自适应多核相关熵卡尔曼滤波器：变分贝叶斯方法

Zhen Zeng，Shile i Li，电气电子工程学会会员，Dawei Shi，电气电子工程学会高级会员，以及 Ling Shi，电气电子工程学会会士

arXiv:2611.32679v1 [stat.AP] 24 Nov 2026

摘要—**本文提出一种自适应多核相关熵卡尔曼滤波器（AMCKF），用于在存在重尾测量噪声和离群值的情况下进行鲁棒状态估计。通过半二次（HQ）对偶，将多核相关熵损失重新表述为带辅助权重的二次型，从而实现高效的类卡尔曼递推。我们在平均场假设下建立了严格的变分贝叶斯框架，并推导出用于联合状态与核宽度估计的闭式坐标上升变分推断更新。为实现在线处理，我们进一步提出AMCKF，作为精确坐标上升变分推断（CAVI）的一种高效单遍近似；其中，状态和辅助变量通过迭代重加权最小二乘（IRLS）迭代更新，随后进行闭式核宽度更新。我们设计了一种遗忘机制，以防止过度自信并确保长期适应性。在目标跟踪和IMU姿态估计方面的仿真结果表明，在存在稀疏离群值、重尾噪声和混合干扰的场景中，所提出的AMCKF显著优于标准KF、VBKF、基于Student t分布的KF和固定核MKCKF，同时在高斯条件下保持最优性能。**

索引词—**半二次分解、重尾噪声、核宽度自适应、鲁棒滤波、变分贝叶斯推断。**

I 引言

TATE 估计是现代信号处理与控制的基础，其中卡尔曼滤波器（KF）在线性高斯假设 [1], [2], [3] 下作为标准解决方案。KF及其扩展已广泛应用于航空航天跟踪、导航和机器人学 [4], [5], [6]领域。

然而，KF的最优性依赖于对噪声协方差的精确了解。在实际应用中，由于环境扰动、传感器退化或未建模动态 [7], [8], [9]，噪声统计特性通常是未知的、时变的，或会发生突变。此外，实际传感器中普遍存在带有离群值的重尾测量噪声，这会严重降低KF的性能，甚至导致发散 [10]。

解决这些问题主要有两个方向：自适应滤波和鲁棒滤波。自适应滤波在线估计随时间变化的噪声统计特性，其中变分贝叶斯（VB）方法日益受到关注。S¨arkk¨a 和 Nummen-maa 首创了用于联合状态和测量噪声协方差估计的递归式 VB 框架 [11]，后来扩展到过程噪声 [12]和非线性系统 [13], [14], [15]。其他自适应方法包括协方差匹配 [16]、基于新息的方法 [17]，和多模型策略 [18]。然而，传统的变分贝叶斯滤波器基于高斯分布，因此对离群值较为敏感 [19], [20]。

鲁棒滤波通过重尾模型或鲁棒代价函数抑制离群值。Student's t 分布滤波器能够自然地降低大误差的权重，并已被纳入变分贝叶斯框架 [21], [22], [23]。最大相关熵准则（MCC）提供了另一种有效解决方案 [24], [25]，[26]。与 Student's t 损失相比，多核相关熵损失在大误差下更快达到饱和，从而提供更强的重尾噪声抑制能力 [27], [8], [28]。因此，基于 MCC 的卡尔曼滤波器对脉冲噪声表现出很强的鲁棒性 [29]。

然而，这些鲁棒滤波器的实际性能在很大程度上取决于核宽度的手动调节。为解决这一问题，研究人员提出了多种自适应策略。Fakoorian 等人 [30] 提出了新息与误差协方差的启发式加权和。Chen 等人 [31] 设计了结合卡方检测的保护窗口，以在 MCC 与最小均方误差（MMSE）之间进行切换。Li 等人 [32] 和 Liu 等人 [33]探索了采用马氏距离或阻抗因子的变分贝叶斯框架。最近，Zhang 等人 [34]引入了带有最优因子的自适应柯西核，以同时处理离群值和卫星姿态的状态突变

估计。尽管取得了这些进展，现有方法仍依赖启发式规则、离散切换阈值或经经验调节的参数，缺乏一个统一的概率框架来实现有原则的在线核宽度学习。这促使我们开发一种能够在线学习核宽度的变分贝叶斯方法。

为克服这些局限，我们建立了一个严谨的变分贝叶斯框架，并推导出用于联合状态与核宽度估计的闭式 CAVI 更新。通过半二次（HQ）对偶性 [35], [36], [37]，我们将非凸相关熵损失改写为带有辅助权重的二次型，从而能够进行高斯状态更新。为实现在线处理，我们进一步提出自适应多核相关熵卡尔曼滤波器（AMCKF），将其作为精确 CAVI 的高效单遍近似；在该方法中，通过迭代重加权最小二乘（IRLS）迭代更新状态和辅助变量，随后以闭式更新核宽度。一种遗忘机制可防止持续出现离群值时核宽度发生塌缩。所提出的框架可自然推广到满足 HQ 条件的更广泛的鲁棒损失函数。主要贡献如下：

1）我们建立了一般的 HQ 对偶性结果（**命题 1**）

Zhen Zeng任职于厦门大学自动化系，中国厦门，361102（电子邮箱：Zengzhen200413@gmail.com）。Shilei Li任职于北京理工大学自动化学院，中国北京，100081（电子邮箱：shileili@bit.edu.cn）。Dawei Shi任职于北京理工大学自动化学院，中国北京，100081（电子邮箱：daweishi@bit.edu.cn）。Ling Shi任职于香港科技大学电子与计算机工程系，中国香港，999077，特别行政区（电子邮箱：eesling@ust.hk）。

这表明，任何特定形式的损失都可以分解为尺度参数在线性形式下出现在对偶势函数中的形式，从而统一一类鲁棒损失。

2）在平均场假设下，我们构建了严格的变分贝叶斯框架，推导出用于联合状态与核宽度估计的闭式 CAVI 更新，并证明了 HQ-诱导似然的伽马共轭性（**定理 1**）。

3）我们提出了 AMCKF，作为精确 CAVI 的一种高效单遍近似方法，其中状态和辅助变量通过迭代重加权最小二乘进行迭代更新，随后执行闭式核宽度更新。

4）蒙特卡洛模拟表明，在各种重尾噪声和离群值场景下，所提出的 AMCKF 始终优于 KF、VBKF、基于 Student'st 的 KF 以及固定核MKCKF，同时达到了与 CAVI 算法相当的精度。

本文其余部分的组织结构如下。第二节对问题进行建模。第三节介绍严格的 VB 框架及其高效近似方法——所提出的 AMCKF。第四节介绍仿真结果。第五节总结全文。

符号说明： $n \times n$ 单位矩阵记为 I_n。矩阵 A 的转置和逆矩阵分别记为 A^T 和A^{-1}。 $\mathbb{E}[\cdot]$表示期望算子。均值为 μ 、协方差为 Σ 的高斯分布记为 $\mathcal{N}(\mu,\ \Sigma)$，而 $\Gamma(\alpha,\ \beta)$ 表示形状参数为 α 、率参数为 β的伽马分布。以向量 v 为对角元素的对角矩阵记为$\mathrm{diag}(v)$。向量的第 i个分量记为 $[\cdot]_i$。

II. 问题表述

考虑以下离散时间线性动态系统：

$$x_{k+1} = Ax_k + w_k, \tag{1}$$
$$y_k = Cx_k + v_k, \tag{2}$$

其中， $x_k \in \mathbb{R}^n$ 为系统状态， $y_k \in \mathbb{R}^m$ 为测量向量， w_k 和 v_k 分别表示过程噪声和测量噪声。矩阵 $A,\ B$ 为系统参数，且矩阵对$(A,\ C)$是可观测的。初始状态 x_0 服从均值为已知值$\hat{x}_0^+$、协方差为 P_0^+的高斯分布，且与 w_k 和 v_k相互独立。

在标准卡尔曼滤波中，假设过程噪声和测量噪声均为零均值高斯白噪声，且其名义协方差已知：

$$w_k \sim \mathcal{N}(0, Q_k), \tag{3}$$
$$v_k \sim \mathcal{N}(0, R_k), \tag{4}$$

其中 Q_k 和 R_k 是正（半）定矩阵。实际应用中，噪声统计特性往往未知、随时间变化，或呈重尾分布 [19]。

将先验残差和测量残差定义为

$$e_k^{\mathrm{pre}} = \hat{x}_k^- - x_k, \qquad e_k^{\mathrm{mea}} = y_k - Cx_k, \tag{5}$$

其中，$\hat{x}_k^- = A\hat{x}_{k-1}^+$ 为预测状态。则增广残差向量为

$$e_k = \begin{bmatrix} e_k^{\mathrm{pre}} \\ e_k^{\mathrm{mea}} \end{bmatrix} = \begin{bmatrix} \hat{x}_k^- - x_k \\ y_k - Cx_k \end{bmatrix} = z_k - H_k x_k, \tag{6}$$

with

$$z_k = \begin{bmatrix} \hat{x}_k^- \\ y_k \end{bmatrix}, \qquad H_k = \begin{bmatrix} I \\ C \end{bmatrix}. \tag{7}$$

增广残差的协方差为

$$\Sigma_k = \mathrm{Cov}(e_k) = \begin{bmatrix} P_k^- & 0 \\ 0 & R_k \end{bmatrix}, \tag{8}$$

其中 P_k^- 是 $\hat{x}_k^- - x_k$ 的协方差， R_k 是测量噪声协方差。由于 Σ_k 是块对角矩阵，其逆平方根为

$$\Sigma_k^{-1/2} = \begin{bmatrix} (P_k^-)^{-1/2} & 0 \\ 0 & R_k^{-1/2} \end{bmatrix}, \tag{9}$$

其中 $(P_k^-)^{-1/2}$ 和 $R^{-1/2}$ 分别是相应逆协方差矩阵的平方根（例如乔列斯基因子）。白化残差定义为

$$\tilde{e}_k = \Sigma_k^{-1/2} e_k = \Sigma_k^{-1/2}(z_k - H_k x_k) = \tilde{z}_k - \tilde{H}_k x_k. \tag{10}$$

引理 1 （半二次对偶性）。 设 $\rho(e)$是鲁棒损失函数，且 $\phi(u) = \rho(\sqrt{u})$在 $u \in [0, \infty)$上是凹的、连续可微且单调递增的。则存在一个对偶势函数 $\psi(w)$，使得：

$$\rho(e) = \inf_{w>0} \left[w\frac{e^2}{2} + \psi(w) \right].$$

下确界在 $w = \rho'(e)/e$ 处取得。

详细证明见 [35], [36]。半二次（HQ）对偶性已广泛用于将非凸损失函数重构为适用于迭代重加权最小二乘（IRLS）算法的形式。

命题1 （半二次分解中的线性缩放）。 设 $\rho(e;\ \theta) = \theta\,\Phi\left(\frac{e}{\sqrt{\theta}}\right)$ 是一个鲁棒损失函数，其中 Φ 满足半二次对偶条件（引理 1）。则 $\rho(e;\theta)$ 的半二次分解形式为

$$\rho(e;\theta) = \min_{w>0} \left[w\frac{e^2}{2} + \theta\,\psi(w) \right],$$

其中， $\psi(w)$ 是由 Φ 的分解得到的对偶势函数，即 $\Phi(u) = \min_{w>0}\left[wu^2/2 + \psi(w)\right]$。因此， $\psi(w)$ 不依赖于 θ，且尺度参数 θ 在线性形式下出现在对偶势函数中。

