arXiv:2304.00392v1 [math.OC] 1 Apr 2023

最优传输粒子滤波器

Mohammad Al-Jarrah*、Bamdad Hosseini†、Amirhossein Taghvaei*

摘要—本文探讨一种称为最优传输粒子滤波器（OTPF）的新型非线性滤波算法的理论与计算发展。该算法基于近期提出的贝叶斯法则变分形式，旨在通过求解随机优化问题，寻找先验分布与后验分布之间的 Brenier最优传输映射。在理论方面，我们结合现有的粒子滤波器误差分析方法与最优传输映射估计的稳定性结果，得到了滤波器性能关于求解变分问题时优化间隙的一致误差界。误差分析揭示了偏差-方差权衡，这最终可用于理解在这些滤波器中是否以及何时能够避免维数灾难。在计算方面，我们将所提出的算法应用于一个非线性滤波示例，并与集合卡尔曼滤波器（EnKF）和序贯重要性重采样（SIR）粒子滤波器进行比较。

I. 引言

最优传输（OT）理论近年来受到了广泛关注，因为它为概率分布的分析与操控提供了自然的几何和数学工具 [1], [51], [36]。其中，两个几何概念尤为重要：（i）用于衡量概率分布之间相似性或差异性的度量，即Wasserstein度量；以及（ii）用于将一个分布传输至另一个分布的映射，即 OT 或 Monge映射。与信息论中的对应方法（例如 Kullback-Liebler（KL）散度）相比，这些最优传输度量能够尊重数据的几何结构，并且通常对扰动和误差更加稳健。得益于这些独特特性，最优传输度量和映射已成功应用于各种领域，包括生成建模和采样 [4], [47],领域自适应 [11],以及图像处理 [24], [15], [31], [21], [41]。

鉴于最优传输理论的重要性，越来越多的研究开始将最优传输工具应用于非线性滤波和贝叶斯推断 [17], [9], [35], [46]。其核心思想是将滤波／贝叶斯更新步骤视为一个传输问题，即把（当前状态的）先验分布传输到（未来状态的）后验分布。这一视角促成了新的贝叶斯推断算法的发展，包括采用多项式、径向基函数和神经网络参数化来学习三角传输映射，以从后验分布中采样 [17],[28], [25], 集合变换粒子滤波器 [34], 以及最优传输对反馈粒子滤波算法的解释 [52],[44], [45], [46]。

本文以作者近期的研究为基础 [43] ，该研究提出了一种关于贝叶斯定律的基于最优传输的变分形式，用于学习任意观测信号值下从先验分布到后验分布的最优传输映射。在该形式中，给定观测 Y时隐藏随机变量 X 的条件分布 $P_{X|Y}$被表示为$P_{X|Y}(\cdot|y) = \nabla_x f(\cdot; y)\#P_X(\cdot)$ $\forall y$，即先验分布 P_X 关于一个具有¯ ¯形式的映射∇f 的推前，其中 f 是一个实值函数，满足以下优化问题：

$$\bar{f} = \operatorname*{arg\,min}_{f \in \mathrm{CVX}_x} \mathbb{E}[f(\bar{X}; Y) + f^*(X; Y)]. \tag{1}$$

¯这里， X 是 X 的一个独立副本，CVX_x 表示函数 $f(x; y)$的集合；对任意固定的 y，这些函数关于x 变量是凸的；而 f^* 是 f 关于 x 变量的凸共轭。

上述变分形式具有三个关键特征，使其有别于以往的研究：(i) 它基于模拟，即可以通过从联合分布中采样来近似目标函数 P_{XY}，且不需要似然的显式表达式；(ii) 通过选择 CVX 集合的不同子集/参数化方式，变分形式能够为计算后验分布提供新的近似方法$_x$；(iii) 问题（1）具有随机性，可以利用近期的机器学习技术高效求解，例如， f 可以参数化为深度神经网络，并使用随机梯度下降进行训练。问题（1）可以视为独立耦合 $P_X \otimes P_Y$ 与联合分布 P_{XY}之间的分块三角 Monge 问题的对偶形式。在条件生成模型背景下，分布的分块三角传输也会产生类似的变分形式；例如 [40], [25], [32], [38], [39]。

本文旨在利用表述（1）开发一种新的非线性滤波算法，称为最优传输粒子滤波器（OTPF），并对该算法进行初步的理论分析和数值验证。所提出的算法可视为集成卡尔曼滤波算法（EnKF）的非线性、非高斯泛化 [18], [7] ，也是反馈粒子滤波器（FPF）算法的离散时间对应算法 [53], [52] （OTPF 通过求解所提出的变分问题（1），同时解决了 FPF 中的增益函数近似问题和数值时间离散化问题。）

*华盛顿大学航空航天系，西雅图；mohd9485@uw.edu,amirtag@uw.edu。

†华盛顿大学应用数学系，西雅图bamdadh@uw.edu。

本文的理论分析聚焦于所提出算法的误差分析。具体而言，我们研究在每个时间步求解问题（1）时产生的误差如何影响滤波算法的整体性能。为此，我们改进了现有的粒子滤波器（PF）误差分析方法，以得到滤波误差关于最优传输映射 [14],[49], [8], [13]近似误差的一致上界。这些结果基于一种强意义下的一致几何滤波稳定性 [6],概念，该概念常用于粒子滤波器的分析。接下来，我们将这些结果与最优传输映射估计的稳定性结果结合起来 [22] ，后者将映射的近似误差与问题（1）的优化间隙联系起来（见引理 2）。误差分析针对算法的平均场极限以及粒子系统的一种变体展开；该变体包含一个额外的重采样步骤，使粒子彼此独立，从而大幅简化了分析。

数值实验从定性和定量两个方面评估了 OTPF 相较于 EnKF 算法 [18], [7] 和序贯重要性重采样（SIR）粒子滤波器 [16]的性能。具体而言，我们考虑了一个具有三种不同观测函数的线性稳定动力系统：线性、二次和三次观测函数。数值结果展示了 OTPF 相较于另外两种方法的多样适用性。

本文其余部分的组织结构如下：第 II节回顾滤波问题及其方程，并介绍滤波稳定性的概念；第 III 节详细介绍 OTPF 算法；第 IV 节给出误差分析；第V 节包含数值实验。

II. 问题表述

A. 滤波问题

考虑由以下更新方程给出的离散时间随机动力学系统

$$X_t \sim a(\cdot|X_{t-1}), \quad X_0 \sim \pi_0 \tag{2a}$$
$$Y_t \sim h(\cdot|X_t) \tag{2b}$$

