_{\mathcal{Y}})$  是波兰空间。

**注 (关于术语“Proper”的使用说明)** 为避免歧义，我们在此区分假设 5.4 与定义 5.2 中所使用的术语 *proper* 在数学上的两种标准含义。

- 若函数  $f: \mathcal{X} \rightarrow [0, +\infty]$  不恒等于  $+\infty$ ，则称其为**适当的 (proper) 函数**。
- 当将  $(\mathcal{Y}, d_{\mathcal{Y}})$  称为一个**适当的 (proper) 波兰空间**时，我们采用的是其拓扑学意义下的定义，即任意闭且有界的子集都是紧的。该性质已在前述评注中详细说明，并且对于建立测度序列的紧性至关重要。

需要强调的是，上述两种“适当性”条件彼此独立，并且在各自的研究领域中均为标准用法。

强制性假设确保了当  $y$  远离原点时，失真函数增长得足够快，从而防止了概率测度在优化过程中向无穷远处“逃逸”或“消散”，这是保证极小化序列紧性的物理本质。此外，此外，下半连续性 (lower semi-continuity) 条件保证了：当对一系列极小化测度序列取弱极限或相应极限点时，变分问题的下确界是可达的 (attainable)，这为问题的适定性与极小值的存在性提供了必要的分析基础。

**引理 5.5 (率失真泛函的下半连续性)** 设  $\mathcal{X}$  与  $\mathcal{Y}$  为 Polish 空间， $\mu \in \mathcal{P}(\mathcal{X})$  为给定的源分布。假设失真函数  $\rho: \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty]$  满足以下条件：

1. 对任意  $y \in \mathcal{Y}$ ，映射  $x \mapsto \rho(x, y)$  是可测的；
2. 对  $\mu$ -几乎处处的  $x \in \mathcal{X}$ ，映射  $y \mapsto \rho(x, y)$  在  $\mathcal{Y}$  上是**下半连续的**。

则对任意  $\beta \geq 0$ ，定义在  $\mathcal{P}(\mathcal{Y})$  上的对偶泛函

$$J_{\beta}(\nu) := - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu(y) \right) d\mu(x) \quad (5.34)$$

在  $\mathcal{P}(\mathcal{Y})$  的弱拓扑下是下半连续的。亦即，若  $\nu_m \xrightarrow{w} \nu$ ，则有

$$J_\beta(\nu) \leq \liminf_{m \rightarrow \infty} J_\beta(\nu_m). \quad (5.35)$$

证明见第 5.4.9 小节。

**定理 5.8 (极小化序列的紧性)** 设  $\{\nu_m\}$  为率失真泛函  $R(D)$  的一个极小化序列，即  $\nu_m \in \mathcal{P}(\mathcal{Y})$  且

$$R^{(m)}(D) := - \int_{\mathcal{X}} \log \left( \int e^{-\beta^* \rho(x,y)} d\nu_m \right) d\mu - \beta^* D, \quad (5.36)$$

其中我们假设  $\mathcal{Y}$  是具有无界度量的波兰空间，且假设 5.4 成立。则存在一个紧子序列  $\{\nu_{m_k}\}$ ，进而最优重构分布存在。

证明见第 5.4.10 小节。

**注 (与紧空间情形的关系)** 在再生空间  $\mathcal{Y}$  为紧空间的经典设定下，存在性证明将显著简化。其主要简化来源于如下事实：在紧空间上，由 Prokhorov 定理可知，任意概率测度序列都是紧的。因此，浓缩-紧性原理以及强制性条件（假设 5.4 中的第 3 项）在此情形下均变得冗余。

然而，我们仍需确保极小化子序列的弱极限确实是一个极小化解。为此，需要证明率失真泛函的下半连续性。该性质可以通过失真函数  $\rho(x, y)$  关于变量  $y$  的下半连续性来建立，从而放宽了经典处理方法中通常假设的更强的连续性条件。

## 5.4 本章定理证明

### 5.4.1 引理 5.2 的证明

对于任意固定的满足  $\mathbb{E}_{\pi} \rho \leq D$  的条件分布族（即正则分解） $\pi_x \in \mathcal{P}(\mathcal{Y})$ ，我们总能找到对应的  $\nu \in \mathcal{P}(\mathcal{Y})$  和  $\pi \in \Pi(\mu, \nu)$ ，使得对于任意 Borel 可测函数  $f: \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty]$ ，满足：

$$\int_{\mathcal{X} \times \mathcal{Y}} f(x, y) d\pi(x, y) = \int_{\mathcal{X}} d\mu(x) \int_{\mathcal{Y}} f(x, y) d\pi_x(y), \quad (5.37)$$

$$d\nu(y) = \int_{\mathcal{X}} d\pi(x, y). \quad (5.38)$$

由此可得：

$$I(X; Y)|_{\pi_x} \geq \inf_{\nu, \pi: \mathbb{E}_{\pi} \rho \leq D} I(X; Y), \quad (5.39)$$

即

$$\inf_{\nu, \pi: \mathbb{E}\pi\rho \leq D} I(X; Y) \geq \inf_{\nu, \pi: \mathbb{E}\pi\rho \leq D} I(X; Y). \quad (5.40)$$

反之, 对于所有固定的满足  $\mathbb{E}\pi\rho \leq D$  的  $\nu$  和  $\pi$ , 我们也总能确定  $\pi$  的正则分解  $\pi_x$ , 从而有:

$$I(X; Y)|_{\nu, \pi} \geq \inf_{\pi_x: \mathbb{E}\pi\rho \leq D} I(X; Y), \quad (5.41)$$

即

$$\inf_{\nu, \pi: \mathbb{E}\pi\rho \leq D} I(X; Y) \geq \inf_{\pi_x: \mathbb{E}\pi\rho \leq D} I(X; Y). \quad (5.42)$$

结合上述两方面, 我们证明了:

$$\inf_{\nu, \pi: \mathbb{E}\pi\rho \leq D} I(X; Y) = \inf_{\pi_x: \mathbb{E}\pi\rho \leq D} I(X; Y). \quad (5.43)$$

进一步地, 设  $\{\nu_k, \pi_k\}$  为  $\inf_{\nu, \pi: \mathbb{E}\pi\rho \leq D} I(X; Y)$  的极小化序列, 则有:

$$\begin{aligned} \inf_{\nu} \inf_{\pi \in \Pi(\mu, \nu): \mathbb{E}\pi\rho \leq D} I(X; Y) &\leq \inf_{\nu} \inf_{\pi \in \Pi(\mu, \nu_k): \mathbb{E}\pi\rho \leq D} I(X; Y) \\ &= \inf_{\pi \in \Pi(\mu, \nu_k): \mathbb{E}\pi\rho \leq D} I(X; Y) \\ &\leq I(X; Y)|_{(\nu_k, \pi_k)}. \end{aligned} \quad (5.44)$$

因此, 取极限可得:

$$\inf_{\nu} \inf_{\pi \in \Pi(\mu, \nu): \mathbb{E}\pi\rho \leq D} I(X; Y) \leq \lim_{k \rightarrow \infty} I(X; Y)|_{(\nu_k, \pi_k)} = \inf_{\nu, \pi: \mathbb{E}\pi\rho \leq D} I(X; Y). \quad (5.45)$$

另一方面, 由于对于任意固定的  $\tilde{\nu} \in \mathcal{P}(\mathcal{Y})$ , 都有

$$\inf_{\nu, \pi: \mathbb{E}\pi\rho \leq D} I(X; Y) \leq \inf_{\pi \in \Pi(\mu, \tilde{\nu}): \mathbb{E}\pi\rho \leq D} I(X; Y), \quad (5.46)$$

我们在两边对  $\tilde{\nu}$  取下确界, 即可得:

$$\inf_{\nu, \pi: \mathbb{E}\pi\rho \leq D} I(X; Y) \leq \inf_{\tilde{\nu}} \inf_{\pi \in \Pi(\mu, \tilde{\nu}): \mathbb{E}\pi\rho \leq D} I(X; Y) = \inf_{\nu} \inf_{\pi \in \Pi(\mu, \nu): \mathbb{E}\pi\rho \leq D} I(X; Y). \quad (5.47)$$

综上所述, 等式成立。

## 5.4.2 引理 5.3 的证明

**步骤 1: 联合下半连续性。** 考虑在乘积拓扑意义下的收敛序列  $(x_k, p_k) \rightarrow (x, p)$ , 即当  $k \rightarrow \infty$  时, 满足  $x_k \rightarrow x \in \mathcal{X}$  且  $p_k \rightarrow p \in \mathcal{P}(\mathcal{Y})$ 。我们需要验证目标函数满足下半连续性条件, 即  $C(x, p) \leq \liminf_{k \rightarrow \infty} C(x_k, p_k)$ 。

一方面, 依据文献 [177] 中的 *Theorem 1* 或文献 [178] 中的 *Theorem 4.9*, 我们有:

$$\int_{\mathcal{Y}} \log \frac{dp}{d\nu} dp \leq \liminf_{k \rightarrow \infty} \int_{\mathcal{Y}} \log \frac{dp_k}{d\nu} dp_k. \quad (5.48)$$

另一方面, 由假设 5.1 可知, 项  $\beta \int_{\mathcal{Y}} (\rho(x, y) - D) dp$  关于  $(x, p)$  是下半连续的。综上所述, 在此情形下  $C(x, p)$  是联合下半连续的。

**步骤 2: 下有界性。** 根据定义, 我们有:

$$\begin{aligned} C(x, p) &:= \int_{\mathcal{Y}} \left[ \log \frac{dp}{d\nu} + \beta(\rho(x, y) - D) \right] dp \\ &\geq \int_{\mathcal{Y}} \log \frac{dp}{d\nu} dp - \beta D \\ &\geq -\beta D. \end{aligned} \quad (5.49)$$

注意到相对熵恒为非负, 这是因为:

$$\begin{aligned} \int_{\mathcal{Y}} \log \frac{dp}{d\nu} dp &= \int_{\mathcal{Y}} \varphi \left( \frac{dp}{d\nu} \right) d\nu \\ &\geq \varphi \left( \int_{\mathcal{Y}} \frac{dp}{d\nu} d\nu \right) \\ &= \varphi(1) = 0. \end{aligned} \quad (5.50)$$

其中,  $\varphi(x) = x \log x$  是定义在  $(0, +\infty)$  上的凸函数。

**步骤 3: 关于第二个变量的凸性。**

对于任意  $p_1, p_2 \in \mathcal{P}(\mathcal{Y})$  以及  $\lambda \in [0, 1]$ , 我们展开如下计算:

$$\begin{aligned} &C(x, \lambda p_1 + (1 - \lambda) p_2) - \lambda C(x, p_1) - (1 - \lambda) C(x, p_2) \\ &= \int_{\mathcal{Y}} \left[ \log \frac{d(\lambda p_1 + (1 - \lambda) p_2)}{d\nu} + \beta \rho(x, y) \right] d(\lambda p_1 + (1 - \lambda) p_2) \\ &\quad - \lambda \int_{\mathcal{Y}} \left[ \log \frac{dp_1}{d\nu} + \beta \rho(x, y) \right] dp_1 - (1 - \lambda) \int_{\mathcal{Y}} \left[ \log \frac{dp_2}{d\nu} + \beta \rho(x, y) \right] dp_2 \\ &= \int_{\mathcal{Y}} \log \frac{\lambda dp_1 + (1 - \lambda) dp_2}{d\nu} (\lambda dp_1 + (1 - \lambda) dp_2) \\ &\quad - \lambda \int_{\mathcal{Y}} \log \frac{dp_1}{d\nu} dp_1 - (1 - \lambda) \int_{\mathcal{Y}} \log \frac{dp_2}{d\nu} dp_2 \\ &= \int_{\mathcal{Y}} \varphi \left( \lambda \frac{dp_1}{d\nu} + (1 - \lambda) \frac{dp_2}{d\nu} \right) d\nu \\ &\quad - \lambda \int_{\mathcal{Y}} \varphi \left( \frac{dp_1}{d\nu} \right) d\nu - (1 - \lambda) \int_{\mathcal{Y}} \varphi \left( \frac{dp_2}{d\nu} \right) d\nu \\ &\leq 0. \end{aligned} \quad (5.51)$$

### 5.4.3 假设 5.2 不成立的例子

设  $\mathcal{X} = \mathcal{Y} = \mathbb{R}$  且赋予标准拓扑, 则它们均为波兰空间 (Polish spaces)。定义  $\rho(x, y) = \mathbf{1}_{\{y=x\}}$  为指示函数, 即当  $y = x$  时取值为 1, 否则为 0。此时, 我们要考察如下映射:

$$(x, p) \in \mathcal{X} \times \mathcal{P}_t(\mathcal{Y}) \mapsto \int_{\mathcal{Y}} \rho(x, y) dp(y) = p(\{x\}). \quad (5.52)$$

具体而言, 该映射描述了概率测度  $p$  在点  $x$  处分配的质量。现考察一个概率测度序列  $p_n = \delta_{y_n + \frac{1}{n}}$ , 其中  $y_n \rightarrow y$ 。

**步骤 1:  $\mathcal{P}_t(\mathcal{Y})$  中的弱收敛性。**

根据  $\mathcal{P}_t(\mathcal{Y})$  中弱收敛的定义 (参见文献 [17] 中的 *Definition 6.8*), 对于任意连续函数  $\varphi$ , 若其满足对于某  $y_0 \in \mathcal{Y}$  及  $C \in \mathbb{R}$  有  $|\varphi(y)| \leq C(1 + d(y_0, y)^p)$ , 则由连续性可知:

$$\int \varphi dp_n = \varphi\left(y_n + \frac{1}{n}\right) \rightarrow \varphi(y) = \int \varphi d\delta_y. \quad (5.53)$$

由此证明了在  $\mathcal{P}_t(\mathcal{Y})$  意义下  $p_n \rightharpoonup \delta_y$ 。

**步骤 2: 非下半连续性。**

定义目标函数为:

$$f(x, p) := p(\{x\}). \quad (5.54)$$

此时我们有  $f(y_n, p_n) = p_n(\{y_n\}) = 0$ , 而  $f(y, \delta_y) = 1$ 。因此可得:

$$\liminf_{n \rightarrow \infty} f(y_n, p_n) = 0 < 1 = f(y, \delta_y). \quad (5.55)$$

这与下半连续性的定义相矛盾。

### 5.4.4 定理 5.2 的证明

关于  $C(x, p)$  在  $\mathcal{P}_t(\mathcal{Y})$  上的下有界性及其关于第二个变量的凸性, 其证明方法与引理 5.3 一致, 此处不再赘述。因此, 结合变分法中的基本存在性定理, 只需验证  $C(x, p)$  在  $\mathcal{X} \times \mathcal{P}_t(\mathcal{Y})$  上的联合下半连续性, 即可保证极小值点的存在。

一方面, 依据文献 [177] 中的 *Theorem 1* 或文献 [178] 中的 *Theorem 4.9*, 我们有:

$$\int_{\mathcal{Y}} \log \frac{dp}{d\nu} dp \leq \liminf_{k \rightarrow \infty} \int_{\mathcal{Y}} \log \frac{dp_k}{d\nu} dp_k. \quad (5.56)$$

另一方面, 由假设 5.2 可知, 项  $\beta \int_{\mathcal{Y}} (\rho(x, y) - D) dp$  关于  $(x, p) \in \mathcal{X} \times \mathcal{P}_t(\mathcal{Y})$  是下半连续的。综上所述, 在此情形下  $C(x, p)$  满足联合下半连续性。

