基于最优传输映射序列的渐进式贝叶斯粒子流

Uwe D. Hanebeck

智能传感器-执行器系统实验室（ISAS）人体工程学
与机器人研究所卡尔斯鲁厄理工学院（KIT），德
国 Uwe.Hanebeck@kit.edu

arXiv:2303.02412v1 [stat.ML] 4 Mar 2023

摘要——**我们提出了一种使用确定性粒子进行最优贝叶斯滤波的方法。为了避免粒子退化，滤波步骤不是一次性完成的。相反，粒子从先验分布逐步流向后验分布。这是通过将滤波步骤拆分为一系列子步骤来实现的。在每个子步骤中，通过一个映射执行最优重采样，用等权重粒子替换非等权重粒子。该过程不需要对映射求逆或施加单调性约束，从而大大简化了流程。映射网络的参数根据粒子集距离进行优化。该距离具有可微性，并能比较非等权重粒子与等权重粒子。映射序列的复合为先验粒子到后验粒子提供了最终映射。我们使用径向基函数神经网络作为映射。重要的是，不需要任何中间的连续密度表示。整个流直接基于粒子表示运行，从而避免了代价高昂的密度估计。**

一、引言

我们研究基于概率密度函数（pdf）粒子表示的贝叶斯滤波。粒子滤波器是一个非常活跃的研究领域，已有数千篇论文致力于这一主题。其中大多数聚焦于大量随机粒子和渐近有效的方法。粒子数量必然有限，这会导致各种问题，包括：（i）粒子退化；（ii）由于粒子的随机性而产生的不可重复性。

应对退化问题的标准方法采用重采样。通过这种方式，加权样本被一组未加权样本所替代，以降低样本方差。然而，标准的重采样过程只是删除小权重样本，并复制大权重样本。在下一次预测步骤之前，它们的位置保持不变。这显然不是最优的，并且不允许在没有中间预测步骤的情况下连续执行多个滤波步骤。此外，还会使用提议密度和重要性采样来减轻退化效应。这使得在预测过程中考虑当前测量值，并将粒子置于更接近似然的位置。然而，这需要额外的工程设计，并会增加计算复杂度。

我们关注采用不同形式贝叶斯滤波更新的方法。这些方法通过使用传输映射或流，以一种本质上避免退化的间接更新取代直接更新。

II. 现有技术

将研究计算传输映射或流的最新技术。重点在于实现贝叶斯滤波步骤转换的方法。然而，有些方法仅针对两个给定密度之间映射的情形推导，未明确考虑贝叶斯更新。

A. 连续密度流滤波器

我们首先介绍连续密度的流，因为它们的提出早于粒子流。首个渐进式贝叶斯滤波器 [12] 采用连续密度表示后验分布。逐步引入似然函数，便形成了一种同伦延拓方法。由此得到一种高斯混合流，可使其与真实后验分布之间的平方积分偏差最小。该方法产生一组显式常微分方程组，并在从 0 到 1 的虚拟时间范围内求解。在 [8] 中，该方法被推广至 Kullback-Leibler散度和平方海林格距离。

B. 基于变换映射的滤波器

通过已知的非线性映射 $\boldsymbol{y} = g(\boldsymbol{x})$ 对随机向量 $\boldsymbol{x}$ 进行变换，是一个经典问题。目标是在给定输入密度 $f_x(x)$ 的情况下，计算输出密度 $f_y(y)$。为此，需要另一种将密度映射为密度的映射，即 $f_y = G(f_x)$。它可以根据原始映射 $g(\cdot)$、其根及其雅可比矩阵推导出来。对于单调映射，推导过程更为简单。

在这里，我们面临一个更复杂的问题。给定两个密度 $f_x(y)$ 和 $f_y(y)$，我们希望找到它们对应的随机向量 和 之间的映射 $g(.)$。当两个密度连续时，这样的映射存在、唯一且单调 [19]。找到该映射颇具挑战，尤其是在单调性约束下。在单变量情形中，可以由输入的累积分布和输出的分位数函数构造该映射。然而，累积分布和分位数函数的计算可能很困难，因为只有在特殊情况下才有解析表达式。在多变量情形中，这一问题更加突出。

如果只给出两个概率密度 $f_x(y)$ 和 $f_y(y)$ 的样本，问题就会更具挑战性。在这种情况下，无法使用概率密度之间的标准距离度量（例如 KL 散度）。

最复杂的情况是贝叶斯更新：此时只给出 $f_x(y)$ 的样本。于是，$f_y(y)$ 仅以样本与似然的乘积形式隐式给出，_我们既无法获取后验分布的取值，也无法从中采样。

基于 *Knothe-Rosenblatt* 三角映射的滤波器：在 [7], 中，假设给定了一个连续先验密度，可以从中轻松抽取样本。此外，还假设给定了一个可计算到常数因子的似然。映射通过先验与后验之间的海灵格度量或 KL 散度的最小值来刻画，参见 [7,第 7818–7819]页。该映射通过最小化计算得到，这是一个非凸问题，其中的距离通过从先验进行蒙特卡洛模拟来评估。Knothe-Rosenblatt 重排 [17] 被用于此目的，因为它们具有三角结构且单调，因此易于求逆[1]。需要注意的是，就传输距离而言，这些映射并非最优。为降低复杂度，文献 [25],研究了传输映射的可分解性，从而得到稀疏三角映射。文献 [18], 对映射的构造进行了推广，其中考虑了两类映射。第一类映射[2] （正向传输）将参考测度变换为目标测度 [18, p。 6]。参考密度是已知的，目标密度可以计算，但未经过归一化。第二类映射（逆向传输）将目标测度变换为给定的参考测度 [18, p。 13]。当目标密度未知且仅给定其样本时，这种方法非常有用。寻找该映射等价于最大似然估计，并且可以通过凸优化求解。这两类映射在 [24]中被用于推导非线性集合滤波器[24, p。 20] ，适用于只能获得先验密度样本的情况：构造一个逆向映射，将先验样本变换为方便处理的参考测度。使用正向映射将参考测度变换为后验分布。

基于归一化流的滤波器：归一化流在机器学习领域广受欢迎，并用于变分推断问题 [22] 中的密度估计，以对复杂的数据分布建模。归一化流是一系列可逆变换，可将参考测度映射为来自所需目标密度 [21, p 的一组样本 3]。这些流对应于 [18] 中的逆向传输，并通过最大似然估计确定。