备注 1。 该命题的重要性在于，它是定理1的结构性前提。 θ 在线性对偶势中的线性出现，正是保证伽马先验在变分贝叶斯更新中具有共轭性的条件。

证明见补充材料中的附录 A。上述命题适用于一类鲁棒损失函数，包括柯西损失、Charbonnier

（伪 Huber）损失、公平损失和多核相关熵损失，详情亦见附录 A。

在这些方法中，多核相关熵损失通过核宽度明确体现了鲁棒性–准确性权衡，因此将作为本文的研究重点。

A. 最大相关熵准则与多核相关熵损失

设 $e=[e_1,\dots,e_N]^\top$ 为一个随机向量。多核相关熵损失（MCKL）按元素定义为：

$$\rho(e_i)=\sigma_i^2\left(1-\exp\left(-\frac{e_i^2}{2\sigma_i^2}\right)\right), \tag{11}$$

其中，$\sigma_i^2>0$ 是第 i个分量的核宽度。

最大相关熵准则（MCC）通过最大化真实变量与估计变量之间的相关熵来实现。对于加性误差估计任务，最小化MCKL等价于最大化MCC。除一个缩放因子外，ρ（e_i）与MCC损失一致：

$$J_{\mathrm{MCC}}(e)=\sum_{i=1}^{N}\left(1-\exp\left(-\frac{e_i^2}{2\sigma_i^2}\right)\right)=\frac{1}{\sigma_i^2}\rho(e_i). \tag{12}$$

MCKL 的关键性质，包括其有界性、重新下降的影响函数和渐近行为，已在补充材料中概述。与 Huber（非重新下降）和 Student-t （多项式衰减）不同，MKCL 具有指数重新下降影响函数和有界性，且仅用一个参数 σ^2 控制鲁棒性。

B. 多核相关熵损失的半二次分解

回顾 MCKL 在式（11）中的定义，可得：

推论 1. MKCL 满足半二次对偶定理的条件，因此可作如下精确分解：

$$\rho(e_i)=\min_{w_i\in(0,1]}\left[w_i\frac{e_i^2}{2}+\sigma_i^2(w_i\log w_i-w_i+1)\right]. \tag{13}$$

唯一的极小值点为：

$$w_i=\exp\left(-\frac{e_i^2}{2\sigma_i^2}\right), \tag{14}$$

该推论的证明见补充材料的附录B。

备注2（概率解释）。引理 1 中建立的 HQ 对偶性具有直接的概率解释。对于辅助变量 w_i，它纯粹是为优化而引入的，且没有可用的先验信息，因此在（$0,1$] ）上采用均匀先验 $p(w_i)\propto$ const 是一种自然的非信息性选择。将 HQ 诱导似然定义为

$$p(e_i\mid w_i,\sigma_i^2)\propto\exp\left(-w_i\frac{e_i^2}{2}-\sigma_i^2\psi(w_i)\right), \tag{15}$$

那么，w_i 的后验分布满足

$$p(w_i\mid e_i,\sigma_i^2)\propto\exp\left(-w_i\frac{e_i^2}{2}-\sigma_i^2\psi(w_i)\right). \tag{16}$$

最大化该后验分布可得

$$\hat{w}_i^{\mathrm{MAP}}=\arg\min_{w_i}\left[w_i\frac{e_i^2}{2}+\sigma_i^2\psi(w_i)\right], \tag{17}$$

这恰好就是引理 1 中的 HQ最小化。因此，HQ 最小化器是辅助变量 w_i 的最大后验估计。对于 MCKL，推论 1 给出：

$$\hat{w}_i^{\mathrm{MAP}}=w_i^*=\exp\left(-\frac{e_i^2}{2\sigma_i^2}\right). \tag{18}$$

C. HQ 诱导的不可归一化广义似然

将 MKCL 应用于式 (10) 中定义的白化残差 $\tilde{e}_k$ 。利用 HQ 分解 (13)，组合损失函数可写为

$$\min_{x_k,\{w_{i,k}\}}\sum_{i=1}^{n+m}\left[w_{i,k}\frac{[\tilde{e}_k]_i^2}{2}+\sigma_i^2(w_{i,k}\log w_{i,k}-w_{i,k}+1)\right], \tag{19}$$

其中，前 n 个分量对应于预测误差，其余 m 个分量对应于测量新息。令 $W=\mathrm{diag}(w_1,\dots,w_{n+m})$ 为权重矩阵。于是，式 (19) 可简洁地表示为

$$\min_{x_k,\{w_{i,k}\}}\left[\frac{1}{2}\tilde{e}_k^\top W\tilde{e}_k+\sum_{i=1}^{n+m}\sigma_i^2(w_{i,k}\log w_{i,k}-w_{i,k}+1)\right]. \tag{20}$$

项 $-\frac{1}{2}\tilde{e}_k^\top W\tilde{e}_k$ 关于 $\tilde{e}_k$ 是二次的，并且由于 $\tilde{e}_k$ 关于 x_k 是仿射的，其指数定义了一个广义似然：

$$p(\tilde{e}_k\mid\{w_{i,k}\},\{\sigma_{i,k}^2\})=p(y_k,\hat{x}_k^-\mid x_k,\{w_{i,k}\},\{\sigma_{i,k}^2\})\propto\exp\left(-\frac{1}{2}\tilde{e}_k^\top W\tilde{e}_k-\sum_{i=1}^{n+m}\sigma_{i,k}^2(w_{i,k}\log w_{i,k}-w_{i,k}+1)\right). \tag{21}$$

在固定 $\sigma_{i、k}^2$ 的情况下，最小化(20)等价于对 x_k 和 $\{w_{i,k}\}$ 最大化该似然，这等价于一个条件最大后验问题。

注 3. (21)中的广义似然隐含地纳入了 x_k 的先验分布。根据白化残差的定义可得

$$p(y_k,\hat{x}_k^-\mid x_k)=p(y_k,x_k\mid\hat{x}_k^-)\propto\exp\left(-\frac{1}{2}\tilde{e}_k^\top W\tilde{e}_k\right).$$

当 $W=I$ 时，展开 $\tilde{e}_k$ 可得

$$-\frac{1}{2}\tilde{e}_k^\top\tilde{e}_k=-\frac{1}{2}(x_k-\hat{x}_k^-)^T(P_k^-)^{-1}(x_k-\hat{x}_k^-)-\frac{1}{2}(y_k-Cx_k)^TR^{-1}(y_k-Cx_k).$$

因此，对 x_k 最大化 $p(y_k、\hat{}x_k^-\mid x_k)$ 等价于最大化后验 $p(x_k\mid y_k、\hat{}x_k^-)=p(x_k\mid y_{1:k})$。

备注4。式（11）中由 MKCL 诱导的似然不可归一化：其尾部衰减不足以产生有限积分。由于 HQ 分解在其极小值点处保持原始损失函数不变，式（21）中的 HQ 分解形式也具有相同性质。这不会影响后续的变分贝叶斯推断，因为该推断作用于联合分布，而不要求似然经过归一化；所得的变分后验分布仍然是适当的分布。

D. 估计问题的贝叶斯公式化

基于 HQ 分解的概率解释（备注2）以及上述推导的广义似然（21），现在我们在贝叶斯框架下对 x_k 和核宽度 $\{\sigma_{i,k}^2\}$ 进行联合估计。状态先验为 $p(x_k) = \mathcal{N}(x_k; \hat{x}_k^-, P_k^-)$。对于核宽度，我们指定伽马先验：

$$p(\sigma_{i,k}^2) \sim \Gamma\left(\frac{\epsilon_1^{i,k}}{2} + 1, \frac{\epsilon_2^{i,k}}{2}\right), \tag{22}$$

其中 $\epsilon_{,1}^{i,k}, \epsilon_{,2}^{i,k} > 0$ 是共轭伽马先验的超参数。该先验的对数形式为：

$$\log p(\sigma_{i,k}^2) = \frac{\epsilon_1^{i,k}}{2} \log \sigma_{i,k}^2 - \frac{\epsilon_2^{i,k}}{2} \sigma_{i,k}^2 + \text{const.} \tag{23}$$

这些超参数与下文所建立的 HQ 诱导似然共轭。

定理1（半二次似然下的伽马共轭性）**.** 设 $\rho(e;\theta)$ 是满足命题1的鲁棒损失函数，因此其半二次分解形式为

$$\rho(e;\theta) = \min_{w>0} \left[w\frac{e^2}{2} + \theta\,\psi(w) \right],$$

其中 $\psi(w)$与 θ无关。考虑 HQ 诱导似然

$$p(e \mid \theta) \propto \exp\left(-\frac{we^2}{2} - \theta\,\psi(w) \right)$$

在给定 $\theta > 0$ 且 w 固定的情况下。对于 θ 上的伽马先验 $\Gamma(\epsilon_1/2 + 1,\ \epsilon_2/2)$，后验分布 $p(\theta \mid e)$ 同样服从伽马分布，其更新后的超参数为

$$\epsilon_1^* = \epsilon_1, \qquad \epsilon_2^* = \epsilon_2 + 2\psi(w).$$

证明见补充材料的附录C。

因此，y_k、x_k、$\{\sigma_{i,k}^2\}_{i=1}^{n+m}$ 和 $\{w_{i,k}\}_{i=1}^{n+m}$ 的完整联合分布为：

$$p(y_k, x_k, \{\sigma_{i,k}^2\}, \{w_{i,k}\} \mid \hat{x}_k^-) \propto \exp\left(-\frac{1}{2}\tilde{e}_k^\top W \tilde{e}_k - \sum_{i=1}^{n+m} \sigma_{i,k}^2 \psi(w_{i,k}) \right) \prod_{i=1}^{n+m} \Gamma\left(\sigma_{i,k}^2; \epsilon_1^{i,k}, \epsilon_2^{i,k}\right) \tag{24}$$

E. 目标

我们的目标是构建一种能够自适应地抵抗异常测量误差的递归滤波算法。为此，我们需要通过各自后验分布的均值获得状态 x_k 和核宽度 $\{\sigma_{i,k}^2\}_{i=1}^{n+m}$ 的点估计：

$$\hat{x}_k \triangleq \mathbb{E}_{p(x_k \mid y_{1:k})}[x_k], \qquad \hat{\sigma}_{i,k}^2 \triangleq \mathbb{E}_{p(\sigma_k^2 \mid y_{1:k})}[\sigma_{i,k}^2]. \tag{25}$$

直接从联合分布 (24) 进行计算是不可行的。因此，我们采用均值场变分近似

$$q(x_k, \{\sigma_{i,k}^2\}, \{w_{i,k}\}) = q(x_k) \prod_{i=1}^{n+m} q(\sigma_{i,k}^2) q(w_{i,k}), \tag{26}$$

并最小化其与真实后验分布之间的 Kullback-Leibler散度，这等价于最大化证据下界。随后，根据所得变分分布计算后验均值 $\hat{x}_k$ 和 $\hat{\sigma}_{i,\,k}^2$，并通过相应的预测方程进行传播，从而得到一种递归滤波算法。

