arXiv:2301.01339v1 [math.NA] 3 Jan 2023

Oja 在线主成分分析的扩散近似

JIAN-GUO LIU 和 Zibu Liu

Abstract. Oja 的主成分分析（PCA）算法是实践中用于降维的方法之一。本文关注离散算法的收敛性质。为此，我们将该算法视为参数空间上的一个随机过程和半群，并用随机微分方程进行近似，再证明随机微分方程的大时间收敛性，以确保算法的性能。整个过程分为三个步骤。首先，离散算法可以视为一个半群： $S^k\varphi = \mathbb{E}[\varphi(\mathrm{W}(k))]$。其次，我们在斯蒂费尔流形上构造随机微分方程（SDE），即扩散近似，用以近似该半群。通过证明弱收敛，我们验证该算法与随机微分方程"接近"。最后，我们利用随机微分方程的可逆性证明其长时间收敛性。

1. Introduction

主成分分析（PCA）是降维中的一种基本工具。由于数据的爆炸式增长，对高效 PCA 算法的需求日益增加。本文聚焦于 Oja 在 [14], 提出的在线主成分分析算法，该算法也称为随机梯度上升（SGA）方法。

设 $\mathbf{x} \in \mathbb{R}^n$ 是一个均值为零的随机变量（R.V.）。令

$$\mathbf{A} := \mathbb{E}[\mathbf{x}\mathbf{x}^T] \tag{1.1}$$

为协方差矩阵。传统的 PCA 算法将 A 对角化，以求得 A 的主特征向量（即主成分）。然而，由于存储空间的限制，以及近年来深度学习等领域中数据的高维性，稠密矩阵 A 的显式形式可能无法获得。因此，实际应用者更倾向于采用"在线"算法：该算法在每次迭代中只需要 x 的有限数量样本。为解决这一问题，Oja 在 [14]提出了如下 SGA 方法:

$$\begin{aligned}\mathbf{w_j}(k) = \mathbf{w_j}(k-1) &+ \eta(k)\mathbf{x}^T(k)\mathbf{w_j}(k-1)[\mathbf{x}(k) - (\mathbf{x}^T(k)\mathbf{w_j}(k-1))\mathbf{w_j}(k-1) \\ &- 2\sum_{i=1}^{j-1}(\mathbf{x}^T(k)\mathbf{w_i}(k-1))\mathbf{w_i}(k-1)],\ j = 1,\ 2,\ ...,\ p.\end{aligned} \tag{1.2}$$

该算法迭代前 p 个主成分 $\mathrm{w_j} \in \mathbb{R}^n, j = 1$、2、...、$p$。其中，$\mathrm{x}(k)$ $k = 1$、2、... 是 x 的独立样本，$\eta(k)$ $k = 1$、2、... 是学习率。

算法 (1.2) 确定了一个离散时间马尔可夫过程，即 $\mathrm{W}(k) = [\mathrm{w_1}(k),\ \mathrm{w_2}(k)\ .....,\ \mathrm{w_P}(k)]$。本文的主要目标是从半群、扩散近似和随机微分方程的角度，深入理解这一随机过程（R.P.）。

关键词和短语：机器学习、降维、在线主成分分析、梯度流、随机微分方程、随机矩阵。

首先，如 $\eta \to 0$，在 (1.2) 中用 A 替换 $\mathbf{x}\mathbf{x}^T$，我们推导出相应的常微分方程：

$$\begin{cases} \dot{\mathbf{q_1}} = \mathbf{A}\mathbf{q_1} - (\mathbf{q_1} \cdot \mathbf{A}\mathbf{q_1})\mathbf{q_1}, \\ \dot{\mathbf{q_j}} = \mathbf{A}\mathbf{q_j} - (\mathbf{q_j} \cdot \mathbf{A}\mathbf{q_j})\mathbf{q_j} - 2\sum_{i=1}^{j-1}(\mathbf{q_i} \cdot \mathbf{A}\mathbf{q_j})\mathbf{q_i},\ j = 2,\ 3,\ ...,\ n. \\ \mathbf{q_i}(0) = \mathbf{q_{i,0}},\ i = 1,\ 2,\ ...,\ p. \end{cases} \tag{1.3}$$

我们在先前的工作 [10] 中深入研究了收敛性质，包括全局收敛、稳定流形和指数收敛。具体而言，我们证明，对于几乎所有初始值 $\mathbf{Q_0} \in O(n)$，解都会指数收敛到特征向量基（相差一个符号）。此外，特征向量按特征值降序排列。参见 [10] 中的定理 5.2。据我们所知，这是首个针对 PCA 流给出全局指数收敛性及稳定流形闭式公式的完整结果 [1]。

鉴于相应常微分方程的收敛性，本文旨在证明离散算法 (1.2) 具有类似的结果。我们分三个步骤研究这一问题：将 SGA 迭代视为一个半群，构造适当的扩散近似，并证明扩散近似的收敛性，以确保算法的性能。

首先，我们将 SGA 方法视为一个半群。它可以改写为以下形式：

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta \cdot \mathbf{G}(\mathbf{x}(k+1)\mathbf{x}^T(k+1), \mathbf{W}(k)).$$

这里，$\mathbf{G} \in \mathbb{R}^{n\times p}$ 定义于 (2.1) 中。对于任意测试函数 $\varphi \in C(\mathbb{R}^{n\times p})$，定义

$$S\varphi(\mathbf{W}) := \mathbb{E}\varphi(\mathbf{W} + \eta\mathbf{G}(\mathbf{x}\mathbf{x}^T, \mathbf{W})). \tag{1.4}$$

在此记号下，若 SGA 方法的初始数据为 $\mathbf{W_0}$，，则马尔可夫性质给出

$$S^k\varphi(\mathbf{W}_0) = \mathbb{E}\varphi(\mathbf{W}(k)). \tag{1.5}$$

因此，SGA 方法的收敛也可以解释为半群 $\{S^k\}, k = 1, 2, ...$ 的收敛。

其次，我们构造适当的扩散近似。尽管 SGA 方法并不保持 $\mathbf{W}(k)$ 始终位于斯蒂费尔流形 $O(n \times p)$ 上，但我们所期望的结果，即特征向量基，位于 $O(n \times p)$ 中。因此，我们旨在推导半群 S 与 SGA 方法的良好扩散近似。它应当是一个始终位于斯蒂费尔流形上的随机微分方程。在某个流形上推导随机微分方程的经典方法，是将 Stratonovich 随机微分方程投影到目标流形 [5] 上：

$$\mathrm{d}\mathbf{W} = \mathcal{P}_{T_{\mathbf{W}}\mathcal{M}}(\mathbf{G}(\mathbf{A}, \mathbf{W})\mathrm{d}t + \boldsymbol{\sigma} \circ \mathrm{d}\mathbf{B}).$$

这里，$\mathcal{P}_{T_{\mathbf{Z}}\mathcal{M}}$ 是在 $\mathcal{M}$ 上 W 处切空间上的投影算子。如果半群接近扩散过程，那么通过以某种意义证明扩散过程的收敛，我们也能保证算法的性能。

最后，我们证明该扩散过程的收敛性。证明方法是寻求“可逆性”。事实上，如果随机微分方程的福克-普朗克方程可以改写为以下形式：

$$\partial_t\rho = \nabla \cdot (\rho\nabla U + \nabla\rho) = \nabla \cdot \left(e^{-U}\nabla\left(\frac{\rho}{e^U}\right)\right)$$

对于某个势函数 U，扩散过程满足细致平衡条件，即该过程具有可逆性。随后，庞加莱不等式可以确保其指数收敛

在某种 L^2 意义下。这证明了随机微分方程的收敛性，也完成了我们的分析。

在这一分析框架下，我们将给出主要结果，并修正以往文献中的结论。

1.1. 以往结果与尚未解决的问题。(1.2) 的一个重要特性是其他算法所不具备的半解耦性：$\mathbf{w_j}$ 的迭代不依赖于 $\mathbf{w_i}$, $i > j$。这一特性便于在神经网络 [13],中实现，因此研究者对此颇为关注。这一特性也被推广到了相应的常微分方程，即 (2.5)。基于这种半解耦性质，我们还在 [10]中证明了 (2.5) 的所有收敛结果。

然而，关于 (1.2) 的令人满意的收敛结果仍然缺失。自 Oja 和 Karhunen 在 [14], 中提出 (1.2) 以来，其收敛行为一直是在线 PCA 分析的重点。Oja 和 Karhunen 使用随机逼近，在关于 $\mathbf{x}$ 分布的一个隐含条件下，推导出了 (1.2) 几乎必然收敛的结论 [14, 12, 13]。这一隐含条件要求迭代过程无限多次地访问一个包含平衡点的紧集。然而，这一条件在实践中难以验证。

为改进 Oja 的结果，近期 [7] 的作者推导出 (1.2) 的第一个分量弱收敛到一个多维奥恩斯坦-乌伦贝克过程。沿着 [7], 的思路，他们考虑了执行完整正交归一化的算法，并推导出所有分量的弱收敛 [9]。

我们先前的工作 [3] 也研究了 (1.2) 第一个分量的扩散近似，并推导出一阶和二阶近似。作为推论，该工作验证了 SGA 方法第一个分量的弱收敛。然而，整个 SGA 迭代方法 (1.2) 的扩散近似仍是一个开放问题。

[3] 使用的主要工具是 Lax 等价定理。Milstein [11], 提出了一种用于证明收敛性的随机分析替代方法，该方法被用于推导随机梯度下降（SGD）方法的扩散近似 [8]。

1.2. 主要结果。首先，我们研究了半群 S^k, $k = 1$,2...,. 的性质。特别是，我们证明了它的稳定性和正则性。关于稳定性，我们证明，对于固定的终止时间 T，若 $\|\mathbf{W_0}\|_F \leq r$，则存在常数 C 和 η_0，它们取决于 r、T 以及 $\mathbf{x}$ 的分布，使得

$$\|\mathbf{W}(k)\|_F \leq C$$

对所有 $k = 1, 2, ..., \left[\frac{T}{\eta}\right]$ 均成立。这里，$\eta \in (0, \eta_0)$ 是 (1.2) 中的学习率。参见引理 3.1。

我们证明了稳定性，因为它是应用 Lax 等价定理的必要条件。关于正则性，我们证明 $S^k\varphi$, $k = 1, 2, ...$ 具有与 φ 相同阶的正则性，即

$$\|S^k\varphi\|_{C^m(B(\mathbf{0},r))} \leq C\|\varphi\|_{C^m(B(\mathbf{0},r'))}.$$

这里，C 和 r' 是依赖于 m, r, T 以及 $\mathbf{x}$ 的分布的常数。详情参见定理 3.1。

其次，我们构造了所需的扩散近似。我们证明，以下随机微分方程族

$$\dot{\mathbf{W}} = \mathbf{G}(\mathbf{A}, \mathbf{W}) + \eta\mathcal{P}_{T_{\mathbf{W}}O(n)}\mathbf{F}(\mathbf{W}) + \sqrt{\eta}\mathcal{P}_{T_{\mathbf{W}}O(n)}\dot{\mathbf{Z}},\ \mathbf{W}(0) = \mathbf{W_0} \in O(n)$$

对于所有 $t>0$ 都将保持在斯蒂费尔流形上。参见引理 3.5。这里

$$\dot{\mathbf{Z}}=(\dot{Z}_{ij})_{n\times n},\ \dot{Z}_{ij}(\mathbf{W})=H_{ijkl}(\mathbf{W})\circ\dot{B}_{kl},$$

其中 $\mathbf{H}=(H_{ijkl})_{n\times n\times n\times n}$ 是系数，B_{kl} 是白噪声。详情请参见 (3.13)。

事实上，随机微分方程 (3.13) 是 SGA 方法的一阶扩散近似。我们证明，在测试函数 φ 满足适当正则性条件的情况下，存在常数 $C_1=C_1(\mathbf{x}\ T,\eta)$，使得

$$\sup_{\mathbf{W}\in O(n),k\eta\le T}|S^k\varphi(\mathbf{W})-u(\mathbf{W},k\eta)|\le C_1(\mathbf{x},T,\varphi)\eta.$$

这里，$u(\mathbf{W}\ t)$ 是由 (3.13) 确定的柯尔莫哥洛夫方程的解，其初始值为 φ。参见定理 3.2。证明的主要思路源自 Lax 等价定理 [6]：稳定性和相容性等价于收敛性。相容性由泰勒展开保证，详情请参见第 6 节。

一个自然的问题是，是否存在更高阶近似。遗憾的是，答案是否定的。我们证明，可能的二阶近似（它是一个随机微分方程）并不保持在斯蒂费尔流形上。参见引理 3.6。这种不稳定性可能是由于 SGA 算法中省略了高阶项：在进行格拉姆-施密特正交化时，(1.2) 中忽略了二阶及更高阶（关于 η）的项。

最后，对于两种特殊情形，我们证明了随机微分方程的指数收敛。正如前文所述，我们通过寻求可逆性来证明指数收敛。

首先，我们考虑斯蒂费尔流形上的过阻尼朗之万方程：

$$\mathrm{d}\mathbf{Q}(t)=\mathcal{P}_{T_{\mathbf{Q}}O(n)}\circ(-\nabla U(\mathbf{Q})\mathrm{d}t+\sigma\mathrm{d}\mathbf{W}(t)).$$

其指数收敛性已在第 3 节中得到证明。若选择势函数 U 为加权瑞利商（见 [10]），并令 $\sigma=0$，则可恢复 Oja-Brockett 流 [2]。需要强调的是，过阻尼朗之万方程不是 (3.13) 的特例，因为 (1.2) 中的 G 并非某个势函数的梯度。

其次，对于 (3.13) 中的 $n=2$，随机微分方程可改写为

$$\mathrm{d}\mathbf{W}=\mathbf{F_1}(\mathbf{W})\mathrm{d}t+\sqrt{\eta}\cdot c(\mathbf{W})\mathbf{W}\circ\mathrm{d}\mathbf{Z}.$$

这里，$\mathbf{F_1}$ 定义于 (2.5)，而 $c(\mathbf{W})$ 是一个标量。详情见 (4.11)。该情形的指数收敛性已在定理 4.1 中得到证明。主要方法是考察 $\mathbf{W}\in O(2)$ 的旋转角的动力学，该动力学是一个一维随机微分方程，且可逆性自然成立。

2. 前提

首先，我们用矩阵重写 (1.2)。对于 $\mathbf{\Lambda}$，$\mathbf{Q}\in\mathbb{R}^{n\times n}$，定义

$$\begin{aligned}\mathbf{\Sigma}(\mathbf{\Lambda},\mathbf{Q})&:=\sum_{j=1}^{n}\sum_{k=1}^{j-1}\mathbf{E_j}\mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}\mathbf{E_k}-\mathbf{E_k}\mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}\mathbf{E_j},\\ \mathbf{G}(\mathbf{\Lambda},\mathbf{Q})&:=\mathbf{\Lambda}\mathbf{Q}-\mathbf{Q}\mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}+\mathbf{Q}\mathbf{\Sigma}(\mathbf{\Lambda},\mathbf{Q}).\end{aligned}\tag{2.1}$$

那么，(1.2) 也可写作

$$\begin{cases}\mathbf{A}(k)=\mathbf{x}(k)\mathbf{x}^T(k),\\ \mathbf{W}(k)=\mathbf{W}(k-1)+\eta_k\mathbf{G}(\mathbf{A}(k),\mathbf{W}(k-1)).\end{cases}\tag{2.2}$$

在我们前一篇论文 [10], 中，我们深入研究了相应的常微分方程，可将其写为:

$$\begin{cases} \dot{\mathbf{Q}} = \mathbf{Q}\sum_{j=1}^{n}\sum_{k=1}^{j-1}(\mathbf{E_j}\mathbf{Q}^T\mathbf{A}\mathbf{Q}\mathbf{E_k} - \mathbf{E_k}\mathbf{Q}^T\mathbf{A}\mathbf{Q}\mathbf{E_j}), \\ \mathbf{Q}(0) = \mathbf{Q_0} \in O(n). \end{cases} \tag{2.3}$$

我们定义

$$\mathbf{F_1}(\mathbf{Q}) := \mathbf{Q}\mathbf{\Sigma}(\mathbf{A}, \mathbf{Q}). \tag{2.4}$$

因此，可以将 (2.3) 改写为

$$\begin{cases} \dot{\mathbf{Q}} = \mathbf{F_1}(\mathbf{Q}), \\ \mathbf{Q}(0) = \mathbf{Q_0} \in O(n). \end{cases} \tag{2.5}$$

从现在起，我们将在所有证明中使用 (2.5)。

2.1. 记号与假设。我们将沿用我们上一篇论文 [10]中的记号约定。

我们假设 ν 是紧致的，即存在一个常数 $M > 0$ 使得

$$\|\mathbf{x}\|_2 \leq M. \tag{2.6}$$

在接下来的章节中，我们将同时采用式（2.5）的矩阵表示和分量表示，因此在此说明相关记号。$\mathbf{q_i}$, $i = 1, 2, ..., n$ 依次表示 Q 的列向量，即

$$\mathbf{Q} = [\mathbf{q_1},\ \mathbf{q_2},\ ...,\ \mathbf{q_n}], \tag{2.7}$$

而 $\tilde{\mathbf{q}}_\mathbf{i}$, $i = 1, 2, ..., n$ 依次表示 Q 的行向量，即

$$\mathbf{Q}^T = [\tilde{\mathbf{q_1}}^T,\ \tilde{\mathbf{q_2}}^T,\ ...,\ \tilde{\mathbf{q_n}}^T]. \tag{2.8}$$

对于每个元素，$q_{i,\ j}$, $i, j = 1, 2, ..., n$ 表示矩阵 Q 中第 i 行、第 j 列的元素，即

$$\mathbf{q_j} = (q_{1,j},\ q_{2,j},\ ...,\ q_{n,j})^T. \tag{2.9}$$

$\mathbb{R}^n$ 中的标准正交基记为 $\mathbf{e_i}$, $j = 1, 2, ..., n$，它们以列向量形式写出，即

$$\mathbf{I_n} = [\mathbf{e_1},\ \mathbf{e_2},\ ...,\ \mathbf{e_n}]. \tag{2.10}$$

这里 $\mathbf{I_n}$ 是 n 阶单位矩阵。

对于 $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{n\times n}$, $\|\mathbf{M}\|_F$ 表示 M 的弗罗贝尼乌斯范数，而 $\langle \mathbf{M},\ \mathbf{N}\rangle_F$ 表示弗罗贝尼乌斯意义下的内积:

$$\|\mathbf{M}\| = \sqrt{\mathrm{tr}(\mathbf{M}\mathbf{M}^T)},\ \langle \mathbf{M},\ \mathbf{N}\rangle_F = \mathrm{tr}(\mathbf{M}^T\mathbf{N}). \tag{2.11}$$