对于 $t = 1, 2, \ldots\ldots$其中 $X_t \in \mathbb{R}^n$ 是系统的状态，$Y_t \in \mathbb{R}^m$ 是观测，π_0 是初始状态的概率分布 X_0，$a(x'|x)$ 是从状态 x 到状态 x'的转移概率核，且 $h(y|x)$ 是在给定状态 y 的情况下观测的似然分布x。我们假设更新方程（2a）通过随机映射实现

$$X_t = \bar{a}(X_{t-1}, V_t) \tag{2c}$$

其中，$\{V_t\}_{t=1}^{\infty}$ 是一个独立同分布序列，并且对于所有 v，$\bar{a}(x, v)$关于 x 满足利普希茨条件。本文始终假设所有概率测度都具有密度，并使用相同的记号表示分布或相应的测度。如有需要，将根据上下文区分这两个概念。

滤波问题是推断在给定观测历史的情况下，状态的条件分布 X_t $\{Y_1, Y_1, \ldots, Y_t\}$，即分布

$$\pi_t := \mathbb{P}(X_t \in \cdot|Y_1, \ldots, Y_t), \quad \text{for} \quad t = 1, 2, \ldots,$$

通常称为后验分布。

B. 滤波器递归更新

后验分布 π_t 具有一个递归更新方程，这对于滤波算法的设计至关重要。为介绍这一递归更新，我们先引入以下算子：

$$\text{(propagation)} \quad \pi \mapsto \mathcal{A}\pi := \int_{\mathbb{R}^n} a(\cdot|x)\pi(x)dx \tag{3a}$$
$$\text{(conditioning)} \quad \pi \mapsto \mathcal{B}_y\pi := \frac{h(y|\cdot)\pi(\cdot)}{\int_{\mathbb{R}^n} h(y|x)\pi(x)dx} \tag{3b}$$

第一个算子根据动态模型（2a）更新状态的分布。第二个算子表示贝叶斯法则，根据观测模型（2b）进行条件化。利用这两个算子，后验分布的更新规律可表示为（例如，参见 [8]）：

$$\pi_t = \mathcal{T}_t\pi_{t-1} = \mathcal{B}_{Y_t}\mathcal{A}\pi_{t-1}. \tag{3c}$$

其中，我们引入了 $\mathcal{T}_t := \mathcal{B}_{Y_t}\mathcal{A}$。稍微滥用一下记号，我们进一步将转移算子定义为

$$\mathcal{T}_{t,s} := \mathcal{T}_t \circ \cdots \circ \mathcal{T}_{s+1}, \quad \forall \quad t > s \geq 0.$$

于是，对于所有 $t > s \geq 0$，我们有 $\pi_t = \mathcal{T}_{t,s}\pi_s$ 。注意，转移算子 $\mathcal{T}_{t,s}$ 本质上具有随机性，因为它取决于观测信号 $\{Y_{s+1}, \ldots, Y_t\}$ 的实现。为简化表述，我们省略了这种依赖关系。

C. 滤波稳定性

我们在（可能是随机的）概率测度 μ, ν上使用如下度量：

$$d(\mu, \nu) := \sup_{g \in \mathcal{G}} \sqrt{\mathbb{E}\left|\int g d\mu - \int g d\nu\right|^2} \tag{4}$$

其中，期望是对概率测度 μ 和 ν可能具有的随机性取的，并且 $\mathcal{G} := \{g : \mathbb{R}^n \to \mathbb{R};\ |g(x)| \leq 1, |g(x) - g(x')| \leq \|x - x'\|, \quad \forall x, x'\}$ 表示函数空间，其中的函数一致有界于 1，且一致满足利普希茨条件，其常数小于 1（该度量也称为对偶有界-Lipschitz距离）。我们利用这一度量引入滤波器的一种一致几何稳定性概念。

定义 *1*（均匀几何稳定滤波器）：如果滤波器更新（3）满足以下条件，则称其具有均匀几何稳定性：$\exists \lambda \in (0,1)$ 以及正的常数 $C > 0$ 使得对于所有μ，ν 和 $t > s \geq 0$ 均有

$$d(\mathcal{T}_{t,s}\mu, \mathcal{T}_{t,s}\nu) \leq C(1 - \lambda)^{t-s} d(\mu, \nu). \tag{5}$$

注 *1*：一致几何稳定性（5）这一性质也用于 [14], [13] 中 PFs 的误差分析。它可以

在动态转移核满足小集条件时得到验证，即存在一个概率测度 ρ 和一个常数 $\epsilon > 0$，使得 $a(x|x') \geq \epsilon\rho(x)$。小集条件是一种混合条件，可确保马尔可夫过程 X_t [29] 具有几何遍历性。我们承认，这一条件较强，并且只能针对一类受限的系统进行验证，例如，X_t 应属于一个紧集。具有一致几何稳定滤波器的系统的完整刻画，是该领域中一个尚未解决且具有挑战性的问题。对于较弱的滤波器渐近稳定性概念，即 $\lim_{t\to\infty} d(\mathcal{T}_{t,\ s}\mu,\ \mathcal{T}_{t,\ s}\nu) = 0$，已有更多相关认识；当系统按照适用于非线性随机动力系统 [50], [10],[48], [23] 的某种意义“可检测”时，该性质成立。对具有滤波渐近稳定性的系统所作的这一刻画，与现有的卡尔曼滤波器稳定性结果一致；后者在线性系统按照经典意义可检测时成立 [30]。现有滤波稳定性结果的完整综述可见 [12]。

以下引理对我们的误差分析很有用。

引理*1*：设 π 是一个（随机）分布，T 和 S 是两个（随机）可测映射。那么，

$$d(T\#\pi; S\#\pi) \leq \mathbb{E}\left[\|T - S\|^2_{L^2(\pi)}\right]^{\frac{1}{2}},$$

其中，期望是对分布 π 以及映射 T、S 可能具有的随机性取的。

证明：证明可由一个直接的论证得到，该论证使用度量的定义（4）以及测试函数的利普希茨性质 g。 ■

III. 最优传输粒子滤波器

OTPF 的构造依赖于变分形式（1）。考虑目标泛函

$$J(f;\pi) := \mathbb{E}[f(\bar{X};Y) + f^\star(X;Y)], \tag{6}$$

其中 $X \sim \pi$, $Y \sim h(\cdot|X)$，且 $X \sim \pi$ 是 X 的独立副本，同时满足该优化问题

$$\inf_{f\in\mathrm{CVX}_x} J(f,\pi). \tag{7}$$

已在 [43, 命题 1] 中证明，该问题的解给出了贝叶斯算子的一个最优传输刻画（3b）。为完整起见，现将结果重述如下

命题 *1*：假设 π 相对于勒贝格测度存在密度。那么，目标函数（6）有一个唯一的（相差一个常数平移）极小化子 $f \in \mathrm{CVX}_x$ 并且

$$\nabla_x \bar{f}(\cdot; y)\#\pi = \mathcal{B}_y\pi, \quad \text{for a.e. } y. \tag{8}$$

