鲁棒最大相关熵卡尔曼滤波器

Joydeb Saha [1], Shovan Bhaumik[2],{[1]joydeb 2121ee32, [2]shovan.bhaumik}@iitp.ac.in[1,2] 印度理工学院巴特那分校（IITP），印度比哈尔邦巴特那

arXiv:2302.02694v1 [math.OC] 6 Feb 2023

摘要

卡尔曼滤波器可对具有高斯噪声的线性系统进行最优估计。然而，当噪声具有非高斯特性时，其性能会迅速下降。针对非高斯噪声，研究人员提出了最大相关熵卡尔曼滤波器（MCKF），可获得更好的结果。但当系统模型偏离标称设定时，MCKF 的性能会下降。为此，我们提出了一种新的鲁棒滤波技术，该技术通过最大化一个代价函数来实现滤波；此代价函数由加权的过去误差和当前误差的指数项与高斯核函数共同定义。通过求解这一成本准则，我们推导出了先验和后验均值及协方差矩阵的传播方程。通过最大化误差矩阵的相关熵函数，我们在每个时间步选取核带宽值。此外，我们还推导了所提算法的收敛条件。文中给出了两个数值示例，以展示这一新滤波技术的实用性。

I. 引言

状态估计是一项非常重要的技术，广泛应用于各种工业问题和研究领域，例如目标跟踪、导航系统、通信系统、图像处理、系统辨识、数据融合、卫星状态估计等。对于噪声服从高斯分布的线性系统，卡尔曼滤波器能够对状态进行最优估计。但对于非高斯噪声，卡尔曼滤波器的性能会急剧下降。

为克服这一局限，人们提出了若干方法，例如基于最小误差熵的卡尔曼滤波器 [1],贝叶斯推断算法 [2], 最大相关熵卡尔曼滤波器（MCKF） [3] 等。相似性是表达或度量时间信号数量的一个关键概念。相关熵与两个随机变量在由核带宽定义的联合空间邻域内相似的概率直接相关。它是一个二元函数，可生成一个包含二阶及更高阶概率密度函数矩的标量。近年来，基于相关熵 [4] 的滤波被用于非高斯噪声环境下的状态估计，其目标是最大化相关熵；这类滤波器称为最大相关熵卡尔曼滤波器（MCKF） [3]。传统卡尔曼滤波器通过最小化均方误差进行估计，涉及二阶概率密度函数矩；而最大相关熵准则（MCC）除二阶矩外，还考虑所有更高阶的偶数阶矩。这正是基于MCC的滤波器在非高斯噪声环境下比传统卡尔曼滤波器效果更好的主要原因。

核函数在基于相关熵的滤波技术中发挥着重要作用。高斯核在最大相关熵卡尔曼滤波中得到了广泛应用。文献中还提供了少数其他核函数，例如拉普拉斯核 [5],高斯混合核 [6] 也可用。不过，高斯核平滑、对称，且两个高斯函数乘积的积分仍为高斯函数 [4]。由于这些特性，高斯核更受青睐。

在许多实际应用中，我们无法确定过程模型。在这种情况下，模型失配可能会降低 MCKF 的性能。我们需要一种鲁棒算法来应对这一情况。文献中已有考虑高斯噪声的风险敏感滤波器（RSF）。在 [7],[8]和 [9],中，对风险敏感估计问题的构建进行了详细研究。但是，对于

存在系统不确定性的非高斯噪声，文献中尚无相关研究。通过将风险敏感滤波器的概念与相关熵滤波器的思想相结合，我们构建了一个新的代价函数，该函数基于所有过去误差的加权和以及采用具有核带宽的高斯核函数加权的当前误差。通过最大化该函数，我们构建了一种新算法，适用于存在非高斯噪声的系统不确定性模型。

在基于相关熵的滤波器中，核带宽对滤波技术的性能至关重要。如何选择合适的带宽值是研究人员面临的一大挑战。已有少数文献讨论了自适应核带宽选择方法 [10]–[12]，但这些方法无法保证获得最优值。我们提出了一种采用高斯核的替代代价函数，用于数值选取每个时间步的带宽值。我们还推导了所提出滤波器的收敛性和稳定性准则。

II. Problem Formulation

让我们考虑一个具有以下过程方程和测量方程的线性系统：

$$\mathcal{X}_{k+1} = (F_k + \Delta F_k)\mathcal{X}_k + q_k, \tag{1}$$

$$\mathcal{Y}_k = H_k\mathcal{X}_k + r_k, \tag{2}$$

其中，$\mathcal{X}_k = \left[\mathrm{x1}、_k\,\mathrm{x2}、_k \ldots \mathrm{x}_{n,k}\right]^T$ 是系统的状态向量，$\mathcal{Y}_k = \left[\mathrm{y1}、_k\,\mathrm{y2}、_k \ldots \mathrm{y}_{m,k}\right]^T$ 是测量向量。F_k 和 H_k 分别是状态矩阵和测量矩阵。$\Delta\mathrm{F}_k$ 是过程模型中一个任意的确定性未知参数，用于描述系统的不确定性。我们假设 $\Delta\mathrm{F}_k$ 有界，且其范围足以保证受扰动系统保持稳定。如果系统参数已准确知晓，则 $\Delta\mathrm{F}_k = 0$。过程噪声 q_k 和测量噪声 r_k 均值为零，且服从非高斯分布，其等效协方差分别为 Q_k 和 R_k，即 $\mathrm{q}_k = \sum \mathrm{a}_i\mathcal{N}(0, \mathrm{Q}_i)$，$\mathrm{r}_k = \sum \mathrm{a}_i\mathcal{N}(0, \mathrm{R}_i)$。我们还假设噪声彼此不相关，即 $\mathrm{E}[\mathrm{q}_k\mathrm{r}_k{}^T] = 0$。我们的目标是根据测量值 $\mathcal{Y}_{1:k}$，求得后验估计 $\hat{}\,\mathcal{X}_{k|k}$，即为系统（1）和（2）所定义系统的 $\mathrm{p}(\hat{}\,\mathcal{X}_{k|k}|\mathcal{Y}_{1:k})$；其中滤波器假设系统不存在扰动，即 $\Delta\mathrm{F}_k = 0$。

备注*1*：这里的噪声q_k和r_k均为非高斯噪声。它们可以表示为多个高斯噪声的加权和，形式为 $\sum a_i\mathcal{N}(v_i, \Sigma_i)$，其中$a_i$为满足条件$\sum a_i = 1$的权重。$v_i$ 和 Σ_i 分别是第i^{th} 个正态分布的均值和协方差矩阵。