进一步地, 根据定理 5.1 可知,  $J(\nu, \beta)$  的极小值点  $\pi^*$  存在。

### 5.4.5 命题 5.1 的证明

基于 Wasserstein 距离的性质（即三角不等式），我们可以得到如下关系：

$$\begin{aligned}
 W_t(\nu, \delta_{y_0}) &= \inf_{\pi \in \Pi(\nu, \delta_{y_0})} \left( \int_{\mathcal{X} \times \mathcal{Y}} d_{\mathcal{X}}(x, y)^t d\pi \right)^{1/t} \\
 &= \left( \int_{\mathcal{X}} d_{\mathcal{X}}(x, y_0)^t d\nu \right)^{1/t} \\
 &\leq W_t(\mu, \delta_{y_0}) + W_t(\mu, \nu) \\
 &= \left( \int_{\mathcal{X}} d_{\mathcal{X}}(x, y_0)^t d\mu \right)^{1/t} + W_t(\mu, \nu). \tag{5.57}
 \end{aligned}$$

同时，注意到如下不等式成立：

$$\begin{aligned}
 W_t(\mu, \nu) &= \inf_{\pi \in \Pi(\mu, \nu)} \left[ \mathbb{E}_{\pi}(d_{\mathcal{X}}^t) \right]^{1/t} \\
 &\leq c^{-1/t} \inf_{\pi \in \Pi(\mu, \nu)} [\mathbb{E}_{\pi} \rho]^{1/t} \\
 &\leq c^{-1/t} D^{1/t}. \tag{5.58}
 \end{aligned}$$

此外，根据空间  $\mathcal{P}_t(\mathcal{X})$  的定义，已知：

$$\int_{\mathcal{X}} d_{\mathcal{X}}(x, y_0)^t d\mu < +\infty. \tag{5.59}$$

由此推导可得：

$$W_t^t(\nu, \delta_{y_0}) = \int_{\mathcal{X}} d_{\mathcal{X}}(x, y_0)^t d\nu \leq \left( c^{-1/t} D^{1/t} + W_t(\mu, \delta_{y_0}) \right)^t < +\infty. \tag{5.60}$$

### 5.4.6 定理 5.3 的证明

根据引理 5.1 可知，极小值点  $\pi_x^*$  与  $\pi^*$  存在，且存在非负可测函数  $f, g$ ，使得：

$$\frac{d\pi^*}{d\gamma} = f(x)g(y), \quad \gamma\text{-a.e.} \tag{5.61}$$

具体而言，可表示为：

$$d\pi^* = K f(x)g(y) e^{-\beta\rho(x,y)} d\mu(x)d\nu(y). \tag{5.62}$$

此时可推导得：

$$\begin{aligned}
 d\pi_x^* &= K f(x)g(y) e^{-\beta\rho(x,y)} d\nu = \frac{K}{K_x} f(x)g(y) d\gamma_x, \\
 \frac{d\pi_x^*}{d\gamma_x} &= \frac{K f(x)g(y)}{K_x}. \tag{5.63}
 \end{aligned}$$

由于  $\pi^* \in \Pi(\mu, \nu)$ , 满足边缘分布条件  $\int_{\mathcal{Y}} d\pi^* = d\mu$  以及  $\int_{\mathcal{X}} d\pi^* = d\nu$ 。这一条件可被写为 Schrödinger 方程组 [167]:

$$\begin{aligned} K &= \left( \int_{\mathcal{X} \times \mathcal{Y}} e^{-\beta\rho(x,y)} d\mu d\nu \right)^{-1} \\ &= \left( \int_{\mathcal{X} \times \mathcal{Y}} f(x)g(y)e^{-\beta\rho(x,y)} d\mu d\nu \right)^{-1}, \end{aligned} \quad (5.64)$$

$$\begin{aligned} 1 &= K f(x) \int_{\mathcal{Y}} g(y)e^{-\beta\rho(x,y)} d\nu \\ &= K g(y) \int_{\mathcal{X}} f(x)e^{-\beta\rho(x,y)} d\mu, \\ \Rightarrow \int_{\mathcal{X}} \frac{g(y')e^{-\beta\rho(x,y')}}{\int_{\mathcal{Y}} g(y)e^{-\beta\rho(x,y)} d\nu} d\mu &= 1, \quad \forall y' \in \mathcal{Y}. \end{aligned} \quad (5.65)$$

进而, 我们将目标泛函  $J(\nu, \beta)$  展开如下:

$$\begin{aligned} J(\nu, \beta) &= \int_{\mathcal{X} \times \mathcal{Y}} \log \frac{d\pi^*}{d\gamma} d\pi^* + \log K - \beta D \\ &= \int_{\mathcal{X} \times \mathcal{Y}} \log(f(x)g(y)) d\pi^* + \log K - \beta D \\ &= \int_{\mathcal{X}} \log(K f(x)) d\mu + \int_{\mathcal{Y}} \log(g(y)) d\nu - \beta D \\ &= - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} g(y)e^{-\beta\rho(x,y)} d\nu \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu - \beta D \\ &= - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta\rho(x,y)} d\nu \right) d\mu - \beta D + \int_{\mathcal{Y}} \log g(y) d\nu \\ &\quad + \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta\rho(x,y)} d\nu}{\int_{\mathcal{Y}} g(y)e^{-\beta\rho(x,y)} d\nu} \right) d\mu. \end{aligned} \quad (5.66)$$

对于率失真函数  $R(D)$ , 有如下推导:

$$\begin{aligned} R(D) &= \inf_{\nu} \inf_{\pi \in \Pi(\mu, \nu): \mathbb{E}\pi\rho \leq D} I(X; Y) \\ &\geq \inf_{\nu} \sup_{\beta \geq 0} \inf_{\pi \in \Pi(\mu, \nu)} I(X; Y) + \beta(\mathbb{E}\pi\rho - D) \\ &= \inf_{\nu} \sup_{\beta \geq 0} \left\{ - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta\rho(x,y)} d\nu \right) d\mu - \beta D \right. \\ &\quad \left. + \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta\rho(x,y)} d\nu}{\int_{\mathcal{Y}} g(y)e^{-\beta\rho(x,y)} d\nu} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu \right\} \\ &\geq \sup_{\beta \geq 0} \inf_{\nu} \left\{ - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta\rho(x,y)} d\nu \right) d\mu - \beta D \right. \\ &\quad \left. + \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta\rho(x,y)} d\nu}{\int_{\mathcal{Y}} g(y)e^{-\beta\rho(x,y)} d\nu} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu \right\}. \end{aligned} \quad (5.67)$$

现定义辅助函数  $L(\beta)$  如下：

$$L(\beta) := \inf_{\nu} \left\{ - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta \rho(x,y)} d\nu \right) d\mu \right. \\ \left. + \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta \rho(x,y)} d\nu}{\int_{\mathcal{Y}} g(y) e^{-\beta \rho(x,y)} d\nu} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu \right\}.$$

则对于任意  $\beta \geq 0$ ，成立：

$$R(D) \geq L(\beta) - \beta D, \\ R(D) + \beta D \geq L(\beta). \quad (5.68)$$

众所周知， $R(D)$  是关于  $D$  的非增凸函数，因此其二阶导数几乎处处存在。此外，对于任意  $D \geq D_{\min}$ ， $R(D)$  的次梯度存在，记为  $\partial R(D)$ 。对于凸函数  $R(D)$  而言，该次梯度集合总是闭的、凸的且有界的。于是，对于任意  $-\beta_D \in \partial R(D)$ ，我们有：

$$R(D') \geq R(D) - \beta_D(D' - D), \quad \forall D' \in (D_{\min}, D_{\max}) \quad (5.69)$$

即

$$R(D') + \beta_D D' \geq R(D) + \beta_D D \geq L(\beta_D). \quad (5.70)$$

此外，根据  $L(\beta)$  的定义，对于任意足够小的  $\epsilon > 0$ ，存在对应的  $\nu^\epsilon \in \mathcal{P}(\mathcal{Y})$  使得：

$$- \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta_D \rho(x,y)} d\nu^\epsilon \right) d\mu + \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta_D \rho(x,y)} d\nu^\epsilon}{\int_{\mathcal{Y}} g(y) e^{-\beta_D \rho(x,y)} d\nu^\epsilon} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu^\epsilon < L(\beta_D) + \epsilon.$$

另一方面，可得：

$$- \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta_D \rho(x,y)} d\nu^\epsilon \right) d\mu + \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta_D \rho(x,y)} d\nu^\epsilon}{\int_{\mathcal{Y}} g(y) e^{-\beta_D \rho(x,y)} d\nu^\epsilon} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu^\epsilon \\ = (I(X; Y) + \beta_D \mathbb{E}_{\pi} \rho) |_{\pi_x = \pi_x^*(\nu^\epsilon)} \\ \geq R(\mathbb{E}_{\pi^*(\nu^\epsilon)} \rho) + \beta_D \mathbb{E}_{\pi^*(\nu^\epsilon)} \rho \\ \geq R(D) + \beta_D D. \quad (5.71)$$

综合上述式子，对于任意  $\epsilon > 0$ ，有：

$$R(D) + \beta_D D - \epsilon < L(\beta_D) \leq R(D) + \beta_D D. \quad (5.72)$$

进一步地，我们有：

$$L(\beta_D) = R(D) + \beta_D D, \quad (5.73)$$

$$R(D) \geq \sup_{\beta \geq 0} (L(\beta) - \beta D) \geq \sup_{\beta_D \in \partial R(D)} (L(\beta_D) - \beta_D D) = R(D). \quad (5.74)$$

综上所述, 我们证明了如下等式:

$$\begin{aligned}
R(D) &= \inf_{\nu} \inf_{\pi \in \Pi(\mu, \nu): \mathbb{E}_{\pi} \rho \leq D} I(X; Y) \\
&= \inf_{\nu, \pi \in \Pi(\mu, \nu): \mathbb{E}_{\pi} \rho \leq D} I(X; Y) \\
&= \inf_{\nu} \sup_{\beta \geq 0} \inf_{\pi \in \Pi(\mu, \nu)} I(X; Y) + \beta(\mathbb{E}_{\pi} \rho - D) \\
&= \inf_{\nu} \sup_{\beta \geq 0} \left\{ - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu \right) d\mu - \beta D \right. \\
&\quad \left. + \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu}{\int_{\mathcal{Y}} g(y) e^{-\beta \rho(x, y)} d\nu} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu \right\} \\
&= \sup_{\beta \geq 0} \inf_{\nu} \left\{ - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu \right) d\mu - \beta D \right. \\
&\quad \left. + \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu}{\int_{\mathcal{Y}} g(y) e^{-\beta \rho(x, y)} d\nu} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu \right\} \\
&= \inf_{\nu} \max_{\beta \in \partial R(D)} \left\{ - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu \right) d\mu - \beta D \right. \\
&\quad \left. + \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu}{\int_{\mathcal{Y}} g(y) e^{-\beta \rho(x, y)} d\nu} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu \right\} \\
&= \max_{\beta \in \partial R(D)} \inf_{\nu} \left\{ - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu \right) d\mu - \beta D \right. \\
&\quad \left. + \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu}{\int_{\mathcal{Y}} g(y) e^{-\beta \rho(x, y)} d\nu} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu \right\} \\
&= \inf_{\nu} \left\{ - \int_{\mathcal{X}} \log \left( \int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu \right) d\mu - \beta D + L(\nu, \beta) \right\}, \quad \forall \beta \in \partial R(D). \tag{5.75}
\end{aligned}$$

其中,

$$L(\nu, \beta) = \int_{\mathcal{X}} \log \left( \frac{\int_{\mathcal{Y}} e^{-\beta \rho(x, y)} d\nu}{\int_{\mathcal{Y}} g(y) e^{-\beta \rho(x, y)} d\nu} \right) d\mu + \int_{\mathcal{Y}} \log g(y) d\nu. \tag{5.76}$$

设测度  $\mathbb{Q}_{\nu}$  为:

$$d\mathbb{Q}_{\nu}(x, y) := \frac{e^{-\beta \rho(x, y)}}{\int e^{-\beta \rho(x, z)} d\nu(z)} d\mu(x) d\nu(y), \tag{5.77}$$

并定义配分函数:

$$Z_{\nu}(x) := \int e^{-\beta \rho(x, z)} d\nu(z). \tag{5.78}$$

注意到  $L(\nu, \beta)$  实际上是  $\mathbb{Q}_\nu$  相对于  $\pi^*$  的相对熵, 即:

$$\begin{aligned}
D_{KL}(\pi^* \parallel \mathbb{Q}_\nu) &= \int_{\mathcal{X} \times \mathcal{Y}} \log \frac{d\pi^*}{d\mathbb{Q}_\nu} d\pi^* \\
&= \int_{\mathcal{X} \times \mathcal{Y}} \log(Kf(x)g(y)Z_\nu(x)) d\pi^*(x, y) \\
&= \int_{\mathcal{X}} \log Z_\nu(x) d\mu(x) + \int_{\mathcal{X} \times \mathcal{Y}} \log(Kf(x)g(y)) d\pi^*(x, y) \\
&= \int_{\mathcal{X}} \log Z_\nu(x) d\mu(x) + \int_{\mathcal{Y}} \log g(y) d\nu(y) + \int_{\mathcal{X}} \log(Kf(x)) d\mu(x) \\
&= \int_{\mathcal{X}} \log \left( \frac{Z_\nu(x)}{\int_{\mathcal{Y}} g(y) e^{-\beta\rho(x,y)} d\nu(y)} \right) d\mu(x) + \int_{\mathcal{Y}} \log g(y) d\nu(y) \\
&= L(\nu, \beta) \geq 0.
\end{aligned} \tag{5.79}$$

其中我们用到了如下事实:

$$Kf(x) \int_{\mathcal{Y}} g(y) e^{-\beta\rho(x,y)} d\nu(y) = 1, \mu - a.s. \tag{5.80}$$

由此我们完成了定理 5.3 的证明。

#### 5.4.7 定理 5.4 的证明

由前文分析可知, 对于任意  $D > D_{\min}$  以及任意  $\beta \in \partial R(D)$ , 成立:

$$R(D) = \inf_{\nu \in \mathcal{P}_t(\mathcal{Y})} \inf_{\pi \in \Pi(\mu, \nu)} \left\{ \int_{\mathcal{X} \times \mathcal{Y}} \log \frac{d\pi}{d\mu \times d\nu} d\pi + \beta \int_{\mathcal{X} \times \mathcal{Y}} \rho(x, y) d\pi \right\} - \beta D. \tag{5.81}$$

现定义如下辅助泛函与函数:

$$\begin{aligned}
\tilde{J}(\pi, \nu, \beta) &:= D_{KL}(\pi \parallel \mu \times \nu) + \beta \mathbb{E}_\pi(\rho) \\
&= \int_{\mathcal{X} \times \mathcal{Y}} \log \frac{d\pi}{d\mu \times d\nu} d\pi + \beta \int_{\mathcal{X} \times \mathcal{Y}} \rho(x, y) d\pi,
\end{aligned} \tag{5.82}$$

$$f_{\nu, \beta}(x) = \left( \int_{\mathcal{Y}} e^{-\beta\rho(x,y)} d\nu \right)^{-1}, \tag{5.83}$$

$$d\pi_{\nu, \beta}(x, y) = f_{\nu, \beta}(x) e^{-\beta\rho(x,y)} d\mu(x) \times d\nu(y). \tag{5.84}$$

进而可得  $R(D)$  的另一种表示形式:

$$\begin{aligned}
R(D) &= \inf_{\nu} \inf_{\pi \in \Pi(\mu, \nu)} \tilde{J}(\pi, \nu, \beta) - \beta D \\
&= \inf_{\pi \in \Pi(\mu, \cdot)} \tilde{J}(\pi, \pi_1, \beta) - \beta D,
\end{aligned} \tag{5.85}$$

其中  $\pi_1$  表示  $\pi$  的  $Y$ -边缘分布。

注意到，正如定理 5.3 所述， $R(D)$  的值对于任意  $\beta \in \partial R(D)$  的选取保持恒定。换言之， $R(D)$  独立于  $\beta$  在  $\partial R(D)$  中的具体取值。