对于 $\mathbf{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$，$\|x\|_2$ 表示 x 的 ℓ_2 范数，即

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{j=1}^{n}|x_j|^2} \tag{2.12}$$

假设 A 的特征值均为单特征值，即重数均为 1。将它们记作

$$\lambda_1 > \lambda_2 > ... > \lambda_n > 0 \tag{2.13}$$

并按降序排列。不失一般性，假设 A 为对角矩阵：

$$\mathbf{A} = \mathrm{diag}\{\lambda_1, \lambda_2, ..., \lambda_n\}.$$

默认情况下，省略的引理证明以及其他重要但复杂的计算可在第 6 节中查阅。

3. Diffusion approximation of the online PCA algorithm

本节中，我们考虑迭代方案 (1.2)。我们还假设学习率为常数，即 $\eta_k = \eta > 0, k = 1$, 2, ... 在这些假设下，我们推导出 (1.2) 的扩散近似。我们的主要结果表明，当 η 趋近于 0 时，(2.5) 是 (1.2) 的弱极限。我们还推导出一系列矩阵值随机微分方程（在 Steifel 流形上保持不变），可作为一阶弱近似。特别地，取时间步长 $\eta = 0$ 时，(2.5) 可视为这些随机微分方程的一个特例。

3.1. 半群。式 (1.2) 中定义的矩阵值离散时间马尔可夫过程，即 W(k) $k = 1$, 2, ...，是时间齐次的，因为 x(k) 具有相同的分布。

沿用 [3], 中的记号，我们将此马尔可夫链从 $\mathbf{W_0}$ 出发时的分布下的期望记为 $\mathbb{E}_{\mathbf{W_0}}$。在我们的讨论中，假设 $\mathbf{W_0}$ 是确定性的，尽管在一般情形下它也可以是随机变量。将 W(k)（从 $\mathbf{W_0}$ 出发）的分布记为 $\mu^k(\cdot; \mathbf{W_0})$，并将转移概率记为 $\mu(\mathbf{V}\ \cdot)$。于是，根据马尔可夫性质，对于任意博雷尔集 $E \subset \mathbb{R}^{n\times n}$,

$$\mu^{k+1}(E; \mathbf{W_0}) = \int_{\mathbb{R}^{n\times n}} \mu(\mathbf{V}, E)\mu^k(\mathrm{d}\mathbf{V}; \mathbf{W_0}) = \int_{\mathbb{R}^{n\times n}} \mu^k(E; \mathbf{U})\mu(\mathbf{W_0}, \mathrm{d}\mathbf{U}).$$

对于固定的测试函数 $\varphi \in L^\infty(\mathbb{R}^{n\times n})$，定义

$$u^k(\mathbf{W_0}) = \mathbb{E}_{\mathbf{W_0}}\left[\varphi(\mathbf{W}(k))\right] = \int_{\mathbb{R}^{n\times n}} \varphi(\mathbf{V})\mu^k(\mathrm{d}\mathbf{V}; \mathbf{W_0}),\ k = 0, 1, 2, ... \tag{3.1}$$

这里，W(k) 在 (1.2) 中定义。由马尔可夫性质可得

$$\begin{aligned} u^{k+1}(\mathbf{W_0}) &= \mathbb{E}_{\mathbf{W_0}}[\mathbb{E}_{\mathbf{W_0}}[\varphi(\mathbf{W}(k+1))|\mathbf{W}(1)]] \\ &= \mathbb{E}_{\mathbf{W_0}}[u^k(\mathbf{W}(1))] \\ &= \int_{\mathbb{R}^{n\times n}} \mu(\mathrm{d}\mathbf{W_1}, \mathbf{W_0}) \int_{\mathbb{R}^{n\times n}} \varphi(\mathbf{V})\mu^k(\mathrm{d}\mathbf{V}; \mathbf{W_1}). \end{aligned}$$

于是，由 (1.2) 可推得，对于任意 $\mathbf{W} \in \mathbb{R}^{n\times n}$,

$$u^{k+1}(\mathbf{W}) = \mathbb{E}u^k(\mathbf{W} + \eta\mathbf{G}(\mathbf{x}\mathbf{x}^T, \mathbf{W})) := Su^k(\mathbf{W}), \tag{3.2}$$

因此，$u^0(\mathbf{W}) = \varphi(\mathbf{W})$ 和 $\{S^k\}_{k\geq 0}$ 构成一个半群。

在讨论扩散近似之前，我们先推导马尔可夫链和半群的一些基本性质。

引理 3.1。（稳定性）固定实数 $r > 0$ 和终止时间 $T > 0$。设 W(k) k =0, 1, 2, ... 为由 (1.2) 生成的马尔可夫链，其初始数据 $\mathbf{W_0}$ 满足 $\|\mathbf{W_0}\|_F \leq r$。则存在常数 $\eta(r, M, T) > 0$ 和 $C(r, M, T) > 0$，它们依赖于 r、T 以及 M（见 (2.6)），使得对于任意 $0 < \eta \leq \eta(r, M, T)$ 和 $k = 0, 1, 2, ..., \left[\frac{T}{\eta}\right]$,

$$\mathbb{P}\left(\|\mathbf{W}(k)\|_F^2 \leq C(r, M, T)\right) = 1, \tag{3.3}$$

也就是说，对于时间步长小于$\eta(r, M, T)$的任意时间离散，W(k) 都一致有界。

关于该引理的证明，见第 6 节。基于引理 3.1，我们证明 $u^k(\mathbf{W})$ 具有与测试函数 φ 相同的正则性。测试函数的可容许集合为

$$C_b^m(\mathbb{R}^{n\times n}) := \left\{ f \in C^m(\mathbb{R}^{n\times n}) \;\middle|\; \|f\|_{C^m(\mathbb{R}^{n\times n})} := \sum_{|\alpha|\le m} |D^\alpha f|_\infty \right\},\ m = 1, 2, ... \tag{3.4}$$

我们将以 $\mathbf{M} \in \mathbb{R}^{n\times n}$ 为中心、半径为 r 的 $(\mathbb{R}^{n\times n}, \|\cdot\|_F)$ 中的开球记为 $B(\mathbf{M}, r)$，即

$$B(\mathbf{M}, r) := \{\mathbf{N} \in \mathbb{R}^{n\times n} : \|\mathbf{N} - \mathbf{M}\|_F < r\}. \tag{3.5}$$

定理 3.1。（半群的性质）固定一个实数 $r > 0$ 和一个终止时间 $T > 0$。设 $\{u^k\}_{k\ge 0}$, $\{S^k\}_{k\ge 0}$ 为 (3.2) 中定义的半群，M 为 (2.6) 中的上界。假设 $\varphi \in C_b^m(\mathbb{R}^{n\times n})$，其中 $m \ge 1$。则：

(1)（L^∞ 压缩映射）$S:\ L^\infty(\mathbb{R}^{n\times n}) \to L^\infty(\mathbb{R}^{n\times n})$ 是压缩映射；(2)（正则性）存在常数 $\eta(r, m, M, T) > 0$ 和 $C(r, m, M, T) > 0$，使得对任意 $0 < \eta \le \eta(r, m, M, T)$：

$$\|u^k\|_{C^m(B(\mathbf{0},r))} \le C(r, m, M, T)\|\varphi\|_{C^m(B(\mathbf{0},e^{T(2M^2+1)/2}r))}. \tag{3.6}$$

证明。根据引理 3.1，对于任意 $\eta \le \eta(r, M, T)$，存在 $C(r, M, T) > 0$，使得 $\|\mathbf{W}\|_F \le C(r, M, T)$ 几乎必然对 $k = 0, 1, 2, ..., [T/\eta]$ 成立。注意，$\mathbf{G}$ 关于 $\mathbf{W}$ 是一个多项式，因此存在常数 $C'(r, M, T) > 0$，使得对于任意 $\mathbf{W} \in B(\mathbf{0}, r)$ 以及指标 (i, j) 和 (i', j')，

$$\left| \frac{\partial(\mathbf{G}(\mathbf{x}(k)\mathbf{x}(k)^T, \mathbf{W}))_{i',j'}}{\partial w_{i,j}} \right| \le C'(r, M, T). \tag{3.7}$$

其中 $(\mathbf{G})_{i',j'}$ 表示矩阵 $\mathbf{G}$ 中第 i' 行、第 j' 列的元素。一个
ccord根据 t h证明 f of L引理 3.1，我们 k现在 t hat

$$\|\mathbf{W}(k)\|_F^2 \le (1 + (2M^2 + 1)\eta)\|\mathbf{W}(k-1)\|_F^2,\ k = 1, 2, ..., [T/\eta].$$

因此，如果 $\mathbf{W} \in B(\mathbf{0}, r)$，那么

$$\mathbf{W} + \eta\mathbf{G}(\mathbf{x}(k)\mathbf{x}(k)^T, \mathbf{W}) \in B(\mathbf{0}, r\sqrt{1 + (2M^2 + 1)\eta}) \subset B\left(\mathbf{0}, \left(1 + \frac{(2M^2 + 1)\eta}{2}\right) r\right). \tag{3.8}$$

记 $C'' = (2M^2 + 1)/2$。现在我们来证明该定理。

(1) 假设　若 $\psi \in L^\infty(\mathbb{R}^{n\times n})$，则由 (3.2) 可直接推出 L^∞ 压缩映射：

$$\begin{aligned} \|S\psi\|_{L^\infty(\mathbb{R}^{n\times n})} &= \sup_{\mathbf{W}\in\mathbb{R}^{n\times n}} |\mathbb{E}\psi(\mathbf{W} + \eta\mathbf{G}(\mathbf{x}\mathbf{x}^T, \mathbf{W}))| \\ &\le \sup_{\mathbf{W}\in\mathbb{R}^{n\times n}} |\psi(\mathbf{W})| \\ &\le \|\psi\|_{L^\infty(\mathbb{R}^{n\times n})}. \end{aligned} \tag{3.9}$$

(2) 我们用归纳法证明 $m = 1$ 的情形。$m \ge 1$ 的情形证明类似。由于 $u^0 = \varphi$，因此 $k = 0$ 的结论成立。现在，对于 $k \ge 0$，根据主导

收敛定理（DCT），对于任意 W $\in B(\mathbf{0},r)$ 和 $1\le i,j\le n$,

(3.10)
$$\begin{aligned}\frac{\partial u^{k+1}}{\partial w_{i,j}}(\mathbf{W}) &= \sum_{1\le i',j'\le n}\mathbb{E}\left[\frac{\partial u^k}{\partial w_{i',j'}}(\mathbf{W}+\eta\mathbf{G}(\mathbf{x}(k)\mathbf{x}(k)^T,\mathbf{W}))\cdot\frac{\partial(\mathbf{W}+\eta\mathbf{G}(\mathbf{x}(k)\mathbf{x}(k)^T,\mathbf{W}))_{i',j'}}{\partial w_{i,j}}\right]\\ &= \mathbb{E}\left[\frac{\partial u^k}{\partial w_{i,j}}(\mathbf{W}+\eta\mathbf{G}(\mathbf{x}(k)\mathbf{x}(k)^T,\mathbf{W}))\right]\\ &\quad+\eta\sum_{1\le i',j'\le n}\mathbb{E}\left[\frac{\partial u^k}{\partial w_{i',j'}}(\mathbf{W}+\eta\mathbf{G}(\mathbf{x}(k)\mathbf{x}(k)^T,\mathbf{W}))\cdot\frac{\partial(\mathbf{G}(\mathbf{x}(k)\mathbf{x}(k)^T,\mathbf{W}))_{i',j'}}{\partial w_{i,j}}\right].\end{aligned}$$

注意，(3.8) 成立，因此

$$\left|\mathbb{E}\left[\frac{\partial u^k}{\partial w_{i,j}}(\mathbf{W}+\eta\mathbf{G}(\mathbf{x}(k)\mathbf{x}(k)^T,\mathbf{W}))\right]\right|\le\left\|\frac{\partial u^k}{\partial w_{i,j}}\right\|_{L^\infty(B(\mathbf{0},(1+C''\eta)r))}$$

将其代入 (3.10)，得到

$$\begin{aligned}\left|\frac{\partial u^{k+1}}{\partial w_{i,j}}(\mathbf{W})\right| &\le \left\|\frac{\partial u^k}{\partial w_{i,j}}\right\|_{L^\infty(B(\mathbf{0},(1+C''\eta)r))}\\ &\quad+\eta C'(r,M,T)\sum_{1\le i',j'\le n}\left|\frac{\partial u^k}{\partial w_{i',j'}}(\mathbf{W}+\eta\mathbf{G}(\mathbf{x}(k)\mathbf{x}(k)^T,\mathbf{W}))\right|.\end{aligned}$$

然后利用 (3.7)，对两边取 L^∞ 范数，并对 $1\le i$、$j\le n$ 求和，可得

$$\begin{aligned}\sum_{1\le i,j\le n}\left\|\frac{\partial u^{k+1}}{\partial w_{i,j}}\right\|_{L^\infty(B(\mathbf{0},r))} &\le \sum_{1\le i,j\le n}\left\|\frac{\partial u^k}{\partial w_{i,j}}\right\|_{L^\infty(B(\mathbf{0},(1+C''\eta)r))}\\ &\quad+n^2C'(r,M,T)\eta\|u^k\|_{C^1(B(\mathbf{0},(1+C''\eta)r))}.\end{aligned}$$

由此不等式可得

$$\|u^{k+1}\|_{C^1(B(\mathbf{0},r))}\le(1+n^2C'(r,M,T)\eta)\|u^k\|_{C^1(B(\mathbf{0},(1+C''\eta)r))}. \tag{3.11}$$

通过归纳法可得

$$\begin{aligned}\|u^k\|_{C^1(B(\mathbf{0},r))} &\le(1+n^2C'(r,M,T)\eta)^k\|u^0\|_{C^1(B(\mathbf{0},(1+C''\eta)^kr))}\\ &\le(1+n^2C'(r,M,T)\eta)^{T/\eta}\|\varphi\|_{C^1(B(\mathbf{0},(1+C''\eta)^{T/\eta}r))}\\ &\le e^{n^2C'(r,M,T)T}\|\varphi\|_{C^1(B(\mathbf{0},e^{T(2M^2+1)/2}r))}.\end{aligned} \tag{3.12}$$

□

3.2. 扩散近似。本节讨论半群 (3.2) 的扩散近似。

3.2.1. 斯蒂费尔流形上的随机微分方程。由于 SGA 方法旨在求得正确的单位特征向量基，因此我们期望随机微分方程始终位于斯蒂费尔流形上。因此，我们考虑以下随机微分方程族:

$$\dot{\mathbf{W}}=\mathbf{G}(\mathbf{A},\mathbf{W})+\eta\mathcal{P}_{T_{\mathbf{W}}O(n)}\mathbf{F}(\mathbf{W})+\sqrt{\eta}\mathcal{P}_{T_{\mathbf{W}}O(n)}\dot{\mathbf{Z}},\ \mathbf{W}(0)=\mathbf{W_0}\in O(n). \tag{3.13}$$

其中 $\mathcal{P}_{T_{\mathbf{W}}O(n)}$ 表示在 W 处投影到 $O(n)$ 切空间的投影，见 (6.5)；Z 定义为

$$\dot{\mathbf{Z}}(\mathbf{W})=(\dot{Z}_{ij})_{n\times n},\ \dot{Z}_{ij}(\mathbf{W})=H_{ijkl}(\mathbf{W})\circ\dot{B}_{kl}; \tag{3.14}$$

H(W) $= (H_{ijkl}(\mathbf{W}))_{n\times n\times n\times n}$ 是布朗运动的系数张量；B $= (B_{ij})_{n\times n}$是 $\mathbb{R}^{n\times n}$ 中的标准布朗运动；符号'$\circ$'表示 (3.13) 是 Stratonovich 意义下的随机微分方程。

令 (3.13) 中的 $\eta = 0$，随机微分方程退化为常微分方程

$$\dot{\mathbf{W}} = \mathbf{G}(\mathbf{A}, \mathbf{W}),\ \mathbf{W}(0) = \mathbf{W_0} \in O(n).$$

这正是 (2.5)。因此，我们预期 (3.13) 是 SGA 方法的扩散近似。

我们首先验证，对于任意的 F(W) $\in \mathbb{R}^{n\times n}$ 和 H(W) $\in \mathbb{R}^{n\times n\times n\times n}$，若 $\mathbf{W_0} \in O(n)$，则 (3.13) 始终位于斯蒂费尔流形上。作为准备，我们先重写投影算子。

引理 3.2。设 $\mathcal{P}_{T_{\mathbf{W}}O(n)}$ 如 (6.5) 中所定义，其中某个 W $= (w_{ij})_{n\times n} \in O(n)$。令

$$\mathbf{P} := (P_{ijkl})_{n\times n\times n\times n},\ P_{ijkl} := \frac{1}{2}(w_{is}w_{ks}\delta_{jl} - w_{kj}w_{il}). \tag{3.15}$$

则有：

(i)（投影）对于任意 $\mathbf{F} = (f_{ij})_{n\times n}$,

$$(\mathcal{P}_{T_{\mathbf{W}}O(n)}\mathbf{F})_{ij} = P_{ijkl}f_{kl} = \frac{1}{2}(f_{ij} - w_{il}f_{kl}w_{kj}); \tag{3.16}$$

(ii)（对称性） $P_{ijkl} = P_{klij}$，即 P 是对称的；

(iii)（幂等性） $P_{ijkl}P_{klrs} = P_{ijrs}$，即 $\mathbf{P}^2 = \mathbf{P}$。

证明。由于 W $\in O(n)$，所以 $w_{is}w_{ks} = \delta_{ik}$ and

$$P_{ijkl} = \frac{1}{2}(\delta_{ik}\delta_{jl} - w_{kj}w_{il}). \tag{3.17}$$

我们先证明 (i)。由于 $\mathcal{P}_{T_{\mathbf{W}}O(n)}$ F $= \frac{1}{2}$(F $-$ WF^{T}W)，我们有

$$(\mathcal{P}_{T_{\mathbf{W}}O(n)}\mathbf{F})_{ij} = \frac{1}{2}(f_{ij} - w_{il}f_{kl}w_{kj}).$$

由 (3.17)，可得

$$P_{ijkl}f_{kl} = \frac{1}{2}(\delta_{ik}\delta_{jl} - w_{kj}w_{il})f_{kl} = \frac{1}{2}(f_{ij} - w_{il}f_{kl}w_{kj}).$$

Thus

$$(\mathcal{P}_{T_{\mathbf{W}}O(n)}\mathbf{F})_{ij} = P_{ijkl}f_{kl} = \frac{1}{2}(f_{ij} - w_{il}f_{kl}w_{kj}).$$

(ii) 显然成立：

$$P_{ijkl} = \frac{1}{2}(\delta_{ik}\delta_{jl} - w_{kj}w_{il}) = \frac{1}{2}(\delta_{ki}\delta_{lj} - w_{il}w_{kj}) = P_{klij}.$$