C. 粒子流滤波器：连续推导

第一类粒子流滤波器在假设相关密度连续的前提下，根据适当参数化的后验分布推导偏微分方程或常微分方程。随后进行所需的离散化。

Daum-Huang（*DH*）粒子流：粒子流在 [2] 中从先验分布与后验分布之间的对数同伦推导而来。假设密度连续，该流由一个偏微分方程（福克-普朗克方程）表示。因此，从粒子中估计所需的梯度颇具挑战性的问题，见 [4]。已有多个版本的 DH 流被提出 [3], ，其中许多旨在应对流中的刚性。然而，它们通常依赖于并行运行的某种非线性卡尔曼滤波器，从而影响性能 [5]。

基于刘维尔方程的粒子流： [16], 采用了与 [12]类似的同伦延拓方法。假设存在一个常微分方程，使粒子从先验分布移动到后验分布，由此推导出相应的刘维尔偏微分方程[3]。（ [20, 第 5]页也采用了类似方法。）然后，通过求解刘维尔偏微分方程，得到所需常微分方程的速度场。在单变量情形下，可以得到可处理的解。多变量情形则需要吉布斯近似，这需要完整的条件分布，并通过数值方法近似。最后，映射后的粒子“仅仅”用作序贯蒙特卡罗采样器中的提议分布。

具有排斥核的粒子流：可以通过密度估计找到一种中间的连续表示来计算梯度。不过，密度估计通常不可微。在 [9], 中，使用 [10]排斥核来表示粒子周围概率质量的分布。这会产生一个常微分方程，描述粒子位置在从 0 到 1 的人工时间中的变化。

D. 粒子流滤波器：直接离散推导

第二类粒子流滤波器承认，计算机实现无论如何都需要进行离散化，因此直接推导出一系列离散更新。

在 [23], 中，似然函数被自适应地拆分为若干个子似然函数，每个子似然函数都更易于处理。随后，利用这些子似然函数对先验粒子进行顺序更新。每次更新后，都通过最优重采样，将加权粒子替换为权重相等的粒子。重采样通过最小化适当的粒子集距离 [14] 和 [11]来完成。因此，先验粒子会移动到后验密度较高的区域。在 [15], 中，利用平滑性假设，在重采样前进行局部上采样。

III. 问题表述

我们考虑一个状态为 $x \in \mathbb{R}^D$ 、状态维数为 D的动态系统。描述状态演化的转移密度可提供预测。该预测以先验密度 $f_p(x)$的形式给出。

给定先验密度 $f_p(x)$ 和似然函数 $f_L(x)$[4]，我们考虑一个非线性贝叶斯滤波步骤，用于更新先验密度并计算后验密度 $f_e(x)$：

$$f_e(x) \propto f_p(x) \cdot f_L(x) \ . \tag{1}$$

符号 $\propto$ 表示需要进行归一化，因为 $f_p(\circ)$ 与 $f_L(\circ)$ 的乘积不会自动得到归一化结果。

[1]关于一般三角变换，参见 [1]。[2]这是 [7]中使用的映射。

[3]刘维尔偏微分方程或连续性方程，是扩散率为零时福克-普朗克偏微分方程的一个特例。

[4]似然函数通常通过将特定测量值（例如 $\hat{y}$）代入条件密度 $f(y|x)$ 得到，该条件密度描述了测量值 y 与状态 x 之间的关系，使得 $f_L(x) = f(\hat{y}|x)$。

我们考虑一种重要情形：先验密度完全由一组样本（或粒子）给出，形式上可写为狄拉克混合密度。

$$f_p(\underline{x}) = \sum_{i=1}^{L} w_{p,i} \cdot \delta(\underline{x} - \underline{x}_{p,i}) \tag{2}$$

权重为 $w_{p,\,i} > 0$, $\sum_{i=1}^{L} w_{p,\,i} = 1$，样本位置为$x_{p,\,i}$。

对于给定的狄拉克混合先验，贝叶斯滤波步骤变为

$$\begin{aligned} \tilde{f}_e(\underline{x}) &\propto f_L(\underline{x}) \cdot \sum_{i=1}^{L} w_{p,i} \cdot \delta(\underline{x} - \underline{x}_{p,i}) \\ &= \sum_{i=1}^{L} \underbrace{w_{p,i} \cdot f_L(\underline{x}_i)}_{\tilde{w}_{e,i}} \cdot \delta(\underline{x} - \underline{x}_{p,i}) \ . \end{aligned} \tag{3}$$

归一化后，我们得到后验权重 $\tilde{w}_{e,\,i} / \sum_{i=1}^{L} \tilde{w}_{e\,i}$ $_{i} =$ $_{,,}$。

此时，后验狄拉克混合可表示为

$$\tilde{f}_e(\underline{x}) = \sum_{i=1}^{L} \tilde{w}_{e,i} \cdot \delta(\underline{x} - \tilde{\underline{x}}_{e,i}) \ . \tag{4}$$

后验样本位置相对于先验样本保持不变，即 $\tilde{x}_{e,\,i} = x_{p,\,i}$ 对于 $i = 1, \ldots, L$。

(4) 中的后验 $\tilde{f}_e(x)$ 是将贝叶斯滤波步骤直接应用于狄拉克混合后得到的。这会导致一个严重问题：样本不再具有相同的权重，对后验分布表示的贡献也不再相同。通常，某些粒子权重会变为（接近）零，实际上逐渐消失，从而导致前文提到的粒子退化。典型情形是系统噪声较大，使粒子在预测步骤中分散；同时测量噪声较小，导致似然函数较窄。

人们提出了许多方法来解决退化问题，其中一些是系统性的，但许多带有启发式性质。退化问题是一个根本性且棘手的问题。

注释 III.1. 我们的目标是推导出一种能够从根本上避免退化问题、且无需任何启发式方法的贝叶斯滤波器。它应易于理解、实现简单、数值稳定且稳健。

此外，我们提议使用确定性粒子，而非随机粒子。这（i）由于粒子分布更加均匀，可减少所需的粒子数量；（2）可确保结果可复现。本文采用[13] 中使用的采样方法，该方法见于 [27]。

IV. 最优重采样

现在，我们提出一种最优重采样步骤。它将非等权狄拉克混合 $\tilde{f}_e(x)$替换为其等权近似 $f_e(x)$。 $f_e(x)$由非等权重$\tilde{w}_{e,\,i}$ 和位置$\tilde{x}_{e,\,i} = x_{p,\,i}$组成。 $f_e(x)$具有相等的权重，即$w_{e,\,i} = w_{p,\,i} = 1/L$ 以及新的位置$x_{e,\,i}$。

我们将权重和位置分别收集到集合中，其中 $f_e(x)$

$$\begin{aligned} \mathcal{W}_e &= \{w_{e,1}, w_{e,2}, \ldots, w_{e,L}\} \ , \\ \mathcal{X}_e &= \{\underline{x}_{e,1}, \underline{x}_{e,2}, \ldots, \underline{x}_{e,L}\} \ , \end{aligned} \tag{5}$$