III. 相关熵

相关熵与由核带宽定义的联合空间邻域内两个随机变量相似性的概率直接相关。相关熵函数产生一个包含二阶及更高阶概率密度函数矩的标量。我们分别用$p(x_i, \hat{}\,x_i)$和$F(x_i, \hat{}\,x_i)$表示各i^{th} 状态的联合概率密度函数和累积分布函数，其中$i \in [1,n]$。因此，每个状态x_i 的相关熵可定义为

$$\begin{aligned} \mathrm{v}(x_i, \hat{x}_i) &= \mathbf{E}[\mathrm{k}(x_i, \hat{x}_i)] \\ &= \int\int \mathrm{k}(x_i, \hat{x}_i)p(x_i, \hat{x}_i)dx_i d\hat{x}_i \\ &= \int \mathrm{k}(x_i, \hat{x}_i)dF(x_i, \hat{x}_i), \end{aligned} \tag{3}$$

其中$\mathbf{E}$ 是期望算子，$\mathsf{k}(.,.)$是核函数。对于高斯核，$\mathsf{k}(x_i, \hat{x}_i)$可表示为

$$\mathsf{k}(x_i, \hat{x}_i) = G_\sigma(x_i, \hat{x}_i) = \exp(-\frac{(x_i - \hat{x}_i)^2}{2\sigma^2}), \tag{4}$$

其中 σ 定义了核带宽。现在，考虑误差函数，高斯核可写为

$$G_\sigma(x_i, \hat{x}_i) = G_\sigma(e_i) = \exp(-\frac{e_i^2}{2\sigma^2}). \tag{5}$$

因此，对于高斯核，系统每个状态的相关熵为

$$\mathsf{v}_\sigma(x_i, \hat{x}_i) = \int G_\sigma(e_i) dF(x_i, \hat{x}_i). \tag{6}$$

在实际情况下，可用的样本数据有限，因此通常无法获得联合累积分布函数$F(x_i, \hat{x}_i)$。因此，我们可以借助样本均值估计量写出每个状态的相关熵。于是，系统在任意时间步k 的总相关熵为

$$\hat{\mathsf{v}}_\sigma(\mathcal{X}_k, \hat{\mathcal{X}}_k) = \frac{1}{n}\sum_{i=1}^{n} G_\sigma(e_{k,i}). \tag{7}$$

其中 $e_{k,i} = (x_{k,i} - \hat{x}_{k,i})$。

注 *2*： 将式（7）按泰勒级数展开，可以说明相关熵函数的几个重要性质。通过泰勒级数展开，得到 $\hat{\mathsf{v}}_\sigma(x_i, \hat{x}_i) = \sum_{N=0}^{\infty} \frac{(-1)^N}{2^N \sigma^{2N}} N! \mathbf{E}[(x_i - \hat{x}_i)^{2N}]$。可以说，对于高斯核函数，相关熵是两个随机变量之差的所有偶数阶矩之和。这使均方误差（MSE）与相关熵准则有了显著区别，因为均方误差仅涉及二阶矩。

注释 *3*： 对于高斯分布，均方误差可提供最优估计，因为在这种情况下二阶矩已足够。然而，要描述非高斯分布，仅凭二阶矩是不够的。因此，基于均方误差的滤波方法（如卡尔曼滤波器）在非高斯分布情况下无法提供最优解，所以我们转而采用基于相关熵的滤波方法。

注释 *4*： 核带宽 σ 对偶数阶矩起加权参数的作用。因此可以说，与二阶矩相比， σ 的增加对高阶矩的影响更大。当核带宽取较大值时，高阶矩将接近于零。因此，该方法的作用将类似于均方误差准则。

IV. 鲁棒最大相关熵卡尔曼滤波（RMCKF）

A. 代价函数

本节为鲁棒 MCKF 定义一种不同于现有成本函数的代价函数。为描述该代价函数，首先将系统 (1) 和 (2) 增广如下：

$$\begin{bmatrix} \hat{\mathcal{X}}_{k|k-1} \\ \mathcal{Y}_k \end{bmatrix} = \begin{bmatrix} I \\ H_k \end{bmatrix} \mathcal{X}_k + v_k, \tag{8}$$

其中

$$v_k = \begin{bmatrix} -(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}) \\ r_k \end{bmatrix}. \tag{9}$$

So,

$$\begin{aligned}\mathbf{E}[v_k v_k^T] &= \begin{bmatrix} P_{k|k-1} & 0 \\ 0 & R_k \end{bmatrix} \\ &= \begin{bmatrix} B_{p,k|k-1}B_{p,k|k-1}^T & 0 \\ 0 & B_{r,k}B_{r,k}^T \end{bmatrix} \\ &= B_k B_k^T.\end{aligned} \tag{10}$$

$B_{P、k|k-1}$ 和 $B_{r、k}$ 分别是 $P_{k|k-1}$ 和 R_k 的平方根。由式（10）可知，B_k 是矩阵 $\begin{bmatrix} P_{k|k-1} & 0 \\ 0 & R_k \end{bmatrix}$ 的平方根，并且可以表示为一个对角矩阵

$$B_k = \begin{bmatrix} B_{p,k|k-1} & 0 \\ 0 & B_{r,k} \end{bmatrix}. \tag{11}$$

现在，将式（8）的两边左乘$B^{-1}{}_k$，可得

$$D_k = W_k \mathcal{X}_k + e_k, \tag{12}$$

其中，$D_k = B^{-1}{}_k \begin{bmatrix} \hat{\mathcal{X}}_{k|k-1} & \mathcal{Y}_k \end{bmatrix}^T$，$W_k = B^{-1}{}_k \begin{bmatrix} I & H_k \end{bmatrix}^T$，且$e_k = B^{-1}{}_k v_k$。$e_k$ 表示维度为（$n + m$）$\times 1$的误差矩阵。现在，我们定义如下代价函数，

$$J_L(\mathcal{X}_k) = \frac{1}{L}\sum_{i=1}^{L}[\exp\{-\sum_{j=1}^{k-1}\frac{\mu_1 e_{j,i}^2}{2\sigma^2} - \frac{\mu_2 e_{k,i}^2}{2\sigma^2}\}], \tag{13}$$

其中，$e_{j,i}$ 是i^{th} 在j^{th} 时间步的误差矩阵元素（e_j）。μ_1 和μ_2 是两个风险敏感参数，σ 是核带宽，且 $L = n + m$。利用式（12），$e_{k,i}$ 可定义为

$$e_{k,i} = d_{k,i} - w_{k,i}x_{k,i}, \tag{14}$$

其中，$e_{k,i}$、$d_{k,i}$ 和$x_{k,i}$ 分别是e_k、D_k 和 $\mathcal{X}_k$ 的i^{th} 个元素，而i^{th} 是W_k的第$w_{k,i}$ 行。我们的目标是通过最大化代价函数（13），利用接收到的测量值 $\mathcal{Y}_{1:k}$，求得状态$\hat{\mathcal{X}}^*{}_{k|k}$ 的最优后验估计，其表达式为

$$\hat{\mathcal{X}}_{k|k}^* = \arg\max_{\mathcal{X}_k} J_L(\mathcal{X}_k). \tag{15}$$