对于 (iii)，我们有

$$\begin{aligned} P_{ijkl}P_{klrs} &= \frac{1}{4}(\delta_{ik}\delta_{jl} - w_{kj}w_{il})(\delta_{kr}\delta_{ls} - w_{rl}w_{ks}) \\ &= \frac{1}{4}(\delta_{ir}\delta_{js} - w_{rj}w_{is} - w_{rj}w_{is} + w_{kj}w_{il}w_{rl}w_{ks}). \end{aligned}$$

因为 W $\in O(n)$，我们有

$$w_{il}w_{rl} = \delta_{ir}, w_{kj}w_{ks} = \delta_{js}.$$

Thus

$$P_{ijkl}P_{klrs} = \frac{1}{2}(\delta_{ir}\delta_{js} - w_{rj}w_{is}) = P_{ijrs}.$$

□

注记 3.1. 我们将 P_{ijkl} 定义为式 (3.15) 中的形式，而不是式 (3.17) 中的形式，因为在这种情况下，尽管 $\mathbf{W} \notin O(n)$，以下等式仍然成立：

$$w_{ij}P_{tjkl} + w_{tj}P_{ijkl} = 0. \tag{3.18}$$

这有助于证明斯蒂费尔流形的不变性，即引理 3.3。

现在我们可以检验斯蒂费尔流形对于 (3.13) 是否是不变的。

引理 3.3. 考虑 (3.13)。若 $\mathbf{W}_0 \in O(n)$，则对于所有 $t > 0$，$\mathbf{W}(t) \in O(n)$ 均成立。

证明。根据引理 3.2，我们可以将 (3.13) 改写为

$$\mathrm{d}w_{ij} = g_{ij}(\mathbf{A}, \mathbf{W})\mathrm{d}t + \eta P_{ijkl}(\mathbf{W})f_{kl}(\mathbf{W})\mathrm{d}t + P_{ijkl}(\mathbf{W})H_{klrs}(\mathbf{W}) \circ \mathrm{d}B_{rs}.$$

那么，对于 $1 \le i, t \le n$，请记住，该随机微分方程采用 Stratonovich 意义，我们有

$$\begin{aligned}\mathrm{d}(w_{ij}w_{tj}) &= w_{ij} \circ \mathrm{d}w_{tj} + w_{tj} \circ \mathrm{d}w_{ij} \\ &= (w_{ij}g_{tj}(\mathbf{A}, \mathbf{W}) + w_{tj}g_{ij}(\mathbf{A}, \mathbf{W}))\mathrm{d}t + (w_{ij}P_{tjkl} + w_{tj}P_{ijkl})(f_{kl}\mathrm{d}t + H_{klrs} \circ \mathrm{d}B_{rs}\end{aligned}$$

注意，

$$\begin{aligned}w_{ij}P_{tjkl} + w_{tj}P_{ijkl} &= \frac{1}{2}[w_{ij}(w_{ts}w_{ks}\delta_{jl} - w_{kj}w_{tl}) + w_{tj}(w_{is}w_{ks}\delta_{jl} - w_{kj}w_{il})] \\ &= \frac{1}{2}(w_{il}w_{ts}w_{ks} - w_{ij}w_{kj}w_{tl} + w_{is}w_{ks}w_{tl} - w_{il}w_{tj}w_{kj}) = 0.\end{aligned}$$

Thus

$$\mathrm{d}(w_{ij}w_{tj}) = (w_{ij}g_{tj}(\mathbf{A}, \mathbf{W}) + w_{tj}g_{ij}(\mathbf{A}, \mathbf{W}))\mathrm{d}t,$$

这是一个常微分方程组。我们可以将其改写为矩阵形式：

$$\begin{aligned}\frac{\mathrm{d}(\mathbf{W}\mathbf{W}^T)}{\mathrm{d}t} &= \mathbf{W}\mathbf{G}^T(\mathbf{A}, \mathbf{W}) + \mathbf{G}(\mathbf{A}, \mathbf{W})\mathbf{W}^T \\ &= \mathbf{W}\mathbf{W}^T\mathbf{A} + \mathbf{A}\mathbf{W}\mathbf{W}^T - 2\mathbf{W}\mathbf{W}^T\mathbf{A}\mathbf{W}\mathbf{W}^T.\end{aligned}$$

因此，$\mathbf{X} = \mathbf{W}\mathbf{W}^T$ 满足代数黎卡提方程：

$$\frac{\mathrm{d}\mathbf{X}}{\mathrm{d}t} = \mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A} - 2\mathbf{X}\mathbf{A}\mathbf{X},$$

初始值为 $\mathbf{X} = \mathbf{I}_n$。因此，根据 [18], 中的结果，我们知道 $\mathbf{X}(t) = \mathbf{I}_n$ 成立。因此，$\mathbf{W}(t) \in O(n)$。

□

3.2.2. 福克-普朗克方程和柯尔莫哥洛夫方程。在伊藤意义下，(3.13)可写为

$$\dot{\mathbf{W}} = \mathbf{G}(\mathbf{A}, \mathbf{W}) + \eta\mathcal{P}_{T_{\mathbf{W}}O(n)}\mathbf{F}(\mathbf{W}) + \frac{\eta}{2}\mathbf{J}(\mathbf{W}) + \sqrt{\eta}\mathcal{P}_{T_{\mathbf{W}}O(n)}\dot{\mathbf{Y}}(\mathbf{W}), \tag{3.19}$$

这里 $\mathbf{Y}$ 定义为

$$\dot{\mathbf{Y}}(\mathbf{W}) = (\dot{Y}_{ij})_{n\times n},\ \dot{Y}_{ij}(\mathbf{W}) = H_{ijkl}(\mathbf{W})\dot{B}_{kl}. \tag{3.20}$$

修正项 $\mathbf{J}$ 可表示为

$$\begin{aligned}&\mathbf{K} = (K_{ijkl})_{n\times n\times n\times n},\ K_{ijkl} = P_{ijrs}H_{rskl} \\ &\mathbf{J} = (J_{ij})_{n\times n},\ J_{ij}(\mathbf{W}) := K_{rsml}\frac{\partial K_{ijml}}{\partial w_{rs}}.\end{aligned} \tag{3.21}$$

于是，(3.19) 可改写为

$$\begin{aligned} \dot{w_{ij}} &= g_{ij}(\mathbf{A},\mathbf{W}) + \eta P_{ijkl} f_{kl} + \frac{\eta}{2} J_{ij} + \sqrt{\eta} K_{ijkl} \dot{B_{kl}} \\ &= g_{ij}(\mathbf{A},\mathbf{W}) + \eta P_{ijkl} f_{kl} + \frac{\eta}{2} J_{ij} + \sqrt{\eta} P_{ijkl} H_{klrs} \dot{B_{rs}} \end{aligned}$$

(3.19) 的对应后向柯尔莫哥洛夫方程为

$$\partial_t u(\mathbf{W},t) = \mathcal{L} u(\mathbf{W},t),\ u(\mathbf{W},0) = \varphi(\mathbf{W}), \tag{3.22}$$

其中，$\mathcal{L}$ 是定义如下的椭圆算子

$$\begin{aligned} \mathcal{L}u &:= \left(g_{ij}(\mathbf{A},\mathbf{W}) + \eta P_{ijkl} f_{kl} + \frac{\eta}{2} J_{ij} \right) \frac{\partial u}{\partial w_{ij}} + \frac{\eta}{2} K_{ijrs} K_{klrs} \frac{\partial^2 u}{\partial w_{ij} w_{kl}} \\ &= \left(g_{ij}(\mathbf{A},\mathbf{W}) + \eta P_{ijkl} f_{kl} + \frac{\eta}{2} J_{ij} \right) \frac{\partial u}{\partial w_{ij}} + \frac{\eta}{2} P_{ijzv} H_{zvrs} P_{klxy} H_{xyrs} \frac{\partial^2 u}{\partial w_{ij} w_{kl}}. \end{aligned} \tag{3.23}$$

(3.22) 的解为

$$u(\mathbf{W},t) = e^{t\mathcal{L}} \varphi(\mathbf{W}) = \mathbb{E}_{\mathbf{W}}[\varphi(\mathbf{X}(t))], \tag{3.24}$$

这里，X(t) $\in$ O(n) 是由 (3.13)（或等价地，由 (3.19)）确定的随机过程（向量化之前），其初始点为 X(0) $=$ W $\in$ O(n)。更多详情请参阅 [15]。

另一方面，前向柯尔莫哥洛夫方程（福克-普朗克方程）的解记为 ρ(W t) $= e^{t\mathcal{L}^*} \rho_0$(W)：

$$\partial_t \rho = \mathcal{L}^* \rho,\ \rho(\mathbf{W},t) = \rho_0(\mathbf{W}), \tag{3.25}$$

其中 ρ_0 是 W 的初始分布，且 $\mathcal{L}^*$ 定义为 (3.26)

$$\begin{aligned} \mathcal{L}^* \rho &:= -\frac{\partial}{\partial w_{ij}} \left[\left(g_{ij}(\mathbf{A},\mathbf{W}) + \eta P_{ijkl} f_{kl} + \frac{\eta}{2} J_{ij} \right) \rho \right] + \frac{\eta}{2} \frac{\partial^2}{\partial w_{ij} w_{kl}} \left(\rho K_{ijrs} K_{klrs} \right) \\ &= -\frac{\partial}{\partial w_{ij}} \left[\left(g_{ij}(\mathbf{A},\mathbf{W}) + \eta P_{ijkl} f_{kl} + \frac{\eta}{2} J_{ij} \right) \rho \right] + \frac{\eta}{2} \frac{\partial^2}{\partial w_{ij} w_{kl}} \left(\rho P_{ijzv} H_{zvrs} P_{klxy} H_{xyrs} \right). \end{aligned}$$

解 ρ(W t) $= e^{t\mathcal{L}^*} \rho_0$(W) 可解释为式 (3.13) 中随机过程 X 的概率分布，其初始分布为 ρ_0。

注意，$\{e^{t\mathcal{L}}\}_{t\geq 0}$ 和 $\{e^{t\mathcal{L}^*}\}_{t\geq 0}$ 构成两个半群，因此我们有以下基本性质：

引理 3.4。考虑式 (3.23) 中的 $\{e^{t\mathcal{L}}\}_{t\geq 0}$ 和式 (3.26) 中的 $\{e^{t\mathcal{L}^*}\}_{t\geq 0}$。则：

(1)（概率与正性保持）若 $\rho_0 \in L^1$(O(n)) 且 $\rho_0 \geq 0$，则 $e^{t\mathcal{L}^*} \rho_0 \geq 0$且

$$\| e^{t\mathcal{L}^*} \rho \|_{L^1(O(n))} = \| \rho_0 \|_{L^1(O(n))}. \tag{3.27}$$

特别地，$e^{t\mathcal{L}^*}$ 是从 L^1(O(n)) 到 L^1(O(n)) 的压缩映射。

(2)（L^∞ 压缩映射）假设 φ : O(n) $\to$ R 是一个连续函数。$e^{t\mathcal{L}}$:L^∞(O(n)) $\to$ L^∞(O(n)) 是一个压缩映射，即

$$\| e^{t\mathcal{L}} \varphi \|_{L^\infty(O(n))} \leq \| \varphi \|_{L^\infty(O(n))}. \tag{3.28}$$

证明。(1) 的证明可见 [3], , 这里我们只证明 (2)。由于 φ 连续且 $O(n)$ 是紧集，因此有 $\varphi \in L^\infty(O(n))$。设 $\mathbf{W_0} \in O(n)$，令 $\mathbf{W}(t)$ 为由 (3.13) 生成且初始于 $\mathbf{W_0}$ 的随机过程。那么我们有

$$e^{t\mathcal{L}}\varphi(\mathbf{W_0}) = \mathbb{E}_{\mathbf{W_0}}\left[\varphi(\mathbf{W}(t))\right]. \tag{3.29}$$

由引理 3.3 可知，对于所有 $t \geq 0$，$\mathbf{W}(t) \in O(n)$，因此

$$|e^{t\mathcal{L}}\varphi(\mathbf{W_0})| \leq \mathbb{E}_{\mathbf{W_0}}|\varphi(\mathbf{W}(t))| \leq \|\varphi\|_{L^\infty(O(n))}. \tag{3.30}$$

这对任意 $\mathbf{W_0} \in O(n)$ 都成立，因此 (3.28) 成立。 □

[16] 中对压缩映射和保正性的讨论更为详尽。

现在我们准备验证 (3.13) 是否构成 SGA 迭代的弱扩散近似。

3.2.3. 一阶扩散近似。验证 (3.13) 构成扩散近似的方法源于 Lax 等价定理的思想 [6], , 该思想最初在 [3] 中被用于同一目的。

定理 3.2。（一阶扩散近似）固定时间 $T > 0$。考虑在 (3.1) 中定义的 $\{u^k(\mathbf{W})\}_{k\geq 0}$。假设 $\varphi \in C_b^4(\mathbb{R}^{n\times n})$，且 $u(\mathbf{W}\ t)$ 是方程 (3.22) 在初始值 $u(\cdot, 0) = \varphi(\cdot)$ 下的解。那么存在 $\eta_1(M, T) > 0$ 和 $C_1(M, T, \varphi) > 0$，使得对任意 $\eta \in (0, \eta_1(M, T)]$,

$$\sup_{\mathbf{W}\in O(n), k\eta\leq T} |u^k(\mathbf{W}) - u(\mathbf{W}, k\eta)| \leq C_1(M, T, \varphi)\eta. \tag{3.31}$$

其中 M 是 (2.6) 中的常数。

证明。在下文的证明中，C 是一个一般常数，会随方程而变化。

根据定理 3.1，存在常数 $\eta(M, T)$ 和 $C(M, T)$，使得对所有 k =0, 1, 2..., $[T/\eta]$ 且 $\eta \in (0, \eta(M, T)]$,

$$\|u^k\|_{C^4(B(0,\sqrt{n+1}))} \leq C(M, T)\|\varphi\|_{C^4(\mathbb{R}^{n\times n})}. \tag{3.32}$$

然后，根据关于 η 的泰勒展开，对于任意 $\mathbf{W} \in O(n)$，我们有，

$$\begin{aligned}\left|u^{k+1}(\mathbf{W}) - u^k(\mathbf{W}) - \eta g_{ij}(\mathbf{A}, \mathbf{W})\frac{\partial u^k}{\partial w_{ij}}\right| &\leq C(M, T)\|u^k\|_{C^4(B(0,\sqrt{n+1}))}\eta^2 \\ &\leq C(M, T, \varphi)\eta^2.\end{aligned} \tag{3.33}$$

该计算的详细过程见第 6 节。

根据引理 3.4，$e^{t\mathcal{L}}$ 是从 $L^\infty(O(n))$ 到其自身的压缩映射。因此，对于任意 $\mathbf{W} \in O(n)$，存在 $\eta' \in (0, \eta]$，使得

$$\begin{aligned}|e^{\eta\mathcal{L}}u^k(\mathbf{W}) - u^k(\mathbf{W}) - \eta\mathcal{L}u^k(\mathbf{W})| &\leq \frac{\eta^2}{2}|e^{\eta'\mathcal{L}}\mathcal{L}^2u^k(\mathbf{W})| \\ &\leq \frac{\eta^2}{2}\|\mathcal{L}^2u^k\|_{L^\infty(O(n))} \\ &\leq C(M, T, \|u^k\|_{C^4(B(0,\sqrt{n+1}))})\eta^2 \\ &\leq C(M, T, \varphi)\eta^2.\end{aligned}$$

回顾 (3.23) 中 $\mathcal{L}$ 的定义，我们可以将上述估计改写为 a s

$$\left|e^{\eta\mathcal{L}}u^k(\mathbf{W}) - u^k(\mathbf{W}) - \eta g_{ij}(\mathbf{A}, \mathbf{W})\frac{\partial u^k}{\partial w_{ij}}\right| \leq C(M, T, \varphi)\eta^2 \tag{3.34}$$

有关此计算的详细信息，请参见第 6 节。

因此，由 (3.33) 和 (3.34)，我们有

$$\|u^{k+1}(\mathbf{W}) - e^{\eta\mathcal{L}}u^k(\mathbf{W})\|_{L^\infty(O(n))} \le C(M,T,\varphi)\eta^2. \tag{3.35}$$

现在，对于 $k=0$，1, 2..., $[T/\eta]$，定义

$$e_k := \|u^k(\mathbf{W}) - u(\mathbf{W},k\eta)\|_{L^\infty(O(n))},\ r_k := \|u^{k+1}(\mathbf{W}) - e^{\eta\mathcal{L}}u^k(\mathbf{W})\|_{L^\infty(O(n))}. \tag{3.36}$$

因此，对于 $k=1$，2, ..., $[T/\eta]$，有 $r_k \le C(M,T,\varphi)\eta^2$。注意到 $u(\mathbf{W},(k+1)\eta) = e^{\eta\mathcal{L}}u(\mathbf{W}\ k\eta)$，我们有

$$\begin{aligned} e_{k+1} &= \|u^{k+1}(\mathbf{W}) - u(\mathbf{W},(k+1)\eta)\|_{L^\infty(O(n))} \\ &= \|u^{k+1}(\mathbf{W}) - e^{\eta\mathcal{L}}u(\mathbf{W},k\eta)\|_{L^\infty(O(n))} \\ &\le \|u^{k+1}(\mathbf{W}) - e^{\eta\mathcal{L}}u^k(\mathbf{W})\|_{L^\infty(O(n))} + \|e^{\eta\mathcal{L}}(u(\mathbf{W},k\eta) - u^k(\mathbf{W}))\|_{L^\infty(O(n))} \\ &\le r_k + e_k. \end{aligned} \tag{3.37}$$

(3.37) 中的最后一步用到了 $e^{\eta\mathcal{L}}$ 是 $L^\infty(O(n))$ 上的压缩映射。因此，将 (3.37) 对 k 求和可得

$$e_k \le \sum_{l=1}^{k-1} r_l \le \frac{TC''(M,T,\varphi)}{\eta}\cdot\eta^2 = C_1(M,T,\varphi)\eta. \tag{3.38}$$

这对 k 一致成立，因此该结论得证。 □

3.2.4. 不稳定的二阶近似。根据定理 3.2 的证明，确保一阶近似成立的关键步骤是对截断误差进行如下估计，其阶数为二阶：

$$\|S\varphi(\mathbf{W}) - e^{\eta\mathcal{L}}\varphi(\mathbf{W})\|_{L^\infty(O(n))} \le C\eta^2.$$

这里，$\mathcal{L}$ 定义于 (3.23) 中。

为了推导二阶近似，我们需要谨慎选择 (3.13) 中的漂移项 F 和 H，使得

$$\|S\varphi(\mathbf{W}) - e^{\eta\mathcal{L}}\varphi(\mathbf{W})\|_{L^\infty(O(n))} \le C\eta^3. \tag{3.39}$$