假设最优重构分布  $\nu^*$  存在，则有：

$$R(D) = \inf_{\pi \in \Pi(\mu, \nu^*)} \tilde{J}(\pi, \nu^*, \beta) - \beta D \stackrel{OWT}{=} \tilde{J}(\pi^*, \nu^*, \beta) - \beta D. \quad (5.86)$$

另一方面，回顾定义：

$$R(D) = \inf_{\pi \in \Pi(\mu, \cdot)} \tilde{J}(\pi, \pi_1, \beta) - \beta D, \quad \text{其中 } \pi_1 \text{ 为 } \pi \text{ 的 } Y\text{-边缘分布。} \quad (5.87)$$

若下确界  $\inf_{\pi \in \Pi(\mu, \cdot)} \tilde{J}(\pi, \pi_1, \beta)$  无法取到，则存在某分布  $\tilde{\pi}$ ，使得：

$$\tilde{J}(\tilde{\pi}, \tilde{\pi}_1, \beta) < \tilde{J}(\pi^*, \nu^*, \beta). \quad (5.88)$$

此时我们将得到如下不等式链：

$$R(D) + \beta D = \tilde{J}(\pi^*, \nu^*, \beta) > \tilde{J}(\tilde{\pi}, \tilde{\pi}_1, \beta) \geq \inf_{\pi \in \Pi(\mu, \cdot)} \tilde{J}(\pi, \pi_1, \beta) = R(D) + \beta D. \quad (5.89)$$

这显然导出了矛盾。因此，若最优重构分布  $\nu^*$  存在，则下确界  $\inf_{\pi \in \Pi(\mu, \cdot)} \tilde{J}(\pi, \pi_1, \beta)$  必在  $\pi^*, \nu^*$  处达到。

进一步地，依据文献 [13] 中的 *Lemma 1.4*，最优耦合  $\pi^*$  具有如下显式形式：

$$d\pi^* = \alpha(x) e^{-\beta\rho(x,y)} d\mu \times d\nu^* = \frac{e^{-\beta\rho(x,y)}}{\int_Y e^{-\beta\rho(x,y)} d\nu^*} d\mu \times d\nu^*. \quad (5.90)$$

#### 5.4.8 定理 5.7 的证明

我们首先引入测度  $\mu$  的集中函数的概念：

$$Q(r) := \sup_{x \in \mathcal{X}} \int_{B_r(x)} d\mu. \quad (5.91)$$

设  $Q_m$  为与测度  $\mu_m$  关联的集中函数。注意  $\{Q_m\}$  是  $\mathbb{R}_+$  上的一列非减、非负有界函数，且  $\lim_{R \rightarrow \infty} Q_m(R) = 1$ ，因为  $\mu_m$  是概率测度。由 Helly 引理，存在一个子序列  $\{Q_m\}$ （忽略符号的滥用），它对于几乎每个  $R > 0$  逐点收敛到一个非减、非负有界函数  $Q$ 。注意任何单调函数都具有左右极限，且至多有可数个间断点，因此我们将  $Q$  规范化为左连续的，即

$$Q(R) = \lim_{\epsilon \rightarrow 0^+} Q(R - \epsilon). \quad (5.92)$$

对于任意固定的  $R > 0$ , 我们考虑一个序列  $\{R_k\}$  使得  $R_k \nearrow R$ , 且满足:

$$\lim_{m \rightarrow \infty} Q_m(R_k) = Q(R_k). \quad (5.93)$$

由于  $Q_m$  是非减的, 我们有  $Q_m(R_k) \leq Q_m(R)$ , 对于每个  $k, m \in \mathbb{N}$ . 于是有

$$\liminf_{m \rightarrow \infty} Q_m(R_k) = \lim_{m \rightarrow \infty} Q_m(R_k) = Q(R_k) \leq \liminf_{m \rightarrow \infty} Q_m(R). \quad (5.94)$$

利用左连续性, 取极限  $R_k \rightarrow R$  得到

$$\lim_{k \rightarrow \infty} Q(R_k) = Q(R) \leq \liminf_{m \rightarrow \infty} Q_m(R). \quad (5.95)$$

令  $\lambda := \lim_{R \rightarrow \infty} Q(R)$ . 显然  $\lambda \in [0, 1]$ . 如果  $\lambda = 0$ , 则我们得到 **消散**情况:

$$\lim_{m \rightarrow \infty} \left( \sup_{x \in \mathcal{X}} \int_{B_R(x)} d\mu_m \right) = 0, \quad \forall R > 0. \quad (5.96)$$

如果  $\lambda = 1$ , 则对于某个  $R_0 > 0$ , 我们有  $Q(R_0) > \frac{1}{2}$ . 对于任意  $m \in \mathbb{N}$ , 设  $x_m \in \mathcal{X}$  满足:

$$Q_m(R_0) \leq \int_{B_{R_0}(x_m)} d\mu_m + \frac{1}{m}, \quad \lim_{m \rightarrow \infty} Q_m(R_0) = Q(R_0). \quad (5.97)$$

现在对于  $0 < \epsilon < \frac{1}{2}$ , 我们固定  $R$  使得  $Q(R) > 1 - \epsilon > \frac{1}{2}$ , 并设  $y_m \in \mathcal{X}$  满足:

$$Q_m(R) \leq \int_{B_{R_0}(y_m)} d\mu_m + \frac{1}{m}, \quad \lim_{m \rightarrow \infty} Q_m(R) = Q(R), \quad (5.98)$$

其中  $y_m$  的选择依赖于  $\epsilon$ . 那么我们有

$$\begin{aligned} \int_{B_{R_0}(x_m)} d\mu_m + \int_{B_{R_0}(y_m)} d\mu_m &\geq Q_m(R_0) + Q_m(R) - \frac{2}{m} \\ &\rightarrow Q(R_0) + Q(R) \\ &> \frac{1}{2} + \frac{1}{2} = 1 = \int_{\mathcal{X}} d\mu_m, \end{aligned} \quad (5.99)$$

对于足够大的  $m$  成立. 显然,  $B_{R_0}(x_m) \cap B_{R_0}(y_m) \neq \emptyset$ , 否则我们将有

$$\begin{aligned} \int_{B_{R_0}(x_m)} d\mu_m + \int_{B_{R_0}(y_m)} d\mu_m &= \int_{B_{R_0}(x_m) \cup B_{R_0}(y_m)} d\mu_m \\ &\leq \int_{\mathcal{X}} d\mu_m = 1, \end{aligned} \quad (5.100)$$

这产生了矛盾。因此，我们有

$$\text{dist}\{x_m, y_m\} \leq R_0 + R, \quad B_R(y_m) \subset B_{2R+R_0}(x_m), \quad (5.101)$$

$$\int_{B_{2R+R_0}(x_m)} d\mu_m \geq \int_{B_R(y_m)} d\mu_m \geq Q_m(R) - \frac{1}{m} \geq 1 - \epsilon, \quad (5.102)$$

对于足够大的  $m \in \mathbb{N}$  成立，即存在  $M \in \mathbb{N}$  使得对于所有  $m \geq M$ ，上述性质成立。此外，存在  $\tilde{R} > 0$  使得

$$\int_{B_{\tilde{R}}(x_m)} d\mu_m \geq 1 - \epsilon, \quad \forall 1 \leq m \leq M. \quad (5.103)$$

于是我们可以取  $R' = \max\{\tilde{R}, 2R + R_0\}$  从而得到 **紧性** 情况。

如果  $0 < \lambda < 1$ ，给定  $\epsilon > 0$ ，存在  $R(\epsilon) > 0$ ，使得  $Q(R) > \lambda - \frac{\epsilon}{3}$ ，且  $\lim_{m \rightarrow \infty} Q_m(R) = Q(R)$ 。那么也存在依赖于  $\epsilon$  的  $M \in \mathbb{N}$  使得

$$Q_m(R) > Q(R) - \frac{\epsilon}{3} > \lambda - \frac{2\epsilon}{3}, \quad \forall m \geq M(\epsilon). \quad (5.104)$$

随后我们选择一个依赖于  $\epsilon$  的序列  $\{x_m\}$ ，使得

$$Q_m(R) \geq \int_{B_R(x_m)} d\mu_m \geq Q_m(R) - \frac{\epsilon}{3} > \lambda - \epsilon. \quad (5.105)$$

类似地，给定  $\epsilon > 0$ ，存在  $R'(\epsilon) > 0$ ， $M'(\epsilon) \geq M(\epsilon)$ ，使得

$$\lambda - \epsilon < Q_m(R') < \lambda + \epsilon, \quad \forall m \geq M'. \quad (5.106)$$

此外，由于  $\lim_{r \rightarrow \infty} Q(r) = \lambda$ ，对于任意  $\epsilon > 0$ ，存在  $R_0 < \infty$ ，使得对于每个  $r \geq R_0$ ，

$$Q(r) < \lambda + \frac{\epsilon}{2}. \quad (5.107)$$

我们取一系列严格递增趋于无穷大的半径  $\{R_k\}_{k \geq 1}$ ，且  $R_1 \geq R_0$ 。那么对于每个  $k \geq 1$ ，我们有  $Q(R_k) < \lambda + \frac{\epsilon}{2}$ 。这里我们也要求  $\lim_{m \rightarrow \infty} Q_m(R_k) = Q(R_k)$  对于每个  $k \in \mathbb{N}$  成立。

对于  $k = 1$ ，根据极限的定义，存在指标  $M_1 \in \mathbb{N}$  且  $M_1 > M'$ ，使得对于所有  $m \geq M_1$ ，有  $|Q_m(R_1) - Q(R_1)| < \frac{\epsilon}{2}$ 。对于  $k = 2$ ，类似地，存在指标  $M'_2 \in \mathbb{N}$  使得对于所有  $m \geq M'_2$ ，有  $|Q_m(R_2) - Q(R_2)| < \frac{\epsilon}{2}$ 。为了确保我们构造的指标序列是严格递增的，我们定义  $M_2 = \max\{M_1 + 1, M'_2\}$ 。以此类推，我们可以递归地定义一个严格递增的指标序列  $\{M_k\}_{k \geq 1}$ 。关键在于，该序列具有如下性质：对于每个  $k \geq 1$ ，只要  $m \geq M_k$ ，我们就有：

$$Q_m(R_k) < Q(R_k) + \frac{\epsilon}{2} < \lambda + \epsilon. \quad (5.108)$$

现在我们利用序列  $\{R_k\}$  和  $\{M_k\}$  来为每个指标  $m$  定义  $\tilde{R}_m$ 。新序列  $\{\tilde{R}_m\}$  定义如下：

- 对于  $M' \leq m < M_1$ ,  $\tilde{R}_m$  的值设为  $R_0$ 。
- 对于  $m \geq M_1$ , 我们将整数集  $[M_1, \infty)$  划分为不相交区间  $[M_1, M_2 - 1], [M_2, M_3 - 1]$  等等。具体来说, 对于每个  $m \geq 1$ , 如果  $M_k \leq m < M_{k+1}$ , 我们定义:

$$\tilde{R}_m = R_k. \quad (5.109)$$

综上所述, 给定  $\epsilon > 0$ , 我们可以找到  $R > 0, m_0 \in \mathbb{N}$ , 递增趋于无穷大的序列  $\{R_m\}$ , 以及  $\{x_m\} \subset \mathcal{X}$  (均依赖于  $\epsilon$ ), 使得对于每个  $m > m_0(\epsilon)$ , 有以下性质:

$$Q_m(R) \geq \int_{B_R(x_m)} d\mu_m > \lambda - \epsilon, \quad (5.110)$$

$$Q_m(R) \leq Q_m(R_m) < \lambda + \epsilon. \quad (5.111)$$

此外, 给定  $R' > R$ , 我们可以增大  $m_0(\epsilon)$  并假设对于所有  $m \geq m_0$  有  $R_m \geq R'$ 。现在令  $\mu_m^1 = \mu_m \cdot \chi_{B_R(x_m)}$ ,  $\mu_m^2 = \mu_m \cdot \chi_{\mathcal{X} \setminus B_{R_m}(x_m)}$ 。显然, 我们有  $0 \leq \mu_m^1 + \mu_m^2 \leq \mu_m$  且

$$\text{supp}(\mu_m^1) \subset B_R(x_m), \quad \text{supp}(\mu_m^2) \subset \mathcal{X} \setminus B_{R_m}(x_m) \subset \mathcal{X} \setminus B_{R'}(x_m). \quad (5.112)$$

最后, 对于  $m \geq m_0(\epsilon)$ , 我们可以估计

$$\begin{aligned} & \left| \lambda - \int_{\mathcal{X}} d\mu_m^1 \right| + \left| (1 - \lambda) - \int_{\mathcal{X}} d\mu_m^2 \right| \\ &= \left| \lambda - \int_{B_R(x_m)} d\mu_m \right| + \left| \lambda - \int_{B_{R_m}(x_m)} d\mu_m \right| \\ &< 2\epsilon, \end{aligned} \quad (5.113)$$

因此我们得到二分情况。

#### 5.4.9 引理 5.5 的证明

设  $\{\nu_m\}_{m=1}^{\infty} \subset \mathcal{P}(\mathcal{Y})$  为一列弱收敛到  $\nu$  的概率测度序列。证明分三步进行。

步骤一: 内层积分的上半连续性。固定  $x \in \mathcal{X}$  及  $\beta \geq 0$ , 定义函数  $h_x : \mathcal{Y} \rightarrow [0, 1]$  为

$$h_x(y) := e^{-\beta\rho(x,y)}. \quad (5.114)$$

由于  $\rho(x, \cdot)$  为下半连续函数,  $h_x(\cdot)$  为上半连续且有界的函数。由 Portmanteau 定理, 弱收敛  $\nu_m \xrightarrow{w} \nu$  蕴含

$$\limsup_{m \rightarrow \infty} \int_{\mathcal{Y}} h_x(y) d\nu_m(y) \leq \int_{\mathcal{Y}} h_x(y) d\nu(y). \quad (5.115)$$

步骤二：被积函数的下半连续性。令

$$\Phi_m(x) := \int_{\mathcal{Y}} e^{-\beta\rho(x,y)} d\nu_m(y), \quad \Phi(x) := \int_{\mathcal{Y}} e^{-\beta\rho(x,y)} d\nu(y). \quad (5.116)$$

步骤一中的不等式可写为

$$\limsup_{m \rightarrow \infty} \Phi_m(x) \leq \Phi(x). \quad (5.117)$$

注意到函数  $t \mapsto -\log t$  在  $(0, 1]$  上连续且严格单调递减，因此有

$$\liminf_{m \rightarrow \infty} (-\log \Phi_m(x)) = -\log \left( \limsup_{m \rightarrow \infty} \Phi_m(x) \right) \geq -\log \Phi(x). \quad (5.118)$$

当  $\Phi(x) = 0$  时，上述不等式两端同为  $+\infty$ ，仍然成立。因此， $J_\beta$  中的被积函数关于  $\nu$  是逐点下半连续的。

步骤三：应用 *Fatou* 引理。注意到  $-\log \Phi_m(x)$  在下方有界（例如在概率测度情形下有  $-\log \Phi_m(x) \geq 0$ ），因此可以对外层关于  $\mu$  的积分应用 *Fatou* 引理，得到

$$\begin{aligned} \liminf_{m \rightarrow \infty} J_\beta(\nu_m) &= \liminf_{m \rightarrow \infty} \int_{\mathcal{X}} -\log \Phi_m(x) d\mu(x) \\ &\geq \int_{\mathcal{X}} \liminf_{m \rightarrow \infty} (-\log \Phi_m(x)) d\mu(x) \\ &\geq \int_{\mathcal{X}} -\log \Phi(x) d\mu(x) \\ &= J_\beta(\nu). \end{aligned} \quad (5.119)$$