A. 精确平均场过程

我们利用条件分布的最优传输刻画，构造一个（精确的）平均场过程 X_t，其分布 $\bar{\pi}_t$ 与后验分布 π_t 完全相同。考虑一个分布为 $\bar{\pi}_t$ 的过程 X_t，定义如下：

$$\begin{aligned} \bar{X}_t &= \nabla_x \bar{f}_t(\bar{a}(\bar{X}_{t-1}, \bar{V}_t); Y_t), \quad \bar{X}_0 \sim \bar{\pi}_0 \\ \bar{f}_t &= \operatorname*{arg\,min}_{f\in\mathrm{CVX}_x} J(f, \mathcal{A}\bar{\pi}_{t-1}), \end{aligned} \tag{9a}$$

其中 V_t 是 V_t 在动态模型中（2c）的一个独立副本。接下来很容易验证

$$\bar{\pi}_t = \nabla_x \bar{f}_t(\cdot; Y_t)\#\mathcal{A}\bar{\pi}_{t-1} = \mathcal{B}_y\mathcal{A}\bar{\pi}_{t-1}, \tag{9b}$$

其中，第二个等式是命题 1的推论。因此，只要 $\bar{\pi}_0 = \pi_0$ 那么 $\bar{\pi}_t = \pi_t$。因此，平均场过程 X_t 称为精确平均场过程。通过以下两个步骤对精确平均场过程进行近似，即可得到 OTPF_X_t，具体如下。

B. 近似平均场过程

第一步近似是将优化问题（7）的可行集限制为一类参数化的凸函数 $\mathcal{F} \subset \mathrm{CVX}_x$。由此得到的近似分布记为 $\pi^{\mathcal{F}}_t$ 其更新规则如下：

$$\begin{aligned} \pi^{\mathcal{F}}_t &= \nabla_x f^{\mathcal{F}}_t(\cdot, Y_t)\#\mathcal{A}\pi^{\mathcal{F}}_{t-1}, \quad \pi^{\mathcal{F}}_0 = \pi_0 \\ f^{\mathcal{F}}_t &= \operatorname*{arg\,min}_{f\in\mathcal{F}} J(f, \mathcal{A}\pi^{\mathcal{F}}_{t-1}) \end{aligned} \tag{10a}$$

此更新定义了近似平均场过程

$$X^{\mathcal{F}}_t = \nabla_x f^{\mathcal{F}}_t(\bar{a}(X^{\mathcal{F}}_{t-1}; Y_t), \bar{V}_t), \quad X^{\mathcal{F}}_0 \sim \bar{\pi}_0. \tag{10b}$$

由函数 $\pi^{\mathcal{F}}_t$ 的参数化 $\bar{\pi}_t$ 导致的 f_t 与 IV-A中研究的近似误差。

C. 有限粒子系统

第二步近似是用一组粒子 $\{X^1_t, \ldots, X^N_t\}$ 的经验分布替代平均场过程，即 $\pi^{\mathcal{F}}_t \approx \frac{1}{N}\sum_{i=1}^N \delta_{X^i_t}$. 有限-$N$ 离散化可通过两种不同的方法实现，从而得到两种不同的粒子系统。第一种系统（带重采样的粒子系统）更便于进行误差分析，而第二种系统（相互作用粒子系统）则更实用。

(C.I) 带重采样的粒子系统：定义采样算子

$$\pi \mapsto \mathcal{S}^N\pi := \frac{1}{N}\sum_{i=1}^N \delta_{X^i} \quad X^i \overset{\text{i.i.d.}}{\sim} \pi, \tag{11}$$

并通过在更新方程中引入采样算子 $\mathcal{S}^N$ 来近似均值场分布 $\pi^{\mathcal{F}}_t$:

$$\begin{aligned} \tilde{\pi}^{(\mathcal{F},N)}_t &= \nabla_x \tilde{f}^{(\mathcal{F},N)}_t(\cdot, Y_t)\#\mathcal{S}^N\mathcal{A}\tilde{\pi}^{(\mathcal{F},N)}_{t-1}, \quad \pi^{(\mathcal{F},N)}_0 = \pi_0 \\ \tilde{f}^{(\mathcal{F},N)}_t &= \operatorname*{arg\,min}_{f\in\mathcal{F}} J(f, \mathcal{S}^N\mathcal{A}\tilde{\pi}^{(\mathcal{F},N)}_{t-1}), \end{aligned} \tag{12a}$$

采样算子的存在确保分布 $\tilde{\pi}^{(\mathcal{F},N)}_t$ 是由一组粒子构成的经验分布，即 $\tilde{\pi}^{(\mathcal{F},N)}_t = \frac{1}{N}\sum_{i=1}^N \delta_{\tilde{X}^i_t}$. 方程（12a）进一步给出了粒子的更新法则：

$$\tilde{X}^i_t = \nabla_x \tilde{f}^{(\mathcal{F},N)}_t(\bar{a}(\tilde{X}^{\sigma_i}_{t-1}, V^i_t); Y_t) \tag{12b}$$

其中，$\sigma_i \sim \mathrm{Unif}\{1, 2, \ldots, N\}$ 和 $\{V^i_t\}_{i=1}^N$ 是 V_t 的独立副本。该采样过程与粒子滤波器中的重采样阶段类似，不同之处在于

此处的权重是均匀的。重采样步骤使粒子彼此独立，从而大大简化了误差分析，详见第IV-B节。

（C.II）相互作用粒子系统：构造有限N粒子系统的第二种方法，是根据以下方式对平均场过程的更新方程（10b）进行离散化$X_t^{\mathcal{F}}$：

$$\begin{aligned} X_t^i &= \nabla_x f_t^{(\mathcal{F},N)}(\bar{a}(X_t^i, V_t^i); Y_t) \\ f_t^{(\mathcal{F},N)} &= \arg\min_{f\in\mathcal{F}} J(f, \frac{1}{N}\sum_{i=1}^N \delta_{\bar{a}(X_t^i,V_t^i)}). \end{aligned} \tag{13}$$

经验分布 $\pi_t^{(\mathcal{F},N)} := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$ 并不遵循类似于 $\tilde{\pi}_t^{(\mathcal{F},N)}$ 的更新定律，这是由于算子 $\mathcal{A}$ 的性质所致，该算子会平滑经验分布。相反，相互作用粒子系统的更新可表示为

$$\pi_t^{(\mathcal{F},N)} = \nabla f^{(\mathcal{F},N)} \# \mathcal{A}^N \pi_{t-1}^{(\mathcal{F},N)}$$