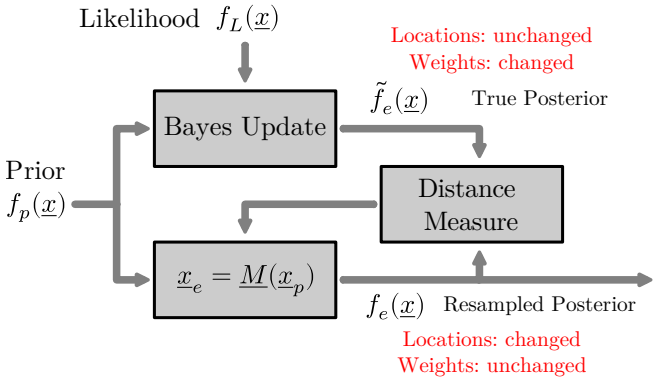


图 1： 给定先验 $f_p(x)$ 和似然 $f_L(x)$ 时生成重采样映射的框图。 $\tilde{f}_e(x)$ 是基于基础更新得到的权重不等的后验分布。 $f_e(x)$ 是重采样后权重相等的后验分布，其中 $f_e(x) \approx \tilde{f}_e(x)$。

以及 $f_e(x)$

$$\begin{aligned} \tilde{\mathcal{W}}_e &= \{\tilde{w}_{e,1}, \tilde{w}_{e,2}, \ldots, \tilde{w}_{e,L}\} \ , \\ \tilde{\mathcal{X}}_e &= \{\tilde{\underline{x}}_{e,1}, \tilde{\underline{x}}_{e,2}, \ldots, \tilde{\underline{x}}_{e,L}\} \ , \end{aligned} \tag{6}$$

这里使用向量和矩阵以外的表述，强调不存在固有的顺序。

注释 IV.1。 我们首先考虑没有发生退化 $\tilde{f}_e(x)$ 的情况。这意味着 $\tilde{w}_{e,i}$ 中的权重方差较小，所有样本都对密度表示有所贡献。第 V 节将讨论退化问题。

寻找 $x_{e,i}$ 的关键思路是使用最优映射 $M(.)$，将先验随机向量 $\boldsymbol{x}_p$ 转换为后验随机向量 $\boldsymbol{x}_e$

$$\boldsymbol{x}_e = \underline{M}(\boldsymbol{x}_p) \ . \tag{7}$$

此映射用于将先验样本 $x_{p,i}$ 映射为后验样本$x_{e,i}$。映射样本不会改变其权重，只会改变其位置。这保证了后验样本$x_{e,i}$的权重相等。现在我们需要找到映射 $M(.)$，使所得后验样本满足 $f_e(x) \approx \tilde{f}_e(x)$。

映射也如图 1 所示。 $f_p(x)$通过贝叶斯更新得到，即将 $f_p(x)$与似然$f_L(x)$相乘（上方路径）。它作为参考密度。其位置$\tilde{x}_{e,i}$ 与 $f_p(x)$的位置相同，只是其权重$\tilde{w}_{e,i}$ 发生了变化。在下方路径中，映射 $M(.)$将$f_p(x)$传播至 $f_e(x)$。 $f_e(x)$的权重与 $f_p(x)$相同，但其位置$x_{e,i}$ 因映射而发生了变化。我们希望 $f_e(x)$相对于适当的距离度量 D，尽可能接近 $\tilde{f}_e(x)$。因此，通过最小化 $D(\tilde{f}_e(x), f_e(x))$来相应地调整映射 $M(.)$。

A. 映射的性质

使用离散映射分别映射L 样本就足够了。然而，这样一来，映射参数的数量将等于样本数量。因此，与直接重新近似相比，我们无法获得复杂度收益。此外，也无法进行平滑。

因此，我们采用连续映射，最好尽可能减少参数数量，以降低复杂度。这会自动

实现样本之间的插值。总之，只需少量先验样本，就能经济地生成平滑映射，用于映射大量样本。

注释 IV.2。 插值可用于增加后验样本的数量 x_i^e。这可以在映射生成后进行。

映射 $M(.)$ 具有以下性质：
性质 1 可能非单调。性质 2 无需逆映射。性质 3 仅用于映射样本值。性质 4 要求关于参数可微。性质 5 不要求关于 x 可微。

B. 特定映射

根据注释 IV.1，本节假设 $f_p(x)$ 到 $f_e(x)$ 的变化很小。因此，映射 $M(.)$ 接近恒等映射。

这里，我们提出将仿射基映射与径向基函数非线性相结合的方法。对于单个输出，其形式为

$$M_i(\underline{x}) = \underbrace{\underline{a}_i^T \cdot \underline{x} + b_i}_{\text{affine part}} + \underbrace{\sum_{r=1}^{R} v_{r,i} \cdot \text{RBF}_r(\underline{x} - \underline{x}_r)}_{\text{radial basis function part}} , \quad (8)$$

权重为 $v_{r,\,i} \in \mathbb{R}$，位置为 x_r，径向基核为 $\text{RBF}_r(\cdot)$。完整的向量值映射 $M(\cdot) : \mathbb{R}^D \to \mathbb{R}^D$ 定义为

$$\underline{M}(\underline{x}) = \left[M_1(\underline{x}), M_2(\underline{x}), \ldots, M_D(\underline{x})\right]^T . \quad (9)$$

对于优化，初始值可直接设为 $a_i = 1$、$b_i = 0$ 和 $v_{r,\,i} = 0$，其中 $r = 1, \ldots, R$、$i = 1, \ldots, D$。径向基函数位置 x_r 可设为 $r = 1, \ldots, R$ 的固定先验位置。

C. 距离度量

我们需要比较加权狄拉克混合 $f_e(x)$ 与其无权对应项 $f_e(x)$。合适的距离度量应满足以下要求：

要求 1：处理狄拉克混合密度[5]。
要求 2：处理非等权重。
要求3：处理支撑集不相等的情况。
要求4：关于位置（以及权重）可微。前面三个性质无法通过要求连续密度的标准距离来处理，包括 KL散度和积分平方距离。另一方面，可以使用 Wasserstein距离，但其复杂度较高。此外，它们不满足性质 4。本文提出使用基于局部累积分布的 Cramér-von Mises 距离[11] [14]。

对于两个包含 L 个分量的狄拉克混合密度 f_x，权重 $w_{x,\,1}$，$w_{x,\,2}$，$\ldots$，$w_{x,\,L}$，位置

$$\underline{x}_i = \left[x_{1,i}, x_{2,i}, \ldots, x_{D,i}\right]^T \in \mathbb{R}^D \quad (10)$$