由此可见， $J_\beta$  关于弱拓扑是下半连续的。

#### 5.4.10 定理 5.8 的证明

不失一般性，我们可以假设  $\beta^* > 0$  且  $\{\nu_m\}$  是一个递减的极小化序列，即对于每个  $m \in \mathbb{N}$ ，有  $R^{(m+1)}(D) \leq R^{(m)}(D)$ 。于是有  $\lim_{m \rightarrow \infty} R^{(m)}(D) = R(D)$ 。根据定理 5.7，存在子序列  $\{\nu_m\}$  满足以下三种情况之一：

1. **紧性 (Compactness)**: 存在序列  $\{y_m\} \subset \mathcal{Y}$ ，使得  $\forall \epsilon > 0, \exists R > 0$ ，满足

$$\int_{B_R(y_m)} d\nu_m \geq 1 - \epsilon, \quad \forall m \in \mathbb{N}. \quad (5.120)$$

2. **消散 (Vanishing)**: 对于每个固定的  $R > 0$ ，有

$$\lim_{m \rightarrow \infty} \left( \sup_{y \in \mathcal{Y}} \int_{B_R(y)} d\nu_m \right) = 0 \quad (5.121)$$

3. **二分 (Dichotomy):** 存在  $\lambda \in (0, 1)$ , 使得  $\forall \epsilon > 0, \exists R > 0$  以及序列  $\{y_m\} \subset \mathcal{Y}$  满足以下性质: 给定  $R' > R$ , 存在非负测度  $\nu_m^1, \nu_m^2$  使得

$$\begin{aligned} 0 &\leq \nu_m^1 + \nu_m^2 \leq \nu_m, \\ \text{supp}(\nu_m^1) &\subset B_R(y_m), \quad \text{supp}(\nu_m^2) \subset B_{R'}^c(y_m), \\ \limsup_{m \rightarrow \infty} \left( \left| \lambda - \int_{\mathcal{Y}} d\nu_m^1 \right| + \left| (1 - \lambda) - \int_{\mathcal{Y}} d\nu_m^2 \right| \right) &\leq \epsilon, \\ \text{dist}(\text{supp}(\nu_m^1), \text{supp}(\nu_m^2)) &\rightarrow \infty, \quad m \rightarrow \infty. \end{aligned} \quad (5.122)$$

如果**消散**情况成立, 对于每个  $\epsilon > 0$  和固定的  $R > 0$ , 存在  $M \in \mathbb{N}$ , 使得

$$\sup_{y \in \mathcal{Y}} \int_{B_R(y)} d\nu_m < \frac{\epsilon}{2}, \quad \forall m \geq M. \quad (5.123)$$

对于每个固定的  $x \in \mathcal{X}$  和大的  $R > 0, m \geq M$ , 我们有如下分段估计:

$$\begin{aligned} \int_{\mathcal{Y}} e^{-\beta^* \rho(x, y)} d\nu_m &= \int_{B_R(y')} e^{-\beta^* \rho(x, y)} d\nu_m + \int_{B_R^c(y')} e^{-\beta^* \rho(x, y)} d\nu_m \\ &\leq \frac{\epsilon}{2} + \int_{B_R^c(y')} e^{-\beta^* \rho(x, y)} d\nu_m \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \end{aligned} \quad (5.124)$$

于是有

$$-\log \left( \int e^{-\beta^* \rho(x, y)} d\nu_m \right) \geq \log(\epsilon^{-1}) \quad (5.125)$$

此外, 由 Fatou 引理, 我们有以下估计:

$$\begin{aligned} R(D) &= \lim_{m \rightarrow \infty} R^{(m)}(D) \\ &\geq \int_{\mathcal{X}} \liminf_{m \rightarrow \infty} \left\{ -\log \left( \int e^{-\beta^* \rho(x, y)} d\nu_m \right) \right\} d\mu - \beta^* D \\ &\geq \int_{\mathcal{X}} \log(\epsilon^{-1}) d\mu - \beta^* D \\ &\rightarrow +\infty, \end{aligned} \quad (5.126)$$

当  $\epsilon \rightarrow 0$  时。这与  $\{\nu_m\}$  是  $R(D)$  的极小化序列的假设矛盾, 因为对于  $D > D_{\min}$ , 有  $R(D) < \infty$ 。因此, **消散**情况不可能成立。

如果**二分**情况成立, 我们可以构造一个子序列  $\{\nu_m\}$  使得以下性质成立:

$$\begin{aligned} 0 &\leq \nu_m^1 + \nu_m^2 \leq \nu_m, \\ \text{supp}(\nu_m^1) &\subset B_R(y_{m_k}), \quad \text{supp}(\nu_m^2) \subset B_{R'}^c(y_{m_k}), \\ \limsup_{k \rightarrow \infty} \left( \left| \lambda - \int_{\mathcal{Y}} d\nu_m^1 \right| + \left| (1 - \lambda) - \int_{\mathcal{Y}} d\nu_m^2 \right| \right) &\leq \epsilon, \\ \text{dist}(\text{supp}(\nu_m^1), \text{supp}(\nu_m^2)) &\rightarrow \infty, \quad k \rightarrow \infty. \end{aligned} \quad (5.127)$$

我们有如下截断：

$$\begin{aligned} \int_{\mathcal{Y}} e^{-\beta^* \rho(x,y)} d\nu_m &= \int_{\text{supp}(\nu_m^1)} e^{-\beta^* \rho(x,y)} d\nu_m + \int_{\text{supp}(\nu_m^2)} e^{-\beta^* \rho(x,y)} d\nu_m \\ &\quad + \int_{\mathcal{Y} \setminus (\text{supp}(\nu_m^1) \cup \text{supp}(\nu_m^2))} e^{-\beta^* \rho(x,y)} d\nu_m \\ &= \int_{\mathcal{Y}} e^{-\beta^* \rho(x,y)} d\nu_m^1 + \int_{\mathcal{Y}} e^{-\beta^* \rho(x,y)} d\nu_m^2 + \int_{\mathcal{Y}} e^{-\beta^* \rho(x,y)} d\nu_m^3, \end{aligned}$$

其中  $\nu_m^3 = \nu_m - \nu_m^1 - \nu_m^2$ 。另一方面，我们通过下式精细地调整  $\nu_m$  的分量：

$$\tilde{\nu}_m := \alpha_m \cdot \nu_m^1 + \beta_m \cdot \nu_m^2 + \nu_m^3, \quad (5.128)$$

其中  $\alpha_m \int d\nu_m^1 + \beta_m \int d\nu_m^2 + \int d\nu_m^3 = 1$ ,  $\alpha_m, \beta_m > 0$ 。如果我们取  $m$  足够大，则根据二分的性质，我们有

$$\alpha_m(\lambda - \epsilon) + \beta_m(1 - \lambda - \epsilon) - \epsilon \leq 1 \leq \alpha_m(\lambda + \epsilon) + \beta_m(1 - \lambda + \epsilon) + \epsilon. \quad (5.129)$$

对于大的  $m$ ，取  $\alpha_m \equiv \frac{1 + \lambda^{-1}}{2} = \frac{1 + \lambda}{2\lambda}$ ，我们可以保证  $\beta$  具有严格正的下界和上界：

$$\frac{1 - \epsilon - \frac{(\lambda+1)(\lambda+\epsilon)}{2\lambda}}{1 - \lambda + \epsilon} \leq \beta_m \leq \frac{1 + \epsilon - \frac{(\lambda+1)(\lambda-\epsilon)}{2\lambda}}{1 - \lambda - \epsilon}. \quad (5.130)$$

如果我们取  $\epsilon$  足够小，则对于所有大的  $m$ ， $\beta_m$  可以被限制在一个紧区间  $[\frac{1}{2} - \delta, \frac{1}{2} + \delta]$  内，其中  $\delta \ll 1$ 。

对于每个大的  $m$  和固定的  $x \in \mathcal{X}$ ，我们可以估计与  $\nu_m$  和  $\tilde{\nu}_m$  相关的积分差值如下：

$$\begin{aligned} & -\log \left( \int e^{-\beta^* \rho(x,y)} d\tilde{\nu}_m \right) - (-1) \cdot \log \left( \int e^{-\beta^* \rho(x,y)} d\nu_m \right) \\ &= \log \frac{\int e^{-\beta^* \rho(x,y)} d\nu_m^1 + \int e^{-\beta^* \rho(x,y)} d\nu_m^2 + \int e^{-\beta^* \rho(x,y)} d\nu_m^3}{\alpha_m \int e^{-\beta^* \rho(x,y)} d\nu_m^1 + \beta_m \int e^{-\beta^* \rho(x,y)} d\nu_m^2 + \int e^{-\beta^* \rho(x,y)} d\nu_m^3} \\ &< \log \frac{\int e^{-\beta^* \rho(x,y)} d\nu_m^1 + 2\epsilon}{\frac{1+\lambda}{2\lambda} \int e^{-\beta^* \rho(x,y)} d\nu_m^1 - (2^{-1} + \delta + 1)\epsilon} \\ &= -\log \left( 1 + \frac{1 - \lambda}{2\lambda} \right) + \log \frac{\int e^{-\beta^* \rho(x,y)} d\nu_m^1 + 2\epsilon}{\int e^{-\beta^* \rho(x,y)} d\nu_m^1 - \frac{(3 + 2\delta)\lambda}{1 + \lambda} \epsilon}. \quad (5.131) \end{aligned}$$

由于失真函数  $\rho$  的适当性假设，不失一般性，我们假设  $\int e^{-\beta^* \rho(x,y)} d\nu_m^1 > 0$ 。那么我们可以取  $\epsilon$  足够小，使得上述不等式中的第二项被一个小常数  $\epsilon'$  界定，即

$$\begin{aligned} -\log \left( \int e^{-\beta^* \rho(x,y)} d\tilde{\nu}_m \right) - (-1) \cdot \log \left( \int e^{-\beta^* \rho(x,y)} d\nu_m \right) &< -\log \left( 1 + \frac{1 - \lambda}{2\lambda} \right) + \epsilon' \\ &< -\log \left( 1 + \frac{1 - \lambda}{3\lambda} \right). \end{aligned}$$

由 Fatou 引理, 我们有以下估计:

$$\begin{aligned}
\liminf_{m \rightarrow \infty} (R^{(m)}(D) - \tilde{R}^{(m)}(D)) &= \liminf_{m \rightarrow \infty} \int_{\mathcal{X}} \left\{ -\log \left( \int e^{-\beta^* \rho(x,y)} d\nu_m \right) \right. \\
&\quad \left. + \log \left( \int e^{-\beta^* \rho(x,y)} d\tilde{\nu}_m \right) \right\} d\mu \\
&\geq \int_{\mathcal{X}} \liminf_{m \rightarrow \infty} \left\{ -\log \left( \int e^{-\beta^* \rho(x,y)} d\nu_m \right) \right. \\
&\quad \left. + \log \left( \int e^{-\beta^* \rho(x,y)} d\tilde{\nu}_m \right) \right\} d\mu \\
&> \log \left( 1 + \frac{1-\lambda}{3\lambda} \right) > 0, \tag{5.132}
\end{aligned}$$

其中

$$\tilde{R}^{(m)}(D) := - \int_{\mathcal{X}} \log \left( \int e^{-\beta^* \rho(x,y)} d\tilde{\nu}_m \right) d\mu - \beta^* D. \tag{5.133}$$

因此, 对于某个大的  $m$ , 我们可以找到测度  $\tilde{\nu}_m$ , 使得

$$\begin{aligned}
\tilde{R}^{(m)}(D) &:= - \int_{\mathcal{X}} \log \left( \int e^{-\beta^* \rho(x,y)} d\tilde{\nu}_m \right) d\mu - \beta^* D \\
&\leq R^{(m)}(D) - \log \left( 1 + \frac{1-\lambda}{3\lambda} \right) \\
&\leq R(D) - \log \left( 1 + \frac{1-\lambda}{4\lambda} \right) \\
&< R(D), \tag{5.134}
\end{aligned}$$

这与  $\{\nu_m\}$  是  $R(D)$  的极小化序列的假设矛盾。因此, 二分情况不可能成立。

由于 **消散 (Vanishing)** 与 **分裂 (Dichotomy)** 两种情形已被排除, 极小化序列  $\{\nu_m\}$  必须满足 **紧性 (Compactness)** 这一情形。具体而言, 存在一列中心点  $\{y_m\} \subset \mathcal{Y}$ , 使得对任意  $\epsilon > 0$ , 都存在半径  $R > 0$  满足

$$\nu_m(B_R(y_m)) \geq 1 - \epsilon, \quad \forall m \in \mathbb{N}.$$

该性质表明测度序列集中于一列一致有界的集合之上, 但这一结论本身并不能排除这些集合整体向无穷远处“逃逸”的可能性。

在此阶段, 我们利用定理中的关键假设来证明中心点序列  $\{y_m\}$  必须是有界的。我们采用反证法。假设  $\{y_m\}$  是无界的, 则存在一个子序列 (为简化记号仍记为  $\{y_m\}$ ), 以及某个固定的参考点  $y_0 \in \mathcal{Y}$ , 使得当  $m \rightarrow \infty$  时,

$$d(y_m, y_0) \rightarrow \infty.$$

下面分析该无界性对泛函  $J_{\beta^*}$  中内层积分的影响。由失真函数  $\rho$  的 **强制性 (coercivity)** (假设 5.4), 对任意满足条件的  $x \in \mathcal{X}$  以及任意给定的足够大的常数  $M > 0$ , 存在一个半径  $K_M > 0$ , 使得

$$\rho(x, y) \geq M, \quad \forall y \in \mathcal{Y} \setminus B_{K_M}(y_0). \quad (5.135)$$

由于  $d(y_m, y_0) \rightarrow \infty$ , 对于紧性情形中给定的固定半径  $R$ , 必然存在一个整数  $N_M$ , 使得当  $m \geq N_M$  时,

$$d(y_m, y_0) > K_M + R.$$

由三角不等式可知, 此时球  $B_R(y_m)$  完全包含于强制性区域中, 即

$$B_R(y_m) \subset \mathcal{Y} \setminus B_{K_M}(y_0), \quad \forall m \geq N_M.$$

因此, 对任意  $y \in B_R(y_m)$ , 不等式 (5.135) 均成立。

对  $m \geq N_M$ , 我们仿照排除**消散**的步骤, 通过分割积分区域来估计内层积分:

$$\int_{\mathcal{Y}} e^{-\beta^* \rho(x, y)} d\nu_m(y) = \int_{B_R(y_m)} e^{-\beta^* \rho(x, y)} d\nu_m(y) + \int_{\mathcal{Y} \setminus B_R(y_m)} e^{-\beta^* \rho(x, y)} d\nu_m(y).$$

对于第一项, 由于  $y \in B_R(y_m)$  蕴含  $\rho(x, y) \geq M$ , 因此  $e^{-\beta^* \rho(x, y)} \leq e^{-\beta^* M}$ ; 对于第二项, 则使用平凡估计  $e^{-\beta^* \rho(x, y)} \leq 1$ . 由此得到

$$\int_{\mathcal{Y}} e^{-\beta^* \rho(x, y)} d\nu_m(y) \leq e^{-\beta^* M} \nu_m(B_R(y_m)) + \nu_m(\mathcal{Y} \setminus B_R(y_m)).$$

由紧性替代情形可知, 对任意  $\epsilon > 0$ , 存在  $R$  使得  $\nu_m(B_R(y_m)) \geq 1 - \epsilon$ , 从而  $\nu_m(\mathcal{Y} \setminus B_R(y_m)) < \epsilon$ . 代入上式可得

$$\int_{\mathcal{Y}} e^{-\beta^* \rho(x, y)} d\nu_m(y) \leq (1 - \epsilon)e^{-\beta^* M} + \epsilon,$$

该不等式对所有  $m \geq N_M$  成立。令  $m \rightarrow \infty$  取上极限, 得到

$$\limsup_{m \rightarrow \infty} \int_{\mathcal{Y}} e^{-\beta^* \rho(x, y)} d\nu_m(y) \leq (1 - \epsilon)e^{-\beta^* M} + \epsilon.$$