注释 *5*：在（13）中，我们提出了一种此前文献中未曾出现过的新代价函数。它可以看作是最大相关熵准则 [3] 与风险敏感代价函数 [7] 的结合。请注意，项 $[\exp\{-\sum_{j=1}^{k-1}\mu_1 e_{j,i}^2 - \mu_2 e_{k,i}^2\}]$ 是（5）中所述高斯核 $G_\sigma(e)$ 的一种改进形式，其中代价函数 $J_L(\mathcal{X}_k)$ 是相关熵函数。$2\sigma^2$

注释 *6*：值得注意的一点是，误差矩阵 e_k 并非真实状态与估计状态之间的直接差值，而是分别针对过程误差和测量误差引入了加权因子 $B^{-1}{}_{p,k|k-1}$ 和 $B^{-1}{}_{r,k}$。这是基于相关熵的滤波器与基于均方误差的滤波器在误差矩阵构造方式上的根本区别。

备注 *7*：对于风险参数 $\mu_1 = 0$、$\mu_2 = 1$ 以及有限核带宽即 $\sigma \neq \infty$，代价函数与 [3] 中提到的最大相关熵代价函数相同。对于 $\mu_1 \neq 0$、$\mu_2 = 1$ 以及无限核带宽即 $\sigma \to \infty$，代价函数与 [7] 第 2 节中定义的风险敏感代价函数相同。对于 $\mu_1 = 0$、$\mu_2 = 1$ 和 $\sigma \to \infty$，该代价函数与均方误差代价函数相同；对其进行最小化可得到卡尔曼滤波。

B. 鲁棒最大相关熵卡尔曼滤波器（*RMCKF*）的构建

后验信息状态密度 ξ_k 定义为 $\xi_k = p(\mathcal{X}_k|\mathcal{I}_k)$，其中 $\mathcal{I}_k = \{\mathcal{Y}_{1:k}, e^{1|1}、\cdots、e_{k-1|k-1}\}$。此外，利用 [13]，的公式 (17)，信息状态密度还可进一步写为

$$\xi_k = p(\mathcal{X}_k|\mathcal{I}_k) = \exp(-\sum_{j=1}^{k-1} \frac{\mu_1 e_{j,i}^2}{2\sigma^2}) p(\mathcal{X}_k|\mathcal{Y}_{1:k}). \tag{16}$$

引理 *1*： ξ_k 的递归更新表达式为

$$\xi_k = \zeta_k p(\mathcal{Y}_k|\mathcal{X}_k) \int p(\mathcal{X}_k|\mathcal{X}_{k-1}) \exp(-\frac{\mu_1 e_{k-1,i}^2}{2\sigma^2}) \xi_{k-1} d\mathcal{X}_{k-1}. \tag{17}$$

其中 $\zeta_k = \frac{1}{p(\mathcal{Y}_k|\mathcal{Y}_{1:k-1})}$。

证明： 根据贝叶斯定理，状态的后验概率密度函数（pdf）$p(\mathcal{X}_k|\mathcal{Y}_{1:k})$可写为

$$\begin{aligned} p(\mathcal{X}_k|\mathcal{Y}_{1:k}) &= \frac{p(\mathcal{Y}_{1:k}|\mathcal{X}_k)p(\mathcal{X}_k)}{p(\mathcal{Y}_{1:k})} \\ &= \frac{p(\mathcal{Y}_k|\mathcal{Y}_{1:k-1}, \mathcal{X}_k)p(\mathcal{Y}_{1:k-1}|\mathcal{X}_k)p(\mathcal{X}_k)}{p(\mathcal{Y}_k|\mathcal{Y}_{1:k-1})p(\mathcal{Y}_{1:k-1})} \\ &= \frac{p(\mathcal{Y}_k|\mathcal{Y}_{1:k-1}, \mathcal{X}_k)p(\mathcal{X}_k|\mathcal{Y}_{1:k-1})p(\mathcal{Y}_{1:k-1})p(\mathcal{X}_k)}{p(\mathcal{Y}_k|\mathcal{Y}_{1:k-1})p(\mathcal{Y}_{1:k-1})p(\mathcal{X}_k)} \\ &= \frac{p(\mathcal{Y}_k|\mathcal{X}_k, \mathcal{Y}_{1:k-1})p(\mathcal{X}_k|\mathcal{Y}_{1:k-1})}{p(\mathcal{Y}_k|\mathcal{Y}_{1:k-1})} \\ &= \zeta_k p(\mathcal{Y}_k|\mathcal{X}_k, \mathcal{Y}_{1:k-1})p(\mathcal{X}_k|\mathcal{Y}_{1:k-1}). \end{aligned} \tag{18}$$

这里， $\zeta_k = \frac{1}{p(\mathcal{Y}_k|\mathcal{Y}_{1:k-1})}$ 是归一化常数，其中 $p(\mathcal{Y}_k|\mathcal{Y}_{1:k-1}) = \int p(\mathcal{Y}_k|\mathcal{X}_k)p(\mathcal{X}_k|\mathcal{Y}_{1:k-1})d\mathcal{X}_k$。考虑到 $\mathcal{Y}_k$ 与之前的测量值 $\mathcal{Y}_{1:k-1}$相互独立，可写为 $p(\mathcal{Y}_k|\mathcal{X}_k, \mathcal{Y}_{1:k-1}) = p(\mathcal{Y}_k|\mathcal{X}_k)$。因此，式（18）变为

$$p(\mathcal{X}_k|\mathcal{Y}_{1:k}) = \zeta_k p(\mathcal{Y}_k|\mathcal{X}_k) p(\mathcal{X}_k|\mathcal{Y}_{1:k-1}). \tag{19}$$

应用查普曼–柯尔莫哥洛夫积分，$p(\mathcal{X}_k|\mathcal{Y}_{1:k})$可表示为

$$p(\mathcal{X}_k|\mathcal{Y}_{1:k}) = \zeta_k p(\mathcal{Y}_k|\mathcal{X}_k) \int p(\mathcal{X}_k|\mathcal{X}_{k-1}) p(\mathcal{X}_{k-1}|\mathcal{Y}_{1:k-1}) d\mathcal{X}_{k-1}. \tag{20}$$