[5]即连续域上的离散密度。

对于 $i = 1, \ldots, L$ 以及具有 M 个分量、权重 $w_{y,1}, w_{y,2}, \ldots, w_{y,M}$ 的 f_y，位置

$$\underline{y}_j = \left[y_{1,j}, y_{2,j}, \ldots, y_{D,j}\right]^T \in \mathbb{R}^D \quad (11)$$

对于 $j = 1, \ldots, M$，Cramér-von Mises 距离 D 定义为

$$D = D_{yy} - 2D_{xy} + D_{xx} + c\,D_E , \quad (12)$$

with

$$D_{yy} = \sum_{i=1}^{M}\sum_{j=1}^{M} w_{y,i} \cdot w_{y,j} \cdot \text{xlog}\left(\sum_{d=1}^{D}(y_{d,i} - y_{d,j})^2\right) , \quad (13)$$

$$D_{xy} = \sum_{i=1}^{L}\sum_{j=1}^{M} w_{x,i} \cdot w_{y,j} \cdot \text{xlog}\left(\sum_{d=1}^{D}(x_{d,i} - y_{d,j})^2\right) , \quad (14)$$

$$D_{xx} = \sum_{i=1}^{L}\sum_{j=1}^{L} w_{x,i} \cdot w_{x,j} \cdot \text{xlog}\left(\sum_{d=1}^{D}(x_{d,i} - x_{d,j})^2\right) , \quad (15)$$

其中 $\text{xlog}(z) = z \cdot \log(z)$。$D_E$ 可视为惩罚项（权重为 c），用于确保均值相等，其定义为

$$D_E = \sum_{d=1}^{D}\left(\sum_{i=1}^{L} w_{x,i} \cdot x_{d,i} - \sum_{i=1}^{M} w_{x,i} \cdot y_{d,i}\right)^2 . \quad (16)$$

当只需要关于参数的 f_x 的极小化解，而不需要对应的值 D 时，可以忽略 D_{yy}。

D. Map Optimization

本文假设 L 与 M 相等，其中 $w_{x,\,i} = w_{e,\,i}$、$w_{y,\,i} = \tilde{w}_{e,\,i}$ 以及 $x_i = x_{e,\,i}$、$y_i = \tilde{x}_{e,\,i}$、$i = 1, \ldots, L$。由此得到以下依赖关系 D

$$D = D\left(\mathcal{W}_e, \mathcal{X}_e, \tilde{\mathcal{W}}_e, \tilde{\mathcal{X}}_e\right) . \quad (17)$$

由于 $\tilde{\mathcal{X}}_e$ 是通过映射 $\mathcal{X}_e = M(\mathcal{X}_e)$ 从 $\mathcal{X}_e$ 得到的，因此可以将其改写为

$$D = D\left(\mathcal{W}_e, \underline{M}(\tilde{\mathcal{X}}_e), \tilde{\mathcal{W}}_e, \tilde{\mathcal{X}}_e\right) . \quad (18)$$

现在通过极小化求得最优映射 M^*

$$\underline{M}^* = \underset{\underline{M} \in \mathcal{M}}{\arg\min}\, D\left(\mathcal{W}_e, \underline{M}(\tilde{\mathcal{X}}_e), \tilde{\mathcal{W}}_e, \tilde{\mathcal{X}}_e\right) , \quad (19)$$

其中，$\mathcal{M}$ 是可行映射的集合。

(12)中 D 的梯度可以用闭式形式表示 [11]。当似然以解析形式给出且可微时，可以推导出关于映射参数的梯度。优化采用BFGS拟牛顿法。

V. 渐进式处理

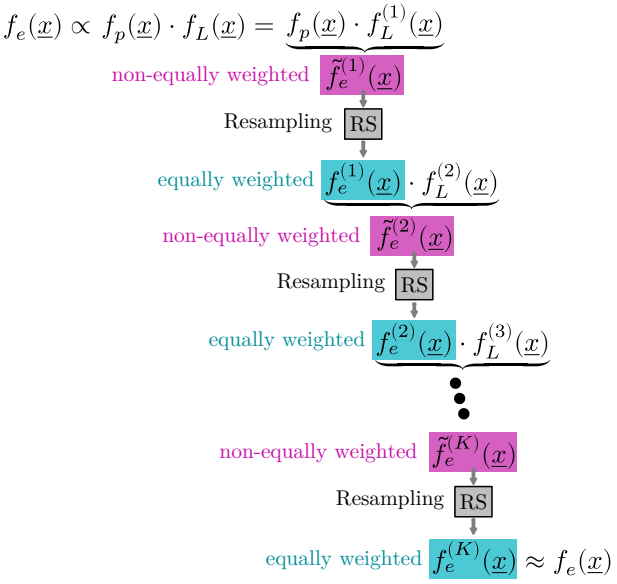
在上一节中，我们假设根据似然 $f_L(x)$ 对粒子进行重加权后，所有粒子都保持“存活”状态。然而，通常一次性执行贝叶斯更新会导致粒子退化。只有少数样本保持“存活”状态，其余样本的权重都接近于零。在这种情况下，由于并非所有粒子都对密度表示作出贡献，分辨率会有所损失。

一种让粒子保持“存活”的有效方法是进行渐进式处理 [23]。似然被分解

为多个子似然的乘积，并仔细选择每个子似然，以避免退化

$$f_L(\underline{x}) = f_L^{(1)}(\underline{x}) \cdot f_L^{(2)}(\underline{x}) \dots f_L^{(K)}(\underline{x}) = \prod_{k=1}^{K} f_L^{(k)}(\underline{x}) \quad . \quad (20)$$

由于我们采用乘积分解，因此每个子似然直观上都比原始似然“更宽”。随后，利用这些子似然进行序贯贝叶斯子更新，即可得到所需的后验分布：



每次子更新后，我们都会得到一个权重不均等的子后验分布 $\tilde{f}_e^{(k)}(x)$。为了准备下一次子更新，我们采用第 IV 节中的最优重采样方法。这包括利用子映射 $M_{(k)}(.)$，将样本 $\tilde{x}_{e\,i}^{(k)}$ 映射为 $x_{e\,i}^{(k)}$。由此得到一个权重均等的子后验分布 $f_{(k)e}(x) \approx \tilde{f}_{(k)e}(x)$。经过 K 次子更新后，得到权重均等的子后验分布 $f_e^{(K)}(x)$，它等于 $f_e(x)$。