其中，$\mathcal{A}^N$是一个随机算子，它接受任意经验分布 $\frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ 并输出 $\frac{1}{N}\sum_{i=1}^N \delta_{a(x^i,V^i)}$。此外，与前一种构造（带重采样的粒子系统）不同，此时粒子之间存在相关性，这使得误差分析颇具挑战性（这通常通过混沌传播分析来研究 [42]）。我们将相互作用粒子系统的误差分析留待未来工作。不过，我们在第 V 节中对这种近似的性能进行了实证验证。

IV. 误差分析

本节旨在研究上文介绍的 OTPF 的近似误差。我们首先分析近似平均场过程，然后再转向带重采样的粒子系统。

A. 均值场分析

精确均值场分布 $\bar{\pi}_t$ 与近似分布 $\pi_t^{\mathcal{F}}$ 之间的距离由以下命题刻画。

命题 *2*：设 $\bar{\pi}_t$ 和 $\pi_t^{\mathcal{F}}$分别如（9）-（10）所示。假设

1）精确滤波器根据定义 1 是稳定的。2）存在 $\epsilon_{\mathcal{F}} > 0$ 使得

$$\inf_{f\in\mathcal{F}} J(f, \mathcal{A}\pi_t^{\mathcal{F}}) - \inf_{f\in \mathrm{CVX}_x} J(f, \mathcal{A}\pi_t^{\mathcal{F}}) \le \epsilon_{\mathcal{F}}, \quad \forall t. \tag{14}$$

3) 对于所有 y和 t，函数 $f_t^{\mathcal{F}}(\cdot; y)$ 是凸的，且$\nabla_x f_t^{\mathcal{F}}(\cdot; y)$ 是 β-利普希茨的。

于是，有

$$d(\pi_t^{\mathcal{F}}, \pi_t) \le \frac{C\sqrt{2\beta\epsilon_{\mathcal{F}}}}{\lambda}, \quad \forall t, \tag{15}$$

其中所有常数均与时间无关。

备注*2*：命题中的第一个假设用于确保算法每一步产生的误差不会随时间增长。第二个假设与函数类 $\mathcal{F}$ 相对于算法 $\mathcal{A}\pi_t^{\mathcal{F}}$引入的概率分布类的表示能力有关。例如，当 $\mathcal{F}$ 是一个凸二次函数类，且滤波问题基于线性高斯动态模型和观测模型时，该误差为零。在这种情况下，概率分布 $\pi_t^{\mathcal{F}}$是高斯分布，对应的最优函数为二次最优函数f_t。一般而言，当分布本身较为简单时，例如问题呈现低维结构或正则性时，误差预计会较小。对这些误差的分析属于表示理论的研究范畴 [3]，[37]。最后一个假设与分布 $\mathcal{A}\pi_t^{\mathcal{F}}$及其所产生的后验分布的正则性有关，可以通过恰当选择函数类$\mathcal{F}$来满足。

证明：为简化表述，我们引入算子 $\pi \mapsto \mathcal{T}_t^{\mathcal{F}}\pi := \nabla f_t^{\mathcal{F}}(\cdot, Y_t)\#\mathcal{A}\pi$ 对所有 t，用以表示式（10a）中近似平均场分布的更新规律。证明的第一步是利用三角不等式和滤波稳定性，对π_t 与 $\pi_t^{\mathcal{F}}$之间的误差作如下界定：

$$\begin{aligned} d(\pi_t, \pi_t^{\mathcal{F}}) &\le \sum_{k=1}^t d(\mathcal{T}_{t,k-1}\pi_{k-1}^{\mathcal{F}}, \mathcal{T}_{t,k}\pi_k^{\mathcal{F}}) \\ &\le \sum_{k=1}^t d(\mathcal{T}_{t,k}\mathcal{T}_k\pi_{k-1}^{\mathcal{F}}, \mathcal{T}_{t,k}\mathcal{T}_k^{\mathcal{F}}\pi_{k-1}^{\mathcal{F}}) \\ &\le \sum_{k=1}^t C(1-\lambda)^{t-k} d(\mathcal{T}_k\pi_{k-1}^{\mathcal{F}}, \mathcal{T}_k^{\mathcal{F}}\pi_{k-1}^{\mathcal{F}}) \\ &\le \frac{C}{\lambda}\max_{k\in\{1,2,\dots,t\}}\{d(\mathcal{T}_k\pi_{k-1}^{\mathcal{F}}, \mathcal{T}_k^{\mathcal{F}}\pi_{k-1}^{\mathcal{F}})\}. \end{aligned}$$

接下来，我们利用引理 1 来界定距离

$$\begin{aligned} &d(\mathcal{T}_k\pi_{k-1}^{\mathcal{F}}, \mathcal{T}_k^{\mathcal{F}}\pi_{k-1}^{\mathcal{F}}) \\ &\quad = d(\nabla\bar{f}_k(\cdot; Y_k)\#\mathcal{A}\pi_{k-1}^{\mathcal{F}}, \nabla f_k^{\mathcal{F}}(\cdot; Y_t)\#\mathcal{A}\pi_{k-1}^{\mathcal{F}}) \\ &\quad \le \mathbb{E}\left[\|\nabla\bar{f}_k(\cdot; Y_k) - \nabla f_k^{\mathcal{F}}(\cdot; Y_k)\|^2_{L^2(\mathcal{A}\pi_{k-1}^{\mathcal{F}})}\right]^{\frac{1}{2}} \end{aligned}$$

对所有 $k \ge 0$。最后，我们利用命题中的第二和第三个假设，并通过引理 $\nabla\bar{f}_k$ 与 $\nabla f_k^{\mathcal{F}}$之间的误差得到一致有界，其中用到引理 2（下文将给出证明）

$$\begin{aligned} &\mathbb{E}\left[\|\nabla\bar{f}_k(\cdot; Y_k) - \nabla f_k^{\mathcal{F}}(\cdot; Y_k)\|^2_{L^2(\mathcal{A}\pi_{k-1}^{\mathcal{F}})}\right] \\ &\quad \le 2\beta(J(f_k^{\mathcal{F}}, \mathcal{A}\pi_{k-1}^{\mathcal{F}}) - J(\bar{f}_k, \mathcal{A}\pi_{k-1}^{\mathcal{F}})) \\ &\quad \le 2\beta\epsilon_{\mathcal{F}} \end{aligned}$$

得到最终界（15）。 ■

引理*2*：考虑优化问题（7），其目标函数（6）。假设 π 具有密度。令 f 为最优函数，$\bar{f}$ 为任意凸且 β-光滑函数。则，

$$J(\pi, f) - J(\pi, \bar{f}) \ge \frac{1}{2\beta}\mathbb{E}[\|\nabla f(\bar{X}; Y) - \nabla\bar{f}(\bar{X}; Y)\|^2].$$