现在，将式 (20) 中的 $p(\mathcal{X}_k|\mathcal{Y}_{1:k})$ 代入式 (16)，得到

$$\begin{aligned} \xi_k = p(\mathcal{X}_k|\mathcal{I}_k) &= \exp(-\sum_{j=1}^{k-1} \frac{\mu_1 e_{j,i}^2}{2\sigma^2}) \times \zeta_k p(\mathcal{Y}_k|\mathcal{X}_k) \\ &\quad \times \int p(\mathcal{X}_k|\mathcal{X}_{k-1}) p(\mathcal{X}_{k-1}|\mathcal{Y}_{1:k-1}) d\mathcal{X}_{k-1} \\ &= \zeta_k p(\mathcal{Y}_k|\mathcal{X}_k) \int p(\mathcal{X}_k|\mathcal{X}_{k-1}) \exp(-\frac{\mu_1 e_{k-1,i}^2}{2\sigma^2}) \\ &\quad \times \exp(-\sum_{j=1}^{k-2} \frac{\mu_1 e_{j,i}^2}{2\sigma^2}) p(\mathcal{X}_{k-1}|\mathcal{Y}_{1:k-1}) d\mathcal{X}_{k-1} \\ &= \zeta_k p(\mathcal{Y}_k|\mathcal{X}_k) \int p(\mathcal{X}_k|\mathcal{X}_{k-1}) \exp(-\frac{\mu_1 e_{k-1,i}^2}{2\sigma^2}) \xi_{k-1} d\mathcal{X}_{k-1}. \end{aligned} \tag{21}$$

■

注释 *8*: 式 (13) 中描述的代价函数也可以用信息状态概率密度函数表示为

$$J_L(\mathcal{X}_k) = \int \exp(-\sum_{j=1}^{k-1} \frac{\mu_1 e_{j,i}^2}{2\sigma^2}) \times \exp(-\frac{\mu_2 e_{k,i}^2}{2\sigma^2}) \times p(\mathcal{X}_k|\mathcal{Y}_{1:k}) d\mathcal{X}_k, \tag{22}$$

or,

$$J_L(\mathcal{X}_k) = \int \exp(-\sum_{j=1}^{k-1} \frac{\mu_1 e_{j,i}^2}{2\sigma^2}) \xi_k d\mathcal{X}_k. \tag{23}$$

注释 *9*: 根据 [13], , 我们定义 $\int p(\mathcal{X}_k|\mathcal{X}_{k-1})\exp(-^{\mu_1}{}_{e^{k-1},_{2\sigma^2}}{}^{i})\xi_{k-1}d\mathcal{X}_{k-1}$ 为先验信息状态密度, 并将其记为 $p(\mathcal{X}_k|\mathcal{I}_{k-1},e^{k-1|k-1})$。由引理 1 可知, 即使初始信息状态为高斯分布, 信息状态概率密度函数也不会保持高斯分布。这是因为非高斯过程噪声和测量噪声的存在会使似然函数和状态转移密度变为非高斯分布。不过, 这里我们将 ξ_k 近似为具有均值和等效协方差的高斯分布。

定理 *1*: 在注释 9 的假设下, 先验均值和先验误差协方差的表达式为

$$\hat{\mathcal{X}}_{k|k-1} = F_k \hat{\mathcal{X}}_{k-1|k-1}, \tag{24}$$

$$P_{k|k-1} = F_{k-1}(P_{k-1|k-1}^{-1} - 2\mu_1 I)^{-1} F_{k-1}^T + Q_{k-1}. \tag{25}$$

证明: 先验 假定为高斯分布的信息状态可表示如下 :

$$\xi_{k-1} = \zeta_k (2\pi)^{-n/2} |P_{k-1|k-1}|^{-1/2} \exp(-\frac{1}{2}(\mathcal{X}_{k-1} - \hat{\mathcal{X}}_{k-1|k-1}) P_{k-1|k-1}^{-1} (\mathcal{X}_{k-1} - \hat{\mathcal{X}}_{k-1|k-1})^T). \tag{26}$$

现在, 将 ξ_{k-1} 的值代入先验信息状态密度中, 我们得到 e

$$\begin{aligned}
& p(\mathcal{X}_k|\mathcal{I}_{k-1}, e_{k-1|k-1}) \\
= & \int p(\mathcal{X}_k|\mathcal{X}_{k-1}) \exp(-\frac{\mu_1 e_{i,k-1}^2}{2\sigma^2}) \zeta_{k-1} (2\pi)^{-n/2} |P_{k-1|k-1}|^{-1/2} \\
& \times \exp(-\frac{1}{2}(\mathcal{X}_{k-1} - \hat{\mathcal{X}}_{k-1|k-1}) P_{k-1|k-1}^{-1} (\mathcal{X}_{k-1} - \hat{\mathcal{X}}_{k-1|k-1})^T) d\mathcal{X}_{k-1} \\
= & \zeta_{k-1} (2\pi)^{-n/2} [|P_{k-1|k-1}|^{-1/2} \exp(\frac{1}{2\sigma^2}) \int p(\mathcal{X}_k|\mathcal{X}_{k-1}) \\
& \times \exp(-\frac{1}{2}(\mathcal{X}_{k-1} - \hat{\mathcal{X}}_{k-1|k-1}) P_{k-1|k-1}^{-1} (\mathcal{X}_{k-1} - \hat{\mathcal{X}}_{k-1|k-1})^T) \\
& \times \exp(-\frac{1}{2}(\mathcal{X}_{k-1} - \hat{\mathcal{X}}_{k-1|k-1})(-2\mu_1 I)(\mathcal{X}_{k-1} - \hat{\mathcal{X}}_{k-1|k-1})^T)] d\mathcal{X}_{k-1} \\
= & \zeta_{k-1} (2\pi)^{-n/2} \exp(\frac{1}{2\sigma^2}) [|P_{k-1|k-1}|^{-1/2} \int p(\mathcal{X}_k|\mathcal{X}_{k-1}) \\
& \times \exp(-\frac{1}{2}(\mathcal{X}_{k-1} - \hat{\mathcal{X}}_{k-1|k-1})(P_{k-1|k-1}^{-1} - 2\mu_1 I)(\mathcal{X}_{k-1} - \hat{\mathcal{X}}_{k-1|k-1})^T)] \\
& d\mathcal{X}_{k-1},
\end{aligned} \tag{27}$$

其中（$P^{-1}{}_{k-1|k-1} - 2\mu_1 I$）应当可逆。显然，（27）表示一个均值为$\hat{\mathcal{X}}_{k-1|k-1}$、协方差为（$P^{-1}{}_{k-1|k-1} - 2\mu_1 I)^{-1}$的高斯分布。因此，得到（24）和（25）。

■

定理 *2*：状态的后验估计和后验误差协方差的表达式为

$$\hat{\mathcal{X}}_{k|k} = \hat{\mathcal{X}}_{k|k-1} + K_k(\mathcal{Y}_k - H_k\hat{\mathcal{X}}_{k|k-1}), \tag{28}$$

$$P_{k|k} = (I - K_kH_k)P_{k|k-1}(I - K_kH_k)^T + K_kR_kK_k^T, \tag{29}$$

其中

$$K_k = \bar{P}_{k|k-1}H_k^T(H_k\bar{P}_{k|k-1}H_k^T + \bar{R}_k)^{-1}, \tag{30}$$