从先验样本 $x_{p,i}$ 到后验样本 $x_{e,i}$ 的总映射由一系列单独映射的复合给出：

$$\underline{M}(\underline{x}) = \underline{M}^{(K)}\left(\underline{M}^{(K-1)}\left(\dots \underline{M}^{(2)}\left(\underline{M}^{(1)}(\underline{x})\dots\right)\right)\right) \quad (21)$$

or

$$\underline{M} = \underline{M}^{(K)} \circ \underline{M}^{(K-1)} \circ \dots \circ \underline{M}^{(2)} \circ \underline{M}^{(1)} \quad . \quad (22)$$

VI. 数值结果

A. 合理性检验：线性系统

作为第一个示例，我们考虑最简单的标量线性测量方程 $y = x + v$。x 是所需的状态，y 是测量值，而 v 是高斯测量噪声，其$v \sim f_v(v) = \mathrm{N}(v; 0,1)$[6]。先验为高斯分布，表示为 $f_p(x) = \mathrm{N}(x; 0,1)$。对于 $y = \hat{y}$，似然表示为$f_L(x) = f(\hat{y}|x) = f_v(\hat{y} - x) = \mathrm{N}(\hat{y};\, x, 1)$。

[6]$\mathrm{N}(x;\, m, \sigma)$ 是定义在实现值 x 上的高斯密度，其均值为 m，标准差为 σ。

我们假设只能获得先验密度 $f_p(x)$ 的样本$x_{p,\,i}$。所提出的方法用于执行似然 $f_L(x)$ 下的贝叶斯滤波步骤。由此得到后验分布 $f_e(x)$ 的样本$x_{e,\,i}$。

在这个简单的情形下，我们可以利用解析先验和似然计算真实后验分布及其累积分布函数。真实后验分布表示为 $f_e^t(x) = \mathrm{N}(x; \hat{y}/2, 1/2)$。样本 $x_{e,i}^t$现可用作来自 $f_e^t(x)$ 的样本 $x_{e,i}$ 的参考。请注意，估计器并不知道解析先验和参考样本。

对于 $L = 10$ 个粒子、先验标准差 $\sigma_p = 1$，以及噪声标准差 $\sigma_v = 1$、$\sigma_v = 0.6$ 和 $\sigma_v = 0.3$，我们得到图 2 所示的结果。真实的连续后验（蓝色）与所提出的贝叶斯粒子流生成的样本（青绿色）非常接近。噪声方差越小，近似效果越差。这是预期之中的，因为仅有 $L = 10$ 个粒子的先验密度分辨率有限。

图 3 展示了相同设置下使用 $L = 30$ 个粒子得到的结果。真实的连续后验仍以蓝色显示。所提出的贝叶斯粒子流生成的样本（青绿色）几乎完美地对其进行了近似。噪声方差较小时，近似效果仅略微变差。作为对比，图中还以红色显示了通过直接从真实连续后验中采样得到的参考样本 $x_{e,\,i}^t$。

先验 $\boldsymbol{x}_p$ 到后验分布 $\boldsymbol{x}_e$的相应总映射如图 6 所示（红色）。我们将其与真实的线性映射（绿色）进行比较。在相关区域内，这些映射几乎完全相同。

B. 立方传感器问题

现在考虑这个（不）知名的立方传感器问题，其测量方程为 $y = x^3 + v$。同样，x 是状态，y是测量，v 是高斯测量噪声，其中$v \sim f_v(v) = \mathrm{N}(v; 0, \sigma_v)$。先验分布为高斯分布，给定为 $f_p(x) = \mathrm{N}(x; 0,1)$。对于 $y = \hat{y}$，似然为$f_L(x) = f_v(\hat{y} - x^3) = \mathrm{N}\left(\hat{y}; x^3, \sigma_v\right)$。真实后验分布为

$$\tilde{f}_e(x) \propto f_p(x) \cdot f_L(x) = \mathrm{N}(0,1) \cdot \mathrm{N}\left(\hat{y}; x^3, \sigma_v\right) \quad . \quad (23)$$

注 VI.1。 从 $f_e(x)$ 中采样很困难。所需的累积分布函数及其逆函数只能通过数值积分计算。

我们仅获得先验密度 $f_p(x)$ 的样本 $x_{p,\,i}$ 以及解析似然 $f_L(x)$。利用所提出的贝叶斯流计算后验分布$f_e(x)$的样本 $x_{e,\,i}$。图 4（4、5）展示了先验 $f_p(x)$、累积分布 $F_p(x)$及其样本$x_{p,\,i}$。图 4（3）展示了粒子的流动过程。图 4（1、2）展示了后验分布 $f_e(x)$、累积分布 $F_e(x)$及其样本 $x_{e,\,i}$。从 $\boldsymbol{x}_p$ 到 $\boldsymbol{x}_e$ 的总映射如图 7 所示（红色），并与数值参考（绿色）进行了对比。同样，在相关区域内，这些映射几乎完全一致。

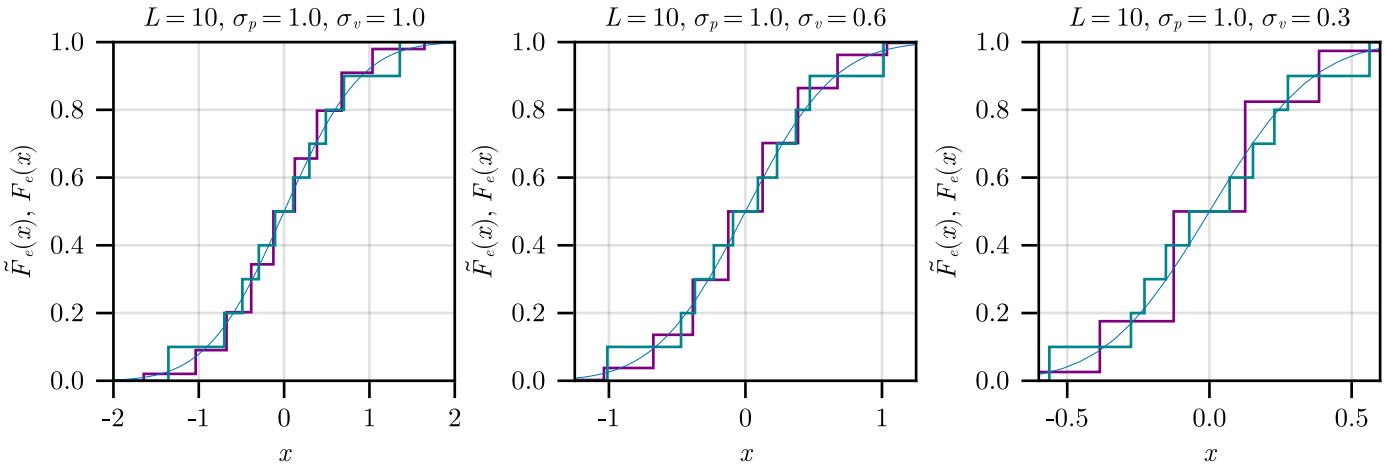


图2： 使用 $L = 10$ 个样本进行线性更新的结果。真实的连续后验以蓝色显示。将先验 $f_p(x)$ 与似然 $f_L(x)$ 相乘后得到的非等权狄拉克混合 $\tilde{f}_e(x)$ 以紫色显示。所提出的贝叶斯粒子流生成的等权狄拉克混合 $f_e(x)$ 以青绿色显示。请注意 x 轴刻度不同。