由于该结论对任意  $\epsilon > 0$  均成立, 令  $\epsilon \rightarrow 0$ , 可得

$$\limsup_{m \rightarrow \infty} \int_{\mathcal{Y}} e^{-\beta^* \rho(x, y)} d\nu_m(y) \leq e^{-\beta^* M}.$$

进一步地, 由于  $M > 0$  是任意大的常数, 可推出极限必为零:

$$\lim_{m \rightarrow \infty} \int_{\mathcal{Y}} e^{-\beta^* \rho(x, y)} d\nu_m(y) = 0.$$

该内层积分在  $\mu$ -几乎处处的  $x$  上逐点收敛到零，从而可以对整体泛函应用 Fatou 引理：

$$\begin{aligned} \lim_{m \rightarrow \infty} J_{\beta^*}(\nu_m) &= \lim_{m \rightarrow \infty} \int_{\mathcal{X}} \left( -\log \int_{\mathcal{Y}} e^{-\beta^* \rho(x,y)} d\nu_m(y) \right) d\mu(x) \\ &\geq \int_{\mathcal{X}} \liminf_{m \rightarrow \infty} \left( -\log \int_{\mathcal{Y}} e^{-\beta^* \rho(x,y)} d\nu_m(y) \right) d\mu(x). \end{aligned}$$

由于对数函数中的积分趋于 0，被积函数趋于  $+\infty$ ，因此

$$\lim_{m \rightarrow \infty} J_{\beta^*}(\nu_m) = +\infty.$$

这与最初假设  $\{\nu_m\}$  是一个极小化序列、其值应收敛到有限值  $R(D)$  相矛盾。因此，关于  $\{y_m\}$  无界的假设必然是错误的，中心点序列  $\{y_m\}$  必须是有界的。

既然  $\{y_m\}$  有界，则存在一个更大的、固定的球  $B_{R'}(y_0)$ ，能够包含所有的  $B_R(y_m)$ 。这意味着对任意  $\epsilon > 0$ ，存在一个固定的有界集合  $B' = B_{R'}(y_0)$ ，使得

$$\nu_m(B') \geq 1 - \epsilon, \quad \forall m \in \mathbb{N}.$$

接下来利用  $\mathcal{Y}$  为适当的波兰空间 (**proper Polish space**) 的假设。根据适当空间的定义，任意闭且有界的集合都是紧的。因此，我们找到了一个固定的紧集  $K = B'$ ，使得  $\nu_m(K) \geq 1 - \epsilon$  对所有  $m$  成立。这正是测度序列  $\{\nu_m\}$  **紧性 (tightness)** 的定义。

由 Prokhorov 定理可知， $\{\nu_m\}$  的紧性蕴含其在概率测度空间  $\mathcal{P}(\mathcal{Y})$  中关于弱拓扑的闭包是紧的。因此，存在一个子序列  $\{\nu_{m_k}\}$  以及某个测度  $\nu \in \mathcal{P}(\mathcal{Y})$ ，使得

$$\nu_{m_k} \xrightarrow{w} \nu, \quad k \rightarrow \infty.$$

最后，由泛函  $J_{\beta^*}$  的下半连续性（见引理 5.5）以及  $\{\nu_{m_k}\}$  为极小化序列这一事实，有

$$J_{\beta^*}(\nu) \leq \liminf_{k \rightarrow \infty} J_{\beta^*}(\nu_{m_k}) = \lim_{m \rightarrow \infty} J_{\beta^*}(\nu_m) = R(D).$$

另一方面，由于  $R(D)$  被定义为所有测度上的下确界，必有  $J_{\beta^*}(\nu) \geq R(D)$ 。综合两者可得  $J_{\beta^*}(\nu) = R(D)$ 。这表明该下确界在  $\nu$  处取得，从而  $\nu$  即为最优的重构分布。

## 5.5 本章小结

本章致力于利用现代分析数学的有力工具，特别是最优弱传输 (OWT) 理论与集中紧性原理，对经典率失真理论进行深度的重构与拓展。通过将几何与分析的双重

视角引入信源编码领域，我们不仅克服了传统理论在非紧空间与非连续失真度量下的局限性，更为理解率失真函数的数学本质提供了全新的范式。

首先，在几何结构层面，我们建立了基于最优弱传输理论的率失真分析框架。不同于传统方法将率失真函数视为单一的条件分布优化问题，本章将其重构为目标测度空间上的变分问题。利用最优弱传输的对偶理论，我们在不依赖 Karush-Kuhn-Tucker (KKT) 条件的情况下，严格导出了率失真函数的显式参数化表示。这一结果深刻揭示了率失真问题与随机控制中的薛定谔桥问题 (Schrödinger Bridge Problem) 之间的内在对偶关系，即率失真函数本质上刻画了从乘积测度到目标联合分布的最小相对熵投影。此外，该几何框架还为香农下界 (SLB) 的可达性条件提供了一个简洁且严谨的证明路径。

其次，针对非紧波兰空间上最优重构分布的存在性这一长期悬而未决的分析难题，本章引入了非线性泛函分析中的集中紧性原理。在缺乏重构空间紧性与失真函数连续性假设的一般设定下，率失真泛函的极小化序列可能会表现出质量“逃逸” (Vanishing) 或“分裂” (Dichotomy) 的病态行为。通过对极小化序列进行精细的渐近分析，我们证明了在失真函数具有“强制性 (Coercivity)”的条件下，上述病态情形可被严格排除，从而确立了序列的“紧性”。这标志着在一般的拓扑设定下，我们成功解决了最优重构分布的存在性问题。

综上所述，本章的工作实现了率失真理论从经典概率论语言向现代分析数学语言的跨越。通过构建这一严密的数学方法，我们不仅夯实了率失真理论在非紧、非光滑设定下的基础，也为处理包括高维几何数据、函数空间信号以及安全约束控制在内的复杂信源编码问题开辟了新的研究路径。

## 第 6 章 总结与展望

### 6.1 全文总结

在经典信息论的框架下，研究对象通常被设置为离散符号序列、无记忆信源或具有良好解析性质的高斯过程。其所依赖的数学工具，主要集中于组合数学、渐近统计学以及经典的概率极限理论。这一以离散建模与统计渐近为核心的分析范式，在传统通信问题中之行之有效。然而，随着大数据与人工智能的发展，现代信息处理的对象已不再局限于离散符号或低维高斯信号，而是扩展至定义在复杂高维空间、且具有强非线性动力学特性的随机过程。在这一背景下，传统基于组合数学与概率极限理论的经典信息论分析框架面临着诸多理论瓶颈：

- 现有对信息量演化的研究多集中于线性高斯信道（如热流或 Ornstein-Uhlenbeck 流），其结论常依赖特定结构或显式计算；而现代随机系统（如生成扩散模型、随机控制）通常由高度非线性的随机微分方程驱动，其对应的 Fokker-Planck 方程缺乏显式解，甚至可能包含非连续的跳跃成分（即跳扩散过程）。在这些更一般的动力学系统中，显式解往往不存在，经典方法无法回答“势函数的几何曲率如何决定信息的耗散速率”以及“互信息的时间凸性是否依然保持”等关键问题。缺乏对这些几何性质的刻画，使得我们难以从根本上理解复杂信道中的信息流变规律。
- 在分析高维采样算法（如 Langevin Dynamics, Proximal Sampler）的收敛性时，经典的数据处理不等式仅提供非增性定性判断，而无法给出量化的收缩系数。更重要的是，现有研究主要集中在较弱的度量（如 KL 散度、TV 距离）上。然而，为了保证算法在对数凹甚至非对数凹分布上的快速混合，往往需要控制更强的几何量（如相对 Fisher 信息或更一般的  $\Phi$ -Fisher 信息）。对于这类涉及概率密度高阶导数的广义信息度量，传统的直接基于 Poincaré 或 Log-Sobolev 不等式的分析框架显得力不从心，难以处理其沿扩散流的演化与收缩问题。
- 在经典框架下，率失真问题的解的存在性往往依赖于字母表的有限性或重构空间的紧致性。然而，现代信源编码的研究对象已扩展至一般的非紧波兰空间（如  $\mathbb{R}^n$  或无穷维函数空间），且失真度量可能具有非连续性（如硬约束或阶梯型代价）。在这些良好性质缺失的拓扑条件下，率失真泛函的极小化序列极易表现出“紧性缺失”的病态行为：即概率质量可能“逃逸”至无穷远处或分裂为互

不相关的碎片。传统的凸分析方法对此束手无策，导致最优编码策略在理论上的存在性长期悬而未决。

因此，重构信息论的数学基础，将其从传统的组合与概率框架，推进到更加深邃的几何与分析框架，已成为当前学术界的迫切需求。本章后续对全文的具体工作的研究重点和创新点进行总结，并提出基于当前研究的后续工作展望。

本文的主要研究工作与创新性成果总结如下：

- **建立了 Fokker-Planck 流下互信息几何性质的完整理论框架。**针对非线性势随机动力系统下的信息演化问题，本文摒弃了对显式解（如热核）的依赖，利用抛物型偏微分方程的正则性理论，严格证明了 Fokker-Planck 方程解的适定性及对数凹性质在时间演化中的保持性。在此基础上，本文系统推导了互信息沿 Fokker-Planck 流的二阶时间导数，建立了一般强凸势函数下互信息凸性的几何判据，将经典的热流与 Ornstein-Uhlenbeck 流结果推广至更一般的非线性系统。此外，本文还将研究边界拓展至非连续的跳-扩散（Jump-diffusion）信道，推导了包含非高斯跳跃项的信息微分恒等式，完善了信道演化理论在非连续路径下的图景。
- **构建了基于对数 Hessian 比较原理的广义 Fisher 信息强数据处理不等式 (SDPI) 理论。**针对高维采样算法收敛性分析中的难题，本文将研究对象从传统的 KL 散度拓展至更强的相对 Rényi-Fisher 信息与相对  $\Phi$ -Fisher 信息。通过精细的演化演算，本文提出了“对数 Hessian 比较原理”（Log-Hessian Comparison Principle），深刻揭示了目标分布的二阶几何结构与信息量收缩系数之间的内蕴联系。基于这一原理，本文建立了一套可验证的判别准则，成功证明了广义 Fisher 信息在一般情形下的最终收缩性及高斯初始化下的永恒收缩性。作为理论应用，本文给出了 Proximal Sampler 算法在更强范数意义下的混合时间上界，为理解朗之万动力学类采样算法的收敛行为提供了新的数学工具。
- **基于最优弱传输与集中紧性原理重构了率失真理论。**本文将视角从动态演化转向静态压缩，利用最优弱传输（OWT）理论揭示了率失真函数与随机控制中薛定谔桥问题（Schrödinger Bridge Problem）的深刻对偶联系，并导出了 RD 函数在一般波兰空间上的参数化表示。针对非紧空间中率失真函数最优重构分布存在性这一长期悬而未决的分析难题，本文创新性地引入了非线性泛函分析中的“集中紧性原理”（Concentration-Compactness Principle）。通过对极小化序列进行“紧性、消散、二分”的三态分析，本文严格排除了质量逃逸与分裂的病态

情形，从而在极宽泛的条件（无界空间、非连续失真度量）下证明了最优解的存在性。此外，本文还利用测度论工具澄清了关于绝对连续性假设的理论争鸣，并指出了本文框架在处理非连续失真度量（如硬约束）方面的普适性，从根本上弥补了经典信息论在处理复杂、非紧致信源时的理论缺口。

## 6.2 未来工作展望

本文虽然在信息论的几何与分析基础方面取得了一定进展，但现代数学与信息科学的交叉融合仍是一片广阔的蓝海。基于本文的研究基础，未来可在以下几个方向展开更深层次的探索：

- **连续时间 Markov 信道下互信息高阶凸性与熵动力学的研究。** 探究 FP 流在初始状态不满足  $V/2$ -相对对数凹时，互信息随时间是否具备“最终凸性”（Eventual Convexity），以及是否能在更弱的矩条件或 Fisher 信息约束下保持凸性。此外，在跳-扩散框架下，进一步探索概率测度演化的几何性质，例如寻找使得 KL 散度指数衰减的 Log-Sobolev 不等式的充分条件，以及建立 Wasserstein 空间中适应非连续路径的度量收缩理论。
- **单调路径与加倍技巧的等价性研究及应用。** 目前有关高斯极值不等式的研究，主要涉及了经典的“加倍技巧”（Doubling Trick）[179] 与“单调路径法”（Monotone Path）[180-181]。传统观点常将 Doubling Trick 视为利用独立副本和旋转不变性的一次性离散操作，而将 Monotone Path 视为沿扩散流的连续微积分分析，将二者完全区分开。未来的工作可以深入挖掘这两者背后统一的几何结构。从算子半群的视角来看，Doubling Trick 本质上验证了目标泛函在 OU 半群演化中特定时间步长（ $t = \ln \sqrt{2}$ ）下的收缩性。基于扩散过程的时间无穷可分性（Infinite Divisibility）与自相似性，这种离散的收缩性足以蕴含沿整个演化路径的单调性。换言之，通过不断对时间区间二分加细，Monotone Path 可以被视为 Doubling Trick 的无穷小微分形式，而 Doubling Trick 则是 Monotone Path 在乘积空间对称方向上的宏观积分体现。深入研究这种在乘积空间中连续旋转路径的几何性质，有望为证明更复杂的高斯极值不等式提供统一的方法论指导。
- **高阶 Otto 演算与高斯完全单调性猜想（GCMC）。** 本文第三章涉及了熵泛函沿 Wasserstein 梯度流的二阶导数，这对应于 Otto 演算中的二阶几何结构。然而，信息论中一个更深远的未解之谜是“高斯完全单调性猜想”（Gaussian Completely Monotone Conjecture, GCMC），即熵沿着热流的任意  $m$  阶时间导数满足

$(-1)^m \frac{d^m}{dt^m} H(\mu_t) \geq 0$ 。目前已有工作探索了 **高阶 Otto 演算** (Higher-order Otto Calculus), 用于推进 GCMC 的研究 [182]。由于 Wasserstein 空间具有非零的黎曼曲率, 直接计算高阶导数极其困难。一个极具潜力的思路是借鉴流体力学中的 **对流导数** (Convective Derivative) 概念, 将其视为 Wasserstein 空间上的一种平坦联络 (Flat Connection)。通过建立 Wasserstein 空间上的单变量 Faa di Bruno 公式, 可以解析计算熵泛函的任意高阶微分, 从而尝试证明 GCMC 在一般维数及对数凹初始分布下的成立。这一方向不仅具有纯数学的理论美感, 也可能揭示信息耗散的深层物理规律。目前已有初步工作在探索这一思路, 将结论推进到了前五阶时间导数, 但更高阶导数甚至完整的 GCMC 是否成立仍有待进一步研究。

- **基于广义信息测度的采样算法理论深化。** 本文第四章建立了广义 Fisher 信息的 SDPI, 并初步应用于 Proximal Sampler 的分析。未来的工作可以进一步挖掘这一框架在更复杂现代采样算法中的潜力:
  - **退化扩散过程与欠阻尼朗之万动力学 (Underdamped Langevin Dynamics):** 现有的分析主要集中在过阻尼 (Overdamped) 情形, 其对应的 Fokker-Planck 方程是抛物型的。然而, 在加速采样算法中, 引入动量项的欠阻尼朗之万动力学 (ULD) [183] 表现出更优的收敛性质。ULD 对应的 Fokker-Planck 方程具有退化 (Degenerate) 的扩散系数矩阵, 这导致标准的几何判据失效。未来的工作可以尝试结合 Villani 等人发展的亚强制性 (Hypocoercivity) 理论 [184], 在相空间 (Phase Space) 中建立广义 Fisher 信息的收缩性界, 从而为动量加速算法提供更精细的收敛保证。
  - **非对数凹 (Non-log-concave) 分布下的收敛分析:** 本文的主要结论依赖于目标分布的对数凹性假设。然而, 现实中的贝叶斯推断与深度学习模型往往涉及高度非凸的势函数 (即非对数凹分布), 呈现出多模态 (Multi-modal) 特征。在这种情形下, 采样算法容易陷入局部极小值 (亚稳态)。未来的研究可以利用广义 Fisher 信息对分布尾部特征的敏感性, 结合 [137, 185] 的研究, 探究非凸景观下的混合时间, 大大拓展当前采样算法在非凸环境中的适用性。
- **基于神经网络与 OT 的率失真问题的求解算法及应用。** 本文在理论上解决了率失真函数在非紧空间的存在性与对偶表示问题, 这为设计新型数据压缩算法提供了理论基石。未来的工作可以致力于将这一理论转化为高效的数值算法, 特