$\bar{P}_{k|k-1} = (B_{p,k|k-1})\Pi^{-1}{}_{p,k}(B_{p,k|k-1})^T$，$\bar{R}_k = (B_{r,k})\Pi^{-1}{}_{r,k}(B_{r,k})^T$，$\Pi_{p,k} = \mathrm{diag}(\Pi_{p,k,1}, \Pi_{p,k,2}, \cdots, \Pi_{p,k,n})$，以及 $\Pi_{r,k} = \mathrm{diag}(\Pi_{r,k,1}, \Pi_{r,k,2}, \cdots, \Pi_{r,k,m})$，其中 $\Pi_{p,k,i} = \exp(\rho_{p,i} - \mu_2 e_{p,k,i}^2/2\sigma^2)$ 和 $\Pi_{r,k,i} = \exp(\rho_{r,i} - \mu_2 e_{r,k,i}^2/2\sigma^2)$，$e_k = \left[e_{p,k}\, e_{r,k}\right]^T$ 其中 $e_{p,k} = -B^{-1}{}_{p,k|k-1}(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1})$ 和 $e_{r,k} = B^{-1}{}_{r,k}(\mathcal{Y}_k - H_k\mathcal{X}_k)$ 分别表示加权过程误差和测量误差。而 $\rho_{p,i} = \sum_{j=1}^{k-1}(-\mu_1 e_{p,j,i}^2/2\sigma^2)$ 和 $\rho_{r,i} = \sum_{j=1}^{k-1}(-\mu_1 e_{r,j,i}^2/2\sigma^2)$ 分别表示加权的过去过程误差和测量误差。

证明： 对$e_{p,k}$ 和$e_{r,k}$ 相对于 $\mathcal{X}_k$ 进行偏微分后，将得到以下展开式

$$\frac{\partial e_{p,k}}{\partial \mathcal{X}_k} = -(B_{p,k|k-1}^{-1})^T, \tag{31}$$

and

$$\frac{\partial e_{r,k}}{\partial \mathcal{X}_k} = -H_k^T(B_{r,k}^{-1})^T. \tag{32}$$

式（13）中描述的代价函数还可进一步写为

$$\begin{aligned} J_L(\mathcal{X}_k) &= \frac{1}{L}\sum_{i=1}^{L}\exp(\rho_i - \frac{\mu_2 e_{k,i}^2}{2\sigma^2}) \\ &= \frac{1}{L}[\sum_{i=1}^{n}\exp(\rho_i - \frac{\mu_2 e_{p,k,i}^2}{2\sigma^2}) + \sum_{i=1}^{m}\exp(\rho_i - \frac{\mu_2 e_{r,k,i}^2}{2\sigma^2})] \end{aligned} \tag{33}$$

其中 $\rho_i = \sum_{j=1}^{k-1}(-\frac{\mu_1 e_{j,i}^2}{2\sigma^2})$。现在，对$J_L(\mathcal{X}_k)$相对于 $\mathcal{X}_k$进行偏微分，可得

$$\begin{aligned} \frac{\partial J_L(\mathcal{X}_k)}{\partial \mathcal{X}_k} &= \frac{1}{L}[\sum_{i=1}^{n}\exp(\rho_{p,i} - \frac{\mu_2 e_{p,k,i}^2}{2\sigma^2})(-\frac{\mu_2 e_{p,k,i}}{\sigma^2})\frac{\partial e_{p,k,i}}{\partial \mathcal{X}_k} + \sum_{i=1}^{m}\exp(\rho_{r,i} - \frac{\mu_2 e_{r,k,i}^2}{2\sigma^2})(-\frac{\mu_2 e_{r,k,i}}{\sigma^2})\frac{\partial e_{r,k,i}}{\partial \mathcal{X}_k}] \\ &= 0. \end{aligned} \tag{34}$$

通过化简（34），可得

$$\sum_{i=1}^{n}\exp(\rho_{p,i} - \frac{\mu_2 e_{p,k,i}^2}{2\sigma^2})e_{p,k,i}\frac{\partial e_{p,k,i}}{\partial \mathcal{X}_k} + \sum_{i=1}^{m}\exp(\rho_{r,i} - \frac{\mu_2 e_{r,k,i}^2}{2\sigma^2})e_{r,k,i}\frac{\partial e_{r,k,i}}{\partial \mathcal{X}_k} = 0. \tag{35}$$

由（31）、（32）和（35）可得以下递归方程:

$$(B_{p,k|k-1}^{-1})^T \Pi_{p,k} B_{p,k|k-1}^{-1}(\hat{\mathcal{X}}_{k|k} - \hat{\mathcal{X}}_{k|k-1}) - H_k^T (B_{r,k}^{-1})^T \Pi_{r,k} B_{r,k}^{-1}(\mathcal{Y}_k - H_k \hat{\mathcal{X}}_{k|k}) = 0, \tag{36}$$

考虑到 $B_{p,\ k|k-1}\Pi^{-1}{}_{p,\ k}B^T{}_{p,\ k|k-1} = \bar{}P_{k|k-1}$ 和 $B_{r,\ k}\Pi^{-1}{}_{r,\ k}B^T{}_{r,\ k} = \bar{}R_k$，式（36）可改写为

$$\bar{P}_{k|k-1}^{-1}(\hat{\mathcal{X}}_{k|k} - \hat{\mathcal{X}}_{k|k-1}) = H_k^T \bar{R}_k^{-1}(\mathcal{Y}_k - H_k \hat{\mathcal{X}}_{k|k}). \tag{37}$$

求解式（37），可得

$$\hat{\mathcal{X}}_{k|k} = \hat{\mathcal{X}}_{k|k-1} + K_k(\mathcal{Y}_k - H_k \hat{\mathcal{X}}_{k|k-1}), \tag{38}$$

其中

$$K_k = (\bar{P}_{k|k-1} + H_k^T \bar{R}_k^{-1} H_k)^{-1} H_k^T \bar{R}_k^{-1}. \tag{39}$$

现在，将 Sherman–Morrison–Woodbury矩阵恒等式 [14] 应用于式（39），即可得到式（30）。利用式（28），后验误差协方差可计算为

$$\begin{aligned} P_{k|k} &= \mathbf{E}[(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k})(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k})^T] \\ &= \mathbf{E}[((\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}) - K_k(H_k\mathcal{X}_k + r_k - H_k\hat{\mathcal{X}}_{k|k-1}))((\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}) - K_k(H_k\mathcal{X}_k + r_k - H_k\hat{\mathcal{X}}_{k|k-1}))^T] \\ &= \mathbf{E}[((\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}) - K_kH_k(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}) - K_kr_k)((\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}) - K_kH_k(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}) - K_kr_k)^T] \\ &= P_{k|k-1} - P_{k|k-1}H_k^TK_k^T - K_kH_kP_{k|k-1} + K_kH_kP_{k|k-1}H_k^TK_k^T + K_kR_kK_k^T \\ &= (I - K_kH_k)P_{k|k-1}(I - K_kH_k)^T + K_kR_kK_k^T. \end{aligned} \tag{40}$$