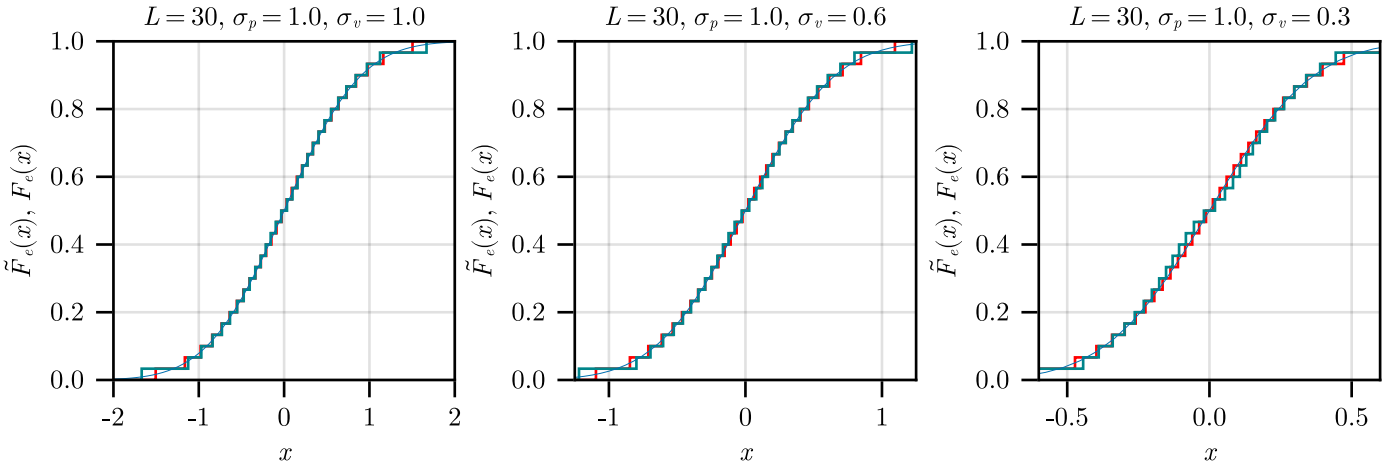


图 3： 使用 $L = 30$ 个样本进行线性更新的结果。真实的连续后验以蓝色显示。所提出的贝叶斯粒子流生成的等权狄拉克混合 $f_e(x)$ 以青绿色显示。通过直接从真实连续后验中抽样得到的参考狄拉克混合以红色显示。请注意 x 轴刻度不同。

C. 与粒子滤波器的比较

现在将所提出的贝叶斯粒子流与标准粒子滤波器进行比较。底层的连续先验表示为 $f_p(x) = \mathrm{N}(x; 0, 1)$。似然

$$f_L(x) = \exp\left(-\frac{1}{2}\big((x-1.2)(x-1.5)(x+1.2)(x+1.5)\big)^2\right) \tag{24}$$

如图8所示，先验 $f_p(x)$ 以蓝色显示，似然 $f_L(x)$ 以绿色显示，后验分布 $f_e(x)$ 以红色显示。这些底层连续密度对于估计器而言是未知的。

对于该仿真，我们仅能获得先验 $f_p(x)$ 的样本 $x_{p,\,i}$，$i = 1, \ldots, L$以及式 (24) 中的解析似然。结果如图5所示。图5 (1) 中所示，所提出的贝叶斯粒子流在 $L = 50$ 个粒子下的后验估计与真实后验分布非常接近。图5 (2)-(5) 展示了标准粒子滤波器在不同L 下各运行十次的结果。对于 $L = 50$，后验分布峰值处的分辨率较低。这是因为粒子滤波器的滤波步骤会产生加权样本。此外，不同运行之间的差异很大。随着 L 增大，分辨率下降的现象会减轻，运行之间的差异也会减小。不过，即使对于 $L = 500$，差异仍然十分明显。

VII. 结论

本文推导出一种新的贝叶斯粒子流，可将粒子从先验确定性地引导至后验分布。无论在其推导还是实现中，均不需要任何连续密度表示。该方法由有限序列的、可能为非单调映射的映射构成，用于传播粒子。该滤波器适用于任意非线性测量方程。然而，假设这些方程已经被转换为可求值的似然函数。该方法易于理解，实现也很直接。