证明：该证明是对 [22, 命题10]结果的扩展，因篇幅所限略去。 ■

B. 带重采样的粒子系统分析

接下来，我们分析粒子系统（12）与精确平均场过程（9）之间的误差。该过程与上一节所述的均值场分析类似，但由于采样算子和经验近似，还会产生额外误差。

命题*3*：考虑精确的均值场分布 $\bar{\pi}_t$ 和粒子分布 $\tilde{\pi}_t^{(\mathcal{F},N)}$ 分别定义于（9）和（12）中。假设

1）精确滤波器根据定义1是稳定的。2）存在常数 $\epsilon_{\mathcal{F},N} > 0$ 使得对于所有 t和 N：

$$\inf_{f\in\mathcal{F}} J(f,\mathcal{S}^N\mathcal{A}\tilde{\pi}_t^{(\mathcal{F},N)}) - \inf_{f\in\mathrm{CVX}_x} J(f,\mathcal{A}\tilde{\pi}_t^{(\mathcal{F},N)}) \le \epsilon_{\mathcal{F},N}$$

3）对于所有 y、t 和 N，函数 $f_t^{(\mathcal{F},N)}(\cdot;y)$ 是凸的，并且 $\nabla_x f_t^{(\mathcal{F},N)}(\cdot;y)$ 是 β-利普希茨的。

则有

$$d(\tilde{\pi}_t^{(\mathcal{F},N)},\pi_t) \le \frac{C}{\lambda}\left(\sqrt{2\beta\epsilon_{\mathcal{F},N}} + \frac{1}{\sqrt{N}}\right), \quad \forall t, \tag{16}$$

其中所有常数均与时间无关。

证明：该证明与命题2的证明类似。定义算子 $\tilde{\mathcal{T}}_{(\mathcal{F},N)t}$：$\pi \mapsto \nabla\tilde{f}_t^{(\mathcal{F},N)}(\cdot,Y_t)\#\mathcal{S}^N\mathcal{A}\pi$。于是，根据三角不等式和滤波稳定性可得

$$d(\pi_t,\pi_t^{(\mathcal{F},N)}) \le \frac{C}{\lambda}\max_{k\in\{1,2,\dots,t\}}\{d(\mathcal{T}_k\pi_{k-1}^{(\mathcal{F},N)},\tilde{\mathcal{T}}_k^{(\mathcal{F},N)}\pi_{k-1}^{(\mathcal{F},N)})\}.$$

再次应用三角不等式，可写为

$$\begin{aligned}
&d(\mathcal{T}_k\pi_{k-1}^{(\mathcal{F},N)},\tilde{\mathcal{T}}_k^{(\mathcal{F},N)}\pi_{k-1}^{(\mathcal{F},N)})\\
&= d(\nabla\bar{f}_k(\cdot,Y_t)\#\mathcal{A}\pi_{k-1}^{(\mathcal{F},N)},\nabla f_k^{(\mathcal{F},N)}\#\mathcal{S}^N\mathcal{A}\pi_{k-1}^{(\mathcal{F},N)})\\
&\le d(\nabla\bar{f}_k(\cdot,Y_t)\#\mathcal{A}\pi_{k-1}^{(\mathcal{F},N)},\nabla f_k^{(\mathcal{F},N)}\#\mathcal{A}\pi_{k-1}^{(\mathcal{F},N)})\\
&+ d(\nabla f_k^{(\mathcal{F},N)}\#\mathcal{A}\pi_{k-1}^{(\mathcal{F},N)},\nabla f_k^{(\mathcal{F},N)}\#\mathcal{S}^N\mathcal{A}\pi_{k-1}^{(\mathcal{F},N)})
\end{aligned}$$

应用引理1 和引理2，第一项的上界为以下各项之和的平方根

$$\begin{aligned}
&\mathbb{E}\left[\|\nabla\bar{f}_k(\cdot;Y_k) - \nabla f_k^{(\mathcal{F},N)}(\cdot;Y_k)\|^2_{L^2(\mathcal{A}\tilde{\pi}_{k-1}^{(\mathcal{F},N)})}\right]\\
&\le 2\beta(J(f_k^{(\mathcal{F},N)},\mathcal{A}\tilde{\pi}_{k-1}^{(\mathcal{F},N)}) - J(\bar{f}_k,\mathcal{A}\tilde{\pi}_{k-1}^{(\mathcal{F},N)}))\\
&\le 2\beta\epsilon_{\mathcal{F},N}
\end{aligned}$$

其中我们使用了第二个和第三个假设。这给出了（16）右端的第一项。第二项源于采样误差，其上界为 $\frac{1}{\sqrt{N}}$，因为度量 d的定义中的测试函数 g 均由1一致有界（例如参见 [33, 引理2。17]）。将这两个误差相加即可得到最终界。 ■

备注*3*：这个命题的假设与命题 2 中的假设相似，只是第二个假设略有不同。第二个假设中的界可以分解为两项：

$$\begin{aligned}
&\inf_{f\in\mathcal{F}} J(f,\mathcal{S}^N\mathcal{A}\tilde{\pi}_t^{(\mathcal{F},N)}) - \inf_{f\in\mathcal{F}} J(f,\mathcal{A}\tilde{\pi}_t^{(\mathcal{F},N)})\\
&+ \inf_{f\in\mathcal{F}} J(f,\mathcal{A}\tilde{\pi}_t^{(\mathcal{F},N)}) - \inf_{f\in\mathrm{CVX}_x} J(f,\mathcal{A}\tilde{\pi}_t^{(\mathcal{F},N)})
\end{aligned}$$

第二项与命题2中使用的项相似，并且与 $\mathcal{F}$的表示能力有关。第一项对应于使用经验样本近似分布所导致的统计泛化误差，也是统计泛化理论 [37], [5], [26],[54]研究的主题。误差预计按 $O(\frac{C_{\mathcal{F}}}{\sqrt{N}})$ 的比例缩放，其中常数 $C_{\mathcal{F}}$ 是函数类 $\mathcal{F}$复杂度的代理量，且与维数d无关。第一项也可以解释为方差，而第二项是偏差。因此，我们的误差分析体现了依赖于函数类 $\mathcal{F}$复杂度的偏差-方差权衡。对局部粒子滤波器的分析中也出现了类似的偏差-方差权衡，见 [33]。

V. 数值实验

我们通过一个数值示例，将所提出的OTPF与另外两种滤波器进行比较，以说明其性能：EnKF [18],和序贯重要性重采样（SIR）PF [16]。

对于 OTPF，我们求解变分问题（7）的极小极大形式，如 [43, 第 III-B 节] 中所述，该形式最初用于 [27] 估计最优传输映射。极小极大形式还涉及对一个额外的凸函数 ψ 进行优化，该函数用于表示凸共轭 f^*，如下所示：

$$\mathbb{E}[f^*(X;Y)] = \max_{\psi\in\mathrm{CVX}_x} \mathbb{E}[X^\top\nabla_x\psi(X;Y) - f(\nabla_x\psi(X;Y);Y)]$$