■

ˆ算法1 计算 $\mathcal{X}_{k|k}$的不动点迭代

$$[\hat{\mathcal{X}}_{k|k}] := FPI[\hat{\mathcal{X}}_{k|k}, \hat{\mathcal{X}}_{k|k-1}, \mathcal{Y}_k]$$

1) 设置初始迭代 $t = 0$，选择 t_{max} 值和 $\hat{\ }\mathcal{X}^{(t)}{}_{k|k} = \hat{\ }\mathcal{X}_{k|k-1}$。

2）**对于** $t = 1 : t_{max}$

• 计算 ep,k = −B^{-1} p,k|k−1($\hat{\ }\mathcal{X}$(t) k|k − $\hat{\ }\mathcal{X}$k|k−1) 和 er,k = B^{-1} r,k ($\mathcal{Y}$ k − Hk $\hat{\ }\mathcal{X}$(t) k|k)。• 计算 Πp,k 和 Πr,k。

• 根据 (30) 计算 Kk。• 根据 (28) 计算 $\hat{\ }\mathcal{X}$(t+1) k|k。• 如果 $\|\hat{\ }\mathcal{X}$(t+1) k|k $-\mathcal{X}$ $\|$ $\hat{\ }$(t) kk $\hat{\ }$(t) k|k $\leq \epsilon$ $\|\mathcal{X}$，其中 ϵ 是阈值——跳出并更新 $\hat{\ }\mathcal{X}$k| k = $\hat{\ }\mathcal{X}$(t+1) k|k。否则——t = t + 1 并继续迭代。

• **结束**

if3) **结束 for**

备注*10*：误差矩阵e_k 在我们的滤波算法中发挥着重要作用。可以看出，矩阵 $\Pi_k = \begin{bmatrix} \Pi_{p,\ k} & 0 \\ 0 & \Pi_{r,\ k} \end{bmatrix}$ 包含误差矩阵中的元素，而这些元素对我们来说是不可用的。由于在实际场景中我们无法获得真实状态，因此无法直接计算该误差矩阵。相反，我们需要采用一种可以通过以下方式计算的近似值

算法2

ˆ 1) 设置 $\mathcal{X}_{0|0}$ 和$P_{0|0}$的初始值。ˆ 2) 使用（24）和（25）计算 $\mathcal{X}_{k|k-1}$ 和$P_{k|k-1}$ 。3) 使用（10）计算 $B_{p,\ k|k-1}$ 和$B_{r,\ k}$ 。4) 计算后验状态：$\hat{\mathcal{X}}_{k|k} = FPI[\hat{\mathcal{X}}_{k|k}, \hat{\mathcal{X}}_{k|k-1}, \mathcal{Y}_k]$5) 根据（40）计算$P_{k|k}$ 。

算法1中介绍的迭代方法。值得注意的是，我们使用后验估计来计算误差，而要计算后验估计，则需要误差矩阵。因此，这构成了一个不动点迭代，我们按照算法1中的说明，选择后验估计的初始值。

注*11*：在（34）中，$\rho_{p,\ i}$ 和 $\rho_{r,\ i}$ 表示加权过去误差。显然，在计算k^{th} 时间步的估计时，过去误差已经过优化。此外，由于 μ_1, 的值较小，加权过去误差与加权当前误差相比要小得多。因此，与加权当前误差相比，可以忽略它。

注*12*：所提出的滤波器对核带宽 σ非常敏感。对于 $\sigma \to \infty$, $\Pi_{p,\ k} = I_n$,$\Pi_{r,\ k} = I_m$ ，所提出的 RMCKF 会变为鲁棒平方根卡尔曼滤波器。对于较小的 σ值，算法可能无法收敛。这将在后续章节中说明。

备注*13*：风险敏感参数 μ_1 在此算法中充当调节参数。增大 μ_1 的值会提高滤波技术的鲁棒性。然而，在选择参数值时，必须确保满足条件 $(P^{-1}{}_{k-1|k-1} - 2\mu_1 I) > 0$，以保证每个时间步的误差协方差矩阵均为正定。进一步选择 μ_2 时，应确保$\Pi_{p,\ k}$ 和 $\Pi_{r,\ k}$ 在任何传播步骤中均不奇异。值得注意的是，尽管我们将 μ_1 和 μ_2 视为常数，但只要上述条件在每个时间步都得到满足，它们也可以随时间变化。

C. 核带宽的选择（σ）

核带宽（σ）是一个重要且敏感的参数，正确选择它对于实现精确估计至关重要。文献中尚无可靠的方法可用于确定核带宽的最优值。尽管有少数论文 [10]–[12] 探讨了核带宽选择方面的难题，并提出了一些基于测量的方程，用于在每个时间步求出最佳值。在[10], 中，观测误差的范数被启发式地视为核带宽值，即$\sigma_k = \|\mathcal{Y}_k - \mathrm{H}_k \hat{\mathcal{X}}_k\|$ ，它表示实际测量值与估计测量值之间的欧几里得距离。在 [11], 中，作者没有采用欧几里得距离，而是采用马哈拉诺比斯距离，即 $\sigma_k = \|(\mathcal{Y}_k - \mathrm{H}_k \hat{\mathcal{X}}_k)^T \mathrm{R}^{-1}\mathrm{k}\,(\mathcal{Y}_k - \mathrm{H}_k \hat{\mathcal{X}}_k)\|$。在 [12], 中，作者考虑了加权新息与加权误差协方差之和，并取 $\sigma_k = (\|\mathrm{r}_k\|_{R_k^{-1}} + \mathrm{H}_k \mathrm{P}_{k|k-1} \mathrm{H}^T \mathrm{k})^{-1}$，其中 $\mathrm{r}_k = (\mathcal{Y}_k - \mathrm{H}_k \hat{\mathcal{X}}_k)$。然而，上述所有方法都无法保证得到最优解。

本文提出了一种基于误差的代价函数，该函数包含受 [15] 中公式（9）启发的相关熵准则，用于在估计的每个时间步确定核带宽值。我们定义

我们的成本准则为

$$\begin{aligned} J_{KB}(e_{k,i}) &= log(\frac{1}{L}\sum_{i=1}^{L} G_{\sigma}(e_{k,i})) \\ &= log(\frac{1}{L}\sum_{i=1}^{L} \exp(-\frac{e_{k,i}^2}{2\sigma_c^2})). \end{aligned} \tag{41}$$