本文重点研究测量通道的核宽度（$i = n+1$、……、$n+m$）的自适应调整；过程噪声核宽度（$i = 1$、……、n）固定为足够大的值，使得相应的权重 w_i effectively unity；在这种情况下，过程噪声更新可简化为标准卡尔曼滤波器的更新。将其扩展到完全自适应的过程噪声是直接的。

III. 主要结果

我们首先从联合分布 (24) 推导出一种精确的变分贝叶斯推断算法。在均值场假设下，坐标上升可为 $q(x_k)$、$q(w_{i,\,k})$ 和 $q(\sigma_{i,\,k}^2)$ 导出闭式 CAVI 更新，如算法 1 所总结。该精确算法要求在每个时间步迭代更新，直至证据下界收敛。

随后，我们提出一种适用于在线实现的高效近似方法。将 $\sigma_{i,\,k}^2$ 固定为其当前均值，并执行一次 CAVI 更新，可将 $q(x_k)$ 和 $q(w_{i,\,k})$ 的更新转化为一个通过迭代重加权最小二乘求解的条件最大后验问题，从而得到 AMCKF 算法。

A. 通过 *CAVI* 进行变分贝叶斯推断

现在，我们基于联合分布 (24) 和均值场假设 (26) 推导精确的变分贝叶斯推断算法。证据下界（ELBO）定义为

$$\mathcal{L}(q) = \mathbb{E}_q \left[\log \frac{p(y_k, x_k, \{\sigma_{i,k}^2\}, \{w_{i,k}\} \mid \hat{x}_k^-)}{q(x_k, \{\sigma_{i,k}^2\}, \{w_{i,k}\})} \right]. \tag{27}$$

后续推导所依据的一个关键结果是：MAP估计与采用退化（Delta）变分分布的变分推断等价。

引理2（退化变分推断与 MAP 估计的等价性）。设 $p(y,\ \theta,\ \zeta)$ 是一个联合分布，其中观测值为 y，目标变量为 θ，其他变量为ζ，并令

$$\mathcal{L}(q) = \mathbb{E}_q \left[\log \frac{p(y, \theta, \zeta)}{q(\theta, \zeta)} \right]$$

表示证据下界（ELBO）。在均值场假设 $q(\theta,\ \zeta) = q(\theta)\,q(\zeta)$ 下，若将 $q(\theta)$ 限定为狄拉克分布族，则关于 $q(\theta)$ 最大化 ELBO 等价于计算条件最大后验估计：

$$\max_{q(\theta) = \delta(\theta - \hat{\theta})} \mathcal{L}(q) \iff \hat{\theta} = \arg\max_{\theta} \mathbb{E}_{q(\zeta)}[\log p(y, \theta, \zeta)].$$

推论 2（HQ 最小化的变分实现）。设 w 是通过鲁棒损失 $\rho(e;\sigma^2)$ 的 HQ 分解引入的辅助变量，其均匀先验为$p(w) \propto$const。在平均场假设下，将 w 的变分分布限制为 Delta 分布族

$$q(w) = \delta(w - \tilde{w})$$

根据引理2，这意味着，就 $\tilde{w}$ 而言最大化证据下界等价于 MAP估计。结合 HQ 分解的概率解释（备注 2），该最大后验估计与 HQ 最小化器一致：

$$\tilde{w} = \arg\min_{w} \left[w\frac{e^2}{2} + \sigma^2\psi(w) \right].$$

因此，对 w 进行变分优化，同时完成了引理 1 中 HQ 对偶性所要求的最小化。

现在我们明确辅助变量 $\{w_{i,k}\}_{i=1}^{n+m}$ 的变分形式。根据推论 2，对于每个 $w_{i,\,k}$，最优变分分布是退化的：

$$q(w_{i,k}) = \delta(w_{i,k} - \tilde{w}_{i,k}). \tag{28}$$

因此，HQ 分解自然地嵌入到变分贝叶斯框架中。其余因子的坐标上升更新如下推导。

a) 更新 $q(w_{i,\,k})$：给定 $q(w_{i,\,k}) = \delta(w_{i,\,k} - \tilde{w}_{i,\,k})$，根据引理 2，对 $\tilde{w}_{i,\,k}$ 最大化证据下界可得

$$\tilde{w}_{i,k} = \arg\min_{w_{i,k}} \left(w_{i,k}\frac{\mathbb{E}_{q(x_k)}[[\tilde{e}_k(x_k)]_i^2]}{2} + \mathbb{E}_{q(\sigma_k^2)}[\sigma_{i,k}^2]\psi(w_{i,k}) \right)$$

其中 $\psi(w_{i,\,k}) = w_{i,\,k}\log w_{i,\,k} - w_{i,\,k} + 1$。求解可得

$$\tilde{w}_{i,k} = \exp\left(-\frac{\mathbb{E}_{q(x_k)}[[e_k(x_k)]_i^2]}{2\mathbb{E}_{q(\sigma_k^2)}[\sigma_{i,k}^2]} \right), \tag{29}$$

b) 更新 $q(x_k)$：固定 $q(\sigma_k^2)$ 和 $q(w_k)$，收集证据下界中所有依赖于 x_k 的项，可得

$$\log q(x_k) = -\frac{1}{2}\tilde{e}_k^\top \tilde{W}_k \tilde{e}_k + \text{const}, \tag{30}$$

其中 $\tilde{W}_k = \text{diag}(\tilde{w}_{1,k}, \ldots, \tilde{w}_{n+m,k})$，且 $\tilde{w}_{i,k} = \mathbb{E}_{q(w_k)}[w_{i,k}]$。该二次型表明 $q(x_k)$服从高斯分布：

$$q(x_k) = \mathcal{N}(x_k; \mu_k, P_k^+)$$

with

$$P_k^+ = \left(\tilde{H}_k^\top \tilde{W}_k \tilde{H}_k \right)^{-1} \tag{31}$$

$$\mu_k = P_k^+ \tilde{H}_k^\top \tilde{W}_k \tilde{z}_k. \tag{32}$$

现在已确定 $q(x_k)$，因此可以显式计算式 (29) 中出现的期望：

$$\mathbb{E}_{q(x_k)}[[e_k(x_k)]_i^2] = [e_k(\mu_k)]_i^2 + [H_k P_k^+ H_k^\top]_{ii}. \tag{33}$$

注 5（固定过程噪声核时的权重结构）。将过程噪声核宽度固定为足够大的值后，相应权重满足 $\tilde{w}_i = 1$，其中$i = 1, \ldots, n$。因此，权重矩阵变为块对角矩阵：

$$\bar{W}_k = \begin{bmatrix} I_n & 0 \\ 0 & \bar{W}_{\text{meas,k}} \end{bmatrix}, \bar{W}_{\text{meas,k}} = \text{diag}(\bar{w}_{n+1,k}, \ldots, \bar{w}_{n+m,k}).$$

将其代入 μ_k 和 P_k^+ 的表达式中，其中

$$\tilde{z}_k = \begin{bmatrix} (P_k^-)^{-1/2}\hat{x}_k^- \\ R_k^{-1/2} y_k \end{bmatrix}, \qquad \tilde{H}_k = \begin{bmatrix} (P_k^-)^{-1/2} \\ R_k^{-1/2} C \end{bmatrix}$$

得到

$$P_k^+ = \left((P_k^-)^{-1} + C^\top R_k^{-1/2} \bar{W}_{\text{meas,k}} R_k^{-1/2} C \right)^{-1}, \tag{34}$$

$$\mu_k = P_k^+ \left((P_k^-)^{-1}\hat{x}_k^- + C^\top R_k^{-1/2} \bar{W}_{\text{meas,k}} R_k^{-1/2} y_k \right). \tag{35}$$

因此，过程噪声通道保留了标准卡尔曼滤波器结构，而测量通道则采用自适应鲁棒加权。

c) 更新 $q(\sigma_{i,k}^2)$：以下推论直接由定理1得出，并确立了本文所用特定损失函数的共轭性。

推论3. 对于多核相关熵损失，命题1适用于 $\psi(w_{i,k}) = w_{i,k}\log w_{i,k} - w_{i,k} + 1$。因此，根据定理1，在伽马先验下，$\sigma_{i,\,k}^2$ 的后验分布也服从伽马分布，其超参数更新为

$$\epsilon_1^{i,k,+} = \epsilon_1^{i,k}, \tag{36}$$

$$\epsilon_2^{i,k,+} = \epsilon_2^{i,k} + 2(w_{i,k}\log w_{i,k} - w_{i,k} + 1). \tag{37}$$

固定 $q(x_k)$ 和 $q(w_{i,k})$ 后，根据推论3，最优的 $q(\sigma_i^2)$ 为

$$q(\sigma_{i,k}^2) = \Gamma\left(\sigma_{i,k}^2; \frac{\epsilon_1^{i,k,+}}{2} + 1, \frac{\epsilon_2^{i,k,+}}{2} \right), \tag{38}$$

with

$$\epsilon_2^{i,k,+} = \epsilon_2^{i,k} + 2(\tilde{w}_{i,k}\log\tilde{w}_{i,k} - \tilde{w}_{i,k} + 1). \tag{39}$$

其后验均值为

$$\mathbb{E}_{q(\sigma_k^2)}[\sigma_{i,k}^2] = \frac{\epsilon_1^{i,k,+} + 2}{\epsilon_2^{i,k,+}}. \tag{40}$$

备注6（活动集）。在实际应用中，某些测量通道可能已知是可靠的，因此不需要自适应鲁棒性。为此，我们引入一个活动集 $\mathcal{A} \subseteq \mathcal{I}$，其中 $\mathcal{I}$ 表示所有测量通道组成的集合。仅对 $i \in \mathcal{A}$ 的核宽度 $\sigma_{i,k}^2$ 进行自适应估计；对于 $i \notin \mathcal{A}$，σ_i^2 固定为标称值。

上述更新构成一个CAVI方案：从变分参数的初始猜测开始，$q(w_i)$、$q(x_k)$ 和 $q(\sigma_i^2)$ 按坐标方式迭代更新，直至证据下界收敛。

B. 算法流程

现在，我们给出基于第III-A节推导的 CAVI 更新的完整递归滤波算法。在每个时间步 k，给定先验分布 $\hat{x}_k^-, P_k^-$ 和超参数 $\epsilon_{\cdot,\,1}^{i\,\,k}$，$\epsilon_{\cdot,\,2}^{i\,\,k}$，执行 CAVI 过程以近似联合后验分布。

a) 初始化：CAVI 迭代通过将辅助权重设为以下值来初始化

$$\tilde{w}_{i,k}^{(0)} = 1, \quad i = 1, \ldots, n+m. \tag{41}$$

然后，通过依次应用第三节 A 中推导的坐标上升变分推断更新，得到其余变分参数：

其中，$\tilde{w}_i^{(0)}=1$，更新 $q(x_k)$ 得到 $\mu_k^{(0)}=\hat{x}_k^-$ 和 $P_k^{(0)}=P_k^-$；随后，更新 $q(\sigma_i^2)$ 得到 $\epsilon_2^{i,k,(0)}=\epsilon_2^{i,k}$ 和 $\sigma_i^{2,(0)}=\epsilon_1^{i,k}+\cdots$，因此得到先验分布均值 $\hat{\cdot}(\cdot)/\epsilon^{i,\cdot,2}{}^{(0)}$。