别是在高维数据压缩领域。结合神经最优传输 (Neural Optimal Transport) 技术 [186], 利用本文推导的参数化表示 (特别是涉及薛定谔桥的对偶变分形式) 设计新型的损失函数, 开发基于深度神经网络的高维率失真估计算法 (Neural RD Estimation) [187-188]。这将有助于突破传统算法在高维空间中的计算瓶颈, 为图像、视频等复杂模态数据的端到端神经压缩 (Neural Compression) 提供理论指导与算法支撑, 实现理论与工程实践的深度融合。

- **基于神经网络与 OT 的率失真问题的求解算法及应用。** 本文在理论上解决了率失真函数在非紧空间的存在性与对偶表示问题, 这为设计新型数据压缩算法提供了理论基石。未来的工作可以致力于将这一理论转化为高效的数值算法, 特别是在高维数据压缩领域。结合神经最优传输 (Neural Optimal Transport) 技术 [186], 利用本文推导的参数化表示 (特别是涉及薛定谔桥的对偶变分形式) 设计新型的损失函数, 开发基于深度神经网络的高维率失真估计算法 (Neural RD Estimation) [187-188]。这将有助于突破传统算法在高维空间中的计算瓶颈, 为图像、视频等复杂模态数据的端到端神经压缩 (Neural Compression) 提供理论指导与算法支撑, 实现理论与工程实践的深度融合。
- **非预测率失真 (Non-anticipative Rate-Distortion) 理论在非紧空间中的适定性研究。** 经典的率失真理论已被本文推广至非紧空间, 然而在涉及实时通信与网络控制系统的非预测 (Non-anticipative) 或因果 (Causal) 率失真问题中, 现有的存在性理论仍主要局限于紧致字母表或需施加较强的连续性假设 [189]。由于因果约束破坏了联合分布的对称性, 使得传统的紧性论证在处理无界状态空间时面临严峻挑战。未来的工作可以将本文发展的集中紧性方法论迁移至因果测度空间。核心思路在于构建适应于带流概率空间 (Filtered Probability Space) 的紧性判据, 证明在强制性等约束下, 因果核序列 (Causal Kernels) 的极小化序列不会发生“质量逃逸” (Vanishing) 或“二分” (Dichotomy)。这将有望移除现有文献中关于重构空间紧性和失真函数一致连续性的严苛假设, 从而建立一般波兰空间上非预测率失真函数的最优解存在性定理, 为实时高斯信源编码及随机控制系统的性能极限分析提供更坚实的数学基础。



## 参考文献

- [1] SHANNON C E. A Mathematical Theory of Communication[J]. The Bell System Technical Journal, 1948, 27(3): 379-423.
- [2] STAM A J. Some Inequalities Satisfied by the Quantities of Information of Fisher and Shannon [J]. Information and Control, 1959, 2(2): 101-112.
- [3] WIBISONO A, JOG V, LOH P L. Information and Estimation in Fokker-Planck Channels[C]// Proc. IEEE Int. Symp. Inf. Theory (ISIT). Aachen, Germany, 2017: 2673-2677.
- [4] BAKRY D, GENTIL I, LEDOUX M. Analysis and Geometry of Markov Diffusion Operators[M]. Springer, 2014.
- [5] WIBISONO A, JOG V. Convexity of Mutual Information Along the Heat Flow[C]// Proc. IEEE Int. Symp. Inf. Theory (ISIT). Vail, CO, USA, 2018: 1615-1619.
- [6] WIBISONO A, JOG V. Convexity of Mutual Information Along the Ornstein-Uhlenbeck Flow [C]// International Symposium on Information Theory and Its Applications (ISITA). Singapore, 2018: 55-59.
- [7] KLARTAG B, ORDENTLICH O. The Strong Data Processing Inequality Under the Heat Flow [J]. IEEE Trans. Inf. Theory, 2025, 71(5): 3317-3333.
- [8] CHEN Y, CHEWI S, SALIM A, et al. Improved Analysis for a Proximal Algorithm for Sampling [J]. Proceedings of the 35th Conference on Learning Theory, 2022, 178: 2984-3014.
- [9] VEMPALA S, WIBISONO A. Rapid Convergence of the Unadjusted Langevin Algorithm: Isoperimetry Suffices[J]. Advances in Neural Information Processing Systems, 2019, 32: 8092-8104.
- [10] WIBISONO A. Mixing Time of the Proximal Sampler in Relative Fisher Information via Strong Data Processing Inequality[J]. Proceedings of the 38th Conference on Learning Theory, 2025, 291: 5716-5717.
- [11] CHEWI S, GERBER P, LEE H, et al. Fisher Information Lower Bounds for Sampling[J]. Proceedings of the 34th International Conference on Algorithmic Learning Theory, 2023: 375-410.
- [12] SHANNON C E. Coding Theorems for a Discrete Source With a Fidelity Criterion[J]. International Convention Record, 1959, 7: 325-350.
- [13] CSISZÁR I. On an Extremum Problem of Information Theory[J]. Studia Scientiarum Mathematicarum Hungarica, 1974, 9(1): 57-71.
- [14] BLAHUT R. Computation of Channel Capacity and Rate-Distortion Functions[J]. IEEE Trans. Inf. Theory, 1972, 18(4): 460-473.
- [15] ARIMOTO S. An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless

- Channels[J]. IEEE Trans. Inf. Theory, 1972, 18(1): 14-20.
- [16] REZAEI F, AHMED N, CHARALAMBOUS C D. Rate Distortion Theory for General Sources With Potential Application to Image Compression[J]. International Journal of Applied Mathematical Sciences, 2006, 3(2): 141-165.
- [17] VILLANI C. Optimal Transport: Old and New[M]. Springer, 2009.
- [18] GRAY R M. Transportation Distance, Shannon Information, and Source Coding[C]//GRETSI Symposium on Signal and Image Processing. 2013.
- [19] GOZLAN N, ROBERTO C, SAMSON P M, et al. Kantorovich Duality for General Transport Costs and Applications[J]. J. Funct. Anal., 2017, 273(11): 3327-3405.
- [20] SCHRÖDINGER E. Über die Umkehrung der Naturgesetze[J]. Sitzungsberichte der Preussischen Akademie der Wissenschaften. Physikalisch-Mathematische Klasse, 1931, 144: 144-153.
- [21] LIONS P L. The Concentration-Compactness Principle in the Calculus of Variations. The Locally Compact Case, Part 1[J]. Annales de l'Institut Henri Poincaré C, Analyse non linéaire, 1984, 1(2): 109-145.
- [22] LIONS P L. The Concentration-Compactness Principle in the Calculus of Variations. The Locally Compact Case, Part 2[J]. Annales de l'Institut Henri Poincaré C, Analyse non linéaire, 1984, 1(4): 223-283.
- [23] LIONS P L. The Concentration-Compactness Principle in the Calculus of Variations. The Limit Case, Part 1[J]. Revista Matemática Iberoamericana, 1985, 1(1): 145-201.
- [24] LIONS P L. The Concentration-Compactness Principle in the Calculus of Variations. The Limit Case, Part 2[J]. Revista Matemática Iberoamericana, 1985, 1(2): 45-121.
- [25] FISHER R A. On the Mathematical Foundations of Theoretical Statistics[J]. Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character, 1922, 222(594-604): 309-368.
- [26] MCKEAN H P. Speed of Approach to Equilibrium for Kac's Caricature of a Maxwellian Gas[J]. Arch. Ration. Mech. Anal., 1966, 21(5): 343-367.
- [27] ARKERYD L. On the Strong  $L^1$  Trend to Equilibrium for the Boltzmann Equation[J]. Stud. Appl. Math., 1992, 87(3): 283-288.
- [28] BOBYLEV A V, CERCIGNANI C. On the Rate of Entropy Production for the Boltzmann Equation[J]. J. Statist. Phys., 1999, 94(3): 603-618.
- [29] CARLEN E A, CARVALHO M C, LU X. On Strong Convergence to Equilibrium for the Boltzmann Equation With Soft Potentials[J]. J. Statist. Phys., 2009, 135(4): 681-736.
- [30] IMBERT C, SILVESTRE L, VILLANI C. On the Monotonicity of the Fisher Information for the Boltzmann Equation[J]. arXiv:2409.01183, 2024.
- [31] GUO D, SHAMAI S, VERDÚ S. Mutual Information and Minimum Mean-Square Error in Gaussian Channels[J]. IEEE Trans. Inf. Theory, 2005, 51(4): 1261-1282.

- [32] GANESH A, TALWAR K. Faster Differentially Private Samplers via Rényi Divergence Analysis of Discretized Langevin MCMC[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 7222-7233.
- [33] ERDOGDU M A, HOSSEINZADEH R, ZHANG S. Convergence of Langevin Monte Carlo in Chi-Squared and Rényi Divergence[J]. *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, 2022, 151: 8151-8175.
- [34] CHEWI S, ERDOGDU M A, LI M, et al. Analysis of Langevin Monte Carlo From Poincaré to Log-Sobolev[J]. *Found. Comput. Math.*, 2025, 25: 1345-1395.
- [35] ARNOLD A, MARKOWICH P, TOSCANI G, et al. On Convex Sobolev Inequalities and the Rate of Convergence to Equilibrium for Fokker-Planck Type Equations[J]. *Communications in Partial Differential Equations*, 2001, 26(1-2): 43-100.
- [36] DHARMADHIKARI S, JOAG-DEV K. *Unimodality, Convexity, and Applications*[M]. Academic Press, Boston, 1988.
- [37] SAUMARD A, WELLNER J A. Log-Concavity and Strong Log-Concavity: A Review[J]. *Statistics surveys*, 2014, 8: 45.
- [38] BAKRY D, ÉMERY M. Diffusions Hypercontractives[J]. *Séminaire de Probabilités*, 1985, 19: 177-206.
- [39] ØKSENDAL B. *Stochastic Differential Equations: An Introduction With Applications*[M]. 6th ed. Springer, 2003.
- [40] KARATZAS I, SHREVE S E. *Brownian Motion and Stochastic Calculus*[M]. 2nd ed. Springer, 1991.
- [41] SHREVE S E. *Stochastic Calculus for Finance II: Continuous-Time Models*[M]. Springer, 2004.
- [42] GARDINER C W. *Stochastic Methods: A Handbook for the Natural and Social Sciences*[M]. 4th ed. Springer, 2009.
- [43] CHEWI S. Log-Concave Sampling[J]. Book draft, available at <https://chewisinho.github.io>, 2025.
- [44] BROWN R. A Brief Account of Microscopical Observations Made in the Months of June, July, and August, 1827, on the Particles Contained in the Pollen of Plants; and on the General Existence of Active Molecules in Organic and Inorganic Bodies[J]. *Philosophical Magazine*, 1828, 4(21): 161-173.
- [45] EINSTEIN A. On the Movement of Small Particles Suspended in Stationary Liquids Required by the Molecular-Kinetic Theory of Heat[J]. *Annalen der Physik*, 1905, 17: 549-560.
- [46] BACHELIER L. Théorie de La Spéculation[J]. *Annales Scientifiques de l'École Normale Supérieure*, 1900, 17: 21-86.
- [47] WIENER N. Differential-Space[J]. *Journal of Mathematics and Physics*, 1923, 2(1-4): 131-174.
- [48] ITÔ K. Stochastic Integral[J]. *Proceedings of the Imperial Academy*, 1944, 20(8): 519-524.
- [49] ITÔ K. On Stochastic Differential Equations[J]. *Memoirs of the American Mathematical Society*,

- 1951(4): 1-51.
- [50] ARNOLD L. Stochastic Differential Equations: Theory and Applications[M]. Wiley, 1974.
- [51] EVANS L C. An Introduction to Stochastic Differential Equations[M]. American Mathematical Society, 2012.
- [52] ETHIER S N, KURTZ T G. Markov Processes: Characterization and Convergence[M]. Wiley, 1986.
- [53] RISKEN H, RISKEN H. The Fokker-Planck Equation: Methods of Solution and Applications: vol. 18[M]. Springer, 1996.
- [54] PAVLIOTIS G A. Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations[M]. Springer, 2014.
- [55] MONGE G. Mémoire sur la théorie des déblais et des remblais[J]. Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, 1781: 666-704.
- [56] KANTOROVICH L V. On the translocation of masses[J]. C.R. (Doklady) Acad. Sci. URSS (N.S.), 1942, 37: 199-201.
- [57] ROCKAFELLAR R T. Convex Analysis: vol. 28[M]. Princeton, NJ: Princeton University Press, 1970.
- [58] BRENIER Y. Décomposition Polaire et Réarrangement Monotone Des Champs de Vecteurs[J]. Comptes Rendus de l'Académie des Sciences-Series I-Mathematics, 1987, 305(19): 805-808.
- [59] PRATELLI A. On the Equality Between Monge's Infimum and Kantorovich's Minimum in Optimal Mass Transportation[J]. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 2007, 43(1): 1-13.
- [60] TAKATSU A. Wasserstein Geometry of Gaussian Measures[J]. Osaka Journal of Mathematics, 2011, 48(4): 1005-1026.
- [61] GRAY R M. Entropy and Information Theory[M]. 2nd. New York: Springer, 2011.
- [62] ZOU J, FAN L, GAO J, et al. Convexity of Mutual Information along the Fokker-Planck Flow[C] // Proc. IEEE Int. Symp. Inf. Theory (ISIT). Ann Arbor, MI, USA, 2025: 01-06.
- [63] JORDAN R, KINDERLEHRER D, OTTO F. The Variational Formulation of the Fokker-Planck Equation[J]. SIAM journal on mathematical analysis, 1998, 29(1): 1-17.
- [64] SHARMA S N, PATEL H G. The Fokker-Planck Equation[G] // Stochastic Control. Rijeka: In-techOpen, 2010: 1-20.
- [65] MAYFIELD W. A Sequence Solution to the Fokker-Planck Equation[J]. IEEE Trans. Inf. Theory, 1973, 19(2): 165-176.
- [66] SHEN Z, WANG Z, KALE S, et al. Self-Consistency of the Fokker-Planck Equation[C] // Conference on Learning Theory. 2022: 817-841.