直接计算可得

$$\begin{aligned} S\varphi(\mathbf{W}) &= \mathbb{E}\left[\varphi(\mathbf{W}+\eta\mathbf{G}(\mathbf{x}\mathbf{x}^T,\mathbf{W}))\right] \\ &= \varphi(\mathbf{W}) + \eta g_{ij}(\mathbf{A},\mathbf{W})\frac{\partial\varphi}{\partial w_{ij}} + \frac{\eta^2}{2}\mathbb{E}\left[g_{ij}(\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{kl}(\mathbf{x}\mathbf{x}^T,\mathbf{W})\right]\frac{\partial^2\varphi}{\partial w_{ij}\partial w_{kl}} + O(\eta^3). \end{aligned}$$

与此同时，

$$e^{\eta\mathcal{L}}\varphi(\mathbf{W}) = \varphi(\mathbf{W}) + \eta\mathcal{L}\varphi(\mathbf{W}) + \frac{1}{2}\eta^2\mathcal{L}^2\varphi(\mathbf{W}) + O(\eta^3).$$

由 (3.23)，我们得到

$$\begin{aligned} \mathcal{L}^2\varphi &= \mathcal{L}\left[g_{ij}(\mathbf{A},\mathbf{W})\frac{\partial\varphi}{\partial w_{ij}} + O(\eta)\right] \\ &= g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial}{\partial w_{kl}}\left(g_{ij}(\mathbf{A},\mathbf{W})\frac{\partial\varphi}{\partial w_{ij}}\right) + O(\eta) \\ &= \left(g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial g_{ij}(\mathbf{A},\mathbf{W})}{\partial w_{kl}}\right)\frac{\partial\varphi}{\partial w_{ij}} + g_{ij}(\mathbf{A},\mathbf{W})g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial^2\varphi}{\partial w_{ij}\partial w_{kl}} + O(\eta). \end{aligned} \tag{3.40}$$

比较 $\frac{\partial^2\varphi}{\partial w_{ij}\partial w_{kl}}$和 $\frac{\partial\varphi}{\partial w_{ij}}$ 的系数，我们推导出

$$
\begin{aligned}
P_{ijkl}f_{kl} &= -\frac{1}{2}J_{ij} - \frac{1}{2}g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial g_{ij}(\mathbf{A},\mathbf{W})}{\partial w_{kl}},\\
P_{ijuv}H_{uvrs}P_{klxy}H_{xyrs} &= \mathbb{E}[g_{ij}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{kl}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})].
\end{aligned}
\tag{3.41}
$$

这里，F $=(f_{ij})_{n\times n}$ 和 H $=(H_{ijkl})_{n\times n\times n\times n}$ 是式 (3.13) 中的系数。式 (3.41) 的推导细节总结于第 6 节。

为了解释式 (3.41)，可以看到，式 (3.41) 中第二个等式右侧恰好是 G(xx^T, W) 的协方差张量。我们将其记为

$$
\mathbf{M}(\mathbf{W}) = (M_{ijkl}(\mathbf{W}))_{n\times n\times n\times n},\ M_{ijkl} = \mathbb{E}\left[g_{ij}(\mathbf{x}\mathbf{x}^T-\mathbf{A},\mathbf{W})g_{kl}(\mathbf{x}\mathbf{x}^T-\mathbf{A},\mathbf{W})\right].
\tag{3.42}
$$

因此，我们希望找到合适的 H，使得

$$
P_{ijuv}H_{uvrs}P_{klxy}H_{xyrs} = M_{ijkl}.
\tag{3.43}
$$

如果 W $\in O(n)$，则可以实现这一点，这足以满足我们的目的。

引理 3.5。考虑式 (3.42) 中的 M。则存在唯一的 N $=(N_{ijkl})_{n\times n\times n\times n}$，满足

(i)（对称性） $N_{ijkl}=N_{klij}$；(ii)（半正定）对于任意 $(m_{ij})_{n\times n}$, $m_{ij}N_{ijkl}m_{kl}\geq 0$；(iii)（M 的平方根） $N_{ijrs}N_{klrs}=M_{ijkl}$。

此外，若 W $\in O(n)$，则

$$
P_{ijrs}N_{rskl} = N_{ijkl},
\tag{3.44}
$$

即 PN = N。

该引理的证明见第 6 节。根据引理 3.5，我们也将 N 记为

$$
\mathbf{N} = \sqrt{\mathbf{M}}.
\tag{3.45}
$$

若我们取

$$
\mathbf{H} = \mathbf{N} = \sqrt{\mathbf{M}},
$$

在 (3.13) 中，则

$$
P_{ijuv}H_{uvrs}P_{klxy}H_{xyrs} = P_{ijuv}N_{uvrs}P_{klxy}N_{xyrs} = N_{ijrs}N_{klrs} = M_{ijkl}
$$

当 W $\in O(n)$ 时成立。因此，(3.43) 得到满足。

现在我们取

$$
\mathbf{F}=\mathbf{0}, \mathbf{H} = \mathbf{N} = \sqrt{\mathbf{M}}
\tag{3.46}
$$

代入 (3.13)。由引理 3.5 可知，(3.13) 保持在 $O(n)$ 上，因此 (3.13) 可以改写为

$$
\begin{aligned}
\mathrm{d}w_{ij} &= g_{ij}(\mathbf{A},\mathbf{W})\mathrm{d}t + \sqrt{\eta}P_{ijkl}(\mathbf{W})N_{klrs}(\mathbf{W})\circ\mathrm{d}B_{rs},\\
&= g_{ij}(\mathbf{A},\mathbf{W})\mathrm{d}t + \sqrt{\eta}N_{ijrs}(\mathbf{W})\circ\mathrm{d}B_{rs}.
\end{aligned}
\tag{3.47}
$$

这个随机微分方程似乎是 SGA 方法的二阶近似：我们在 (1.2) 中有 g_{ij} 作为漂移项；g_{ij} 的协方差也反映在布朗运动中。然而，由于 (3.46) 中的 F 不满足 (3.41)，根据定理 3.2，(3.47) 不是二阶近似，而是一阶近似。

此外，甚至不存在满足（3.41）第一个方程的解。这排除了在斯蒂费尔流形上推导二阶近似的可能性。

即便采用最后一种尝试，我们仍要求 H 满足（3.41），并且只需将（3.13）中的 $\mathcal{P}_{T_{\mathbf{W}}O(n)}$F 替换为

$$\mathbf{L}=(L_{ij})_{n\times n},\ L_{ij}:=-\frac{1}{2}J_{ij}-\frac{1}{2}g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial g_{ij}(\mathbf{A},\mathbf{W})}{\partial w_{kl}}, \tag{3.48}$$

由此得到的随机微分方程仍未保持在斯蒂费尔流形上。因此，我们没有保持在斯蒂费尔流形上的二阶扩散近似。

引理 3.6。（不稳定的二阶近似）考虑 (3.41)、(3.48) 中的 L，以及 (3.21) 中的 J =$(J_{ij})_{n\times n}$。则

(i) 假设 H 满足 (3.41)。则不存在满足以下条件的 F = $(f_{ij})_{n\times n}$

$$P_{ijkl}f_{kl}=L_{ij},$$

也就是说，(3.41) 的第一个方程无解。

(ii) 考虑以下随机微分方程的解 W(t):

$$\mathrm{d}w_{ij}=(g_{ij}(\mathbf{A},\mathbf{W})+\eta L_{ij})\,\mathrm{d}t+\sqrt{\eta}N_{ijrs}(\mathbf{W})\circ\mathrm{d}B_{rs},\ \mathbf{W}(0)=\mathbf{W_0}\in O(n). \tag{3.49}$$

因此，不存在时间区间 $[t_0,t_1]$，使得 W(t) $\in O(n)$ 对 $t\in[t_0,t_1]$ 成立。这里，N = $(N_{ijkl})_{n\times n\times n\times n}$ 是协方差矩阵 M 的平方根（见引理 3.5）。

证明。我们首先用反证法证明 (i)。假设存在一个解 F，那么由 (3.18) 可得，

$$w_{ij}L_{tj}+w_{tj}L_{ij}=w_{ij}P_{tjkl}f_{kl}+w_{tj}P_{ijkl}f_{kl}=0.$$

注意

$$-2(w_{ij}L_{tj}+w_{tj}L_{ij})=\underbrace{w_{ij}J_{tj}+w_{tj}J_{lj}}_{\mathrm{I}}+\underbrace{g_{kl}(\mathbf{A},\mathbf{W})\left(w_{ij}\frac{\partial g_{tj}(\mathbf{A},\mathbf{W})}{\partial w_{kl}}+w_{tj}\frac{\partial g_{ij}(\mathbf{A},\mathbf{W})}{\partial w_{kl}}\right)}_{\mathrm{II}}.$$

根据 (3.21) 中 J、K 的定义，我们有

$$\mathrm{I}=\frac{\partial}{\partial w_{rs}}\left(\left(w_{ij}P_{tjuv}+w_{tj}P_{ijuv}\right)H_{uvml}K_{rsml}\right)-K_{tjml}\frac{\partial}{\partial w_{rs}}(w_{ij}K_{rsml})-K_{ijml}\frac{\partial}{\partial w_{rs}}(w_{tj}K_{rsml}).$$

由 (3.18) 可知，I 的第一项为零，因此我们有

$$\begin{aligned}\mathrm{I}&=-K_{tjml}\frac{\partial}{\partial w_{rs}}(w_{ij}K_{rsml})-K_{ijml}\frac{\partial}{\partial w_{rs}}(w_{tj}K_{rsml})\\&=-2K_{tjml}K_{ijml}-w_{ij}P_{tjuv}H_{uvml}K_{rsml}-w_{tj}P_{ijuv}H_{uvml}K_{rsml}\\&=-2K_{tjml}K_{ijml}\\&=-2\mathbb{E}[g_{tj}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{ij}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})].\end{aligned}$$

对于 II，我们有

$$\begin{aligned}\mathrm{II}=&\frac{\partial}{\partial w_{kl}}[(w_{ij}g_{tj}(\mathbf{A},\mathbf{W})+w_{tj}g_{ij}(\mathbf{A},\mathbf{W}))g_{kl}(\mathbf{A},\mathbf{W})]\\&-g_{tj}(\mathbf{A},\mathbf{W})\frac{\partial}{\partial w_{kl}}(w_{ij}g_{kl}(\mathbf{A},\mathbf{W}))-g_{ij}(\mathbf{A},\mathbf{W})\frac{\partial}{\partial w_{kl}}(w_{tj}g_{kl}(\mathbf{A},\mathbf{W}))\\=&-2g_{ij}(\mathbf{A},\mathbf{W})g_{tj}(\mathbf{A},\mathbf{W})+g_{kl}\frac{\partial}{\partial w_{kl}}(g_{tj}w_{ij}+g_{ij}w_{tj}).\end{aligned}$$

Thus

$$\begin{aligned}-2(w_{ij}L_{tj}+w_{tj}L_{ij}) &= \mathrm{I}+\mathrm{II}\\ &= -2\mathbb{E}[g_{tj}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{ij}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})]-2g_{ij}(\mathbf{A},\mathbf{W})g_{tj}(\mathbf{A},\mathbf{W})\\ &\quad+g_{kl}\frac{\partial}{\partial w_{kl}}(g_{tj}w_{ij}+g_{ij}w_{tj})\\ &= -2\mathbb{E}[g_{tj}(\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{ij}(\mathbf{x}\mathbf{x}^T,\mathbf{W})]+g_{kl}\frac{\partial}{\partial w_{kl}}(g_{tj}w_{ij}+g_{ij}w_{tj}).\end{aligned}$$

对于一般的随机变量 x，这不可能为零，因为第一项取决于四阶动量，而第二项仅取决于二阶动量。这是一个矛盾，因此 $w_{ij}L_{tj}+w_{tj}L_{ij}\neq 0$ 且 (3.41) 无解。

对于 (ii)，我们仍然用反证法证明。假设对于某个时间区间 $[t_0,t_1]$，$\mathsf{W}(t)\in O(n)$。那么

$$\frac{\mathrm{d}(\mathbf{W}\mathbf{W}^T)}{\mathrm{d}t}=0,\ t\in(t_0,t_1).$$

由引理 3.5 可知，若 $\mathsf{W}\in O(n)$，则 $P_{ijrs}N_{rskl}=N_{ijkl}$。因此我们可以改写为 (3.49)

$$\mathrm{d}w_{ij}=(g_{ij}(\mathbf{A},\mathbf{W})+\eta L_{ij})\,\mathrm{d}t+\sqrt{\eta}P_{ijrs}N_{rskl}\circ\mathrm{d}B_{kl}$$

然而，

$$\begin{aligned}\mathrm{d}(w_{ij}w_{tj}) &= w_{ij}\mathrm{d}w_{tj}+w_{tj}\mathrm{d}w_{ij}\\ &= (w_{ij}g_{tj}(\mathbf{A},\mathbf{W})+w_{tj}g_{ij}(\mathbf{A},\mathbf{W}))\mathrm{d}t+\eta(w_{ij}L_{tj}+w_{tj}L_{ij})\mathrm{d}t\\ &\quad+\sqrt{\eta}(w_{ij}P_{tjrs}+w_{tj}P_{ijrs})N_{rskl}\circ B_{kl}.\end{aligned}$$

记住，若 $\mathsf{W}\in O(n)$，则 $\mathsf{G}(\mathsf{A},\mathsf{W})=\mathcal{P}_{T_{\mathbf{W}}O(n)}\mathsf{G}(\mathsf{A},\mathsf{W})$。因此 $g_{ij}=P_{ijkl}g_{kl}$，这意味着对所有 $t\in(t_0,t_1)$，

$$w_{ij}g_{tj}+w_{tj}g_{ij}=(w_{ij}P_{tjrs}+w_{tj}P_{ijrs})g_{rs}=0$$

由 (3.18) 得出。出于同样的原因，

$$\sqrt{\eta}(w_{ij}P_{tjrs}+w_{tj}P_{ijrs})N_{rskl}\circ B_{kl}=0.$$

Thus

$$\mathrm{d}(w_{ij}w_{tj})=\eta(w_{ij}L_{tj}+w_{tj}L_{ij})\mathrm{d}t.$$

然而，根据 (i)，$(w_{ij}L_{tj}+w_{tj}L_{ij})\neq 0$。因此，对于 $t\in(t_0,t_1)$，$\mathsf{d}(w_{ij}w_{tj})\neq 0$，这与

$$w_{ij}w_{tj}=\delta_{it}.$$

这是一个矛盾，证明结束。 □

根据引理 3.6，(3.49) 无法保持在斯蒂费尔流形上。在这种情况下，(3.22) 是一个系数无界的退化抛物型偏微分方程，无法控制斯蒂费尔流形法向方向上的扩散。因此，利用 (3.22) 的解来近似半群 (3.2) 的行为毫无意义。

4. Reversible diffusion approximation: exponential convergence

在定理 3.2 中，我们推导了斯蒂费尔流形上的扩散近似。一个自然的问题是，该随机微分方程是否遍历并收敛。

事实上，可逆性和庞加莱不等式保证了指数收敛。为此，我们可以将 Fokker-Planck 算子 (3.26) 改写为

$$\begin{aligned}\mathcal{L}^*\rho &= \frac{\partial}{\partial w_{ij}}\left[\rho\left(-g_{ij}-\eta P_{ijkl}f_{kl}+\frac{\eta}{2}K_{ijrs}K_{klrs}\frac{\partial\log\rho}{\partial w_{kl}}+\frac{\eta}{2}J_{ij}\right)\right]\\ &= \frac{\partial}{\partial w_{ij}}\left[\frac{\eta}{2}K_{ijrs}K_{klrs}\rho\left(\frac{\eta J_{ij}-2g_{ij}-2\eta P_{ijkl}f_{kl}}{\eta K_{ijrs}K_{klrs}}+\frac{\partial\log\rho}{\partial w_{kl}}\right)\right].\end{aligned}$$

如果存在一个函数 $U(\mathbf{W})$，使得

$$\frac{\partial U}{\partial w_{kl}}=\frac{\eta J_{ij}-2g_{ij}-2\eta P_{ijkl}f_{kl}}{\eta K_{ijrs}K_{klrs}}, \tag{4.1}$$

那么 $\partial_t\rho=\mathcal{L}^*\rho$ 的解满足

$$\partial_t\rho=\frac{\partial}{\partial w_{ij}}\left[\frac{\eta}{2}K_{ijrs}K_{klrs}\rho\frac{\partial(\log\rho+U)}{\partial w_{kl}}\right]=\frac{\partial}{\partial w_{ij}}\left[\frac{\eta}{2}e^{-U}K_{ijrs}K_{klrs}\frac{\partial(e^U\rho)}{\partial w_{kl}}\right].$$

不失一般性，我们假设 $\int_{O(n)}e^{-U(\mathbf{W})}\mathrm{d}\mathbf{W}=1$。然后在两边乘以 $e^U\rho-1$，并根据 (3.21) 中 K 的定义，我们推导得到

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{O(n)}e^{-U}|e^U\rho-1|^2\mathrm{d}\mathbf{W} &= -\eta\int_{O(n)}e^{-U}\left(K_{ijrs}\frac{\partial}{\partial w_{ij}}(e^U\rho-1)\right)\left(K_{klrs}\frac{\partial}{\partial w_{kl}}(e^U\rho-1)\right)\mathrm{d}\mathbf{W}\\ &= -\eta\int_{O(n)}e^{-U}|\mathbf{H}^T\nabla_{\mathbf{W}}(e^U\rho-1)|^2\mathrm{d}\mathbf{W}.\end{aligned}$$

这里，$\nabla_{\mathbf{W}}$ 是（$O(n)$ g_e）上的梯度算子，其中 g_e 是欧几里得度量。由庞加莱不等式可得指数收敛。

因此，我们希望谨慎选择 H 和 F，使其满足势条件（4.1）。下面给出两种确保可逆性的特殊情形。

首先，我们考虑 $O(n)$上的过阻尼朗之万。具体而言，我们将势函数选为 $U(\mathbf{W};\mathbf{A},\mathbf{N})=\mathrm{tr}(\mathbf{N}\mathbf{W}^T\mathbf{A}\mathbf{W})$，其中 N 是对角矩阵。这样便得到带有布朗运动的 Oja-Brockett 流。事实上，Oja-Brockett 流是（$O(n)$ g_e）上$U(\mathbf{W};\mathbf{A},\mathbf{N})=\mathrm{tr}(\mathbf{N}\mathbf{W}^T\mathbf{A}\mathbf{W})$的梯度流。