然而，在我们的数值实验中，我们观察到，放宽约束并对映射 $T(x;y)$ 而不是 $\nabla_x\psi(x;y)$ 进行优化，由于参数化具有更大的自由度，可以得到更好的数值结果。因此，我们采用如下形式：

$$\mathbb{E}[f^*(X;Y)] = \max_{T} \mathbb{E}[X^\top T(X;Y) - f(T(X;Y);Y)]$$

请注意，这不会改变优化问题，因为最优的 T 具有梯度形式，并且等于 $\nabla_x f^*$。最终的目标函数形式为

$$\begin{aligned}
\min_{f\in\mathrm{ICNN}}\max_{T\in\mathrm{ResNet}}\{&\mathbb{E}_{P_{XY}}[f(X,Y)]\\
&+ \mathbb{E}_{P_X\otimes P_Y}[X^T T(X,Y) - f(T(X,Y),Y)]\}
\end{aligned} \tag{17}$$

注 *4*：请注意，与原始形式（$P_X\otimes P_Y$ 和目标 P_{XY}）相比，我们调换了源1的角色，因此 T 表示从先验分布到后验分布的传输映射，而不是 $\nabla_x f$。这种形式使映射的参数化更加方便。

类似于上述方法的松弛也被证明有助于计算 Wasserstein重心 [19] 和 Wasserstein梯度流 [20]。这里，ICNN 表示部分输入凸神经网络的集合 [2]。

为展示滤波器的性能，考虑以下动力学和观测模型：

$$X_t = (1-\alpha)X_{t-1} + 2\sigma V_t, \quad X_0 \sim \mathcal{N}(0,I_n) \tag{18a}$$
$$Y_t = h(X_t) + \sigma W_t \tag{18b}$$

对于 $t = 1, 2, 3, \ldots\ldots$，其中 X_t，$Y_t \in \mathbb{R}^n$、$\{V_t\}_{t=1}^{\infty}$和$\{W_t\}_{t=1}^{\infty}$是n维标准高斯随机变量的独立同分布序列，$\alpha = 0.1$ 和 $\sigma = \sqrt{0.1}$。我们使用三个观测函数：

$$h(x) = x, \quad h(x) = x \odot x, \quad h(x) = x \odot x \odot x$$

其中，$\odot$ 表示逐元素（即 Hadamard）乘积（当 x 是向量时）。

为了求解 (17)，我们将 ICNN 参数化为

$$f(x; y) = \sum_{k=1}^{K} W_k (x^\top W_k^x + y^\top W_k^y + b_k)_+^2$$

其中，$W_k \geq 0$、$W_k^x, W_k^y \in \mathbb{R}^n, b_k \in \mathbb{R}$（对于 $k = 1, \ldots, K$），且 $K = 32$为网络的规模。映射 T 采用标准残差网络建模，该网络包含两个大小为 32 的块，并使用 ReLU 激活函数。

我们使用 Adam 优化器求解极小极大问题，学习率为 10^{-2}，内循环迭代次数为 10，总迭代次数为 1024；在滤波问题的每个时间步之后，总迭代次数减半，直至达到 64。每次迭代都会从总共 $N = 1000$个粒子中随机选取一个大小为 32 的样本批次 $\{(X_t^1, Y_t^1) \ldots, (X_t^N, Y_t^N)\}$。观测样本$Y_t^i$由观测模型生成：$Y_t^i \sim h(\cdot|X_t^i)$。独立耦合 $P_X \otimes P_Y$的样本通过随机打乱生成。所有算法使用相同数量的粒子 N。数值代码的详情可在线获取[1]。

数值结果如图 1 所示，针对的是一个二维问题 $n = 2$，而图中仅展示了第一个分量（我们选择 $n = 2$，因为当 $n = 1$时，SIR 和最优传输方法之间的差异并不显著；而当 $n = 2$时，差异则变得明显。）图中展示了粒子的轨迹以及隐状态的轨迹。第一个实验如面板 (a) 所示，展示了线性观测函数的表现。正如预期的那样，三种算法的表现相似，因为它们都能够捕捉到精确解；在这种情况下，精确解为高斯分布，并通过线性映射得到。第二个实验如面板 (b) 所示，涉及二次观测函数 $h(x) = x \odot x$。这是一个有趣的情形，因为该问题不可观测，我们预期会看到一个（对称的）双峰分布。可以看到，EnKF 未能表示双峰分布，而最优传输和 SIR 均捕捉到了两个峰，尽管 SIR 在 $t \in [2, 3.5]$时间范围内出现了模态坍塌。最后，如面板 (c) 所示，针对三次观测函数 $h(x) = x \odot x \odot x$，SIR 和最优传输的表现都优于 ENKF，这符合预期，因为观测模型具有强非线性。

我们还通过计算对状态函数 ϕ 进行估计时的均方误差（MSE），来量化这三项实验中所有算法的性能：

$$\mathrm{MSE}_t(\phi) = \mathbb{E}\|\frac{1}{N}\sum_{i=1}^{N} \phi(X_t^i) - \phi(X_t)\|^2 \qquad (19)$$

[1]https://github.com/Mohd9485/OT-EnKF-SIR

其中，经验平均值近似于对 100 次独立模拟取期望。结果如图 2所示。对于线性观测模型和三次观测模型，我们使用了 $\phi(x) = x$。对于二次情形，我们使用了 $\phi(x) = \max(0, x)$（对于二次情形，比较估计均值与真实均值并非良好的评判标准，因为分布是双峰的，且均值为零。）

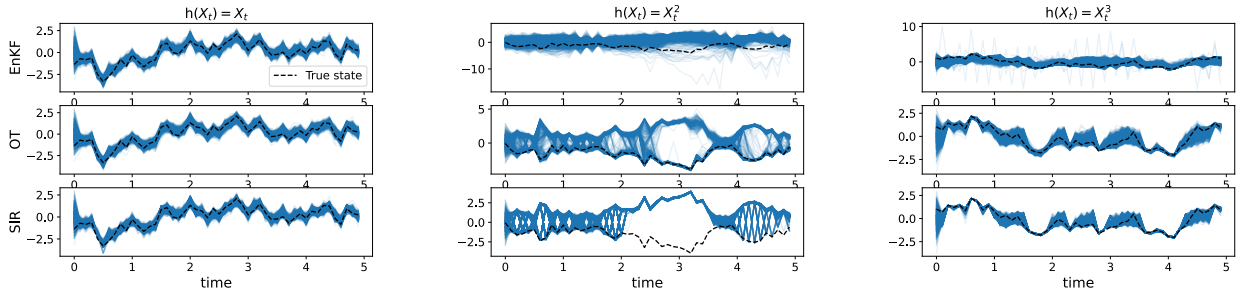
在图 2-(a) 中可以看到，最优传输滤波器和SIR滤波器的结果都接近 EnKF；对于线性情形，EnKF 渐近精确。然而，在图 2-(b) 中，对于二次情形，最优传输方法优于 EnKF 和 SIR；而在图 2-(c) 所示的三次情形中，最优传输方法与 SIR 之间的差异并不显著。通过进一步微调、增加训练迭代次数以及神经网络中的参数数量，预计最优传输滤波器的性能将有所提升，但计算开销也会增加。对最优传输方法效率的适当分析、研究其如何扩展至高维问题，以及将其应用于更真实数据，都是未来工作的研究方向。