映射优化可以采用多种距离度量，参见第 IV-A 小节。不过，我们发现 [11] 中导出的距离度量最为实用，因为它复杂度较低且

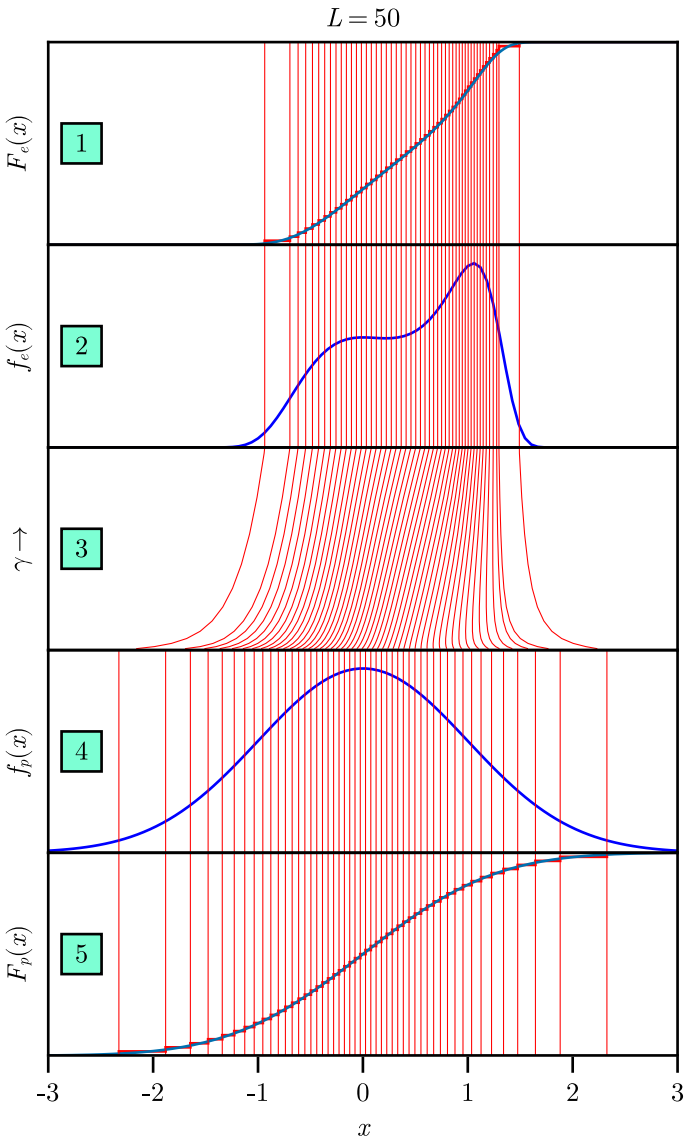


图 4： 立方传感器问题：更新结果。(1,2) 后验分布 $f_e(x)$、累积分布 $F_e(x)$和样本$x_{e,\ i_e}$。(3) 粒子的流动。(4,5) 先验 $f_p(x)$、累积分布 $F_p(x)$和样本$x_{p,\ i_p}$。

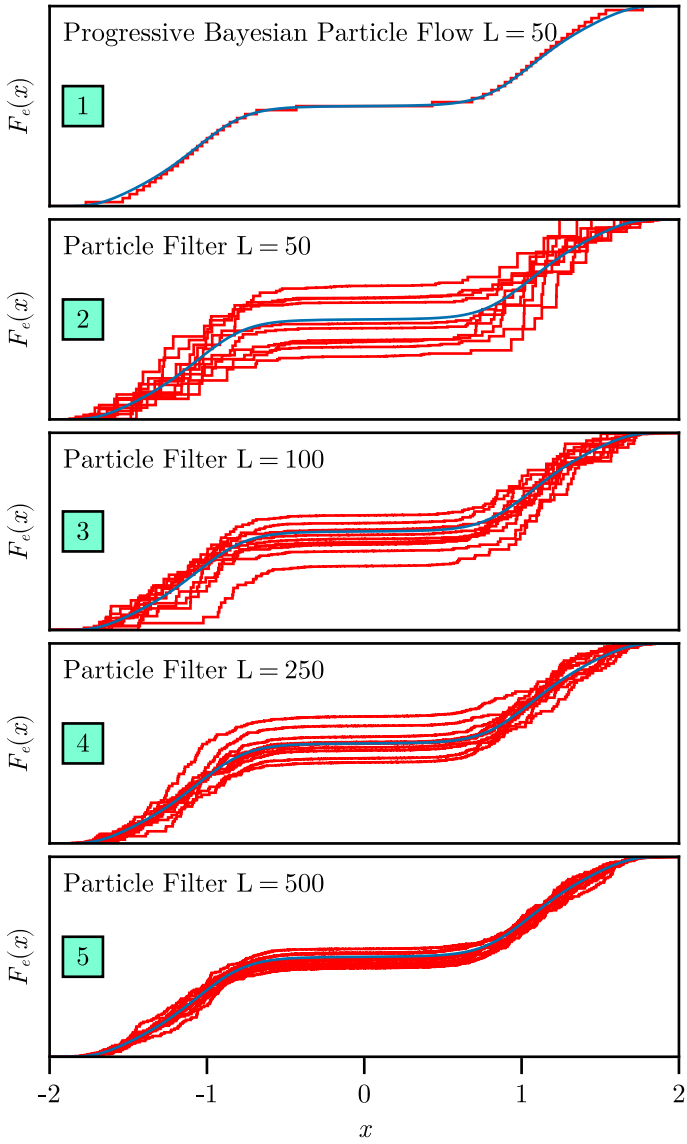


图5： 比较贝叶斯粒子流与标准粒子滤波器在图8所示似然下进行单次测量更新的效果。(1) 所提出的滤波器在 $L = 50$ 下的结果。(2)-(5) 粒子滤波器在不同 L 下的结果，每种情况各运行10次。

可微。其复杂度与粒子数量 L 的平方成正比，与维度数量 D 成正比。

由于篇幅限制，本文省略了若干方面。(i) 我们没有详细说明如何找到合适的子似然，使粒子能够“存活”及其数量。为此，我们采用了 [26]中描述的方法。(ii) 我们假设粒子数量 L 是恒定的。其实不必如此。样本数量可以根据底层密度的复杂度进行调整。我们将进一步开发高效的调整方法。(iii) 我们没有讨论预测步骤，即如何通过系统模型传播粒子。对于确定性粒子，在结合状态和噪声样本方面，预测步骤与随机情形有显著差异。为避免完整的笛卡尔积，可以使用 [6]中的方法。

参考文献

[1] V. I. Bogachev、A. V. Kolesnikov 和 K. V. Medvedev。《测度的三角变换》。《*Sbornik: Mathematics*》196.3（2005年4月30日），第309页。[2] Fred Daum 和 Jim Huang。《带对数同伦的非线性滤波器粒子流》。*SPIE*论文集，小目标信号与数据处理。第6969卷。2008年。[3] Fred Daum 和 Jim Huang。《在非线性滤波器、贝叶斯决策和传输中，缓解具有非零扩散的粒子流刚性的七种可疑方法》。第9092卷。2014年，第90920C–90920C–11页。[4] Frederick Daum 等。《对数同伦诱导的非线性滤波器粒子流梯度估计》。*SPIE*论文集，信号处理、传感器融合与目标识别 *XVIII*。第7336卷。2009年。[5] Tao Ding 和 M.J. Coates。《Daum-Huang精确流粒子滤波器的实现》。*2012*年 *IEEE Statistical*