其次，我们考虑二维情形，即 $n=2$。在这种情形下，正交矩阵由旋转角确定。角度的随机微分方程是 $\mathbb{R}$上的随机微分方程，因此自动满足势条件。

在每种情况下，均可验证庞加莱不等式，因此它们会以指数速度收敛到不变测度。

4.1. $O(n)$ 上的过阻尼朗之万动力学。

设 $U(\mathbf{Q}):O(n)\to\mathbb{R}$是一个光滑函数（作为自由能），考虑 $O(n)$ 上的过阻尼朗之万动力学:

$$\mathrm{d}\mathbf{Q}(t)=\mathcal{P}_{T_{\mathbf{Q}}O(n)}\circ(-\nabla U(\mathbf{Q})\mathrm{d}t+\sigma\mathrm{d}\mathbf{W}(t)). \tag{4.2}$$

这里 $\mathcal{P}_{T_{\mathbf{Q}}O(n)}$ 是到 $T_{\mathbf{Q}}O(n)$ 上的投影，∇ 表示关于 Q 的导数，即

$$(\nabla U(\mathbf{Q}))_{ij}=\frac{\partial u(\mathbf{Q})}{\partial q_{ij}}.$$

W(t) $\in \mathbb{R}^{n\times n}$ 是标准布朗运动， σ 是一个常数。“$\circ$”表示上述随机微分方程采用 Stratonovich 意义。一般而言， σ 可以是一个依赖于 Q 的矩阵。为了说明 $O(n)$ 上的 Langevin动力学，我们在此考虑最简单的情形，这已足以进行扩散近似。

若取 U(Q) $=$ $-$tr(NQ^{T}AQ)，其中 N 为对角矩阵，其对角线上的元素按降序排列，则 (4.2) 可写为

$$\mathrm{d}\mathbf{Q}(t) = (\mathbf{AQN} - \mathbf{QNQ}^T\mathbf{AQ})\mathrm{d}t + \sigma\mathcal{P}_{T_\mathbf{Q}O(n)} \circ \mathrm{d}\mathbf{W}(t). \tag{4.3}$$

于是，(4.3) 应被视为受扰动的 Oja-Brockett 流 [2]。

根据转换规则，(4.2) 应以 Ito意义表述为

$$\mathrm{d}\mathbf{Q}(t) = \left[-\mathcal{P}_{T_\mathbf{Q}O(n)}\nabla U(\mathbf{Q}) - \frac{\sigma^2(n-1)}{4}\mathbf{Q}\right]\mathrm{d}t + \sigma\mathcal{P}_{T_\mathbf{Q}O(n)}\mathrm{d}\mathbf{W}(t). \tag{4.4}$$

By 利用 Ito公式，我们可以推导出 (4.2) 的福克-普朗克方程，即

$$\frac{\partial\rho(\mathbf{Q},t)}{\partial t} = \nabla_\mathbf{Q}\cdot(\rho(\mathbf{Q},t)\nabla_\mathbf{Q}U(\mathbf{Q})) + \frac{\sigma^2}{2}\Delta_\mathbf{Q}\rho(\mathbf{Q},t). \tag{4.5}$$

这里， $\nabla_\mathbf{Q}$、 $\nabla_\mathbf{Q}$ 和 $\Delta_\mathbf{Q}$ 分别是 ($O(n)$ g_e) 上的散度、梯度和 Laplace-Beltrami算子。更多详情请参见第 6 节。

$O(n)$的紧性意味着(4.5)指数收敛。直接计算可得，(4.5)的不变测度为

$$\rho_{eq}(\mathbf{Q}) := \frac{1}{Z}e^{-2U(\mathbf{Q})/\sigma^2},\ Z := \int_{O(n)} e^{-2U(\mathbf{Q})/\sigma^2}\mathrm{d}\mathbf{V}, \tag{4.6}$$

其中，dV是($SO(n)$ g_e)上的体积形式。(4.5)可改写为

$$\frac{\partial\rho(\mathbf{Q},t)}{\partial t} = \frac{\sigma^2}{2}\nabla_\mathbf{Q}\cdot\left(\rho_{eq}(\mathbf{Q})\nabla_\mathbf{Q}\left(\frac{\rho(\mathbf{Q},t)}{\rho_{eq}(\mathbf{Q})}\right)\right) := \mathcal{L}^*\rho(\mathbf{Q},t), \tag{4.7}$$

并将Fokker-Planck算子记为 $\mathcal{L}^*$。这也意味着(4.5)的不变测度是唯一的，因为

$$\begin{aligned}
\mathcal{L}^*\rho = 0 &\iff \nabla_\mathbf{Q}\cdot\left(\rho_{eq}(\mathbf{Q})\nabla_\mathbf{Q}\left(\frac{\rho(\mathbf{Q},t)}{\rho_{eq}(\mathbf{Q})}\right)\right) = 0\\
&\iff \int_{O(n)}\rho_{eq}(\mathbf{Q})\left|\nabla_\mathbf{Q}\left(\frac{\rho(\mathbf{Q},t)}{\rho_{eq}(\mathbf{Q})}\right)\right|^2\mathrm{d}V = 0\\
&\iff \rho(\mathbf{Q}) = c\rho_{eq}(\mathbf{Q}),
\end{aligned}$$

其中 c 是一个常数。若 ρ(Q)是 $SO(n)$上的概率测度，则 $c = 1$ 且 $\rho = \rho_{eq}$。在上述归纳法中，我们利用了 ρ_{eq}(Q)的正性，这是 U(Q)的连续性和 $SO(n)$的紧性的推论。

在等式 (4.5) 两边乘以 ρ(Q t)/ρ_{eq}(Q) -1，可得

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{O(n)}\rho_{eq}(\mathbf{Q})\left|\frac{\rho(\mathbf{Q},t)-\rho_{eq}(\mathbf{Q})}{\rho_{eq}(\mathbf{Q})}\right|^2\mathrm{d}\mathbf{V} = -\sigma^2\int_{O(n)}\rho_{eq}(\mathbf{Q})\left|\nabla_\mathbf{Q}\left(\frac{\rho(\mathbf{Q},t)-\rho_{eq}(\mathbf{Q})}{\rho_{eq}(\mathbf{Q})}\right)\right|^2\mathrm{d}\mathbf{V}. \tag{4.8}$$

如果我们能够在 $L^2(\rho_{eq}(\mathbf{Q})\mathrm{d}\mathbf{V})$ 中证明庞加莱不等式，即存在常数 $C>0$，使得对于所有满足 $\int_{SO(n)} f(\mathbf{Q})\rho_{eq}(\mathbf{Q})\mathrm{d}\mathbf{V}=0$ 的 $f\in H^1(\rho_{eq}(\mathbf{Q})\mathrm{d}\mathbf{V})$，都有

$$\int_{O(n)} \rho_{eq}(\mathbf{Q})\left|\nabla_{\mathbf{Q}}\left(f(\mathbf{Q})\right)\right|^2 \mathrm{d}\mathbf{V} \geq C\int_{O(n)} \rho_{eq}(\mathbf{Q})|f(\mathbf{Q})|^2\mathrm{d}\mathbf{V}, \tag{4.9}$$

则 (4.8) 得出

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{O(n)} \rho_{eq}(\mathbf{Q})\left|\frac{\rho(\mathbf{Q},t)-\rho_{eq}(\mathbf{Q})}{\rho_{eq}(\mathbf{Q})}\right|^2 \mathrm{d}\mathbf{V} \leq -C\int_{O(n)} \rho_{eq}(\mathbf{Q})\left|\frac{\rho(\mathbf{Q},t)-\rho_{eq}(\mathbf{Q})}{\rho_{eq}(\mathbf{Q})}\right|^2 \mathrm{d}\mathbf{V}.$$

根据格朗沃尔不等式，(4.5) 指数收敛，其初始值是一个概率测度。

现在我们证明 $L^2(\rho_{eq}\mathrm{d}\mathbf{V})$ 中的庞加莱不等式。

引理 4.1。（庞加莱不等式）设 $f\in H^1(\rho_{eq}(\mathbf{Q})\mathrm{d}\mathbf{V})$ 满足 $\int_{O(n)} f(\mathbf{Q})\rho_{eq}(\mathbf{Q})\mathrm{d}\mathbf{V}=0$。则 (4.9) 成立。

证明。为证明这一点，我们证明对于任意 $\lambda>0$，$(\lambda-\mathcal{L}^*)^{-1}: L^2(\mathrm{d}\mathbf{V}/\rho_{eq}(\mathbf{Q})) \to L^2(\mathrm{d}\mathbf{V}/\rho_{eq}(\mathbf{Q}))$ 是一个紧算子。设 $\{u_n\}_{n\geq 1}$ 和 $\{g_n\}_{n\geq 1}$ 是 $L^2(\mathrm{d}\mathbf{V}/\rho_{eq}(\mathbf{Q}))$ 中的两个序列，使得

$$(\lambda-\mathcal{L}^*)u_n = g_n,\ n\geq 1,$$

且 $\{g_n\}_{n\geq 1}$ 在 $L^2(\mathrm{d}\mathbf{V}/\rho_{eq}(\mathbf{Q}))$中一致有界。因此，两边乘以 u_n 可得

$$\begin{aligned}\int_{O(n)} \frac{1}{\rho_{eq}}\cdot u_n\cdot(\lambda-\mathcal{L}^*)u_n\mathrm{d}\mathbf{V} &= \int_{O(n)} \rho_{eq}\left|\nabla_{\mathbf{Q}}\left(\frac{u_n}{\rho_{eq}}\right)\right|^2\mathrm{d}\mathbf{V} + \lambda\int_{SO(n)}\rho_{eq}\left|\frac{u_n}{\rho_{eq}}\right|^2\mathrm{d}\mathbf{V}\\ &= \int_{O(n)}\frac{1}{\rho_{eq}}\cdot g_n u_n\mathrm{d}\mathbf{V}.\end{aligned}$$

根据柯西-施瓦茨不等式，我们知道 u_n/ρ_{eq} 在 $H^1(\rho_{eq}\mathrm{d}\mathbf{V})$中一致有界。紧嵌入 $H^1(\rho_{eq}\mathrm{d}\mathbf{V})$ 到 $\hookrightarrow\hookrightarrow L^2(\rho_{eq}\mathrm{d}\mathbf{V})$（见 [17]）意味着，除去子列后，存在 $u^*\in L^2(\rho_{eq}\mathrm{d}\mathbf{V})$，

$$\frac{u_n}{\rho_{eq}}\to\frac{u^*}{\rho_{eq}}\ \text{in}\ L^2(\rho_{eq}\mathrm{d}\mathbf{V}),$$

或者等价地

$$u_n\to u^*\ \text{in}\ L^2(\mathrm{d}\mathbf{V}/\rho_{eq}).$$

因此，$(\lambda-\mathcal{L}^*)^{-1}$ 是紧的。因此，$1/\lambda\neq 0$ 不是 $\sigma((\lambda-\mathcal{L}^*)^{-1})$的聚点，故0不是 $\sigma(\mathcal{L}^*)$的聚点；但它是 $\mathcal{L}^*$的唯一主特征值，其特征向量为 $c\cdot\rho_{eq}$，其中 c 是常数。因此，对于任意 $g\in L^2(\mathrm{d}\mathbf{V}/\rho_{eq})$，我们有

$$-\int_{O(n)}\frac{g}{\rho_{eq}}\mathcal{L}^*g\mathrm{d}\mathbf{V}\geq C\int_{O(n)}\frac{|g|^2}{\rho_{eq}}\mathrm{d}\mathbf{V} \tag{4.10}$$

对于所有满足 $\int_{O(n)} g\mathrm{d}\mathbf{V}=0$ 的 g，令 $g=\rho_{eq}f$，可得（4.9）。

□

4.2. $n=2$ 的情形。如果我们要求（3.13）中的 F、H 满足（3.46），则随机微分方程为

$$\mathrm{d}w_{ij} = g_{ij}(\mathbf{A},\mathbf{W})\mathrm{d}t + \sqrt{\eta}\cdot N_{ijkl}\circ \mathrm{d}B_{kl}. \tag{4.11}$$

这里，M(W) 是 (3.42) 中定义的 G($\mathbf{x}\mathbf{x}^T$, W) 的协方差矩阵。由引理 3.3 可知，(4.11) 的不变集为斯蒂费尔流形。利用这一事实，我们可以将 (4.11) 改写为：

引理 4.2。假设 $n=2$。考虑满足 (3.46) 的 H 和 H 所对应的 (3.13)。那么，(或 (4.11)) 可以改写为 (3.13)

$$\mathrm{d}\mathbf{W} = \mathbf{F_1}(\mathbf{W})\mathrm{d}t + \sqrt{\eta}\cdot c(\mathbf{W})\mathbf{W}\circ \mathrm{d}\mathbf{Z}. \tag{4.12}$$

这里，$\mathbf{F_1}$ 在 (2.4) 中定义，c(W) 是一个标量，定义如下：

$$\begin{aligned} c(\mathbf{W}) &:= \sqrt{c_1(w_{1,1}^4+w_{1,2}^4)+c_2w_{1,1}^2w_{1,2}^2+c_3w_{1,1}w_{1,2}(w_{1,1}w_{2,2}+w_{1,2}w_{2,1})},\\ c_1 &:= \mathbb{E}[x_1^2x_2^2]-(\mathbb{E}[x_1x_2])^2,\\ c_2 &:= \mathbb{E}[x_1^4+x_2^4-4x_1^2x_2^2]+2(\mathbb{E}[x_1x_2])^2+2\mathbb{E}[x_1^2]\mathbb{E}[x_2^2]-(\mathbb{E}[x_1^2])^2-(\mathbb{E}[x_2^2])^2,\\ c_3 &:= 2\mathbb{E}[x_1^3x_2-x_1x_2^3]-\mathbb{E}[x_1^2]\mathbb{E}[x_1x_2]+\mathbb{E}[x_2^2]\mathbb{E}[x_1x_2] \end{aligned} \tag{4.13}$$

而 Z(t) $\in\mathbb{R}^{2\times 2}$ 定义如下：

$$\mathbf{Z}(t) := \begin{pmatrix} 0 & -B(t)\\ B(t) & 0\end{pmatrix}, \tag{4.14}$$

其中，$B(t)$ 是一维标准布朗运动。

此引理的证明见第 6 节。请注意，O(2) 中的每个元素都可以表示为以下两种形式之一：

$$O_1(\theta) = \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix},\ \theta\in[0,2\pi), \tag{4.15}$$

$$O_{-1}(\theta) = \begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{pmatrix},\ \theta\in[0,2\pi). \tag{4.16}$$

由于 (3.13) 中 W(t) 的轨道几乎必然连续，且行列式关于 W 也是连续函数，因此，如果对某个 θ，$\mathbf{W_0}=O_1(\theta)$，那么对某个 $\theta(t)\in[0,2\pi)$，$\mathbf{W}(t)=O_1(\theta(t))$。不失一般性，假设对某个 $\theta_0\in[0,2\pi)$，$\mathbf{W_0}=O_1(\theta_0)$，则对任意 $t\geq 0$，

$$w_{1,1}(t)=w_{2,2}(t),\ w_{1,2}(t)+w_{2,1}(t)=0,\ w_{1,1}^2(t)+w_{1,2}^2(t)=1. \tag{4.17}$$

为证明 (4.12) 的收敛性，我们考虑 $\theta(t)$ 的过程，而不是 W。我们构造如下的一维 Ito意义随机微分方程：

$$\mathrm{d}\theta(t) = f(\theta)\mathrm{d}t + \sqrt{\eta}g(\theta)\mathrm{d}B,\ \theta(t)=\theta_0, \tag{4.18}$$

其中，$B(t)$ 是标准布朗运动，f、g 定义为

$$\begin{aligned} g(\theta) &= -c(\theta)\\ f(\theta) &= (\mathbb{E}[x_2^2]-\mathbb{E}[x_1^2])\cos\theta\sin\theta+\frac{\eta(2c_1(\theta)-c_2(\theta))}{2}(\cos^3\theta\sin\theta-\sin^3\theta\cos\theta)\\ &\quad+\frac{3\eta c_3(\theta)}{2}\cos^2\theta\sin^2\theta. \end{aligned} \tag{4.19}$$

这里， c、c_1、c_2 和 c_3 是式（4.13）中定义的标量函数，而 $c(\theta)$ 是将 W 替换为 $O_1(\theta)$ 后得到的 $c(\mathrm{W})$，即

$$c(\theta) = c\left(\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\right). \tag{4.20}$$

于是，我们可以利用 (4.18) 的解来表示 (4.12) 的解。

引理 4.3. 考虑 (4.12)。假设初始值 $\mathrm{W}(0)$ 满足 $\mathrm{W}(0) = O_1(\theta_0)$（定义见 (4.15)），其中某个 $\theta_0 \in [0, 2\pi)$。假设 $\theta(t)$ 是 (4.18) 的解，则

$$w_{1,1}(t) = w_{2,2}(t) = \cos(\theta(t)), w_{1,2}(t) = \sin(\theta(t)),\ w_{2,1}(t) = -\sin(\theta(t)) \tag{4.21}$$

是满足初始值 $\mathrm{W}(0) = O_1(\theta_0)$ 的 (4.12) 的解。

该引理的证明见第 6 节。

4.2.1. 收敛性分析。令 $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$。现在考虑 (4.18)。我们将该随机微分方程的不变测度记为 $\rho_\infty(x) \in \mathcal{P}(\mathbb{T})$。那么 ρ_∞ 是具有周期边界条件的福克–普朗克方程的平稳解，即

$$\frac{\partial \rho(x,t)}{\partial t} = \mathcal{L}_1\rho(x,t),\ \mathcal{L}_1\rho(x,t) := -\frac{\partial(f(x)\rho(x,t))}{\partial x} + \frac{1}{2}\cdot\frac{\partial^2(\eta g^2(x)\rho(x,t))}{\partial x^2}. \tag{4.22}$$

这里 f, g 定义于 (4.19) 中。直接计算可得

$$\rho_\infty(x) = C\exp\left(\int_0^x \frac{2f(s)}{\eta\cdot g^2(s)}\mathrm{d}s\right),\ C = \left(\exp\left(\int_0^{2\pi}\frac{2f(s)}{\eta\cdot g^2(s)}\mathrm{d}s\right)\right)^{-1}. \tag{4.23}$$

根据引理 4.3 的证明（见第 6 节），我们知道 $g(x)$ 在 $\mathbb{T}$ 上严格为正，因此对于任意 $x \in \mathbb{T}$，$\rho_\infty(x) > 0$，故由庞加莱不等式可得指数收敛。定义加权 L^2 空间 $L^2(\mathbb{T}, \mathrm{d}x/\rho_\infty)$ 如下：

$$L^2(\mathbb{T}, \mathrm{d}x/\rho_\infty) := \left\{p : \int_{\mathbb{T}}\frac{p^2(x)}{\rho_\infty(x)}\mathrm{d}x < \infty\right\}, \langle p, q\rangle_{\mathrm{d}x/\rho_\infty} = \int_{\mathbb{T}}\frac{p(x)q(x)}{\rho_\infty(x)}\mathrm{d}x. \tag{4.24}$$