*b) CAVI*迭代：在每次迭代 t中，按顺序执行以下更新：

1) 更新 $q(w_i)$：使用 $\mu_k^{(t)}$、$P_{(t)k}$和$\hat{\sigma}^{2,(t)}_{\cdot,i}$，通过式(29)计算 $\tilde{w}_i^{(t+1)}$；2) 更新 $q(x_k)$：通过式(34)并使用当前的 $\tilde{w}_i^{(t+1)}$ 计算 $\mu_k^{(t+1)}$、$P_k^{(t+1)}$；3) 更新 $q(\sigma_i^2)$：通过式(39)计算 $\epsilon^{i,k,(t+1)}_{\cdot,2}$，并使用 $\tilde{w}_i^{(t+1)}$ 通过式(40)计算$\hat{\sigma}^{2,(t+1)}_{\cdot,i}$。

迭代持续进行，直至证据下界收敛。

c) 后验提取与传播：在第 T_k次迭代收敛后，提取后验均值如下

$$\hat{x}_k^+=\mu_k^{(T_k)},P_k^+=P_k^{(T_k)},\epsilon_2^{i,k,+}=\epsilon_2^{i,k,(T_k)},\hat{\sigma}_{i,k}^{2,+}=\hat{\sigma}_i^{2,(T_k)}. \tag{42}$$

状态估计通过标准卡尔曼预测进行传播：

$$\hat{x}_{k+1}^-=A\hat{x}_k^+,\qquad P_{k+1}^-=AP_k^+A^\top+Q_k. \tag{43}$$

超参数传播如下：

$$\epsilon_1^{i,k+1}=\epsilon_1^{i,k,+},\qquad \epsilon_2^{i,k+1}=\rho\,\epsilon_2^{i,k,+}, \tag{44}$$

其中 $\rho\in(0,1]$ 是遗忘因子。

d) 遗忘机制的动机：在正常运行时，$\tilde{w}_{i,k}\approx 1$ 且 $\hat{\sigma}^{2,+}_{i,k}$ 保持在标称值。当出现离群值时，$\hat{\sigma}^{2,+}_{i,k}$ 会减小，从而降低 $\tilde{w}_{i,k}$，并降低受污染测量的权重。然而，如果没有遗忘机制，$\hat{\sigma}^{2,+}_{i,k}$ 将单调不增，并可能在离群值持续较长时间时趋于崩溃，导致过度自信，并将真实的状态变化误判为离群值。遗忘因子 ρ 使 $\hat{\sigma}^{2,+}_{i,k}$ 能够在正常运行期间恢复，从而确保长期适应能力。

注释 7（核宽度界限）。为防止核宽度变得过大（从而丧失鲁棒性）或趋于零（从而导致过度自信），对超参数施加明确界限：

$$\epsilon_2^i\in[\epsilon_{2,\min}^i,\epsilon_{2,\max}^i]. \tag{45}$$

由于 $\hat{\sigma}_i^2=(\epsilon_1^i+2)/\epsilon_2^i$，这直接限定了核宽度。实践中，$\epsilon^i_{2,\max}$ 通常设置为 $\epsilon^i_{2,\min}$的 10 到 100 倍，具体取决于所需的鲁棒性水平。

C. 超参数演化的 *EWMA* 滤波

仅根据瞬时新息进行核宽度更新，可能会导致 $\hat{\sigma}^2_{i,k}$ 在瞬态离群值的影响下发生剧烈波动。为利用历史信息，可以对超参数更新应用指数移动平均（EWMA）滤波器：

$$\epsilon_2^{i,k,+}=\beta\,\epsilon_2^{i,k-1,+}+(1-\beta)\,\epsilon_2^{i,k,(T_k)} \tag{46}$$

其中，$\beta\in[0,1)$ 是滤波因子，而在 $\beta=0$ 情况下可恢复瞬时更新。

算法 1 用于联合状态与核宽度估计的坐标上升变分推断

输入： 系统矩阵 A，C，噪声协方差 R_k，Q_k，初始状态 $\hat{x}_{0|0}$ 和 $P_{0|0}$，初始超参数 $\epsilon^{i,0}_1,\epsilon^{i,0}_2$，遗忘因子 ρ，滤波因子 β
输出： 状态估计值 $\hat{x}_k^+$ 和协方差 P_k^+，核宽度估计值 $\hat{\sigma}^2_{i,k}$
对于每个时间步 $k=1,2,\ldots\ldots$**执行**
  预测：
  通过（43）获取 $\hat{x}_k^-$ 和 P_k^-
  对于 $i\in\mathcal{A}$ **执行**
    获取 $\tilde{w}_{i,k}$，$\epsilon^{i,k}_1,\epsilon^{i,k}_2$ 通过式 (44)
  结束循环
  CAVI 初始化： $\tilde{w}^{(0)}_{i,k}=1$, $t=0$
  重复
    $t\leftarrow t+1$
    更新 $q(w_i)$：计算 $\tilde{w}_{i,k}^{(t)}$ 通过（29）使用 $\mu_k^{(t-1)}$，$P_k^{(t-1)}$，$\hat{\sigma}^{2,(t-1)}_{i,\cdot}$
    更新 $q(x_k)$：计算 $\mu_k^{(t)}$，$P_k^{(t)}$ 通过（34）并且使用 $\tilde{w}^{(t)}_{i,k}$ (35)
    对于 $i\in\mathcal{A}$ **执行**
      更新 $q(\sigma_i^2)$：计算 $\epsilon^{i,k,(t)}_{\cdot,2}$，通过（39）和 $\hat{\sigma}^{2,(t)}_{i,\cdot}$ 使用 $\tilde{w}^{(t)}_{i,k}$，通过（40）
    结束循环
  直到 ELBO收敛
  后验提取：
  $\hat{x}_k^+=\mu_k^{(t)}$，$P_k^+=P_k^{(t)}$
  对于 $i\in\mathcal{A}$，**执行**
    获取 $\epsilon^{i,k,+}_2$，通过 (46)
    应用界限： $\epsilon^{i,k,+}_2\in[\epsilon^i_{2,\min},\epsilon^i_{2,\max}]$
    计算核宽度：获取 $\hat{\sigma}^{2,+}_{i,k}$ 通过 (40)
  结束循环
结束循环

这种 EWMA 滤波在响应性与稳定性之间提供了可调节的权衡。它是可选的，不会影响核心的变分贝叶斯推导；在需要更平滑的适应时，可以纳入历史信息。

结合 CAVI 更新、递归传播和可选的 EWMA 滤波，完整且严谨的变分贝叶斯推断算法总结于算法 1 中。

D. *AMCKF* 算法

前文各节推导出的严谨 CAVI 算法提供了一种有理论依据的变分贝叶斯解法，但需要在每个时间步进行迭代坐标更新，直至证据下界收敛。这种计算负担促使我们为在线应用寻求高效的近似方法。

我们引入两项简化。首先，在优化过程中将 $q(x_k)$ 限制为狄拉克分布；在收敛至众数后，再将后验分布恢复为高斯形式，其协方差由 Joseph 更新（52）给出。这为真实后验分布提供了类拉普拉斯近似，仅当残差较小时才是精确的（$e_i\ll\sigma_i^2$）；对于较大的残差，真实的 MCKL 海森矩阵还需要额外的

因子（$1-e_i^2/\sigma_i^2$），在这种情况下，Joseph 更新得到的协方差估计比真实的拉普拉斯近似更保守（更大）。相比之下，众所周知，在重尾噪声下，精确 CAVI 会产生过度自信（低估的）协方差估计 x_k；AMCKF 仅在收敛后恢复协方差，从而缓解了这种过度自信。其次，每个时间步我们只执行一次 M步：通过迭代重加权最小二乘迭代求取 x_k 和 w_i 的点估计，直至收敛（固定 σ_i^2），随后对（39）中的 σ_i^2 执行一次闭式更新计算，因为 x_k 与 w_i 之间的主要耦合已由迭代重加权最小二乘过程捕获，而后续的 σ_i^2 更新通常只是小幅修正。

在这些近似下，权重计算变为

$$\tilde{w}_{i,k}=\exp\left(-\frac{[\tilde{e}_k(\hat{x}_k)]_i^2}{2\hat{\sigma}_{i,k}^2}\right), \tag{47}$$

并且，对 x_k 和 w_i的联合变分贝叶斯估计（根据引理2）简化为条件最大后验问题，其中 σ_i^2固定为其当前均值（如式（21）所述）。由此得到的优化问题正是下文所述迭代重加权最小二乘过程所求解的CMAP公式。 σ_i^2 的更新仍与式（39）相同。因此，AMCKF在实现高效在线实现的同时，保留了严谨的坐标上升变分推断框架的核心鲁棒性和自适应性。

*E. CMAP*的求解

为求解式（21）中的条件最大后验问题，我们直接对状态 x_k进行最小化，以求得MKCL。利用式（10）中定义的白化残差 $\tilde{e}_k$，目标函数为

$$\mathcal{J}(x_k)=\sum_{i=1}^{n+m}\rho([\tilde{e}_k]_i)=\sum_{i=1}^{n+m}\sigma_i^2\left(1-\exp\left(-\frac{[\tilde{e}_k]_i^2}{2\sigma_i^2}\right)\right).$$

对 x_k 求导，并利用 $\rho'(e)=e\exp(-e^2/(2\sigma_i^2))$，可得

$$\frac{\partial\mathcal{J}}{\partial x_k}=-\tilde{H}_k^\top W(\tilde{e}_k)\tilde{e}_k,$$

其中 $W(\tilde{e}_k)=\mathrm{diag}\left(\exp\left(-[\tilde{e}_k]_i^2/(2\sigma_i^2)\right)\right)$。令其为零可得到加权最小二乘解：

$$x_k=\left(\tilde{H}_k^\top W(\tilde{e}_k)\tilde{H}_k\right)^{-1}\tilde{H}_k^\top W(\tilde{e}_k)\tilde{z}_k. \tag{48}$$

命题2（多核相关熵损失的IRLS单调收敛性）。设 $e(x)=z-Hx$ 为线性残差向量。考虑目标函数

$$\mathcal{J}(x)=\sum_{i=1}^{N}\sigma_i^2\left(1-\exp\left(-\frac{[e(x)]_i^2}{2\sigma_i^2}\right)\right).$$

定义不动点迭代

$$x^{(j+1)}=\left(H^\top W^{(j)}H\right)^{-1}H^\top W^{(j)}z,$$

其中 $W^{(j)}=\mathrm{diag}\left(\exp\left(-[e(x^{(j)})]_i^2/(2\sigma_i^2)\right)\right)$。那么，对于任意起始点 $x^{(0)}$，序列 $\{\mathcal{J}(x^{(j)})\}_{j\geq 0}$ 单调不增，并收敛到一个驻点 $\mathcal{J}(x)$。证明见补充材料的附录D。