式（13）和式（41）中定义的两个成本函数在基本结构上存在根本差异。由于我们在该成本准则中引入了相关熵函数，因此式（41）中还出现了另一个带宽参数 σ_c 。但该参数 σ_c 不同于我们之前的核带宽 σ ，它是一个常数。为了求出k^{th} 时间步的 σ 值，我们的目标是最大化$J_{KB}(e_{k,i})$，即

$$\sigma_k^* = \arg\max_{\sigma_k} J_{KB}(e_{k,i}). \tag{42}$$

可以看出，式（42）实际上表示的是误差最小化。因此，为了正确识别 σ_k的最优值， σ_c 应为一个常数。

为了确定所需的核带宽，我们采用数值搜索规则。首先，定义一系列可能的 σ 值，并按照算法1中的说明，计算每个时间步的误差。利用误差值，可针对每个 σ_k求得$J_{KB}(e_{k,i})$。随后，我们比较$J_{KB}(e_{k,i})$的值，选出其中的最大值及其对应的 σ_k。我们在每个时间步重复相同的过程。

V. 收敛性与稳定性分析

本节将证明所提算法的收敛性和稳定性。此外，还将证明算法-1中提到的不动点迭代算法的收敛性。

A. 滤波器收敛

引理2：若 $\tilde{\theta}_{\bar{n}} = \mathcal{X}_{\bar{n}} - \hat{\mathcal{X}}_{\bar{n}|\bar{n}}$ 表示状态误差，则$\mathrm{E}[\|\tilde{\theta}_{\bar{n}}\|^2] \leq \mathrm{E}[\|\mathrm{S}_{\bar{n}|\bar{n}}\|^2]\mathrm{E}[\mathrm{V}_{\bar{n}}]$，其中$\mathrm{V}_{\bar{n}}$是由$\mathrm{V}_{\bar{n}} = \tilde{\theta}_{\bar{n}}^T \mathrm{P}_{\bar{n}|\bar{n}}^{-1} \tilde{\theta}_{\bar{n}}$定义的李雅普诺夫函数，$\mathrm{P}_{\bar{n}|\bar{n}} = \mathrm{S}_{\bar{n}|\bar{n}} \mathrm{S}^T_{\bar{n}|\bar{n}}$，且 $\bar{n}$是满足条件 $\bar{n} \geq k \geq 0$的最大时间步。

证明： 根据矩阵逆法则，可写为 $S_{\bar{n}|\bar{n}} S^{-1}_{\bar{n}|\bar{n}} = 1$。因此，

$$\begin{aligned} \mathbf{E}[\|\tilde{\theta}_{\bar{n}}\|^2] &= \mathbf{E}[\|S_{\bar{n}|\bar{n}} S^{-1}_{\bar{n}|\bar{n}} \tilde{\theta}_{\bar{n}}\|^2] \\ &\leq \mathbf{E}[\|S_{\bar{n}|\bar{n}}\|^2 \|S^{-1}_{\bar{n}|\bar{n}} \tilde{\theta}_{\bar{n}}\|^2] \\ &\leq \mathbf{E}[\|S_{\bar{n}|\bar{n}}\|^2] \mathbf{E}[\|S^{-1}_{\bar{n}|\bar{n}} \tilde{\theta}_{\bar{n}}\|^2] \end{aligned} \tag{43}$$

利用矩阵范数性质 $\|A\|^2 = \|A^T A\| = \|A A^T\| \ \forall A$，可得

$$\begin{aligned} \mathbf{E}[\|\tilde{\theta}_{\bar{n}}\|^2] &\leq \mathbf{E}[\|S_{\bar{n}|\bar{n}}\|^2] \mathbf{E}[\|\tilde{\theta}_{\bar{n}}^T S^{-1T}_{\bar{n}|\bar{n}} S^{-1}_{\bar{n}|\bar{n}} \tilde{\theta}_{\bar{n}}\|] \\ &\leq \mathbf{E}[\|S_{\bar{n}|\bar{n}}\|^2] \mathbf{E}[\|\tilde{\theta}_{\bar{n}}^T P^{-1}_{\bar{n}|\bar{n}} \tilde{\theta}_{\bar{n}}\|] \\ &\leq \mathbf{E}[\|S_{\bar{n}|\bar{n}}\|^2] \mathbf{E}[V_{\bar{n}}] \end{aligned} \tag{44}$$

引理*3*：当 $\bar{n} \to \infty$时，若满足以下条件，则滤波器收敛 s

$$\mathbf{E}[\|\tilde{\theta}_{\bar{n}}\|] \to 0$$

证明：（见附录A）。

注释 *14:* 符号O 表示兰道符号（也称大 O 符号），用于描述函数衰减或增长的快慢。更多详情请参见[16]中的附录 B。

B. 不动点迭代的收敛性

引理*4:* 对于 $\beta > 0$ 和 $0 < \alpha < 1$，当初始向量 $\|\mathcal{X}_{0|0}\|_1 \le \beta$ 和 $\forall\ \mathcal{X}_{k|k}$满足以下条件时，不动点迭代的收敛性得以保证

$$(i) \quad \|\hat{f}(\hat{\mathcal{X}}_{k|k})\|_1 \le \beta, \tag{45}$$

$$(ii) \quad \|\frac{\partial \hat{f}(\hat{\mathcal{X}}_{k|k})}{\partial \mathcal{X}_k}\|_1 \le \alpha, \tag{46}$$

其中，$\hat{f}(\hat{\mathcal{X}}_{k|k})$ 表示不动点迭代，使得 $\hat{f}(\hat{\mathcal{X}}_{k|k}) = (W^T{}_k \Pi_k W_k)^{-1} W_k \Pi_k D_k$。

证明：（见附录 B）。

C. 稳定性分析

我们假设系统参数 F_k、H_k、Q_k 和 R_k 是随机有界的，并且式（1）中定义的系统不确定性参数 ΔF_k 是有限的。我们认为测量噪声的等效协方差 R_k 非零且有限，因此 $R^{-1}{}_k$ 非奇异且有界。令系统（1）、（2）的可控性格拉姆矩阵和可观测性格拉姆矩阵为[17]

$$C_{k,k-l} = \sum_{i=k-l}^{k-1} (F_{k,i+1} + \Delta F_{k,i+1})^T Q_i (F_{k,i+1} + \Delta F_{k,i+1}), \tag{47}$$

$$O_{k,k-l} = \sum_{i=k-l}^{k} (F_{i,k} + \Delta F_{i,k})^T H_i^T R_i^{-1} H_i (F_{i,k} + \Delta F_{i,k}), \tag{48}$$