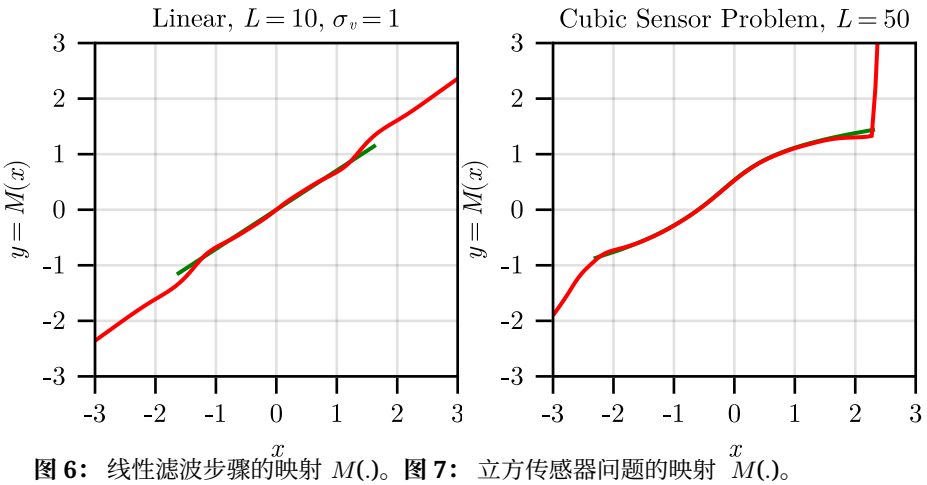


图 6： 线性滤波步骤的映射 $M(.)$。**图 7：** 立方传感器问题的映射 $M(.)$。

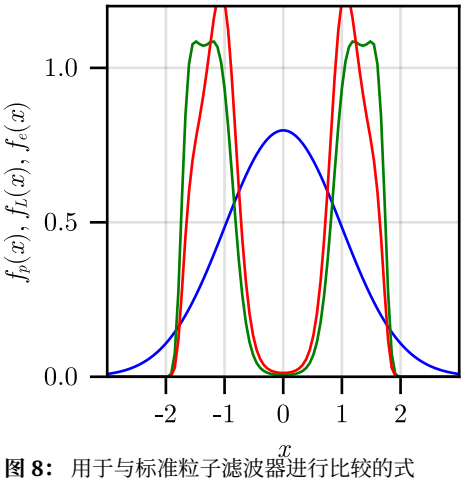


图 8： 用于与标准粒子滤波器进行比较的式 (24) 中的密度。

信号处理研讨会（*SSP*）。2012年 IEEE统计信号处理研讨会（SSP）。2012年8月，第257–260页。[6] Henning Eberhardt、Vesa Klumpp 和 Uwe D. Hanebeck。《利用独立性实现最优狄拉克近似》。*2010*年美国控制会议论文集（*ACC 2010*）。美国马里兰州巴尔的摩，2010年6月。[7] Tarek A. El Moselhy 和 Youssef M. Marzouk。《利用最优映射进行贝叶斯推断》。《计算物理学杂志》231.23（2012年10月1日），第7815–7850页。[8] J. Hagmar 等。《利用同伦对后验密度进行最优参数化》。第*14*届国际信息融合会议论文集（*Fusion 2011*）。第14届国际信息融合会议（Fusion 2011）。IEEE，2011年7月5–8日，第1–8页。[9] Uwe D. Hanebeck。《FLUX：基于Zakai型分布式常微分方程的渐进式状态估计》。*arXiv*预印本：系统与控制（*cs.SY*）（2018年8月）。[10] Uwe D. Hanebeck。《任意概率密度函数的基于核的确定性蓝噪声采样》。第*48*届信息科学与系统年度会议论文集（*CISS 2014*）。美国新泽西州普林斯顿，2014年3月。[11] Uwe D. Hanebeck。《多元狄拉克混合密度的最优约简》。《自动化技术》，*Oldenbourg*出版社63.4（2015年4月），第265–278页。[12] Uwe D. Hanebeck、Kai Briechle 和 Andreas Rauh。《渐进式贝叶斯：非线性状态估计的新框架》。*SPIE*论文集，*AeroSense*研讨会。第5099卷。美国佛罗里达州奥兰多，2003年5月，第256–267页。[13] Uwe D. Hanebeck、Marco F. Huber 和 Vesa Klumpp。《多元高斯密度的狄拉克混合近似》。*2009*年 *IEEE*决策与控制会议论文集（*CDC 2009*）。中国上海，2009年12月。[14] Uwe D. Hanebeck 和 Vesa Klumpp。《局部累积分布及克拉梅尔–冯·米塞斯距离的多元推广》。*2008*年 *IEEE*智能系统多传感器融合与集成国际会议论文集（*MFI 2008*）。大韩民国首尔，2008年8月，第33–39页。[16] Jeremy Heng、Arnaud Doucet 和 Yvo Pokern。《用于近似传输的吉布斯流及其在贝叶斯计算中的应用》。《皇家统计学会杂志*B*辑：统计方法论》83.1（2021年2月1日），第156–187页。

[15] Uwe D. Hanebeck 和 Martin Pander。“使用确定性粒子的渐进式贝叶斯估计”。第 *19* 届国际信息融合会议论文集（*Fusion 2016*）。德国海德堡，2016年7月。[17] Herbert Knothe。“凸体理论的贡献”。《密歇根数学期刊》4.1（1957年1月），第 39–52 页。[18] Youssef Marzouk 等。“通过测度传输进行采样：导论”。《不确定性量化手册》。Roger Ghanem、David Higdon 和 Houman Owhadi 编。查姆：施普林格国际出版公司，2016年，第 1–41 页。[19] Robert J. McCann。“单调保测映射的存在性与唯一性”。《杜克数学期刊》80.2（1995年11月1日）。[20] De Melo 等。用于非线性高维滤波问题的随机粒子流。2015年11月4日。[21] George Papamakarios 等。“用于概率建模和推断的归一化流”。《机器学习研究期刊》22.1（2021年1月1日），57:2617–57:2680。[22] Danilo Rezende 和 Shakir Mohamed。“基于归一化流的变分推断”。国际机器学习会议。PMLR，2015年，第 1530–1538 页。[23] Patrick Ruoff、Peter Krauthausen 和 Uwe D. Hanebeck。“确定性狄拉克混合近似的渐进校正”。第 *14* 届国际信息融合会议论文集（*Fusion 2011*）。美国伊利诺伊州芝加哥，2011年7月。[24] Alessio Spantini、Ricardo Baptista 和 Youssef Marzouk。用于非线性集合滤波的耦合技术。2019年6月30日。[25] Alessio Spantini、Daniele Bigoni 和 Youssef Marzouk。“通过低维耦合进行推断”。《机器学习研究期刊》19.1（2018年1月1日），第 2639–2709 页。[26] Jannik Steinbring 和 Uwe D. Hanebeck。“使用显式似然函数的渐进式高斯滤波”。第 *17* 届国际信息融合会议论文集（*Fusion 2014*）。西班牙萨拉曼卡，2014年7月。[27] Jannik Steinbring、Martin Pander 和 Uwe D. Hanebeck。“采用对称样本的智能采样卡尔曼滤波器”。《信息融合进展期刊》11.1（2016年6月），第 71–90 页。