注释8。命题2保证IRLS迭代单调降低目标函数并收敛到一个驻点。实践中，该迭代以先验状态估计$x_{k,0}^+=\hat{x}_k^-=A\hat{x}_{k-1}$进行初始化；在纳入当前《测量》之前，该估计在最小二乘意义下是最优的。

为高效实现IRLS迭代，我们将加权最小二乘解（48）重新表述为递归卡尔曼形式。按照 [27], 中的推导，记作 $x_{k,0}^+=\hat{x}_k^-$。第 t 次迭代的状态更新写为

$$x_{k,t+1}^+=\hat{x}_k^-+K_{k,t}\left(y_k-C\hat{x}_k^-\right), \tag{49}$$

$$K_{k,t}=P_k^-C^T\left(CP_k^-C^T+\tilde{R}_{k,t}\right)^{-1}, \tag{50}$$

$$\tilde{R}_{k,t}=R_k^{1/2}\bar{W}(\tilde{e}_{k,t}^{\mathrm{mea}})^{-1}R_k^{1/2}, \tag{51}$$

其中，$\bar{W}(\tilde{e}_{k,t}^{\mathrm{mea}})=\mathrm{diag}\left(\exp\left(-[\tilde{e}_{k,t}^{\mathrm{mea}}]_i^2/(2\sigma_i^2)\right)\right)$是测量通道的自适应权重矩阵，而 $\tilde{e}_{k,t}^{\mathrm{mea}}=R_k^{-1/2}(y_k-Cx_{k,t}^+)$是白化创新。

这种递归形式在计算上更具优势：直接形式 (48) 需要对完整的 $n\times n$矩阵 $\tilde{H}_k^\top W(\tilde{e}_k)\tilde{H}_k$求逆，而 (50) 只需对 $m\times m$矩阵 $CP_k^-C^T+\tilde{R}_{k,t}$求逆，其中 $m\ll n$在典型应用中成立。更重要的是，与严格的 CAVI 算法不同，后者需要在每次迭代时重新计算后验协方差 P_k^+，以评估权重更新中的方差项 $[H_kP_k^+H_k^\top]_{ii}$；AMCKF 的权重更新仅依赖残差点估计。因此，协方差更新只需在 IRLS 收敛后执行一次，而非在每次迭代时执行，从而显著降低总体计算成本。

IRLS迭代收敛后，使用 Joseph形式更新后验误差协方差：

$$P_k^+=(I-K_{k,t}C)P_k^-(I-K_{k,t}C)^T+K_{k,t}\tilde{R}_{k,t}K_{k,t}^T \tag{52}$$

整体过程总结于算法2中。

*F. AMCKF*算法流程

AMCKF 算法将第三节-E中的基于IRLS的CMAP状态估计与第三节-D中的闭式核宽度更新相结合。在每个时间步，首先通过算法2中的CMAP过程获得状态估计$\hat{x}_k^+$ 和协方差 P_k^+。利用CMAP迭代中收敛的权重w_i，随后通过式（39）更新核宽度超参数，并将核宽度估计计算为其后验均值（40）。完整过程总结于算法3中。

IV. 仿真

本节通过数值仿真评估所提出的 AMCKF 的性能。我们考虑一个标准的目标跟踪问题，仿真时长为 20 s，采样间隔为 $\Delta t=0.01$ s（共 2000 个时间步）。所有结果均取 100 次独立蒙特卡洛运行的平均值，以确保统计显著性。

算法2 迭代CMAP状态估计器

输入： $\hat{x}_k^-$, P_k^-, y_k, C, R_k, $\{\sigma_i^2\}$, m_{iter}, ϵ
输出： $\hat{x}_k^+$, P_k^+
初始化： $x_{k,0}^+ = \hat{x}_k^-$
对于 $t = 0, 1, 2, \ldots, m_{\text{iter}}$ **执行**
  计算白化测量新息 $e_{k,t-1}^{\text{whe}} = R_k^{-1/2}(y_k - Cx_{k,t}^+)$
  按照（14）更新权重 $w_i = \exp(-[\tilde{e}_{k,t}^{mea}]_i^2/(2\sigma_i^2))$
  构造 $W = \text{diag}(w_i)$
  计算 $R_{k,t}$ 如式（51）所示
  计算卡尔曼增益 $K_{k,t}$ 如式（50）所示
  更新状态 $x_{k,t+1}^+$ 如式（49）所示
  如果 $\|x_{k,t+1}^+ - x_{k,t}^+\|/\|x_{k,t+1}^+\| < \epsilon$ **然后**
    跳出
  结束条件判断
结束循环
计算后验协方差 P_k^+，如式（52）所示
令 $\hat{x}_k^+ = x_{k,t}^+$

算法3 AMCKF

1：输入：系统矩阵 A、C，噪声协方差 R，初始状态 $\hat{x}_{0|0}$ 和 $P_{0|0}$，初始核宽度 $\sigma_{i,0|0}^2$，初始超参数 ϵ^i，01、ϵ^i,02，遗忘因子 ρ，滤波因子 β，活动集 $\mathcal{A}$，收敛阈值 ϵ，最大迭代次数 m_{iter}。2：输出：状态估计值 $\hat{x}_k^+$ 和协方差 P_k^+，核宽度估计值 $\hat{\sigma}^2$、$_{i\,k}^{+}$。3：对于每个时间步 $k = 1$、2、.....，执行 4：状态与超参数预测：5：通过式（43）获得 $\hat{x}_k^-$ 和 P_k^- 6：对于 $i \in \mathcal{A}$，执行 7：通过式（44）获得 ϵ^i、$_{1}^{k}$、ϵ^i、$_{2}^{k}$ 8：结束循环 9：CMAP迭代（状态更新）：10：执行算法2中描述的CMAP算法，获得 $\hat{x}_k^+$ 和 P_k^+ 11：核宽度更新（M步）：12：对于 $i \in \mathcal{A}$，执行 13：计算瞬时更新：14：按照式（36）获得 ϵ^i、$_{k+}$ 15：按照式（39）获得 ϵ^i、$_{k+}$ 16：EWMA 滤波：

$$\epsilon_2^{i\,k+} = \beta\,\epsilon_2^{i\,k-1+} + (1-\beta)\,\epsilon_2^{i\,k+}$$

17：应用限界处理：.....、

$$\epsilon_2^{i\,k+} \in [\epsilon_{2,\min}, \epsilon_{2,\max}]$$

18：计算核宽度：通过式（40）获得 $\hat{\sigma}^2$、$_{i\,k}^{+}$ 19：结束循环 20：递推：设置 $k \leftarrow k+1$ 21：结束循环

A. 仿真设置

我们考虑一个二维目标跟踪问题。状态向量定义为 $x_k = [p_x, v_x, p_y, v_y]^T \in \mathbb{R}^4$，其中 p_x 和 p_y 表示位置，v_x 和 v_y 分别表示沿 x 和 y 方向的速度。系统矩阵为：

$$A = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{2}\Delta t^2 & 0 \\ \Delta t & 0 \\ 0 & \frac{1}{2}\Delta t^2 \\ 0 & \Delta t \end{bmatrix}, \tag{53}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{54}$$

$\Delta t = 0.01$ s 为采样间隔。过程噪声协方差设为 $Q_w = I_2$（即两个相互独立的加速度噪声分量的单位协方差），名义测量噪声协方差设为 $R_v = I_2$。初始状态为 $x_0 = [0, 0, 0, 0]^T$，初始协方差为 $P_0 = 0.1 \cdot I_4$。

仿真时长为 20 s，对应 $T = 2000$ 个时间步。所有结果均取 100 次独立蒙特卡洛运行的平均值。

实现以下滤波器进行比较：*KF*、*VBKF* [11], *STKF* ($\nu = 3$)、*MKCKF* ($\sigma^2 = 10$)，以及所提出的 *AMCKF*。

对于所提出的 AMCKF，每个观测通道的初始核宽度均设置为约 200，遗忘因子为 $\rho = 0.98$，阈值为 0.99，EWMA 滤波因子为 $\beta = 0.1$。物理边界设置为 $\epsilon_{2,\min} = 1$ 和 $\epsilon_{2,\max} = 200$。当状态估计的相对变化低于 10^{-6}，或达到最多 10 次迭代时，CMAP迭代停止。

均方根误差和改进率的定义见补充材料。

B. 针对离群值的鲁棒性

为评估鲁棒性，我们考虑三种场景：纯高斯噪声（场景 *1*）、含有 *10%* 均匀离群值的高斯噪声（ $[-20, 20]$ 场景 *2*），以及具有 t分布的 *Student* 噪声（ $\nu = 2$ 场景 *3*）。

表 I 总结了均方根误差结果。表 II 展示了三种场景下各方法相对于标准卡尔曼滤波的改进率。

在高斯噪声（场景 1）下，所提出的 AMCKF 达到了标准KF的最优性能，而 VBKF、STKF 和 MKCKF 由于采用固定的鲁棒性假设，准确率均有所下降。这表明 AMCKF 能够自动适应噪声环境。

在稀疏离群值（场景 2）和重尾噪声（场景 3）下，AMCKF 在所有方法中取得了最低的均方根误差，证实了其卓越的鲁棒性。自适应核宽度机制能够有效抑制离群值，同时保持跟踪能力。

在确立 AMCKF 相较于现有方法的优势后，接下来我们将验证其是否保留了理论前身——严谨的坐标上升变分推断算法——的估计精度。

C. 与坐标上升变分推断算法的比较

为验证所提出的 AMCKF 在降低计算成本的同时保持坐标上升变分推断算法的估计精度，我们比较了二者在耦合

表 I 三种场景下的性能比较（均方根误差、均值 ±标准差）。

方法	场景 1 (高斯分布)	场景 2 (稀疏离群值)	场景 3 (Student-t, $\nu = 2$)
KF	0.1192 ± 0.0114	0.3185 ± 0.0280	0.3065 ± 0.1386
VBKF	0.1951 ± 0.0372	0.3394 ± 0.1079	0.2860 ± 0.0881
STKF	0.2601 ± 0.0595	0.2362 ± 0.0594	0.2637 ± 0.0727
MKCKF	0.2422 ± 0.0523	0.2211 ± 0.0536	0.2521 ± 0.0677
AMCKF	0.1192 ± 0.0114	0.2075 ± 0.0202	0.1776 ± 0.0174

三个场景下各方法相对于标准KF的改进率（正值表示有所改进）。

方法	提升（%）		
	场景 1	场景 2	场景 3
KF	—	—	—
VBKF	-63.61%	-6.59%	$+6.70\%$
STKF	-118.18%	$+25.82\%$	$+13.98\%$
MKCKF	-103.10%	$+30.57\%$	$+17.77\%$
AMCKF	-0.00%	$+34.83\%$	$+42.07\%$

质量-弹簧-阻尼器系统上的性能。该系统由通过弹簧和阻尼器以链式配置连接的 $N = 10$ 个质量块组成，两端均固定在墙上。每个质量块在二维平面内运动；第 i 个质量块的状态为 $[p_{ix}, v_{ix}, p_{iy}, v_{iy}]^\top$，因此总状态维数为 40。测量向量包含所有质量块在两个方向上的位置，因此测量维数为 20。系统矩阵根据耦合振子网络的离散时间动力学构造，参数为：质量 $m = 1$、弹簧常数 $k_s = 2$、阻尼系数 $c = 2$ 和采样间隔 $\Delta t = 0.005$ s，对应的阻尼比为 $\zeta = c/(2\sqrt{k_s m}) \approx 0.707$（欠阻尼）。测量噪声服从 Student-$t$ 分布，其 $\nu = 3$。