定理 4.1. 假设 $\theta(t)$ $t \geq 0$ 是方程 (4.18) 的解，其初始值为 θ_0，这是一个概率密度函数为 $\rho_0 \in \mathcal{P}(\mathbb{T})$ 的随机变量。令 $\rho(x,t) \in \mathcal{P}(\mathbb{T})$ 为 $\theta(t)$ 的分布，即对于任意博雷尔集 $B \in \mathbb{T}$，有 $\mathbb{P}(\theta(t) \in B) = \int_B \rho(x,t)\mathrm{d}x$。考虑 $\rho_\infty(x) \in \mathcal{P}(\mathbb{T})$，即 (4.23) 中的不变测度。那么存在一个常数 $c > 0$，它仅取决于 η 和 x 的矩，使得

$$\|\rho(\cdot,t) - \rho_\infty\|_{L^2(\mathbb{T},\mathrm{d}x/\rho_\infty)} \leq e^{-ct}\|\rho_0 - \rho_\infty\|_{L^2(\mathbb{T},\mathrm{d}x/\rho_\infty)}. \tag{4.25}$$

定理 4.1 的证明见第 6 节。

4.2.2. 一个例子。本节中，我们在此明确计算一个例子，以说明主要结果。假设 $\mathrm{x} = (x_1, x_2)^T$，其中 x_1 和 x_2 是相互独立的随机变量，且它们具有概率密度函数

$$\rho_1 = \frac{1}{4}\cdot\mathbb{I}_{(-2,2)},\ \rho_2 = \frac{1}{2}\cdot\mathbb{I}_{(-1,1)},$$

i.e. $x_1 \sim \mathrm{Uni}(-2, 2)$ 和 $x_2 \sim \mathrm{Uni}(-1, 1)$。那么，x i 的协方差矩阵 s

$$\mathbf{A} = \begin{pmatrix} 4/3 & 0 \\ 0 & 1/3 \end{pmatrix}.$$

那么，由 (4.13) 可得

$$c_1 = \frac{4}{9},\ c_2 = \frac{8}{45},\ c_3 = 0,$$

and

$$c = \sqrt{\frac{4}{9}(\cos^4\theta + \sin^4\theta) + \frac{8}{45}\cos^2\theta\sin^2\theta} = \sqrt{\frac{4}{9} - \frac{32}{45}\cos^2\theta\sin^2\theta}.$$

因此 $c \geq \sqrt{4/15}$ 对于任意 $\theta \in [0, 2\pi)$。于是 (4.18) 可写为

$$\mathrm{d}\theta = \left(-\cos\theta\sin\theta + \frac{16\eta}{45}(\cos^3\theta\sin\theta - \sin^3\theta\cos\theta)\right)\mathrm{d}t + \sqrt{\frac{4}{9} - \frac{32}{45}\cos^2\theta\sin^2\theta}\cdot\mathrm{d}B.$$

根据 (4.23)，不变测度为

$$\rho_\infty(x) = \exp\left(\frac{1}{\eta}\int_0^x \frac{-45\cos\theta\sin\theta + 16\eta(\cos^3\theta\sin\theta - \sin^3\theta\cos\theta)}{10 - 16\cos^2\theta\sin^2\theta}\mathrm{d}\theta\right),\ x \in [0, 2\pi).$$

直接计算可得

$$\rho_\infty(x) = \exp\left(-\frac{15\sqrt{6}}{16\eta}\arctan\frac{2\sqrt{6}\sin^2 x}{5 - 4\sin^2 x}\right)\sqrt{\frac{5}{8\sin^4 x - 8\sin^2 x + 5}}, x \in [0, 2\pi). \tag{4.26}$$

不变测度的显式公式，即 (4.26)，表明若 $\eta << 1$，则不变测度的质量集中在 $x = 0$ 和 $x = \pi$附近；或者用矩阵表示，即集中在 $-\mathrm{I}_2$ 和 I_2,附近，这给出了正确的主成分分解。

5. The case of $p < n$

通过谨慎处理张量的大小并改写投影算子，$p = n$ 情形下的所有结果都可以推广到 $p < n$情形。

首先，关于半群的所有正则性和稳定性结果仍然成立。证明与引理 3.1 和定理 3.1 完全相同，只需将列指标的末端索引 n 替换为 p 。

其次，扩散近似现在表述为

$$\dot{\mathbf{W}} = \mathbf{G}(\mathbf{A}, \mathbf{W}) + \eta\mathcal{P}_{T_{\mathbf{W}}O(n\times p)}\mathbf{F}(\mathbf{W}) + \sqrt{\eta}\mathcal{P}_{T_{\mathbf{W}}O(n\times p)}\dot{\mathbf{Z}},\ \mathbf{W}(0) = \mathbf{W_0} \in O(n\times p). \tag{5.1}$$

其中，$\mathcal{P}_{T_{\mathbf{W}}O(n\times p)}$ 是投影到 W 处 $O(n\times p)$ 的切空间上的投影，见 (5.3)；Z 定义为

$$\dot{\mathbf{Z}}(\mathbf{W}) = (\dot{Z}_{ij})_{n\times p},\ \dot{Z}_{ij}(\mathbf{W}) = H_{ijkl}(\mathbf{W}) \circ \dot{B}_{kl}; \tag{5.2}$$

$\mathrm{H(W)} = (H_{ijkl}(\mathrm{W}))_{n\times p\times n\times p}$ 是布朗运动的系数张量；$\mathrm{B} = (B_{ij})_{n\times p}$是 $\mathbb{R}^{n\times p}$ 中的标准布朗运动；符号 '∘' 表示 (5.1) 是一个处于 Stratonovich 意义下的随机微分方程。

到 $T_{\mathbf{W}}O(n\times p)$ 上的投影算子于是写作

$$\mathcal{P}_{T_{\mathbf{W}}O(n\times p)}\mathbf{M} = (\mathbf{I_n} - \mathbf{W}\mathbf{W}^T)\mathbf{M} + \frac{1}{2}\mathbf{W}(\mathbf{W}^T\mathbf{M} - \mathbf{M}^T\mathbf{W}). \tag{5.3}$$

当 $p = n$ 时，$\mathrm{WW}^T = \mathrm{I_n}$ so $\mathcal{P}_{T_{\mathbf{W}}O(n\times p)}$ 恰好是对 $O(n)$ 的投影；当 $p = 1$ 时，$\mathrm{W}^T\mathrm{M} = \mathrm{M}^T\mathrm{W}$ 都是标量，那么

$$\mathcal{P}_{T_{\mathbf{W}}O(n\times p)}\mathbf{M} = \mathcal{P}_{T_{\mathbf{W}}\mathbb{S}^{n-1}}\mathbf{M} = (\mathbf{I_n} - \mathbf{W}\mathbf{W}^T)\mathbf{M}, \tag{5.4}$$

即 $\mathcal{P}_{T_{\mathbf{W}}O(n\times p)}$ 退化为投影到单位球面上。

使用与引理 3.3 中相同的技术和方法，可以证明 (5.1) 保持在$O(n \times p)$上；此外，扩散近似也与定理 3.2 中一样是一阶的。

关于可逆性，过阻尼朗之万过程仍然是可逆的。第 4.1 节中的所有证明在 $O(n \times p)$ 上一般均成立。

ACKNOWLEDGEMENT

Jian-Guo Liu 获得了美国国家科学基金会（NSF）DMS-2106988 项目资助。

6. 附录

6.1. 黎曼流形和斯蒂费尔流形。我们将流形 $\mathcal{M}$ 上 m 点处的切空间记为 $T_{\mathbf{m}}\mathcal{M}$，将 $\mathcal{M}$ 上的切向量场记为 $\Gamma(T\mathcal{M})$。切丛（切空间的不交并）记为 $T\mathcal{M}$。

定义 6.1.（黎曼流形）设 $\mathcal{M}$ 是一个光滑流形。黎曼流形（$\mathcal{M}$, g）是一个配备内积的光滑流形；在每个 $\mathbf{m} \in \mathcal{M}$ 处，内积 $g_{\mathbf{m}}$ 定义在 $T_{\mathbf{m}}\mathcal{M}$ 上。此外，对于任意切向量场 $\dot{\mathbf{x}}$ 和 $\dot{\mathbf{y}}$，函数

$$\langle \dot{\mathbf{x}}(\mathbf{m}), \dot{\mathbf{y}}(\mathbf{m}) \rangle_{g_{\mathbf{m}}} : \mathcal{M} \to \mathbb{R} \tag{6.1}$$

是光滑的。

给定 M 上的黎曼度量 g， $\mathcal{M}$ 上光滑函数 $\mathcal{E}$ 的梯度定义为

定义 6.2.（黎曼流形上的梯度）若切向量场 $\nabla_g \mathcal{E}$ 满足：对于 M 上的每个切向量场 $\dot{\mathbf{x}}$，均有.....则称其为关于度量 g 的 $\mathcal{E}$ 的梯度

$$\langle \mathcal{E}', \dot{\mathbf{x}} \rangle_F = \langle \nabla_g \mathcal{E}, \dot{\mathbf{x}} \rangle_g. \tag{6.2}$$

这里， $\mathcal{E}'$ 是 $\mathcal{E}$ 的导数，它是一个余切向量场。

现在，我们考虑配备欧几里得度量 g_e 的斯蒂费尔流形，并使用全局坐标 $\mathbf{Q} \in \mathbb{R}^{n \times n}$。我们首先介绍 $O(n)$ 的几个重要性质。更多详情请参见 [4] 。

引理 6.1.（斯蒂费尔流形）斯蒂费尔流形 $O(n)$ 是一个光滑紧致流形，其维数为 $n(n-1)/2$。Q 处的切空间为

$$T_{\mathbf{Q}}O(n) = \{\mathbf{Q}\boldsymbol{\Omega} \mid \boldsymbol{\Omega} \in \mathbb{R}^{n \times n},\ \boldsymbol{\Omega} + \boldsymbol{\Omega}^T = \mathbf{0}\}, \tag{6.3}$$

而 Q 处的法空间为

$$T_{\mathbf{Q}}O(n)^{\perp} = \{\mathbf{Q}\boldsymbol{\Omega} \mid \boldsymbol{\Omega} \in \mathbb{R}^{n \times n},\ \boldsymbol{\Omega} = \boldsymbol{\Omega}^T\}. \tag{6.4}$$

证明见 [4] 。由引理 6.1 可知，对于任意 $\mathbf{M} \in \mathbb{R}^{n \times n}$，切空间和法空间上的投影分别为：

$$\mathcal{P}_{T_{\mathbf{Q}}O(n)}\mathbf{M} := \frac{1}{2}(\mathbf{M} - \mathbf{Q}\mathbf{M}^T\mathbf{Q}),\ \mathcal{P}_{T_{\mathbf{Q}}O(n)^{\perp}}\mathbf{M} := \frac{1}{2}(\mathbf{M} + \mathbf{Q}\mathbf{M}^T\mathbf{Q}). \tag{6.5}$$

引理 6.2.（($O(n)$ g_e)上的梯度）假设 $\varphi(\mathbf{Q}) : O(n) \to \mathbb{R}$ 是定义在 $O(n)$ 上的一个光滑函数（仍记作 φ: $\mathbb{R}^{n \times n} \to \mathbb{R}$）的限制。那么，点 Q 处 φ 关于 g_e 的梯度为

$$\nabla_{g_e}\varphi := \mathcal{P}_{T_{\mathbf{Q}}O(n)}(\nabla\varphi) = \frac{1}{2}(\nabla\varphi - \mathbf{Q}(\nabla\varphi)^T\mathbf{Q}), \tag{6.6}$$

这里， $\nabla\varphi \in \mathbb{R}^{n\times n}$ 是 φ 在 $\mathbb{R}^{n\times n}$中的梯度，即 $(\nabla\varphi)_{ij} = \dfrac{\partial\varphi}{\partial q_{ij}}$。

证明。我们只需证明，对于任意 $\mathbf{Q} \in O(n)$ 以及 $\mathbf{Q}$ 处的任意切向量，(6.2) 都成立。根据引理 6.1，$\mathbf{Q}$ 处的切向量可表示为 $\mathbf{Q\Omega}$，其中 $\mathbf{\Omega}$ 是反对称矩阵。因此，对于任意反对称矩阵 $\mathbf{\Omega}$，我们有

$$\langle\nabla\varphi, \mathbf{Q\Omega}\rangle_F = \langle\mathcal{P}_{T_{\mathbf{Q}}O(n)}(\nabla\varphi),\ \mathbf{Q\Omega}\rangle_F + \langle\mathcal{P}_{T_{\mathbf{Q}}O(n)^\perp}(\nabla\varphi),\ \mathbf{Q\Omega}\rangle_F. \tag{6.7}$$

注意，

$$\begin{aligned}\langle\mathcal{P}_{T_{\mathbf{Q}}O(n)^\perp}(\nabla\varphi),\ \mathbf{Q\Omega}\rangle_F &= \langle\mathbf{Q}(\mathbf{Q}^T\nabla\varphi + (\nabla\varphi)^T\mathbf{Q}),\ \mathbf{Q\Omega}\rangle_F\\ &= \langle\mathbf{Q}^T\nabla\varphi + (\nabla\varphi)^T\mathbf{Q},\ \mathbf{\Omega}\rangle_F\\ &= 0,\end{aligned}$$

since $\langle \mathbf{M}, \mathbf{N}\rangle_F = 0$ ，当 $\mathbf{M}$ 为对称矩阵而 $\mathbf{N}$ 为反对称矩阵时。因此 e,

$$\langle\nabla\varphi, \mathbf{Q\Omega}\rangle_F = \langle\mathcal{P}_{T_{\mathbf{Q}}O(n)}(\nabla\varphi),\ \mathbf{Q\Omega}\rangle_F = \langle\mathcal{P}_{T_{\mathbf{Q}}O(n)}(\nabla\varphi),\ \mathbf{Q\Omega}\rangle_{g_e}.$$

因此 $\nabla_{g_e}\varphi = \mathcal{P}_{T_{\mathbf{Q}}O(n)}(\nabla\varphi)$。 □

在这个全局坐标下，$(O(n),\ g_e)$ 上的散度也可以显式计算。 ij 的 $\nabla_{g_e}\varphi$ 项为

$$(\nabla_{g_e}\varphi)_{ij} = \frac{1}{2}\left(\frac{\partial\varphi}{\partial q_{ij}} - \sum_{k,l} q_{ik}q_{lj}\frac{\partial\varphi}{\partial q_{lk}}\right). \tag{6.8}$$

因此，给定一个切向量场 $\mathbf{H}(\mathbf{Q}) = (h_{ij}(\mathbf{Q}))$，其散度定义为

$$\nabla_{g_e}\cdot\mathbf{H}(\mathbf{Q}) := \sum_{i,j}(\nabla_{g_e}(h_{ij}(\mathbf{Q})))_{ij}. \tag{6.9}$$

Laplace-Beltrami算子定义如下:

$$\Delta_{g_e}\varphi := \nabla_{g_e}\cdot\nabla_{g_e}\varphi. \tag{6.10}$$

Laplace-Beltrami算子的显式表达式由以下引理给出:

引理 6.3。（Laplace-Beltrami算子）$(O(n),\ g_e)$ 上的 Laplace-Beltrami算子为

$$\Delta_{g_e}\varphi = \frac{1}{2}\left(\sum_{i,j}\frac{\partial^2\varphi}{\partial q_{ij}^2} - (n-1)\sum_{i,j}q_{ij}\frac{\partial\varphi}{\partial q_{ij}} - \sum_{i,j,k,l}q_{ik}q_{lj}\frac{\partial^2\varphi}{\partial q_{ij}\partial q_{lk}}\right). \tag{6.11}$$

证明。根据散度和梯度的表达式，我们有

$$\Delta_{g_e}\varphi = \sum_{i,j}\left(\frac{\partial h_{ij}}{\partial q_{ij}} - \sum_{k,l}q_{ik}q_{lj}\frac{\partial h_{ij}}{\partial q_{lk}}\right),\ h_{ij} = \sum_{i,j}\left(\frac{\partial\varphi}{\partial q_{ij}} - \sum_{k,l}q_{ik}q_{lj}\frac{\partial\varphi}{\partial q_{lk}}\right).$$

Thus

$$\begin{aligned}\frac{\partial h_{ij}}{\partial q_{ij}} &= \frac{1}{2}\left(\frac{\partial^2\varphi}{\partial q_{ij}^2} - \sum_{k,l}\delta_{kl}q_{lj}\frac{\partial\varphi}{\partial q_{lk}} - \sum_{k,l}\delta_{il}q_{ik}\frac{\partial\varphi}{\partial q_{lk}} - \sum_{k,l}q_{ik}q_{lj}\frac{\partial^2\varphi}{\partial q_{ij}q_{lk}}\right)\\ &= \frac{1}{2}\left(\frac{\partial^2\varphi}{\partial q_{ij}^2} - \sum_{l}q_{lj}\frac{\partial\varphi}{\partial q_{lj}} - \sum_{k}q_{ik}\frac{\partial\varphi}{\partial q_{ik}} - \sum_{k,l}q_{ik}q_{lj}\frac{\partial^2\varphi}{\partial q_{ij}q_{lk}}\right),\end{aligned}$$

and

$$\frac{\partial h_{ij}}{\partial q_{lk}} = \frac{1}{2}\left(\frac{\partial^2\varphi}{\partial q_{ij}\partial q_{lk}} - \sum_{k',l'}\delta_{il}\delta_{kk'}q_{l'j}\frac{\partial\varphi}{\partial q_{l'k'}} - \sum_{k',l'}\delta_{ll'}\delta_{kj}q_{ik'}\frac{\partial\varphi}{\partial q_{l'k'}} - \sum_{k',l'}q_{ik'}q_{l'j}\frac{\partial^2\varphi}{\partial q_{l'k'}\partial q_{lk}}\right).$$

将上述公式代入拉普拉斯算子，可得（6.11）。 □

6.2. 引理证明与省略的计算。

6.2.1. 第 3 节。

引理 3.1 的证明。根据 (2.1) 和 (2.2)，直接计算可得 6.12

$$\begin{aligned}
\|\mathbf{W}(k)\|_F^2 = {} & \|\mathbf{W}(k-1)\|_F^2 + 2\eta\cdot\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)) \\
& - 2\eta\cdot\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)) \\
& + \eta^2\cdot\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{A}(k)^2\mathbf{W}(k-1)) \\
& - 2\eta^2\cdot\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)) \\
& + \eta^2\cdot\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)) \\
& - 2\eta^2\cdot\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))) \\
& + \eta^2\cdot\mathrm{tr}(\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))^T\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))).
\end{aligned}$$