- [67] YAU S S T. Computation of Fokker-Planck Equation[J]. Quarterly of Applied Mathematics, 2004, 62(4): 643-650.
- [68] WU Y, GUO D, VERDÚ S. Derivative of Mutual Information at Zero SNR: The Gaussian-Noise Case[J]. IEEE Trans. Inf. Theory, 2011, 57(11): 7307-7312.
- [69] GUO D, WU Y, SHITZ S S, et al. Estimation in Gaussian Noise: Properties of the Minimum Mean-Square Error[J]. IEEE Trans. Inf. Theory, 2011, 57(4): 2371-2385.
- [70] EMAMIRAD H, ROUGIREL A. De Bruijn Identities in Different Markovian Channels[J/OL]. Electronic Journal of Differential Equations, 2023, 2023(12): 1-11. DOI: 10.58997/ejde.2023.12.
- [71] COVER T M, THOMAS J A. Elements of Information Theory[M]. 2nd. New York, NY, USA: Wiley, 2006.
- [72] VILLANI C. Topics in Optimal Transportation: vol. 58[M]. Graduate Studies in Mathematics. Amer. Math. Soc., 2021.
- [73] FIGALLI A, GLAUDO F. An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows: vol. 26[M]. 2023.
- [74] SANTAMBROGIO F. Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling: vol. 87[M]. Birkhäuser, 2015.
- [75] AMBROSIO L, GIGLI N, SAVARÉ G. Gradient Flows: In Metric Spaces and in the Space of Probability Measures[M]. Springer, 2008.
- [76] FAN L, ZOU J, GAO J, et al. Differential Properties of Information in Jump-Diffusion Channels [C]//Proc. IEEE Int. Symp. Inf. Theory (ISIT). Ann Arbor, MI, USA, 2025: 01-06.
- [77] MACKEY M C. Time's Arrow: The Origins of Thermodynamic Behavior[M]. Springer, 1992.
- [78] BOGACHEV V I, KRYLOV N V, RÖCKNER M, et al. Fokker-Planck-Kolmogorov Equations: vol. 207[M]. Amer. Math. Soc., 2022.
- [79] ARNOLD A, CARLEN E, JU Q. Large-Time Behavior of Non-Symmetric Fokker-Planck Type Equations[J]. Communications on Stochastic Analysis, 2008, 2(1): 11.
- [80] JÜNGEL A. Entropy Methods for Diffusive Partial Differential Equations[M]. Springer, 2016.
- [81] LI X D. Perelman's W-Entropy for the Fokker-Planck Equation Over Complete Riemannian Manifolds[J]. Bulletin des Sciences Mathématiques, 2011, 135(6-7): 871-882.
- [82] AVRAMIDI I G, et al. Heat Kernel Method and Its Applications[M]. Birkhäuser Verlag AG, 2015.
- [83] BRASCAMP H J, LIEB E H. On Extensions of the Brunn-Minkowski and Prékopa-Leindler Theorems, Including Inequalities for Log Concave Functions, and With an Application to the Diffusion Equation[J]. Journal of Functional Analysis, 1976, 22(4): 366-389.
- [84] LEE K A, VÁZQUEZ J L. Geometrical Properties of Solutions of the Porous Medium Equation for Large Times[J]. Indiana University Mathematics Journal, 2003: 991-1016.
- [85] ISHIGE K, NAKAGAWA K, SALANI P. Spatial Concavity of Solutions to Parabolic Systems[J]. Ann. Scuola Norm. Sup. Pisa Cl. Sci., 2020: 291-313.

- [86] ISHIGE K, SALANI P, TAKATSU A. Characterization of  $F$ -Concavity Preserved by the Dirichlet Heat Flow[J]. arXiv:2207.13449, 2022.
- [87] CRANDALL M G, ISHII H, LIONS P L. User's Guide to Viscosity Solutions of Second Order Partial Differential Equations[J]. Bull. Amer. Math. Soc., 1992, 27(1): 1-67.
- [88] ISHIGE K, SALANI P, TAKATSU A. To Logconcavity and Beyond[J]. Communications in Contemporary Mathematics, 2020, 22(02).
- [89] HANSON F B. Applied Stochastic Processes and Control for Jump-diffusions: Modeling, Analysis and Computation[M]. SIAM, 2007.
- [90] TANKOV P. Financial modelling with jump processes[M]. Chapman, 2003.
- [91] BORODIN A N. Stochastic Processes[M]. Springer, 2017.
- [92] GILLESPIE D T. Markov Processes: An Introduction for Physical Scientists[M]. San Diego: Academic Press, 1992.
- [93] RYDIN GORJÃO L, WITTHAUT D, LEHNERTZ K, et al. Arbitrary-order Finite-time Corrections for the Kramers–Moyal Operator[J]. Entropy, 2021, 23(5): 517.
- [94] SHOHAT J A, TAMARKIN J D. The Problem of Moments[M]. American Mathematical Soc., 1943.
- [95] WANG F Y. Log-Sobolev Inequalities: Different Roles of Ric and Hess[J]. The Annals of Probability, 2009, 37(4): 1587-1604.
- [96] FRIEDMAN A. Partial Differential Equations of Parabolic Type[M]. Courier Dover Publications, 2008.
- [97] RUBIO G. Existence and Uniqueness to the Cauchy Problem for Linear and Semilinear Parabolic Equations With Local Conditions[C]//ESAIM: Proceedings: vol. 31. 2011: 73-100.
- [98] GUT A. Probability: A Graduate Course: vol. 75[M]. Springer Texts in Statistics. Springer, 2013.
- [99] FRIEDMAN A. Stochastic Differential Equations and Applications: Volume 1. Vol. 28[M]. Probability, 1975.
- [100] KENNINGTON A U. Power Concavity and Boundary Value Problems[J]. Indiana University Mathematics Journal, 1985, 34(3): 687-704.
- [101] SCHNEIDER R. Convex Bodies: The Brunn–Minkowski Theory: vol. 151[M]. Encyclopedia of Mathematics, 2013.
- [102] OTTO F. The Geometry of Dissipative Evolution Equations: The Porous Medium Equation[J]. Communications in Partial Differential Equations, 2001, 26(1-2): 101-174.
- [103] LEDOUX M. The Concentration of Measure Phenomenon[M]. Amer. Math. Soc., 2001.
- [104] CARLEN E, CARVALHO M. Entropy Production Estimates for Boltzmann Equations With Physically Realistic Collision Kernels[J]. J. Stat. Phys., 1994, 74(3): 743-782.
- [105] VILLANI C. Fisher Information Estimates for Boltzmann's Collision Operator[J]. J. Math. Pures Appl., 1998, 77(8): 821-837.

- [106] VILLANI C. A Review of Mathematical Topics in Collisional Kinetic Theory[J]. Handbook of mathematical fluid dynamics, 2002, 1: 71-74.
- [107] VILLANI C. Fisher Information in Kinetic Theory[J]. arXiv:2501.00925, 2025.
- [108] CARRILLO J A, GUO S. From Fisher Information Decay for the Kac Model to the Landau-Coulomb Hierarchy[J]. arXiv:2502.18606, 2025.
- [109] GUILLÉN N, SILVESTRE L. The Landau Equation Does Not Blow Up[J]. Acta Math., 2025, 234(2): 315-375.
- [110] TOSCANI G. New A Priori Estimates for the Spatially Homogeneous Boltzmann Equation[J]. Cont. Mech. Thermodyn., 1992, 4(2): 81-93.
- [111] LANDAU L D. Die kinetische Gleichung für den Fall Coulombscher Wechselwirkung[J]. Phys. Z. Sowjet., 1936, 10: 154-164.
- [112] DESVILLETES L, VILLANI C. On the Trend to Global Equilibrium for Spatially Inhomogeneous Kinetic Systems: The Boltzmann Equation[J]. Invent. Math., 2005, 159(2): 245-316.
- [113] SHANNON C E. A Mathematical Theory of Communication[J]. Bell Syst. Tech. J., 1948, 27(3): 379-423.
- [114] BLACHMAN N. The Convolution Inequality for Entropy Powers[J]. IEEE Trans. Inf. Theory, 1965, 11(2): 267-271.
- [115] PLASTINO A R, DAFFERTSHOFER A. Liouville Dynamics and the Conservation of Classical Information[J]. Physical Review Letters, 2004, 93(13), 138701.
- [116] YAMANO T. De Bruijn-Type Identity for Systems With Flux[J]. The European Physical Journal B, 2013, 86(8): 363.
- [117] SHEN Y, GAN L, LING C. Generalized Score Matching: Bridging  $f$ -Divergence and Statistical Estimation Under Correlated Noise[J]. arXiv:2504.19288, 2025.
- [118] CSISZÁR I. A Class of Measures of Informativity of Observation Channels[J]. Periodica Mathematica Hungarica, 1972, 2(1-4): 191-213.
- [119] FAN J, YUAN B, CHEN Y. Improved Dimension Dependence of a Proximal Algorithm for Sampling[J]. Proceedings of the 36th Conference on Learning Theory, 2023, 195: 1473-1521.
- [120] KOOK Y, VEMPALA S S, ZHANG M S. In-and-Out: Algorithmic Diffusion for Sampling Convex Bodies[J]. Advances in Neural Information Processing Systems, 2024, 37: 108354-108388.
- [121] MITRA S, WIBISONO A. Fast Convergence of  $\Phi$ -Divergence Along the Unadjusted Langevin Algorithm and Proximal Sampler.[J]. Proceedings of the 36th International Conference on Algorithmic Learning Theory, 2025, 272: 846-869.
- [122] KLARTAG B, ORDENTLICH O. The Strong Data Processing Inequality Under the Heat Flow [J]. IEEE Trans. Inf. Theory, 2025, 71(5): 3317-3333.
- [123] CHEWI S, LU C, AHN K, et al. Optimal Dimension Dependence of the Metropolis-Adjusted Langevin Algorithm[J]. Proceedings of the 34th Conference on Learning Theory, 2021: 1260-

- 1300.
- [124] LU J, WANG L. Complexity of Zigzag Sampling Algorithm for Strongly Log-Concave Distributions[J]. arXiv:2012.11094, 2020.
- [125] WU H, YU L. Entropic Isoperimetric and Cramér–Rao Inequalities for Rényi–Fisher Information [J]. arXiv:2504.01837, 2025.
- [126] DEMBO A, COVER T M, THOMAS J A. Information Theoretic Inequalities[J]. IEEE Trans. Inf. Theory, 1991, 37(6): 1501-1518.
- [127] CHAFAÏ D. Entropies, Convexity, and Functional Inequalities, On  $\Phi$ -Entropies and  $\Phi$ -Sobolev Inequalities[J]. Journal of Mathematics of Kyoto University, 2004, 44(2): 325-363.
- [128] GOZLAN N, LI X M, MADIMAN M, et al. Log-Hessian Formula and the Talagrand Conjecture [J]. Potential Analysis, 2022.
- [129] CHEWI S, POOLADIAN A A. An Entropic Generalization of Caffarelli’s Contraction Theorem via Covariance Inequalities[J]. Comptes Rendus. Mathématique, 2023, 361(G9): 1471-1482.
- [130] LEE Y T, SHEN R, TIAN K. Structured Logconcave Sampling With a Restricted Gaussian Oracle [J]. Proceedings of the 34th Conference on Learning Theory, 2021, 134: 2993-3050.
- [131] CAFFARELLI L A. Monotonicity Properties of Optimal Transportation and the FKG and Related Inequalities[J]. Communications in Mathematical Physics, 2000, 214(3): 547-563.
- [132] KIM Y H, MILMAN E. A Generalization of Caffarelli’s Contraction Theorem via (Reverse) Heat Flow[J]. Mathematische Annalen, 2012, 354(3): 827-862.
- [133] STROGATZ S H. Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering[M]. 3rd ed. Chapman, 2024.
- [134] BURKHOLDER D L, GUNDY R F. Extrapolation and Interpolation of Quasi-Linear Operators on Martingales[J]. Acta Math., 1970, 124: 249-304.
- [135] DAVIS B. On the Intergrability of the Martingale Square Function[J]. Israel Journal of Mathematics, 1970, 8(2): 187-190.
- [136] HOLLEY R, STROOCK D. Logarithmic Sobolev Inequalities and Stochastic Ising Models[J]. J. Statist. Phys., 1987, 46(5-6): 1159-1194.
- [137] BALASUBRAMANIAN K, CHEWI S, ERDOGDU M A, et al. Towards a Theory of Non-Log-Concave Sampling: First-Order Stationarity Guarantees for Langevin Monte Carlo[J]. Proceedings of the 35th Conference on Learning Theory, 2022: 2896-2923.
- [138] ZOU J, FAN L, GAO J, et al. A Revisit to Rate-Distortion Theory via Optimal Weak Transport [J]. arXiv:2501.09362, 2025.
- [139] BERGER T. Rate-Distortion Theory: A Mathematical Basis for Data Compression[M]. Englewood Cliffs: Prentice-Hall, 1971.
- [140] BLAHUT R E. Principles and Practice of Information Theory[M]. Addison-Wesley Longman Publishing Co., Inc., 1987.

- [141] RIEGLER E, BÖLCSKEI H, KOLIANDER G. Rate-Distortion Theory for General Sets and Measures[C]//Proc. IEEE Int. Symp. Inf. Theory (ISIT). Vail, CO, USA, 2018: 101-105.
- [142] RIEGLER E, KOLIANDER G, BÖLCSKEI H. Lossy Compression of General Random Variables [J]. Information and Inference: A Journal of the IMA, 2023, 12(3): 1759-1829.
- [143] KAWABATA T, DEMBO A. The Rate-Distortion Dimension of Sets and Measures[J]. IEEE Trans. Inf. Theory, 1994, 40(5): 1564-1572.
- [144] KOSTINA V, TUNCEL E. Successive Refinement of Abstract Sources[J]. IEEE Trans. Inf. Theory, 2019, 65(10): 6385-6398.
- [145] DONSKER M D, VARADHAN S S. Asymptotic Evaluation of Certain Markov Process Expectations for Large Time, I[J]. Commun. Pure Appl. Math., 1975, 28(1): 1-47.
- [146] EQUITZ W H, COVER T M. Successive Refinement of Information[J]. IEEE Trans. Inf. Theory, 1991, 37(2): 269-275.
- [147] YANG Y, ECKSTEIN S, NUTZ M, et al. Estimating the Rate-Distortion Function by Wasserstein Gradient Descent[J]. Advances in Neural Information Processing Systems, 2024, 36.
- [148] LÉONARD C. A Survey of the Schrödinger Problem and Some of Its Connections With Optimal Transport[J]. arXiv:1308.0215, 2013.
- [149] BACKHOFF-VERAGUAS J, BEIGLBÖCK M, PAMMER G. Existence, Duality, and Cyclical Monotonicity for Weak Transport Costs[J]. Calculus of Variations and Partial Differential Equations, 2019, 58(203).
- [150] BACKHOFF-VERAGUAS J, PAMMER G. Applications of Weak Transport Theory[J]. Bernoulli, 2022, 28(1): 370-394.
- [151] ROSE K. A Mapping Approach to Rate-Distortion Computation and Analysis[J]. IEEE Trans. Inf. Theory, 1994, 40(6): 1939-1952.
- [152] COURTY N, FLAMARY R, TUIA D, et al. Optimal Transport for Domain Adaptation[J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2016, 39(9): 1853-1865.
- [153] HAMRI M E, BENNANI Y. Regularized Optimal Transport for Dynamic Semi-Supervised Learning[J]. arXiv:2103.11937, 2021.
- [154] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein Generative Adversarial Networks[C]//International Conference on Machine Learning. 2017: 214-223.
- [155] CHEN L, SONG Y, CAI Y, et al. Multi-Prototype Space Learning for Commonsense-Based Scene Graph Generation[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 38: 2. 2024: 1129-1137.
- [156] TALAGRAND M. Transportation Cost for Gaussian and Other Product Measures[J]. Geometric & Functional Analysis (GAFA), 1996, 6(3): 587-600.
- [157] CAO H, GUO X, LAURIÈRE M. Connecting GANs, Mean-Field Games, and Optimal Transport [J]. SIAM Journal on Applied Mathematics, 2024, 84(4): 1255-1287.