为考察可扩展性，我们将质量数从 2 调整至 25，对应的状态维数从 8 变为 100，同时保持其他所有参数不变。图 1 展示了两种算法在不同状态维数下的均方根误差和CPU时间。AMCKF 与坐标上升变分推断（CAVI）的均方根误差曲线在所有测试维数下几乎无法区分，表明无论问题规模如何，该近似都不会造成显著的精度损失。同时，两种算法的计算时间均呈二次增长，这与矩阵求逆的$O(n^3)$成本一致。AMCKF 在所有测试维数下都能持续降低计算成本，并且随着状态维数增加，其优势差距进一步扩大。

D. 自适应核宽度跟踪

为展示所提出 AM-CKF 的自适应行为，我们设计了一项仿真，其中测量噪声
环境在连续四个 5 秒阶段内随时间变化（每个阶段 500 步，每步 $\Delta t = 0.01$ 秒）：高斯分布、稀疏离群值（$[-100, 100]$中有 10% 为均匀分布）、重尾（Student'st、$\nu = 1.8$）以及恢复为高斯分布。整个仿真过程中均记录核宽度 σ^2；由于 $R_v = I_2$，x 和 y 的估计核宽度几乎相同，因此仅展示 x方向的结果。

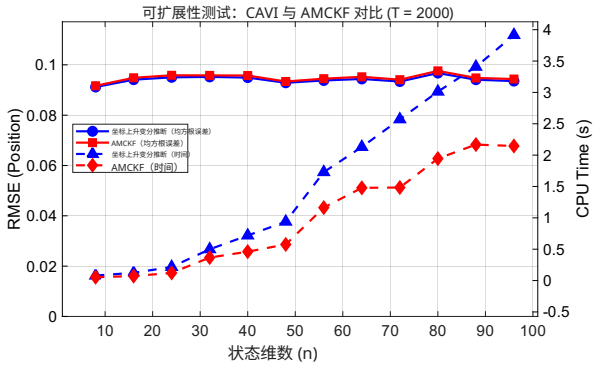

图 1. CAVI和AMCKF的均方根误差（左轴）与CPU时间（右轴）随状态维数的变化。

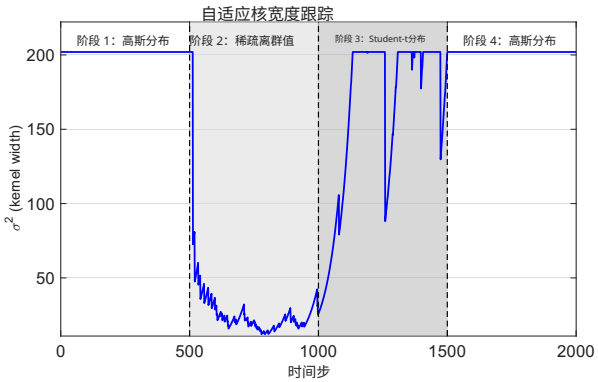

图 2. x方向测量通道的估计核宽度$\hat{\sigma}^2$ 的变化。阴影背景表示不同噪声阶段：白色为 = 高斯分布，浅灰色为 = 稀疏离群值，深灰色为 = 重尾（Student-t, $\nu = 1.8$）。

图 2 展示了 $\hat{\sigma}^2$ 的演化过程。在高斯阶段，$\hat{\sigma}^2$保持在其上限，从而恢复标准卡尔曼性能。在存在离群值和重尾噪声时，$\hat{\sigma}^2$ 降低以抑制受污染的测量值，并通过 EWMA 滤波恢复。总体均方根误差为 0.1481（位置）和 0.1010（速度），表明仍能保持准确跟踪。这些结果表明，AMCKF 能够根据变化的噪声条件自适应调整其核宽度，而无需手动调参。

E. 综合性能比较

为进一步评估所提出的 AMCKF，我们设计了一个包含 $T = 2000$个时间步（20 秒，$\Delta t = 0.01$ 秒）的混合挑战场景，由五个连续阶段组成：高斯分布（阶段 1）、稀疏离群值（阶段 2）、*Student-t* 重尾噪声，且包含 $\nu = 2$（阶段 3）、突发离群值（阶段 4）以及混合离群值（阶段 5）。所有测试均通过 100 次独立的蒙特卡洛运行进行。将所提出的 AMCKF 与标准KF、VBKF [11]，STKF（固定$\nu = 3$）以及 MKCKF（固定 $\sigma^2 = 10$）进行比较。

图 3 和表 III 总结了整体均方根误差结果。所提出的 AMCKF 的均方根误差中位数和均值最低，且在所有方法中四分位距最小，表明其精度和一致性更优。图 4 展示了分阶段性能。在阶段 1（高斯分布）中，所有方法的表现相近。从阶段 2 开始，AMCKF

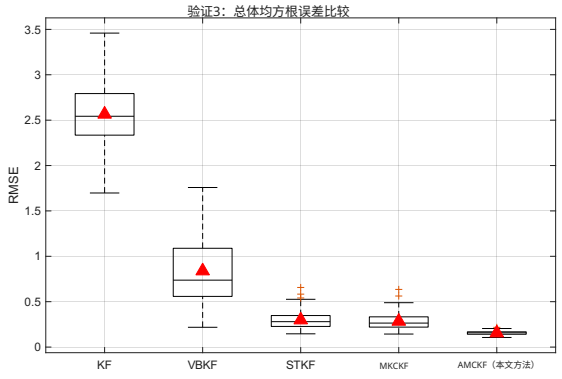

图 3. 100 次蒙特卡洛运行的总体均方根误差箱线图。红色三角形表示均值。所提出的 AMCKF 达到了最低的中位数和均值均方根误差，展现了其在复杂噪声条件下卓越的鲁棒性和精度。

表 III 总体均方根误差比较（100 次蒙特卡洛运行的均值 ± 标准差）。

方法	均方根误差（均值 ± 标准差）	相较于卡尔曼滤波的改进
KF	2.5951 ± 0.3264	—
VBKF	0.9396 ± 0.4209	+63.79%
STKF	0.2845 ± 0.0773	+89.04%
MKCKF	0.2699 ± 0.713	+89.60%
AMCKF	$\mathbf{0.1531 \pm 0.0215}$	**+94.10%**

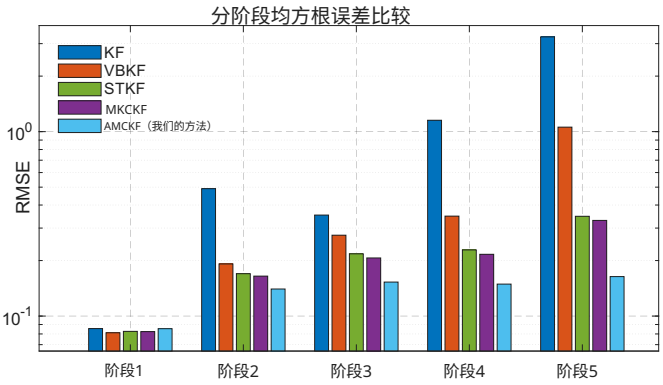

图4. 五个噪声阶段的逐阶段均方根误差比较。

始终保持较低的均方根误差，且波动极小，而其他方法的性能则显著下降。这种稳定性证实了自适应核宽度机制的有效性。总体而言，AMCKF 能够有效处理各种离群值类型，同时在高斯分布噪声下保持最佳性能——这是固定参数鲁棒滤波器（STKF、MKCKF）或传统自适应滤波器（VBKF）无法实现的。

F. 基于*IMU*的姿态估计：真实数据集上的真实步态实验

我们在 [38] 的 Circle-Fast-Both-Disturbances 步态数据集（其中第 6 个索引）上验证了所提出的 AMCKF。遵循 [38] 中确立的采用解耦磁力计更新的右不变SO(3) IEKF建模框架，我们用所提出的自适应变分贝叶斯机制取代原始 MKCIEKF 中固定的核宽度。

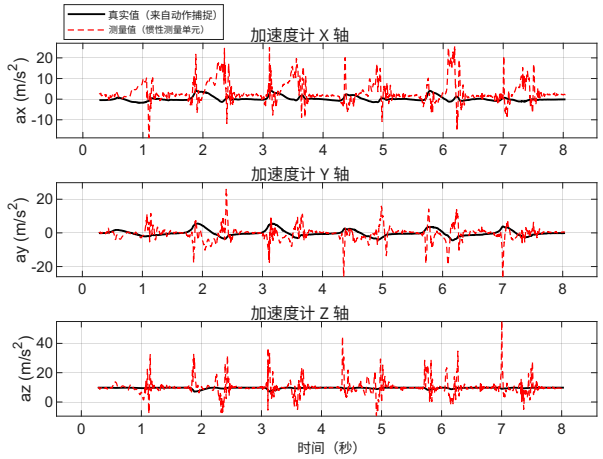

图5. 动作捕捉得出的测量加速度计信号（红色）与真实重力（黑色）的比较。两者的偏差表明，快速行走时存在运动诱发加速度。

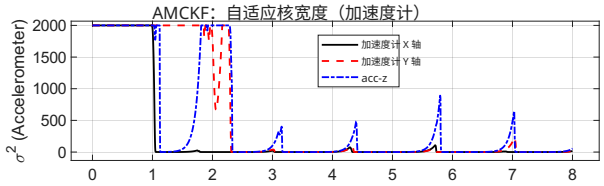

图6. 加速度计通道的自适应核宽度 σ^2 的变化。在运动诱发加速度较强时，核宽度会下降，随后恢复，体现了自动信任度调整。

表IV Circle-Fast-Both-Disturbances场景下的均方根误差比较（度）。

方法	Yaw	Roll	俯仰角
Xsens（商用）	0.7111	0.7130	0.6072
MKCKF	0.6355	0.7269	0.6049
AMCKF（本文方法）	**0.5984**	**0.7093**	**0.6071**

该数据集记录了以 2.5 m 半径沿圆形路径快速行走（≈1.5 m/s）的情况，铁磁板和足部着地振动共同干扰了IMU测量值。Xsens MTi-670 IMU（400 Hz）安装在足部，光学动作捕捉则提供姿态真值。我们将 AMCKF 与商用Xsens滤波器以及核宽度为 $\sigma^2 = 2000$ 的固定核宽度 MKCKF 进行比较。

图5对比了测量加速度与动作捕捉得到的重力；快速行走时的水平偏差表明存在运动诱发的干扰。对此，加速度计通道的自适应核宽度 σ^2 会减小，以降低对受干扰测量值的依赖，并在之后恢复，如图6所示。

表IV报告了欧拉角均方根误差。采用 $\sigma^2 = 2000$ 的固定核MKCKF性能与商用Xsens滤波器相当。所提出的AMCKF取得了最低的偏航误差（0.5984°），优于Xsens（0.7111°）和MKCKF（0.6355°），同时横滚精度和俯仰精度也与二者相当。这一改进表明，自适应核宽度机制具有更好的鲁棒性