根据 (2.1) 中 $\mathbf{\Sigma}$ 的定义，我们有

$$\begin{aligned}
\|\mathbf{\Sigma}(\mathbf{A}(\mathbf{k}),\mathbf{W}(\mathbf{k-1}))\|_F &= \sqrt{\sum_{i\neq j}(\mathbf{w}_i(k-1)^T\mathbf{A}(k)\mathbf{w}_j(k-1))^2} \\
&= \sqrt{\sum_{i\neq j}(\mathbf{w}_i(k-1)^T\mathbf{x}(k))^2(\mathbf{w}_j(k-1)^T\mathbf{x}(k))^2} \\
&\leq \sum_{i=1}^{n}(\mathbf{w}_i(k-1)^T\mathbf{x}(k))^2 \\
&\leq M^2\sum_{i=1}^{n}\|\mathbf{w}_i(k-1)\|_2^2 \\
&= M^2\|\mathbf{W}(k-1)\|_F^2.
\end{aligned} \tag{6.13}$$

根据以下范数不等式： $\|\mathbb{M}\|_2 \leq \|\mathbb{M}\|_F \leq \sqrt{n}\|\mathbb{M}\|_2$、(2.6) 和 (6.13)，我们几乎必然地得到上述等式中各项的以下估计：

$$\begin{aligned}
\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)) = \|\mathbf{W}(k-1)^T\mathbf{x}(k)\|_2^2 &\leq \|\mathbf{W}(k-1)^T\|_2^2\|\mathbf{x}(k)\|_2^2 \\
&\leq M^2\|\mathbf{W}(k-1)\|_F^2.
\end{aligned} \tag{6.14}$$

$$\begin{aligned}
\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{A}(k)^2\mathbf{W}(k-1)) &= \|\mathbf{x}(k)\|_2^2\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)) \\
&\leq M^4\|\mathbf{W}(k-1)\|_F^2.
\end{aligned} \tag{6.15}$$

(6.16)
$$\begin{aligned}
&\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1))\\
&=\mathrm{tr}((\mathbf{W}(k-1)^T\mathbf{x}(k))(\mathbf{x}(k)^T\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)))\\
&\le\|\mathbf{W}(k-1)^T\mathbf{x}(k)\|_2\|\mathbf{x}(k)^T\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)\|_2\\
&\le M^2\|\mathbf{W}(k-1)\|_2^6\|\mathbf{A}(k)\|_2\\
&\le M^4\|\mathbf{W}(k-1)\|_F^6.
\end{aligned}$$

(6.17)
$$\begin{aligned}
&|\mathrm{tr}(\mathbf{W}(k-1)^T\mathbf{A}(k)\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1)))|\\
&=|\mathrm{tr}((\mathbf{W}(k-1)^T\mathbf{x}(k))(\mathbf{x}(k)^T\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))))|\\
&\le\|(\mathbf{W}(k-1)^T\mathbf{x}(k)\|_2\|\mathbf{x}(k)^T\mathbf{W}(k-1)\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))\|_2\\
&\le M^2\|\mathbf{W}(k-1)\|_2^4\|\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))\|_2\\
&\le M^2\|\mathbf{W}(k-1)\|_F^4\|\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))\|_F\\
&\le M^4\|\mathbf{W}(k-1)\|_F^6.
\end{aligned}$$

(6.18)
$$\begin{aligned}
&\mathrm{tr}(\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))^T\mathbf{W}(k-1)^T\mathbf{W}(k-1)\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1)))\\
&=\|\mathbf{W}(k-1)\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))\|_F^2\\
&\le n\|\mathbf{W}(k-1)\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))\|_2^2\\
&\le n\|\mathbf{W}(k-1)\|_2^2\|\mathbf{\Sigma}(\mathbf{A}(k),\mathbf{W}(k-1))\|_2^2.\\
&\le nM^4\|\mathbf{W}(k-1)\|_F^6.
\end{aligned}$$

将 (6.14) $\sim$ (6.18) 代入 (6.12) 得

(6.19)
$$\|\mathbf{W}(k)\|_F^2\le(1+2M^2\eta)\|\mathbf{W}(k-1)\|_F^2+\eta^2(M^4\|\mathbf{W}(k-1)\|_F^2+(n+3)M^4\|\mathbf{W}(k-1)\|_F^6).$$

选择 $\eta(T)$ 作为

(6.20)
$$\eta(T)=\frac{1}{M^4(1+(n+3)r^4e^{2T(2M^2+1)})}.$$

然后我们可以证明，对于任意 $\eta\le\eta(T)$，由 (??) 生成且从 $\mathbf{W_0}$ 开始的马尔可夫链满足

(6.21)
$$\|\mathbf{W}(k)\|_F^2\le r^2e^{2M^2+1},k=1,2,...,[T/\eta],\ \text{a.s.}.$$

我们通过归纳法证明。给定任意 $\eta\le\eta(T)$，若 $\|\mathbf{W}(k-1)\|_F^2\le r^2e^{T(2M^2+1)}$，则

$$\begin{aligned}
\|\mathbf{W}(k)\|_F^2&\le(1+2M^2\eta)\|\mathbf{W}(k-1)\|_F^2+\eta^2(M^4\|\mathbf{W}(k-1)\|_F^2+(n+3)M^4\|\mathbf{W}(k-1)\|_F^6)\\
&\le(1+2M^2\eta)\|\mathbf{W}(k-1)\|_F^2+(\eta M^4(1+(n+3)r^4e^{2T(2M^2+1)}))\cdot\eta\cdot\|\mathbf{W}(k-1)\|_F^2\\
&\le(1+(2M^2+1)\eta)\|\mathbf{W}(k-1)\|_F^2.
\end{aligned}$$

记住 $\|\mathbf{W_0}\|_F^2\le r^2$，因此 $\|\mathbf{W_0}\|_F^2\le r^2e^{2M^2+1}$，于是对于任意 $k=1,2,...,[T/\eta]$,

$$\|\mathbf{W}(k)\|_F^2\le(1+(2M^2+1)\eta)^kr^2\le(1+(2M^2+1)\eta)^{T/\eta}r^2\le r^2e^{T(2M^2+1)}.$$

因此，取 $C(T)=r^2e^{T(2M^2+1)}$ 即可完成证明。 □

引理 3.5 的证明。由于 M $= (M_{ijkl})$ 是 G 的协方差矩阵，因此它是半正定的。因此，M 的平方根是良定义的，并满足上述 (i)、(ii) 和 (iii)。将其记为 N。

当 W $\in O(n)$ 时，我们已证明，对于任意对称矩阵 B $\in \mathbb{R}^{n\times n}$，G(B, W) $\in T_{\mathbf{W}}O(n)$，因此

$$\mathcal{P}_{T_{\mathbf{W}}O(n)}\mathbf{G}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W}) = \mathbf{G}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W}),$$

即 $P_{ijkl}g_{kl} = g_{ij}$ 根据引理 3.2。因此，根据期望的线性性和对称性

$$\begin{aligned} M_{ijkl} &= \mathbb{E}[g_{ij}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{kl}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})] \\ &= \mathbb{E}[P_{ijrs}g_{rs}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{xy}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})P_{klxy}] \\ &= P_{ijrs}\mathbb{E}[g_{rs}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{xy}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})]P_{xykl} \\ &= P_{ijrs}M_{rsxy}P_{xykl}, \end{aligned}$$

或者用矩阵记法，M = PMP。因此根据引理 3.2 中的 (iii)，我们有 e

$$P_{ijrs}M_{rskl} = P_{ijrs}P_{rsuv}M_{uvxy}P_{xykl} = P_{ijrs}P_{uvrs}M_{uvxy}P_{xykl} = P_{ijuv}M_{uvxy}P_{xykl},$$

或者 PM =PMP。同理,

$$M_{ijrs}P_{rskl} = P_{ijuv}M_{uvxy}P_{xyrs}P_{rskl} = P_{ijuv}M_{uvxy}P_{xykl},$$

或 MP =PMP。因此

$$M_{ijrs}P_{rskl} = P_{ijrs}M_{rskl}.$$

因此，P 与 M 可交换。由于 P 是对称的，所以 P 也与 M 的平方根可交换。因此

$$N_{ijrs}P_{rskl} = P_{ijrs}N_{rskl}.$$

Thus

$$P_{ijrs}N_{rsxy}P_{xyuv}N_{uvkl} = P_{ijrs}N_{rsxy}N_{xyuv}P_{uvkl} = P_{ijrs}M_{rsuv}P_{uvkl} = M_{ijkl},$$

或 PNPN = PNNP = PMP = M。因此 PN 也是 M 的平方根，根据唯一性,

$$P_{ijrs}N_{rskl} = N_{ijkl},$$

即 PN =N。 □

定理 3.2 中的计算。直接计算可得

(6.22)

$$\begin{aligned} S\varphi(\mathbf{W}) &= \mathbb{E}\left[\varphi(\mathbf{W}+\eta\mathbf{G}(\mathbf{x}\mathbf{x}^T,\mathbf{W}))\right] \\ &= \varphi(\mathbf{W}) + \left(g_{ij}(\mathbf{A},\mathbf{W})\frac{\partial\varphi}{\partial w_{ij}}\right)\eta + \left(\frac{1}{2}\mathbb{E}\left[g_{ij}(\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{kl}(\mathbf{x}\mathbf{x}^T,\mathbf{W})\right]\frac{\partial^2\varphi}{\partial w_{ij}\partial w_{kl}}\right)\eta^2 + O(\eta^3). \end{aligned}$$

与此同时,

$$e^{\eta\mathcal{L}}\varphi(\mathbf{W}) = \varphi(\mathbf{W}) + \eta\mathcal{L}\varphi(\mathbf{W}) + \frac{1}{2}\eta^2\mathcal{L}^2\varphi(\mathbf{W}) + O(\eta^3).$$

由式（3.23）可得

$$\begin{aligned}\mathcal{L}^2\varphi &= \mathcal{L}\left[g_{ij}(\mathbf{A},\mathbf{W})\frac{\partial\varphi}{\partial w_{ij}}+O(\eta)\right]\\ &= g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial}{\partial w_{kl}}\left(g_{ij}(\mathbf{A},\mathbf{W})\frac{\partial\varphi}{\partial w_{ij}}\right)+O(\eta)\\ &= \left(g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial g_{ij}(\mathbf{A},\mathbf{W})}{\partial w_{kl}}\right)\frac{\partial\varphi}{\partial w_{ij}}+g_{ij}(\mathbf{A},\mathbf{W})g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial^2\varphi}{\partial w_{ij}\partial w_{kl}}+O(\eta).\end{aligned} \tag{6.23}$$

Thus

$$\begin{aligned}e^{\eta\mathcal{L}}\varphi(\mathbf{W}) &= \varphi(\mathbf{W})+\eta\left[\left(g_{ij}(\mathbf{A},\mathbf{W})+\eta P_{ijkl}f_{kl}+\frac{\eta}{2}J_{ij}\right)\frac{\partial\varphi}{\partial w_{ij}}+\frac{\eta}{2}K_{ijrs}K_{klrs}\frac{\partial^2\varphi}{\partial w_{ij}w_{kl}}\right]\\ &\quad+\frac{\eta^2}{2}\left[\left(g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial g_{ij}(\mathbf{A},\mathbf{W})}{\partial w_{kl}}\right)\frac{\partial\varphi}{\partial w_{ij}}+g_{ij}(\mathbf{A},\mathbf{W})g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial^2\varphi}{\partial w_{ij}\partial w_{kl}}\right]+O(\eta^3).\end{aligned} \tag{6.24}$$

将 (6.24) 按 η 的顺序重新表述，可得

$$\begin{aligned}e^{\eta\mathcal{L}}\varphi(\mathbf{W}) &= \varphi(\mathbf{W})+\left(g_{ij}(\mathbf{A},\mathbf{W})\frac{\partial\varphi}{\partial w_{ij}}\right)\eta+\\ &\quad\left[\left(P_{ijkl}f_{kl}+\frac{1}{2}J_{ij}+\frac{1}{2}g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial g_{ij}(\mathbf{A},\mathbf{W})}{\partial w_{kl}}\right)\frac{\partial\varphi}{\partial w_{ij}}\right]\eta^2\\ &\quad+\left[\left(K_{ijrs}K_{klrs}+g_{ij}(\mathbf{A},\mathbf{W})g_{kl}(\mathbf{A},\mathbf{W})\right)\frac{\partial^2\varphi}{\partial w_{ij}\partial w_{kl}}\right]\eta^2+O(\eta^3).\end{aligned} \tag{6.25}$$

比较 (6.22) 和 (6.25) 中的 η^2 项，可得

$$\begin{aligned}&\frac{\partial\varphi}{\partial w_{ij}}:P_{ijkl}f_{kl}+\frac{1}{2}J_{ij}+\frac{1}{2}g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial g_{ij}(\mathbf{A},\mathbf{W})}{\partial w_{kl}}=0;\\ &\frac{\partial^2\varphi}{\partial w_{ij}\partial w_{kl}}:K_{ijrs}K_{klrs}+g_{kl}(\mathbf{A},\mathbf{W})g_{ij}(\mathbf{A},\mathbf{W})=\mathbb{E}\left[g_{ij}(\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{kl}(\mathbf{x}\mathbf{x}^T,\mathbf{W})\right].\end{aligned}$$

因此，$\mathbf{F}=(f_{ij})_{n\times n}$ 和 $\mathbf{H}=(H_{ijkl})_{n\times n\times n\times n}$ 应满足

$$\begin{aligned}P_{ijkl}f_{kl} &= -\frac{1}{2}J_{ij}-\frac{1}{2}g_{kl}(\mathbf{A},\mathbf{W})\frac{\partial g_{ij}(\mathbf{A},\mathbf{W})}{\partial w_{kl}},\\ P_{ijuv}H_{uvrs}P_{klxy}H_{xyrs} &= M_{ijkl}=\mathbb{E}[g_{ij}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})g_{kl}(\mathbf{A}-\mathbf{x}\mathbf{x}^T,\mathbf{W})],\end{aligned}$$

即 (3.41)

□

6.2.2. 第 4 节。

引理 4.2 的证明。我们首先计算 H(W)。根据命题 3.3，我们知道 $\mathbf{W}\mathbf{W}^T=\mathbf{I_2}$。因此

$$\begin{aligned}\mathbf{G}(\mathbf{x}\mathbf{x}^\mathbf{T}-\mathbf{A},\mathbf{W}) &= (\mathbf{x}\mathbf{x}^T-\mathbf{A})\mathbf{W}-\mathbf{W}\mathbf{W}^T(\mathbf{x}\mathbf{x}^T-\mathbf{A})\mathbf{W}+\mathbf{W}\mathbf{\Sigma}(\mathbf{x}\mathbf{x}^T-\mathbf{A},\mathbf{W})\\ &= \mathbf{W}\mathbf{\Sigma}(\mathbf{x}\mathbf{x}^T-\mathbf{A},\mathbf{W}).\end{aligned}$$

记 $\mathbf{B}=\mathbf{x}\mathbf{x}-\mathbf{A}$。由于 $n=2$，我们有

$$\mathbf{\Sigma}(\mathbf{B},\mathbf{W})=\begin{pmatrix}0 & -\mathbf{w_1}\cdot\mathbf{B}\mathbf{w_2}\\ \mathbf{w_1}\cdot\mathbf{B}\mathbf{w_2} & 0\end{pmatrix}.$$

记 $b = \mathbf{w_1}^T \mathbf{B} \mathbf{w_2}$，则

$$\mathbf{G}(\mathbf{x}\mathbf{x}^\mathbf{T} - \mathbf{A}, \mathbf{W}) = \mathbf{W} \begin{pmatrix} 0 & -b \\ b & 0 \end{pmatrix} = \begin{pmatrix} w_{1,2}b & -w_{1,1}b \\ w_{2,2}b & -w_{2,1}b \end{pmatrix}.$$

Thus

$$\mathbf{vec}(\mathbf{G}(\mathbf{B}, \mathbf{W})) = (w_{1,2}b,\ -w_{1,1}b,\ w_{2,2}b,\ -w_{2,1}b)^T.$$

记 $\mathbf{u} = (w_{1,2}, \ -w_{1,1}, \ w_{2,2}, \ -w_{2,1})^T$。则式 (??) 中定义的协方差矩阵为

$$\begin{aligned} \mathbf{M}(\mathbf{W}) &= \mathbb{E}[\mathbf{vec}(\mathbf{G}(\mathbf{B}, \mathbf{W}))\mathbf{vec}(\mathbf{G}(\mathbf{B}, \mathbf{W}))^T] \\ &= \mathbb{E}[b^2]\mathbf{u}\mathbf{u}^T. \end{aligned} \tag{6.26}$$

因此，

$$\mathbf{H}(\mathbf{W}) = \sqrt{\mathbf{M}(\mathbf{W})} = \frac{\sqrt{\mathbb{E}[b^2]}}{\|\mathbf{u}\|_2}\mathbf{u}\mathbf{u}^T.$$

记住，$\mathbf{W} \in O(2)$，因此 $\|\mathbf{u}\|_2 = \sqrt{2}$。所以我们有

$$\mathbf{H}(\mathbf{W}) = \frac{\sqrt{\mathbb{E}[b^2]}}{\sqrt{2}}\mathbf{u}\mathbf{u}^T. \tag{6.27}$$

将 (6.27) 代入 (4.12)，得到

$$\mathrm{d}\mathbf{vec}(\mathbf{W}) = \mathbf{vec}(\mathbf{G}(\mathbf{A}, \mathbf{W}))\mathrm{d}t + \frac{\sqrt{\eta\mathbb{E}[b^2]}}{\sqrt{2}}\mathbf{u}\mathbf{u}^T \circ \mathrm{d}\mathbf{B}. \tag{6.28}$$

此外，由于 $\|\mathbf{u}\|_2 = \sqrt{2}$，且 $\mathbf{B}(t)$ 是 $\mathbb{R}^4$ 中的标准布朗运动，因此 $\mathbf{u}^T\mathrm{d}\mathbf{B}$ 与 $\sqrt{2}\mathrm{d}B$ 具有相同的分布，其中 $B(t)$ 是一维标准布朗运动。因此，(4.12) 也可以改写为

$$\mathrm{d}\mathbf{vec}(\mathbf{W}) = \mathbf{vec}(\mathbf{G}(\mathbf{A}, \mathbf{W}))\mathrm{d}t + \sqrt{\eta\mathbb{E}[b^2]}\mathbf{u} \circ \mathrm{d}B(t). \tag{6.29}$$

此外，由于 $\mathbf{W}(t) \in O(2)$，因此 $\mathbf{G}(\mathbf{A}, \mathbf{W}) = \mathbf{F}(\mathbf{W})$，其中 $\mathbf{F}$ 在 (2.4) 中定义。因此，(4.12) 可直接转换为矩阵形式：

$$\mathrm{d}\mathbf{W} = \mathbf{F}(\mathbf{W})\mathrm{d}t + \sqrt{\eta\mathbb{E}[b^2]} \cdot \mathbf{Q} \circ \mathrm{d}\mathbf{Z}, \tag{6.30}$$