VI. 讨论

本文介绍了 OTPF 算法，并提供了初步的理论误差分析和数值结果，证明了我们的方法在存在非线性观测和非高斯状态时具有竞争力。我们提出了若干未来研究方向：验证动力系统的几何稳定性，例如形式为（18）的系统；分析求解变分问题时的优化间隙误差，包括备注3）中讨论的表示和泛化方面；分析不进行重采样时相互作用粒子系统的误差；以及在真正的高维环境下开展广泛的数值实验和比较。

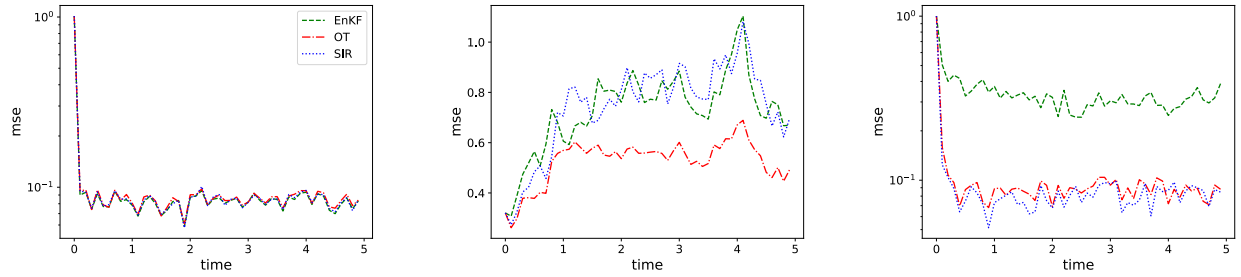
参考文献

[1] Luigi Ambrosio、Nicola Gigli 和 Giuseppe Savar´é。《度量空间与概率测度空间中的梯度流》。Springer Science & Business Media，2005年。[2] Brandon Amos、Lei Xu 和 J Zico Kolter。输入凸神经网络。*arXiv*预印本 *arXiv:1609.07152*，2016年。[3] Martin Anthony 和 Peter L Bartlett。《神经网络学习：理论基础》，第9卷。剑桥大学出版社，剑桥，1999年。[4] Martin Arjovsky、Soumith Chintala 和 L´eon Bottou。Wasserstein生成对抗网络。载于国际机器学习会议，第214–223页。PMLR，2017年。[5] Sanjeev Arora、Rong Ge、Yingyu Liang、Tengyu Ma 和 Yi Zhang。生成对抗网络（GAN）的泛化与均衡。载于国际机器学习会议，第224–232页。PMLR，2017年。[6] Rami Atar 和 Ofer Zeitouni。非线性滤波的指数稳定性。载于亨利·庞加莱研究所年鉴（*B*辑）：概率与统计，第33卷，第697–725页。Elsevier，1997年。[7] Edoardo Calvello、Sebastian Reich 和 Andrew M Stuart。集合卡尔曼方法：平均场视角。*arXiv*预印本 *arXiv:2209.11371*，2022年。[8] Olivier Capp´é、Eric Moulines 和 Tobias Ryd´en。隐马尔可夫模型中的推断。载于*EUSFLAT*会议论文集，第14–16页，2009年。[9] Yuan Cheng 和 Sebastian Reich。Feynman-Kac公式的McKean最优传输视角及其在数据同化中的应用。*arXiv*预印本 *arXiv:1311.6300*，2013年。



(a) (b) (c)

图1. 集合卡尔曼滤波器、最优传输粒子滤波器和序贯重要性重采样粒子滤波器在数值示例（18）中的应用结果，图中分别记为 EnKf、OT 和 SIR。图中显示了粒子 $\{X_t^1, \ldots, X_t^N\}$ 的轨迹，以及真实状态 X_t的轨迹。结果包含三个观测函数：(a) $h(x) = x$，(b) $h(x) = x \odot x$，以及 (c) $h(x) = x \odot x \odot x$。



(a) (b) (c)

图2. 在与图 1类似的设置下，将滤波器应用于示例（18）所得的数值结果。图中显示了对状态函数进行估计时的均方误差（19）。面板 (a) 和 (c) 使用函数 $\phi(x) = x$，而面板 (b) 使用 $\phi(x) = \max(0, x)$。均方误差通过对 100 次独立模拟取经验平均值进行评估。