，因为它能够根据观测噪声统计量自动调整核宽度，而固定核宽度则无法做到这一点。

V. 结论

本文提出了一种用于鲁棒状态估计的自适应多核相关熵卡尔曼滤波器（AMCKF）。通过半二次对偶性，我们建立了一个严格的变分贝叶斯框架，用于联合状态与核宽度估计，并推导出了闭式 CAVI 更新；同时表明，由 HQ 诱导的辅助变量具有退化的变分后验。随后，将 AMCKF 引入为精确 CAVI 的一种高效单遍近似，其中状态变量和辅助变量通过迭代重加权最小二乘（IRLS）进行迭代更新，随后执行闭式核宽度更新。遗忘机制确保了长期适应性。蒙特卡洛仿真表明，在各种离群值和重尾噪声场景下，AMCKF 始终优于 KF、VBKF、STKF 和固定核 MKCKF；同时，在高斯噪声下保持最优性能，并达到与严格 CAVI 相当的精度。

未来的工作可探讨过程噪声自适应、非线性扩展、自调节遗忘以及核宽度动态的理论分析。

参考文献

[1] R. E. Kalman，“线性滤波与预测问题的新方法”，基础工程杂志，第82卷，45，1960 1，第35期，页码–。[2] R. E. Kalman和R. S. Bucy，“线性滤波与预测理论的新成果”，基础工程杂志，第83卷，第1期，页码95–108，1961。[3] R. K. Mehra，“方差辨识与自适应卡尔曼滤波”，*IEEE*自动控制汇刊，第15卷，第2期，页码175–184，1970。[4] Y. C. Ho和R. C. K. Lee，“随机估计与控制问题的贝叶斯方法”，*IEEE*自动控制汇刊，第9卷，第4期，页码333–339，1964。[5] R. J. Fitzgerald，“卡尔曼滤波器的发散”，*IEEE*自动控制汇刊，第16卷，第6期，页码736–747，1971。[6] M. H. Gusrial、N. A. Othman、H. Ahmad和M. H. A. Hassan，“移动机器人学中采用线性化与协方差初始化的同步定位与建图卡尔曼滤波器变体综述”，机电一体化、电力与车辆技术杂志，第16卷，第1期，页码69–83，2025。[7] R. Mur-Artal、J. M. M. Montiel和J. D. Tard´os，“ORB-SLAM：一种通用且精确的单目同步定位与建图系统”，*IEEE Transactions onRobotics*，第31卷，第5期，页码1147–1163，2015。[8] S. Li、D. Shi、Y. Lou、W. Zou和L. Shi，“用于扰动估计的广义多核最大相关熵卡尔曼滤波器”，*IEEE*自动控制汇刊，第69卷，第6期，页码3732–3747，2024。[9] W. Xia和Y. Zhang，“针对虚假数据注入攻击的弹性分布式估计：一种基于相关熵的方法”，信息科学，第635卷，页码236–256，2023。[10] R. J. Fitzgerald，“卡尔曼滤波器的发散”，*IEEE*自动控制汇刊，第16卷，第6期，页码736–747，1971。[11] S. S¨arkk¨a和A. R. Nummenmaa，“基于变分贝叶斯近似的递归噪声自适应卡尔曼滤波”，*IEEE*自动控制汇刊，第54卷，第3期，页码596–600，2009。[12] Y. Huang、Y. Zhang、Z. Wu、N. Li和J. A. Chambers，“一种针对过程噪声与测量噪声协方差矩阵不准确情况的新型自适应卡尔曼滤波器”，*IEEE*自动控制汇刊，第63卷，第2期，页码594–601，2018。[13] H. Lan、J. Hu、Z. Wang和Q. Cheng，“未知过程噪声协方差下的变分非线性卡尔曼滤波”，*IEEE*航空航天与电子系统汇刊，第59卷，第6期，页码9177–9190，2023。[14] X. Lv、P. Duan、Z. Duan、G. Chen和L. Shi，“随机事件触发的变分贝叶斯滤波”，*IEEE*自动控制汇刊，第68卷，第7期，页码4321–4328，2022。[15] V. ˇSm´ıdl 和 A. Quinn，“变分贝叶斯滤波”，IEEE信号处理汇刊，第 56 卷，第 10 期，第 5020–5032 页，2008 年。[16] Y. Meng、S. Gao、Y. Zhong、G. Hu 和 A. Subic，“用于惯性导航系统/全球导航卫星系统组合直接滤波的基于协方差匹配的自适应无迹卡尔曼滤波器”，《宇航学报》，第 120 卷，第 171–181 页，2016 年。[17] A. Mohamed 和 K. P. Schwarz，“用于惯性导航系统/全球定位系统的自适应卡尔曼滤波”，《大地测量学杂志》，第 73 卷，第 4 期，第 193–203 页，1999 年。[18] X. R. Li 和 Y. Bar-Shalom，“一种用于噪声识别的递归多模型方法”，IEEE航空航天与电子系统汇刊，第 30 卷，第 3 期，第 671–684 页，1994 年。[19] J. He、C. Sun、B. Zhang 和 P. Wang，“兼具自适应性和鲁棒性的基于变分贝叶斯的最大相关熵容积卡尔曼滤波器”，IEEE传感器杂志，第 21 卷，第 2 期，第 1982–1992 页，2021 年。[20] S. C. Chan 和 Y. X. Zou，“一种用于脉冲噪声下鲁棒自适应滤波的递归最小M估计算法：快速收敛与低复杂度”，IEEE信号处理汇刊，第 63 卷，第 21 期，第 5686–5699 页，2015 年。[21] M. Rojas、J. M. Ramirez 和 J. R. Iglesias，“一种适用于信号处理应用的鲁棒基于学生t分布的卡尔曼滤波器”，IEEE信号处理汇刊，第 67 卷，第 15 期，第 4010–4023 页，2019 年。[22] Y. Huang、Y. Zhang、N. Li 和 J. A. Chambers，“一种适用于具有重尾过程噪声和测量噪声系统的鲁棒基于学生t分布的卡尔曼滤波器”，IEEE航空航天与电子系统汇刊，第 53 卷，第 5 期，第 2485–2499 页，2017 年。[23] G. Agamennoni、J. I. Nieto 和 E. M. Nebot，“一种离群值鲁棒卡尔曼滤波器”，载于 2011 年 IEEE机器人与自动化国际会议（ICRA），第 1551–1558 页，2011 年。[24] B. Chen、X. Wang、N. Zheng 和 J. C. Pr´ıncipe，“最大相关熵卡尔曼滤波器”，《自动化》，第 76 卷，第 70–77 页，2017 年。[25] W. Liu、P. P. Pokharel 和 J. C. Pr´ıncipe，“相关熵：在非高斯信号处理中的性质与应用”，IEEE信号处理汇刊，第 55 卷，第 11 期，第 5286–5298 页，2007 年。[26] J. C. Pr´ıncipe，《信息论学习：Rényi 熵与核方法视角》。信息科学与统计，纽约，美国：施普林格，2010 年。[27] S. Li、D. Shi、W. Zou 和 L. Shi，“多核最大相关熵卡尔曼滤波器”，IEEE控制系统快报，第 6 卷，1495，2021 1490pp. – 。[28] M. V. Kulikova，“用于动态状态估计的最大相关熵卡尔曼滤波器”，《自动化》，第 77 卷，第 217–224 页，2017 年。[29] V. Azimi 和 S. Onori，“基于稳定核尺寸自适应的最大相关熵卡尔曼滤波器”，《系统与控制快报》，第 204 卷，第 106198 页，2025 年。[30] S. Fakoorian、R. Izanloo、A. Shamshirgaran 和 D. Simon，“具有自适应核尺寸的最大相关熵准则卡尔曼滤波器”，载于 2019 年 IEEE全国航空航天与电子会议（NAECON），第 581–584 页，2019 年。[31] W. Chen、Z. Li、Z. Chen、Y. Sun 和 Y. Liu，“用于 GPS/INS组合导航的具有自适应核宽度的基于多重相似性度量的最大相关熵准则卡尔曼滤波器”，《测量》，第 222 卷，第 113666 页，2023 年。[32] Z. Li、S. Ouyang、Y. Cheng、H. Wang 和 K. Chen，“具有自适应核带宽的变分贝叶斯最大相关熵容积卡尔曼滤波器”，载于 2023 年 IEEE信号处理、通信与计算国际会议（ICSPCC），第 1–6 页，2023 年。[33] B. Liu、X. Zhang、S. Jia、S. Zou 和 D. Tian，“变分贝叶斯自适应最大相关熵广义高阶容积卡尔曼滤波器”，《电路、系统与信号处理》，第 42 卷，第 7073–7098 页，2023 年。[34] L. Zhang、Y. Xiao、G.-W. Fan、Q.-Z. Liu 和 X.-F. Wu，“用于卫星姿态估计的改进的自适应核尺寸最大相关熵容积卡尔曼滤波器”，IEEE仪器与测量汇刊，第 1–1 页，2026 年。[35] D. Geman 和 C. Yang，“采用半二次正则化的非线性图像恢复”，IEEE图像处理汇刊，第 4 卷，第 7 期，第 932–946 页，1992 年。[36] R. T. Rockafellar，《凸分析》。美国新泽西州普林斯顿：普林斯顿大学出版社，1970 年。[37] M. Nikolova 和 M. K. Ng，“信号与图像恢复的半二次最小化方法分析”，《SIAM科学计算杂志》，第 27 卷，第 3 期，第 937–966 页，2005 年。[38] Y. Liang、S. Li、L. Kuang、C. Zhong、D. Shi 和 L. Shi，“用于惯性测量单元的解耦多核相关熵不变扩展卡尔曼滤波器”，IEEE仪器与测量汇刊，第 1–1 页，2026 年。

VI. 自适应多核相关熵卡尔曼滤波器：变分贝叶斯方法的补充材料

多核相关熵损失的性质

主稿件中的式（11）所定义的多核相关熵损失（MCKL）具有以下性质：

1) 齐次性：

$$\rho\ e\ = \sigma^2\Phi\ e/\sigma \qquad \Phi\ u\ = 1 - (\ )(\ ),$$

$e^{-u^2/2}$ σ^2

其中 ( ) ，因此 仅作为尺度参数。**2) 有界性和对称性：** $\rho(e) \in [0, \sigma^2)$，$\rho(-e) = \rho(e)$。**3) 渐近极限：**

$$\lim_{\sigma^2\to\infty} \rho(e) = \frac{e^2}{2}, \qquad \lim_{\sigma^2\to 0} \rho(e) = \sigma^2 \cdot \mathbf{1}_{e\neq 0}.$$

4) ρ（$\sqrt{u}$）**的凹性：** $\bar{}\phi$（u）$= \rho$（$\bar{\sqrt{}}u$）在 $u \geq 0$ 上是严格凹的。**5) 再下降导数：** ρ'（e）$= e\exp(-e^2/$（$2\sigma^2$），其中 $\lim_{|e|\to\infty} \rho'$（$e$）$= 0$。

这些性质共同确保 MCKL 是一种有界、对称且具有红降特性的损失函数。有界性可防止较大的残差主导估计，而红降特性则确保随着残差幅值增大，极端离群值的影响逐渐消失。这些特性使 MCKL 特别适用于存在重尾噪声和离群值时的稳健估计。