其中 $\mathbf{Z}$ 在 (4.14) 中定义。

现在我们计算 $\mathbb{E}[b^2]$。直接计算可得

$$b = \mathbf{w_1}^T\mathbf{B}\mathbf{w_2} = b_{1,1}w_{1,1}w_{1,2} + b_{1,2}(w_{1,1}w_{2,2} + w_{2,1}w_{1,2}) + b_{2,2}w_{2,1}w_{2,2}.$$

Thus

$$\begin{aligned} b^2 = {} & b_{1,1}^2 w_{1,1}^2 w_{1,2}^2 + b_{1,2}^2(w_{1,1}w_{2,2} + w_{1,2}w_{2,1})^2 + b_{2,2}^2 w_{2,1}^2 w_{2,2}^2 + 2b_{1,1}b_{2,2}w_{1,1}w_{1,2}w_{2,1}w_{2,2} \\ & + 2b_{1,1}b_{1,2}w_{1,1}w_{2,2}w_{1,2} + 2b_{1,1}b_{1,2}w_{1,1}w_{1,2}^2w_{2,1} + 2b_{1,2}b_{2,2}w_{1,1}w_{2,1}w_{2,2}^2 + 2b_{1,2}b_{2,2}w_{1,2}w_{2,1}w_{2,2}。 \end{aligned}$$

同时，我们知道 $b_{i,j} = x_i x_j - \mathbb{E}[x_i x_j]$ 对于 $1 \le i, j \le 2$，因此对于任意 $i, j, i', j' = 1, 2$，我们有

$$\mathbb{E}[b_{i,j}b_{i',j'}] = \mathbb{E}[x_i x_j x_{i'} x_{j'}] - \mathbb{E}[x_i x_j]\mathbb{E}[x_{i'} x_{j'}].$$

将其代入 b^2 的表达式，得到

$$\begin{aligned} \mathbb{E}[b^2] = {} & \mathrm{var}(x_1^2)w_{1,1}^2 w_{1,2}^2 + \mathrm{var}(x_1x_2)(w_{1,1}^2w_{2,2}^2 + w_{2,1}^2w_{1,2}^2 + 2w_{1,1}w_{1,2}w_{2,1}w_{2,2}) + \mathrm{var}(x_2^2)w_{2,1}^2w_{2,2}^2 \\ & + 2(\mathbb{E}[x_1^2x_2^2] - \mathbb{E}[x_1^2]\mathbb{E}[x_2^2])w_{1,1}w_{1,2}w_{2,1}w_{2,2} + 2(\mathbb{E}[x_1^3x_2] - \mathbb{E}[x_1^2]\mathbb{E}[x_1x_2])w_{1,1}w_{1,2}(w_{1,1}w_{2,2} + w_{1,2}w_{2,1}) \\ & + 2(\mathbb{E}[x_1x_2^3] - \mathbb{E}[x_2^2]\mathbb{E}[x_1x_2])w_{2,1}w_{2,2}(w_{1,1}w_{2,2} + w_{1,2}w_{2,1}). \end{aligned}$$

利用 $w^2_{1,\ 1} = w^2_{2,\ 2}$，$w^2_{1,\ 2} = w^2_{2,\ 1}$，我们进一步推导得

$$\mathbb{E}[b^2] = c_1(\mathbf{W})(w_{1,1}^4 + w_{1,2}^4) + c_2(\mathbf{W})w_{1,1}^2 w_{1,2}^2 + c_3(\mathbf{W})w_{1,1}w_{1,2}(w_{1,1}w_{2,2} + w_{1,2}w_{2,1}), \tag{6.31}$$

其中 c_1、c_2 和 c_3 在 (4.13) 中定义。因此，取 $c(\mathbf{W}) = \sqrt{\mathbb{E}[b^2]}$，可得，从而证明得证。□(4.12)

引理 4.3 的证明。我们首先用 It´o 意义下的形式写出 (4.12)。由于我们已假设 $|\mathbf{W}(t)| = 1$，因此只需考虑关于 $w_{1,1}$ 和 $w_{1,2}$ 的方程。在 It´o 意义下，有

$$\begin{aligned}
\mathrm{d}w_{1,1} &= [(\mathbf{F}(\mathbf{W}))_{1,1} + h_1(\mathbf{W})]\mathrm{d}t + \sqrt{\eta}\cdot c(\mathbf{W})w_{1,2}\mathrm{d}B \\
\mathrm{d}w_{1,2} &= [(\mathbf{F}(\mathbf{W}))_{1,2} + h_2(\mathbf{W})]\mathrm{d}t - \sqrt{\eta}\cdot c(\mathbf{W})w_{1,1}\mathrm{d}B,
\end{aligned}$$

其中 h_1 和 h_2 在 3.21 中定义。记 $g_1 = \sqrt{\eta}c(\mathbf{W})w_{1,2}$、$g_2 = -\sqrt{\eta}c(\mathbf{W})w_{1,1}$，则 h_1 和 h_2 可计算为

$$\begin{aligned}
h_1(\mathbf{W}) &= \frac{1}{2}\left(g_1\frac{\partial g_1}{\partial w_{1,1}} + g_2\frac{\partial g_1}{\partial w_{1,2}}\right) = \frac{\eta}{2}\left(\frac{1}{2}w_{1,2}^2\frac{\partial c(\mathbf{W})^2}{\partial w_{1,1}} - \frac{1}{2}w_{1,1}w_{1,2}\frac{\partial c(\mathbf{W}^2)}{\partial w_{1,2}} - w_{1,1}c(\mathbf{W})^2\right) \\
h_2(\mathbf{W}) &= \frac{1}{2}\left(g_1\frac{\partial g_2}{\partial w_{1,1}} + g_2\frac{\partial g_2}{\partial w_{1,2}}\right) = \frac{\eta}{2}\left(\frac{1}{2}w_{1,1}^2\frac{\partial c(\mathbf{W})^2}{\partial w_{1,2}} - \frac{1}{2}w_{1,1}w_{1,2}\frac{\partial c(\mathbf{W}^2)}{\partial w_{1,1}} - w_{1,2}c(\mathbf{W})^2\right).
\end{aligned} \tag{6.32}$$

这里用到了链式法则：$c\dfrac{\partial c}{\partial w} = \dfrac{1}{2}\dfrac{\partial c^2}{\partial w}$。将上述方程代入 It´o 意义下的随机微分方程，得到

$$\begin{aligned}
\mathrm{d}w_{1,1} = &\left((\mathbb{E}[x_2^2] - \mathbb{E}[x_1^2])w_{1,1}w_{1,2} + \frac{\eta}{4}w_{1,2}^2\frac{\partial c^2}{\partial w_{1,1}} - \frac{\eta}{4}w_{1,1}w_{1,2}\frac{\partial c^2}{\partial w_{1,2}} - \frac{\eta}{2}w_{1,1}c(\mathbf{W})^2\right)\mathrm{d}t \\
&+ \sqrt{\eta}c(\mathbf{W})w_{1,2}\mathrm{d}B.
\end{aligned} \tag{6.33}$$

现在考虑 过程 θ。假设 $\theta(t)$满足以下 Ito 意义下的随机微分方程 :

$$\mathrm{d}\theta = f(\theta)\mathrm{d}t + \sqrt{\eta}\cdot g(\theta)\mathrm{d}B.$$

于是，由伊藤等距可得

$$\begin{aligned}
\mathrm{d}\cos\theta &= -\sin\theta\cdot\mathrm{d}\theta - \frac{\eta g^2(\theta)}{2}\mathrm{d}t \\
&= \left(-\sin\theta\cdot f(\theta) - \frac{\eta\cos\theta}{2}g^2(\theta)\right)\mathrm{d}t - \sqrt{\eta}\cdot\sin(\theta)g(\theta)\mathrm{d}B.
\end{aligned}$$

在 (6.33) 中用 $\cos\theta$ 替换 $w_{1,1}$，用 $\sin\theta$ 替换 $w_{1,2}$，并将其系数与上式比较，可得

$$\begin{aligned}
&g(\theta) = -c(\theta), \\
&f(\theta) = (\mathbb{E}[x_2^2] - \mathbb{E}[x_1^2])\sin\theta\cos\theta + \eta\cdot\frac{2c_1(\theta) - c_2(\theta)}{2}(\cos^3\theta\sin\theta - \cos\theta\sin^3\theta) \\
&\qquad + \eta\cdot\frac{3c_3(\theta)}{2}\cos^2\theta\sin^2\theta.
\end{aligned} \tag{6.34}$$

这里我们使用了 (6.31)，因为 $c^2(\mathbf{W}) = \mathbb{E}[b^2]$。这正是 (4.18) 中的随机微分方程。 □

定理 4.1 的证明。在下面的证明中，C 只是一个一般常数，在不同的方程中可能有所变化。因为 $\rho(x,t)$ $t \geq 0$ 是 $\theta(t)$ 的分布，所以 ρ 在 $\mathbb{T} \times \mathbb{R}^+$ 上满足 (4.22)。因此，根据周期边界条件，

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{T}} \rho(x,t)\mathrm{d}x = \frac{1}{2}\int_{\mathbb{T}} \left[-\frac{\partial(f(x)\rho(x,t))}{\partial x} + \frac{1}{2}\cdot\frac{\partial^2(\eta g^2(x)\rho(x,t))}{\partial x^2}\right]\mathrm{d}x = 0.$$

因此 $\int_{\mathbb{T}} \rho(x,t)\mathrm{d}x = 1,\ t \geq 0$。

现在考虑 Fokker-Planck算子 $\mathcal{L}_1^*$，它在 $L^2(\mathbb{T}, \mathrm{d}x/\rho_\infty)$ 中是自伴的：

$$\mathcal{L}_1^* : D(\mathcal{L}_1^*) \subset L^2(\mathbb{T}, \mathrm{d}x/\rho_\infty) \to L^2(\mathbb{T}, \mathrm{d}x/\rho_\infty),\ \mathcal{L}_1^* p := -\frac{\mathrm{d}}{\mathrm{d}x}\left(\rho_\infty(x)\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{p(x)}{\rho_\infty(x)}\right)\right).$$

这里 $D(\mathcal{L}_1^*) = H^2(\mathbb{T}, \mathrm{d}x/\mu)$。直接计算可得

$$\langle \mathcal{L}1^* p, q\rangle_{\mathrm{d}x/\rho_\infty} = \int_{\mathbb{T}} \rho_\infty(x)\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{p(x)}{\rho_\infty(x)}\right)\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{q(x)}{\rho_\infty(x)}\right)\mathrm{d}x,$$

因此 $\mathcal{L}_1^*$ 是半正定的，并且

$$\mathcal{L}_1^* p = 0 \iff p(x) = c\rho_\infty(x).$$

所以 0 是 $\mathcal{L}_1^*$ 的单重主特征值，其特征向量为 $\rho_\infty(x)$。

此外，0 是孤立的。对于任意 $\lambda > 0$，我们证明 $(\lambda + \mathcal{L}_1^*)^{-1}$ 是紧的。令 $\{g_n\}_{n=1}^\infty \in L^2(\mathbb{T}, \mathrm{d}x/\rho_\infty)$ 为一个有界序列，其中

$$(\lambda + \mathcal{L}^*)u_n = g_n, n = 1, 2, ...$$

要证明 $(\lambda+\mathcal{L}^*)^{-1}$ 是紧的，只需证明 $\{u_n\}_{n=1}^\infty$ 存在一个子序列，该子序列在 $L^2(\mathbb{T}, \mathrm{d}x/\rho_\infty)$ 中是柯西序列。由于 $\mathcal{L}^*$ 是半正定的，因此 $(\lambda + \mathcal{L}^*)$ 有界，进而 $\{u_n\}_{n=1}^\infty$ 在 $L^2(\mathbb{T}, \mathrm{d}x/\rho_\infty)$ 中有界。根据柯西-施瓦茨不等式，我们有

$$\langle \mathcal{L}_1^* u_n, u_n\rangle_{\mathrm{d}x/\rho_\infty} = \langle u_n, g_n\rangle_{\mathrm{d}x/\rho_\infty} - \lambda\|u_n\|^2_{L^2(\mathbb{T},\mathrm{d}x/\rho_\infty)} \leq C. \tag{6.35}$$

这里 C 是一个常数。因此

$$\begin{aligned}\|u_n/\rho_\infty\|^2_{H^1(\mathbb{T},\rho_\infty \mathrm{d}x)} &= \int_{\mathbb{T}} \rho_\infty(x)\left(\frac{\mathrm{d}}{\mathrm{d}x}\frac{u_n(x)}{\rho_\infty(x)}\right)^2\mathrm{d}x + \int_{\mathbb{T}} \rho_\infty(x)\left(\frac{u_n(x)}{\rho_\infty(x)}\right)^2\mathrm{d}x \\ &= \|u_n\|_{L^2(\mathbb{T},\mathrm{d}x/\rho_\infty)} + \langle \mathcal{L}_1^* u_n, u_n\rangle_{\mathrm{d}x/\rho_\infty} \leq C.\end{aligned}$$

因此 u_n/ρ_∞ 在 $H^1(\mathbb{T}, \rho_\infty \mathrm{d}x)$ 中有界。根据紧嵌入 $H^1(\mathbb{T}, \rho_\infty \mathrm{d}x) \subset\subset L^2(\mathbb{T}, \rho_\infty \mathrm{d}x)$，我们知道 u_n 存在一个子序列（仍记为 u_n），使得

$$\frac{u_n}{\rho_\infty} \to \frac{u^*}{\rho_\infty} \text{ in } L^2(\mathbb{T}, \rho_\infty \mathrm{d}x),$$

或等价地

$$u_n \to u^* \text{ in } L^2(\mathbb{T}, \mathrm{d}x/\rho_\infty).$$

因此 $(\lambda + \mathcal{L}^*)^{-1}$ 是一个紧算子。因此，$(\lambda + \mathcal{L}^*)^{-1}$ 的谱仅以 0 为聚点。所以 0 是 $\mathcal{L}^*$ 的谱中的孤立点。因此，对于任意$p \in L^2(\mathbb{T}, \mathrm{d}x/\rho_\infty)$，只要满足 $\int_{\mathbb{T}} p(x)\mathrm{d}x = 0$，就有以下庞加莱不等式：存在一个常数 $c > 0$ 使得

$$\int_{\mathbb{T}} \rho_\infty(x)\left(\frac{\mathrm{d}}{\mathrm{d}x}\frac{p(x)}{\rho_\infty(x)}\right)^2\mathrm{d}x = \langle \mathcal{L}_1^* p, p\rangle_{L^2(\mathbb{T},\mathrm{d}x/\rho_\infty)} \geq c\|p\|^2_{L^2(\mathbb{T},\mathrm{d}x/\rho_\infty)} = c\int_{\mathbb{T}} \frac{|p(x)|^2}{\rho_\infty(x)}\mathrm{d}x. \tag{6.36}$$

将 $\rho(x,t)/\rho_\infty(x) - 1$ 乘到 (4.22) 两边，并将 $p = \rho(x,t) - \rho_\infty(x)$ 代入 (6.36)，可得

(6.37)
$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{T}}\frac{(\rho(x,t)-\rho_\infty(x))^2}{\rho_\infty(x)}\mathrm{d}x = -\int_{\mathbb{T}}\rho_\infty(x)\left(\frac{\mathrm{d}}{\mathrm{d}x}\frac{\rho(x,t)}{\rho_\infty(x)}\right)^2\mathrm{d}x \le -c\int_{\mathbb{T}}\frac{|\rho(x,t)-\rho_\infty(x)|^2}{\rho_\infty(x)}\mathrm{d}x.$$

因此，根据格朗沃尔不等式，式（4.25）成立。 □

References

[1] Vincent D Blondel、Alexandre Megretski 和 Vincent DD Blondel。《数学系统与控制理论中的未解问题》。普林斯顿大学出版社，新泽西州普林斯顿，2004 年。[2] Roger W Brockett。对列表进行排序、对角化矩阵并求解线性规划问题的动力系统。《线性代数及其应用》，146：79–91，1991 年。[3] Yuanyuan Feng、Lei Li 和 Jian-Guo Liu。随机梯度下降与在线主成分分析的半群：性质与扩散近似。《数学科学通讯》，16(3)，2018 年。[4] Uwe Helmke 和 John B Moore。《优化与动力系统》。施普林格科学与商业媒体，2012 年。[5] Elton P Hsu。《流形上的随机分析》。第 38 卷。美国数学会，2002 年。[6] Peter D Lax 和 Robert D Richtmyer。线性有限差分方程稳定性综述。《纯粹与应用数学通讯》，9(2)：267–293，1956 年。[7] Chris Junchi Li、Mengdi Wang、Han Liu 和 Tong Zhang。在线主成分估计的扩散近似与全局收敛。《神经信息处理系统进展》，第 30 卷，2017 年。[8] Qianxiao Li、Cheng Tai 和 Weinan E。随机修正方程与自适应随机梯度算法。载于《国际机器学习会议》，第 2101–2110 页。PMLR，2017 年。[9] Xin Liang、Zhen-Chen Guo、Ren-Cang Li 和 Wen-Wei Lin。用于在线主子空间估计的近乎最优随机逼近。arXiv 预印本 arXiv:1711.06644，2017 年。[10] Jian-Guo Liu 和 Zibu Liu。Oja 在线主成分流的收敛性。arXiv 预印本 arXiv:2202.11308，2022 年。[11] Grigorii Noikhovich Milstein。《随机微分方程的数值积分》，第 313 卷。施普林格科学与商业媒体，1994 年。[12] Erkki Oja。作为主成分分析器的简化神经元模型。《数学生物学杂志》，15(3)：267–273，1982 年。[13] Erkki Oja。主成分、次成分与线性神经网络。《神经网络》，5(6)：927–935，1992 年。[14] Erkki Oja 和 Juha Karhunen。随机矩阵期望的特征向量和特征值的随机逼近。《数学分析与应用杂志》，106(1)：69–84，1985 年。[15] Bernt Oksendal。《随机微分方程：导论及其应用》。施普林格科学与商业媒体，2013 年。[16] Christian Soize。《随机动力系统的福克-普朗克方程及其显式稳态解》，第 17 卷。世界科学出版社，1994 年。[17] Michael Eugene Taylor。《偏微分方程。第 1 卷：基础理论》。施普林格，1996 年。[18] Wei-Yong Yan、Uwe Helmke 和 John B Moore。神经网络 Oja 流的全局分析。《IEEE 神经网络汇刊》，5(5)：674–683，1994 年。

Department of Mathematics and Department of Physics, Duke University, Durham, NC电子邮箱地址：jliu@math.duke.edu

Department of Mathematics, Duke University, Durham, NC电子邮箱地址：zibu.liu@duke.edu