- [158] GUNSILIUS F, XU Y. Matching for Causal Effects via Multimarginal Unbalanced Optimal Transport[J]. arXiv:2112.04398, 2021.
- [159] COURTADE T A. Transportation Proof of an Inequality by Anantharam, Jog and Nair[J]. arXiv:1901.10893, 2019.
- [160] BAI Y. Optimal Transport Meets Information Science: From Measure Concentration, to Information Theory, to Machine Learning[D]. University of Delaware, 2022.
- [161] BAI Y, WU X, ÖZGÜR A. Information Constrained Optimal Transport: From Talagrand, to Marton, to Cover[J]. IEEE Trans. Inf. Theory, 2023, 69(4): 2059-2073.
- [162] GOZLAN N, ROBERTO C, SAMSON P M, et al. Characterization of a Class of Weak Transport-Entropy Inequalities on the Line[J]. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 2018, 54(3): 1667-1693.
- [163] GOZLAN N, JUILLET N. On a Mixture of Brenier and Strassen Theorems[J]. Proc. Lond. Math. Soc., 2020, 120(3): 434-463.
- [164] BACKHOFF-VERAGUAS J, BEIGLBÖCK M, PAMMER G. Weak Monotone Rearrangement on the Line[J]. Electronic Communications in Probability, 2020, 25: 1-16.
- [165] BACKHOFF-VERAGUAS J, PAMMER G. Stability of Martingale Optimal Transport and Weak Optimal Transport[J]. Ann. Appl. Probab., 2022, 32(1): 721-752.
- [166] ACCIAIO B, BEIGLBÖCK M, PAMMER G. Weak Transport for Non-Convex Costs and Model-Independence in a Fixed-Income Market[J]. Math. Finance, 2021, 31(4): 1423-1453.
- [167] NUTZ M. Introduction to Entropic Optimal Transport[J]. Lecture notes, Columbia University, 2021.
- [168] LEI E, HASSANI H, SAEEDI BIDOKHTI S. On a Relation Between the Rate-Distortion Function and Optimal Transport[C]//International Conference on Learning Representations (ICLR), Tiny Papers Track. 2023.
- [169] PEYRÉ G, CUTURI M. Computational optimal transport: With applications to data science[J]. Foundations and Trends® in Machine Learning, 2019, 11(5-6): 355-607.
- [170] DUPUIS P, ELLIS R S. A Weak Convergence Approach to the Theory of Large Deviations[M]. John Wiley & Sons, 2011.
- [171] STRUWE M. Variational methods: vol. 991[M]. Springer, 2000.
- [172] CHOW W L. Über systeme von linear partiellen differentialgleichungen erster ordnung[J]. Mathematische Annalen, 1940, 117(1): 98-105.
- [173] RASHEVSKY P K. About Connecting Two Points of a Completely Nonholonomic Space by Admissible Curve[J]. Uch. Zapiski Ped. Inst. Libknechta, 1938, 2: 83-94.
- [174] MONTGOMERY R. A Tour of Subriemannian Geometries, Their Geodesics and Applications: vol. 91[M]. Providence, RI: American Mathematical Soc., 2002.
- [175] LE DONNE E. Lecture Notes on Sub-Riemannian Geometry[Z]. Centro di Ricerca Matematica

- Ennio de Giorgi, Scuola Normale Superiore. Available online: <https://cvgmt.sns.it/paper/5339/>. 2023.
- [176] DO CARMO M P, FLAHERTY FRANCIS J. Riemannian Geometry: vol. 2[M]. Springer, 1992.
- [177] POSNER E. Random Coding Strategies for Minimum Entropy[J]. IEEE Trans. Inf. Theory, 1975, 21(4): 388-391.
- [178] POLYANSKIY Y, WU Y. Information Theory: From Coding to Learning[M]. Cambridge University Press, 2025.
- [179] GENG Y, NAIR C. The Capacity Region of the Two-Receiver Gaussian Vector Broadcast Channel With Private and Common Messages[J/OL]. IEEE Trans. Inf. Theory, 2014, 60(4): 2087-2104. DOI: 10.1109/TIT.2014.2304457.
- [180] WANG J, CHEN J. A Monotone Path Proof of an Extremal Result for Long Markov Chains [J/OL]. Entropy, 2019, 21(3). <https://www.mdpi.com/1099-4300/21/3/276>. DOI: 10.3390/e21030276.
- [181] XU Y, CHEN G, CHEN J, et al. New Proofs of Gaussian Extremal Inequalities With Applications [J]. IEEE Trans. Inf. Theory, 2024, 70(5): 3082-3099.
- [182] WANG G. A Higher-Order Otto Calculus Approach to the Gaussian Completely Monotone Conjecture[J/OL]. IEEE Transactions on Information Theory, 2025, 71(12): 9135-9162. DOI: 10.1109/TIT.2025.3565173.
- [183] ALTSCHULER J M, CHEWI S, ZHANG M S. Shifted Composition IV: Underdamped Langevin and Numerical Discretizations with Partial Acceleration[J]. arXiv:2506.23062, 2025.
- [184] VILLANI C. Hypocoercivity: vol. 202[M]. American Mathematical Society, 2009.
- [185] ZHANG K S, PEYRÉ G, FADILI J, et al. Wasserstein Control of Mirror Langevin Monte Carlo [C]//Conference on learning theory. 2020: 3814-3841.
- [186] KOROTIN A, SELIKHANOYCH D, BURNAEV E. Neural Optimal Transport[J]. arXiv:2201.12220, 2022.
- [187] YANG Y, ECKSTEIN S, NUTZ M, et al. Estimating the Rate-distortion Function by Wasserstein Gradient Descent[J]. Advances in Neural Information Processing Systems, 2023, 36: 2768-2794.
- [188] LEI E, HASSANI H, BIDOKHTI S S. Neural Estimation of the Rate-distortion Function with Applications to Operational Source Coding[J]. IEEE Journal on Selected Areas in Information Theory, 2023, 3(4): 674-686.
- [189] CHARALAMBOUS C D, STAVROU P A, AHMED N U. Nonanticipative Rate Distortion Function and Relations to Filtering Theory[J]. IEEE Transactions on Automatic Control, 2014, 59(4): 937-952.



## 致 谢

光阴荏苒，忽焉六载有半；岁月如流，倏尔两度春秋。承恩师之提携，如沐春风化雨；赖椿萱之劬劳，若倚北斗南箕。同窗砥砺，共剪西窗之烛；挚友扶持，长歌东篱之谊。行文至此，心绪驳杂。谨以此篇，赠谢所有为我点亮明灯的人。限于篇幅，恐无法完全提及，望见谅！

感谢我的导师王嘉教授，在学术研究和生活以及博士申请方面给予我全面的指导和无私的帮助。王老师是一位卓越的学者，他严谨的治学态度和渊博的知识令我敬佩不已，他独特的个人魅力与处世之风让人“欲罢不能”：我们始终好奇，如何才能做到如此般纯粹。我从本科阶段加入王老师的小组，近五年来他的指导令我受益匪浅，不仅为我指明了数学的学习道路，更帮助我打牢了数理基础。尽管王老师非常符合大众对大学教授的刻板印象——生活朴素、醉心学术、作息规律、知识渊博，但他完全不是一个古板的人。他在学术上保持一丝不苟，生活中则亦师亦友、热心慷慨。我可以随时去他的办公室畅谈，从喝咖啡聊到咖啡拉花技巧<sup>①</sup>再到庞加莱猜想的“咖啡杯”。他的办公室里堆满了外文专著，每次进去都感觉空气中弥漫着知识的微粒，随着布朗运动撞击在我的脑袋上，我想待久了大概真的会变聪明！此外，王老师给予了我完全自由的科研环境，没有强制的打卡和组会，任由我恣意妄为、野蛮生长。有任何想法只需去他办公室讨论，这逐渐提升了我的独立研究能力，也让我萌生了进一步深造的想法。王老师的教诲和关怀，我将永远铭记。

感谢曾经的 414 实验室<sup>②</sup>的伙伴们，虽然实验室地方不大，只有 8 个工位，但实验室温馨的氛围和大家的热心帮助让我感到无比幸运。感谢实验室的万犇万三牛师兄、郑天翊师兄、熊旭锦师兄、曾伟力师兄等在科研与生活上的帮助与支持。感谢实验室的同门与师弟们：曾卓、范路遥、高佳阳，与大家共同成长的日子历历在目。同时，我要感谢我的本科室友与硕士室友：王钧溥、郑宇航、赵泽、刘鹤洋等同学，感谢你们在生活和学习上给予的支持与陪伴，与你们的相识是我人生中的一大幸事，我仍怀念与大家一同在庭院里与昆虫搏斗的日子、在一战战场上一同开飞艇的日子、

① 王嘉教授目前掌握了爱心、天鹅等图案的拉花技巧，以下为聊天内容的回忆，可能有误，需要大家试验考证：首先牛奶要选取蛋白质含量  $\geq 3.6g$  的，否则奶泡基本打不出来；其次拉花牛奶杯需要选择杯嘴扁一些的，像鸭子嘴那样，拉花更容易；最后就是大量练习，初学者建议使用过期牛奶，因为练习的过程会浪费掉大量的咖啡和牛奶。

② 在我研二阶段，实验室搬到了三楼

一同当瓦学弟的日子、一同在三角洲征战 or 捡垃圾的日子<sup>①</sup>。此外我还要感谢我的高中同学以及老乡朋友们：修彦民、邹德邻 derlin、迟程、沈陆、于春桐、陈实立石粒、葛峪海海哥、侯力广侯哥、孙天一♡赵方一、刘沅鹭鹭神、王湜岚岚神、周志楷楷神、于浩宁宁神、牟迪木神、张城悦 e 君、孙启文 with 张瀚文文神等，感谢你们一路以来的陪伴与支持，尤其是作为同乡兼饭搭子的你们，让我在异乡求学的日子里感受到家的温暖，祝各位前程似锦、广结善缘。在科研以及博士申请方面，我特别感谢谭宇、忻岳对我的鼎力支持与帮助、与他们的交流让我受益匪浅，极大促进了我的脑细胞生长，尤其感谢忻岳同学对我英语口语的帮助、每天都会与我一起练习英语对话，让我从最初的结结巴巴到现在可以流利地说出几句完整的英语句子，尽管我仍然是英语苦手，但已经有了很大进步。希望各位在未来的事业或学术道路上都能一帆风顺，苟富贵勿相忘！

感谢 Prof. Jun Chen in McMaster University, Prof. Andre Wibisono and Prof. Sinho Chewi<sup>②</sup> in Yale University, 以及目前就读于帝国理工学院的沈毅荣博士，在我申请博士期间给予的宝贵建议与帮助。感谢 Prof. Wibisono, Prof. Chewi 在百忙之中抽出时间与我进行线上交流，耐心解答了我诸多问题，并为我提供了许多有价值的参考文献，极大地拓宽了我的学术视野。尤其感谢 Prof. Chen, Prof. Chewi 以及我的导师王嘉教授愿意抽出时间帮我写推荐信，尽管我申请的学校数量非常多，但他们依然慷慨地答应了我的请求，并且 Prof. Chewi 提出 “You can freely apply to as many programs as you want”。他们的支持和鼓励对我意义重大，衷心感谢大家的慷慨帮助与支持。

感谢我的女朋友潘彦茜女士，从本科毕设致谢陪伴我到硕士毕设致谢，一路走来风雨同舟。感谢你在我低谷时给予的鼓励与支持，在我迷茫时给予的指引与陪伴。你的理解与包容让我能够专注于学业与科研，你的智慧与坚韧激励着我不断前行。

君のためにつばさになる君を守りつづける  
我想做你身后的羽翼，许你一世护持  
やわらかく君をつつむあの風になる  
亦想化作那温柔的晚风，与你缱绻相依  
あの日あの時あの場所で君に会えなかったら  
倘若那日未曾在那处与你惊鸿一瞥

① 这里提到的游戏有：《禁闭求生 (Grounded)》、《战地 1 (BATTLEFIELD 1)》、《Valorant》以及《三角洲行动》，其中《禁闭求生 (Grounded)》最近出了第二部，游戏内容完整度及可玩性都大幅提升

② 此处附上 Prof. Chewi 的个人主页，他从本科开始用 L<sup>A</sup>T<sub>E</sub>X 做了大量的笔记，包含了多门课程，非常适合学生阅读参考。

僕等はいつまでも見知らぬ二人のまま

这茫茫人海中，你我大概终将擦肩而过，不复相知。

——小田和正《ラブ・ストーリーは突然に》

——小田和正《突如其来的爱情》

衷心感谢我的父母和我的所有家人们，正因为有你们无私的支持与鼓励，我才能专心致志地投入到学业与科研中，才能走到今天这一步。父母的养育之恩，终生难报；家人的关怀备至，铭记于心。

我从未想过硕士两年半的时光如此转瞬即逝，总是不自觉回到本科毕业的那个暑假，回到硕士入学前的那一晚，即使是待过四年非常熟悉的环境也变得陌生、变得让人彷徨无措，无法预测硕士生活会是怎样的一番景象，无法预测自己将面对何种挑战，自己又能有何进步或倒退。如今硕士生活即将结束，面临的不确定性比以往大得多，我不清楚会在哪里读博、不清楚博士期间由于各种动荡因素会诱发什么问题、不清楚未来会从事何种工作。但无论未来如何变化，我都将铭记这段时光，铭记那些帮助过我的人，以及他们教会我的一切。事已至此，没有大罗洞观的能力，空想未来也无济于事，手头还有不少要紧事没做完，吃点小熊饼干<sup>①</sup>继续干活吧。

此去云程发轫，岂敢忘饮水思源？是以捫管含毫，聊申拳拳之悃愫。

① 此处指香港的“珍妮曲奇聪明小熊”饼干，门店位于香港市尖沙咀。这款饼干由华裔 Jenny 女士在 2005 年于香港创办，店主因喜欢收藏泰迪熊，所以将饼干包装设计成了小熊图案，且每个月的图案还会更新，因其口味出众而成名，如今已成为香港旅游的必买伴手礼之一。笔者尤其喜欢咖啡味的曲奇，奶香、黄油香、咖啡香浓郁而缠绵，口感绵密松软，甜度适中。注：目前店铺只收取现金（港币），不能使用支付宝、微信或八达通，希望后续可以更新支付方式。另：此处并非广告，笔者未收取任何用于广告宣传的推广费用。



## 学术论文和科研成果目录

### 发表学术论文

- [1] **Jiayang Zou**, Luyao Fan, Jiayang Gao, Jia Wang. Convexity of Mutual Information along the Fokker-Planck Flow. IEEE International Symposium on Information Theory (**ISIT**), 2025.
- [2] Luyao Fan, **Jiayang Zou**, Jiayang Gao, Jia Wang. Differential Properties of Information in Jump-diffusion Channels. IEEE International Symposium on Information Theory (**ISIT**), 2025.
- [3] Tianyi Zheng, **Jiayang Zou**, Peng-Tao Jiang, Hao Zhang, Jinwei Chen, Jia Wang, Bo Li. Bidirectional Beta-tuned Diffusion Model. IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2025.

### 投稿学术论文

- [1] **Jiayang Zou**, Luyao Fan, Jiayang Gao, Tianyi Zheng and Jia Wang. On Strong Data Processing Inequalities of Generalized Relative Fisher Information. *Submitted to IEEE Transactions on Information Theory (TIT). Under Review.*
- [2] **Jiayang Zou**, Luyao Fan, Jiayang Gao, and Jia Wang. Rate Distortion Theory for General Sources: Existence of Optimal Reconstruction via the Concentration-Compactness Principle. *Manuscript in preparation for submission to IEEE ISIT 2026.*
- [3] Jiayang Gao\*, Tianyi Zheng\*, **Jiayang Zou**, Fengxiang Yang, Shice Liu, Luyao Fan, Zheyu Zhang, Hao Zhang, Jinwei Chen, Peng-Tao Jiang, Bo Li, Jia Wang. C<sup>2</sup>FG: Control Classifier-Free Guidance via Score Discrepancy Analysis. *Submitted to The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2026. Under Review.*

### 预印本

- [1] **Jiayang Zou**, Luyao Fan, Jiayang Gao, Jia Wang. A Revisit to Rate-distortion Theory via Optimal Weak Transport. *Arxiv:2501.09362, 2025.*