性能指标

均方根误差（RMSE）分别针对每个方向上的位置和速度进行计算：

$$\text{RMSE}_{\text{pos}} = \sqrt{\frac{1}{2N_{\text{MC}}T} \sum_{m=1}^{N_{\text{MC}}} \sum_{k=1}^{T} \left[(p_{x}^{(m)}{}_{,\ k} - \hat{}p_{x\ k}^{(m)}{}_{,\ })^2 + (\ ^{m)}_{y,k})^2 \right]},$$

$p_{y}^{(m)}{}_{,\ k} - \hat{}p^{(}{}_{m=1\ k=1}$

(55)

速度同理：

$$\text{RMSE}_{\text{vel}} = \sqrt{\frac{1}{2N_{\text{MC}}T} \sum_{m=1}^{N_{\text{MC}}} \sum_{k=1}^{T} \left[(v_{x,k}^{(m)} - \hat{v}_{x,k}^{(m)})^2 + (v_{y,k}^{(m)} - \hat{v}_{y,k}^{(m)})^2 \right]},$$

(56)

其中，$N_{\text{MC}} = 100$ 表示蒙特卡洛运行次数，T =2000表示时间步数，$p_{x,\ k}$ 和 $p_{y,\ k}$ 表示真实位置，而 $v_{x,\ k}$ 和 $v_{y,\ k}$ 表示真实速度。

给定方法相对于标准卡尔曼滤波器的改进率定义如下：

$$\text{Improvement} = \frac{\text{RMSE}_{\text{KF}} - \text{RMSE}_{\text{Method}}}{\text{RMSE}_{\text{KF}}} \times 100\%, \quad (57)$$

其中，RMSE_{KF} 表示标准卡尔曼滤波器的均方根误差，$\text{RMSE}_{\text{Method}}$ 表示待评估方法的均方根误差。

附录

A. 命题1的证明及满足该命题的鲁棒损失函数

由于 Φ满足 HQ条件，根据引理 1，存在函数 $\psi(w)$，使得

$$\Phi(u) = \inf_{w>0} \left[w\frac{u^2}{2} + \psi(w) \right],$$

且下确界在某个 $w^* > 0$处取得。现在将$u = e/\sqrt{\theta}$ 代入 ρ 的表达式：

$$\rho(e;\theta) = \theta\,\Phi\left(\frac{e}{\sqrt{\theta}}\right) = \theta \inf_{w>0} \left[w\frac{(e/\sqrt{\theta})^2}{2} + \psi(w) \right].$$

由于 $\theta > 0$ 是常数因子，可以将其乘入下确界：

$$\rho(e;\theta) = \inf_{w>0} \left[w\frac{e^2}{2} + \theta\,\psi(w) \right].$$

因此，我们得到了所需的形式。注意，使 Φ 的表达式最小化的相同 w 在变换后的表达式（缩放后）中也能实现最小化，因此范围 $w > 0$ 保持不变。所以，inf 等于 min。证明完毕。

表 V 是一些满足命题1的示例性损失函数。我们省略尺度参数，重点关注这些损失函数的核。

表 V 满足命题1的代表性鲁棒损失函数及其对偶势函数。

Loss	$\Phi(u)$	$w^*(u)$	$\psi(w)$
Welsch	$1 - e^{-u^2/2}$	$e^{-u^2/2}$	$w\log w - w + 1$
柯西	$\log(1+u^2)$	$\frac{2}{1+u^2}$	$\log\frac{2}{w} - 1 + \frac{w}{2}$
Charbonnier	$\sqrt{1+u^2} - 1$	$\frac{1}{\sqrt{1+u^2}}$	$-\frac{1}{2}\log\ w + \frac{w}{2} - \frac{1}{2}$
Fair	$\lvert u\rvert - \log(1+\lvert u\rvert)$	$\frac{1}{1+\lvert u\rvert}$	对数 $w + \frac{1}{2w} - \frac{w}{2}$

B. 推论 1 的证明

根据 MCKL 的性质，所有条件均已满足，从而保证 HQ 分解的存在性。将半二次对偶定理（引理 1）应用于$\rho(e)$，可得：

$$\rho(e) = \inf_{w>0} \left[w\frac{e^2}{2} + \psi(w) \right]. \quad (58)$$

The 最优 w 满足 $w = \rho'(e)/e$。直接计算得出 $\rho'(e) = ee^{-e^2/}$ $(2\sigma^2)$，因此：

$$w^* = \exp\left(-\frac{e^2}{2\sigma^2}\right). \quad (59)$$

要确定 ψ，请使用极小值点 w^*处的等式：

$$\rho(e) = w^*\frac{e^2}{2} + \psi(w^*) \quad (60)$$

$$= -\sigma^2 w^* \ln w^* + \psi(w^*) = \sigma^2(1 - w^*). \quad (61)$$

求解 $\psi(w^*)$ 得到 $\psi(w^*) = \sigma^2(1 - w^* + w^* \ln w^*)$。推广到任意 w 可得：

$$\psi(w) = \sigma^2(w\log w - w + 1). \quad (62)$$

由此建立了多核相关熵损失的半二次分解：

$$\rho(e) = \min_{w\in(0,1]} \left[w\frac{e^2}{2} + \sigma^2(w\log w - w + 1) \right]. \tag{63}$$

证明完毕。

C. 定理1的证明

对于满足命题1的损失函数，HQ 诱导似然的形式为

$$p(e \mid \theta, w) \propto \exp\left(-\frac{we^2}{2} - \theta\psi(w) \right).$$

结合伽马先验 $p(\theta) \propto \theta^{\epsilon_1/2}\exp(-\epsilon_2\theta/2)$，后验分布为

$$\begin{aligned} p(\theta \mid e, w) &\propto \theta^{\epsilon_1/2} \exp\left(-\frac{\epsilon_2}{2}\theta \right) \cdot \exp\left(-\theta\psi(w) \right) \\ &= \theta^{\epsilon_1/2} \exp\left(-\left(\frac{\epsilon_2}{2} + \psi(w) \right)\theta \right). \end{aligned}$$

这与形状为 $\alpha = \epsilon_1/2 + 1$、率参数为 $\beta = \epsilon_2/2 + \psi(w)$ 的伽马分布成正比。因此，更新后的超参数为

$$\epsilon_1^* = \epsilon_1, \qquad \epsilon_2^* = \epsilon_2 + 2\psi(w).$$

因此，伽马分布与 HQ 诱导似然共轭。

D. 命题2的证明

令 $e_i(x) = [z - Hx]_i$ 表示第 i 个残差元素。目标函数可改写为：

$$\mathcal{J}(x) = \sum_{i=1}^{N} \phi_i\left(e_i^2(x) \right), \tag{64}$$

其中，标量函数 $\phi_i : \mathbb{R}_{\geq 0} \to \mathbb{R}$ 定义为：

$$\phi_i(u) = \sigma_i^2 \left(1 - e^{-u/(2\sigma_i^2)} \right). \tag{65}$$

1) $\phi_i(u)$的性质：对于每个 i，$\phi_i(u)$ 在 $u \geq 0$ 上连续可微。其前两阶导数为：

$$\phi_i'(u) = \frac{1}{2} e^{-u/(2\sigma_i^2)}, \tag{66}$$

$$\phi_i''(u) = -\frac{1}{4\sigma_i^2} e^{-u/(2\sigma_i^2)} < 0, \quad \forall u \geq 0. \tag{67}$$

由于对于所有 $u \geq 0$，$\phi_i''(u) < 0$，因此 $\phi_i(u)$ 在$[0, \infty$上严格凹。

2）代理函数的构造：对于凹函数，一阶泰勒展开提供了一个全局上估计量。具体而言，对于任意凹函数 $\phi(u)$ 以及任意展开点 $u_0 \geq 0$：

$$\phi(u) \leq \phi(u_0) + \phi'(u_0)(u - u_0). \tag{68}$$

令 $u_i^{(j)} \triangleq e_i^2(x^{(j)})$ 表示第 j 次迭代时的残差平方。将凹性应用于每一项，我们构造代理函数 $Q(x, x^{(j)})$，作为 $\mathcal{J}(x)$ 的全局上界：

$$\mathcal{J}(x) \leq Q(x, x^{(j)}) \triangleq \sum_{i=1}^{N} \left[\phi_i(u_i^{(j)}) + \phi_i'(u_i^{(j)}) \left(e_i^2(x) - u_i^{(j)} \right) \right]. \tag{69}$$

3）代理函数的最小化：关于 x 最小化 $Q(x, x^{(j)})$ 等价于最小化一个加权最小二乘项，因为 $\phi_i(u_i^{(j)})$ – $\phi_i' u_i^{(j)} u_i^{(j)}$ x ( ) 中的各项相对于 均为常数。

定义权重：

$$\omega_i^{(j)} = \phi_i'(u_i^{(j)}) = \frac{1}{2} \exp\left(-\frac{[e_i^{(j)}]^2}{2\sigma_i^2} \right). \tag{70}$$

忽略常数因子 $1/2$（该因子可以吸收到权重矩阵中），我们得到权重矩阵：

$$W^{(j)} = \operatorname{diag}\left(\exp\left(-\frac{[e_i^{(j)}]^2}{2\sigma_i^2} \right) \right). \tag{71}$$

待最小化的代理函数为：

$$Q(x, x^{(j)}) = \frac{1}{2}(z - Hx)^\top W^{(j)}(z - Hx) + \text{const.} \tag{72}$$

这是关于 x 的二次型。令梯度为零，可得不动点迭代：

$$x^{(j+1)} = \left(H^\top W^{(j)} H \right)^{-1} H^\top W^{(j)} z. \tag{73}$$

这正是命题2中定义的迭代。

*4)*单调下降与收敛：根据极大化最小化（MM）原理，代理函数 $Q(x, x^{(j)})$ 满足：

$$Q(x, x^{(j)}) \geq \mathcal{J}(x), \quad \forall x, \qquad Q(x^{(j)}, x^{(j)}) = \mathcal{J}(x^{(j)}). \tag{74}$$

由于 $x^{(j+1)}$ 最小化了 $Q(x, x^{(j)})$，因此有：

$$Q(x^{(j+1)}, x^{(j)}) \leq Q(x^{(j)}, x^{(j)}) = \mathcal{J}(x^{(j)}). \tag{75}$$

结合这些不等式：

$$\mathcal{J}(x^{(j+1)}) \leq Q(x^{(j+1)}, x^{(j)}) \leq \mathcal{J}(x^{(j)}). \tag{76}$$

因此，序列 $\{\mathcal{J}(x^{(j)})\}_{j\geq 0}$ 是单调非增的。

*5)*收敛到驻点：由于 $\mathcal{J}(x)$ 的下界为零（多核相关熵损失是非负的），该单调序列收敛于有限极限 $\mathcal{J}^*$。此外，在收敛时，有 $x^{(j+1)} = x^{(j)} \triangleq x^*$，其满足驻点条件：

$$H^\top W(x^*)(z - Hx^*) = 0. \tag{77}$$

这正是梯度条件 $\nabla\mathcal{J}(x^*) = 0$。因此，极限点是 $\mathcal{J}(x)$ 的驻点。至此，证明完成。