[10] Pavel Chigansky、Robert Liptser 和 Ramon Van Handel。滤波稳定性的内在方法。《非线性滤波手册》，2009年。[11] Nicolas Courty、R´emi Flamary、Devis Tuia 和 Alain Rakotoma-monomy。用于领域自适应的最优传输。*arXiv*预印本 *arXiv:1507.00504*，2015年。[12] Dan Crisan 和 Boris Rozovskii。《牛津非线性滤波手册》。牛津大学出版社，2011年。[13] Pierre Del Moral 和 Pierre Del Moral。《Feynman-Kac公式》。Springer，2004年。[14] Pierre Del Moral 和 Alice Guionnet。相互作用过程的稳定性及其在滤波和遗传算法中的应用。载于亨利·庞加莱研究所年鉴´（*B*辑）：概率与统计，第37卷，第155–194页。Elsevier，2001年。[15] Ayelet Dominitz 和 Allen Tannenbaum。通过最优质量传输进行纹理映射。*IEEE*可视化与计算机图形学汇刊，16(3):419–433，2010年。[16] Arnaud Doucet 和 Adam M Johansen。粒子滤波与平滑教程：十五年后。《非线性滤波手册》，12(656-704):3，2009年。[17] Tarek A El Moselhy 和 Youssef M Marzouk。基于最优映射的贝叶斯推断。《计算物理学杂志》，231 7850，2012 23 7815( ): – 。[18] Geir Evensen。《数据同化：集合卡尔曼滤波器》。Springer Science & Business Media，2006年。[19] Jiaojiao Fan、Amirhossein Taghvaei 和 Yongxin Chen。通过输入凸神经网络高效计算Wasserstein重心。*arXiv*预印本 *arXiv:2007.04462*，2020年。[20] Jiaojiao Fan、Amirhossein Taghvaei 和 Yongxin Chen。变分Wasserstein梯度流。*arXiv*预印本 *arXiv:2112.02424*，2021年。[21] Sira Ferradans、Nicolas Papadakis、Gabriel Peyr´e 和 Jean-Franc¸ois Aujol。正则化离散最优传输。《*SIAM*成像科学杂志》，7(3):1853–1882，2014年。[22] Jan-Christian H¨utter 和 Philippe Rigollet。极小极大速率的光滑最优传输映射的估计。*arXiv*预印本 *arXiv:1905.05828*，2019年。[23] Jin Won Kim 和 Prashant G Mehta。非线性滤波的对偶性 I：可观测性。*arXiv*预印本 *arXiv:2208.06586*，2022年。[24] Soheil Kolouri、Se Rim Park、Matthew Thorpe、Dejan Slepcev 和 Gustavo K Rohde。最优质量传输：信号处理与机器学习应用。*IEEE*信号处理杂志，34(4)：43–59，2017年。[25] Nikola Kovachki、Ricardo Baptista、Bamdad Hosseini 和 Youssef Marzouk。使用单调生成对抗网络进行条件采样。*arXiv*预印本 *arXiv:2006.06755*，2020年。[26] Shuang Liu、Olivier Bousquet 和 Kamalika Chaudhuri。生成对抗学习的逼近与收敛性质。神经信息处理系统进展，30，2017年。[27] Ashok Makkuva、Amirhossein Taghvaei、Sewoong Oh 和 Jason Lee。通过输入凸神经网络实现最优传输映射。在国际机器学习会议上，第6672–6681页。PMLR，2020年。[28] Youssef Marzouk、Tarek Moselhy、Matthew Parno 和 Alessio Spantini。基于测度传输的采样简介。*arXiv*预印本 *arXiv:1602.05023*，2016年。[29] Sean P Meyn 和 Richard L Tweedie。马尔可夫链与随机稳定性。Springer Science & Business Media，2012年。[30] Daniel Ocone 和 Etienne Pardoux。最优滤波器相对于其初始条件的渐近稳定性。*SIAM*控制与优化杂志，34(1)：226–243，1996年。[31] Julien Rabin、Gabriel Peyr´e、Julie Delon 和 Marc Bernot。Wasserstein重心及其在纹理混合中的应用。在尺度空间与计算机视觉变分方法国际会议上，第435–446页。Springer，2011年。[32] Deep Ray、Harisankar Ramaswamy、Dhruv V Patel 和 Assad A Oberai。条件生成对抗网络在基于物理的逆问题后验推断中的有效性与泛化能力。*arXiv*预印本 *arXiv:2202.07773*，2022年。

[33] Patrick Rebeschini 和 Ramon Van Handel。局部粒子滤波器能否克服维数灾难？《应用概率年刊》，25(5)：2809–2866，2015年。[34] Sebastian Reich。用于贝叶斯推断的非参数集合集合变换方法。SIAM科学计算杂志，35(4)：A2013–A2024，2013年。[35] Sebastian Reich。数据同化：Schr¨odinger视角。《数值学报》，28：635–711，2019年。[36] Filippo Santambrogio。面向应用数学家的最优传输。Birk¨auser，纽约，55(58-63)：94，2015年。[37] Shai Shalev-Shwartz 和 Shai Ben-David。理解机器学习：从理论到算法。剑桥大学出版社，2014年。[38] Yuyang Shi、Valentin De Bortoli、George Deligiannidis 和 Arnaud Doucet。使用扩散Schr¨odinger桥进行条件模拟。在《人工智能中的不确定性》会议上，第1792–1802页。PMLR，2022年。[39] Ali Siahkoohi、Gabrio Rizzuti、Mathias Louboutin、Philipp A Witte 和 Felix J Herrmann。用于逆问题变分推断的归一化流预条件训练。arXiv预印本 arXiv:2101.03709，2021年。[40] Alessio Spantini、Ricardo Baptista 和 Youssef Marzouk。非线性集成滤波的耦合技术。arXiv预印本 arXiv:1907.00389，2019年。[41] Zhengyu Su、Yalin Wang、Rui Shi、Wei Zeng、Jian Sun、Feng Luo 和 Xianfeng Gu。用于形状匹配与比较的最优质量传输。IEEE模式分析与机器智能汇刊，37(11)：2246–2259，2015年。[42] Alain-Sol Sznitman。混沌传播专题。Saint-Flour概率论暑期学校 XIX—1989，1464：165–251，1991年。[43] Amirhossein Taghvaei 和 Bamdad Hosseini。贝叶斯定律的最优传输表述及其在非线性滤波算法中的应用。在2022年IEEE第61届决策与控制会议（CDC）上，第6608–6613页。IEEE，2022年。[44] Amirhossein Taghvaei 和 Prashant G Mehta。线性反馈粒子滤波器的最优传输表述。在2016年美国控制会议（ACC）上，第3614–3619页。IEEE，2016年。[45] Amirhossein Taghvaei 和 Prashant G Mehta。集合卡尔曼滤波器的最优传输表述。IEEE自动控制汇刊，66(7)：3052–3067，2020年。[46] Amirhossein Taghvaei 和 Prashant G Mehta。非线性滤波中的最优传输方法：反馈粒子滤波器。arXiv预印本 arXiv:2102.10712，2021年。[47] Ilya Tolstikhin、Olivier Bousquet、Sylvain Gelly 和 Bernhard Schoelkopf。Wasserstein自编码器。arXiv预印本 arXiv:1711.01558，2017年。[48] Ramon Van Handel。可观测性与非线性滤波。《概率论及相关领域》，145：35–74，2009年。[49] Ramon Van Handel。隐马尔可夫模型的统一可观测性与不稳定信号的滤波稳定性。2009年。[50] Ramon Van Handel。非线性滤波与系统理论。在第19届国际网络与系统数学理论研讨会论文集（MTNS半 plenary论文）中，2010年。[51] C´edric Villani。最优传输：新旧理论，第338卷。Springer，2009年。[52] Tao Yang、Richard S Laugesen、Prashant G Mehta 和 Sean P Meyn。多变量反馈粒子滤波器。Automatica，71：10–23，2016年。[53] Tao Yang、Prashant G Mehta 和 Sean P Meyn。反馈粒子滤波器。IEEE自动控制汇刊，58 2480，2013 10 2465( )：–。[54] Pengchuan Zhang、Qiang Liu、Dengyong Zhou、Tao Xu 和 Xiaodong He。关于生成对抗网络中的判别-泛化权衡。arXiv预印本 arXiv:1711.02771，2017年。