其中，l 为正整数。现在，如果可观测性格拉姆矩阵$O_{k、k-l}$和可控性格拉姆矩阵$C_{k、k-l}$ 有限且有界，即 *i.e.* $0 < k_1 I < O_{k、k-l} < k_2 I$ 以及 $0 < k_3 I < C_{k、k-l} < k_4 I$，则系统（1）、（2）是一致完全可观测且一致完全可控的，其中 k_1, k_2, k_3 和 k_4 为实数且为正。我们认为，对于任意k，等效后验误差协方差 $P_{k|k}$ 是正定的。

引理*5*：如果系统(1)、(2)一致完全可观测且一致完全可控，并且对于所有$k\ l$，满足P

$_0 > 0$ $_{-1|}{}_{-1} - 2\ _1 > 0$ $>$ $-$ $<$ and $(P^{-1}{}_{kk}\mu I)$，且$I_{k,kl}, I$，其中，$O^{-1}{}_{-}\Delta O_{k\,k-l} <$ $\Delta O_{k\,k-l} = \sum_{=-}^{k}{}_{i}^{T}{}_{i}^{T}{}_{i}^{-1}{}_{i}^{-1}\Delta_{i\,k}$ $_{ikl}F_{i,k}HRHF_{,}$，则滤波器稳定。

证明： 考虑上述假设，并遵循 [13], 附录 B，即可证明滤波器的稳定性。

■

VI. 仿真结果

本节将通过仿真一些数值示例来验证我们所提算法的有效性。我们将把结果与文献中已有的滤波器进行比较。我们还将比较不确定性参数取不同值时的结果。

A. 问题*1*

考虑一个系统，其模型为

$$\mathcal{X}_{k+1} = (F + \Delta F)\mathcal{X}_k + Gq_k,$$

$$\mathcal{Y}_k = H\mathcal{X}_k + r_k,$$

其中，$F = \begin{bmatrix} 0 & 1 \\ -0.99 & 0.01 \end{bmatrix}$，$\Delta F = \begin{bmatrix} 0 & 0 \\ \delta & 0 \end{bmatrix}$，$G = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$，$H = \begin{bmatrix} 1 & -1 \end{bmatrix}$。$\delta$ 表示给定系统中的不确定性，该不确定性有界且不会影响系统的稳定性。我们考虑 $|\delta| \leq 0.5$。过程噪声 q_k 和测量噪声 r_k 均为零均值、非高斯噪声，并建模为高斯分布之和。我们考虑 $q_k = 0.8\mathcal{N}(0, 0.01) + 0.2\mathcal{N}(0, 1)$，以及 $r_k = 0.8\mathcal{N}(0, 1) + 0.2\mathcal{N}(0, 1000)$。真实值的初始状态取为 $x_0 = \begin{bmatrix} 10 & 20 \end{bmatrix}^T$，初始误差协方差为 $P_0 = \begin{bmatrix} 35^2 & 0 \\ 0 & 70^2 \end{bmatrix}$。为进行估计，我们采用均值为 x_0、协方差为 P_0 的随机初始状态。风险参数 $\mu_{1,k-1}$ 的选取使得 $(P^{-1}_{k-1|k-1} - \mu_{1,k-1}I) > 0$ 在每个时间步 k 均满足相应条件，而 $\mu_{2,k-1}$ 为任意标量值。滤波时，我们采用过程噪声和测量噪声的等效协方差。

图1比较了卡尔曼滤波器（KF）、风险敏感卡尔曼滤波器（RSKF）、最大相关熵卡尔曼滤波器（MCKF）、采用固定核带宽的鲁棒最大相关熵卡尔曼滤波器（RMCKF-FK）、采用选定核带宽的最大相关熵卡尔曼滤波器（MCKF-SK），以及采用选定核带宽的鲁棒最大相关熵卡尔曼滤波器（RMCKF-SK）的状态2均方根误差（RMSE）。由于状态1具有相似特征，因此未展示其结果。可以看出，与其他滤波器相比，RMCKF-SK的结果更好。图2展示了每个时间步的核带宽选定值（σ）$_k$。

表I 不同滤波器的平均均方根误差比较

δ	KF	RSKF	MCKF	RMCKF-FK	MCKF-SK	RMCKF-SK
0	0.55	0.59	0.52	0.55	0.43	0.50
0.1	2.29	2.23	2.12	2.08	2.07	1.97
0.2	4.45	4.26	3.97	4.11	3.85	3.15
0.3	6.64	6.32	5.47	5.05	4.22	3.41
0.4	8.85	8.49	5.64	4.96	4.68	2.77
0.5	10.65	10.43	6.23	4.96	4.06	2.67

对KF、RSKF、MCKF、RMCKF-FK、MCKF-SK和RMCKF-SK的性能进行了详细研究，结果如表1所示。可以观察到，RMCKF-SK优于所有其他滤波器。从表1中可以得出以下几点观察结果：

1) 当系统模型不存在不确定性时，即 $\delta = 0$，鲁棒滤波器的精度低于普通滤波器。这表明，在 $\delta = 0$时，KF、MCKF 和 MCKF-SK 分别优于 RSKF、RMCKF 和 RMCKF-SK。2) 随着不确定性参数 δ增大，所有滤波器的均方根误差也会增大。此外，值得注意的是，对于非零的 δ，RMCKF-FK 和 RMCKF-SK 始终优于 MCKF 和 MCKF-SK。3) MCKF-SK 和 RMCKF-SK 始终分别优于 MCKF 和 RMCKF-FK。这表明，核带宽是基于相关熵的滤波器中一个敏感且非常重要的参数。恰当地选择核带宽总能带来更好的结果。

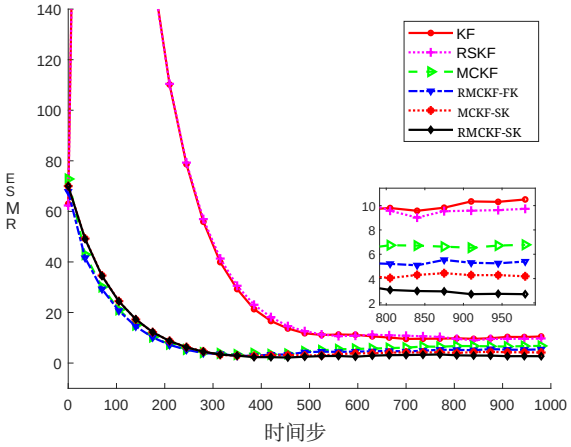

图1. 均方根误差图

B. 问题*2*

考虑一个以恒定加速度运动的系统。同时，我们还考虑到，由于某些外力或内部扰动，模型存在不确定性。将状态定义为 $\mathcal{X}_k = \left[s_k\ v_k\ a_k\right]^T$，其中$s_k$、$v_k$ 和a_k 分别表示系统的位置、速度和加速度。在离散时间下，系统可建模为

$$\mathcal{X}_{k+1} = (F + \Delta F)\mathcal{X}_k + q_k,$$

$$\mathcal{Y}_k = H\mathcal{X}_k + r_k,$$

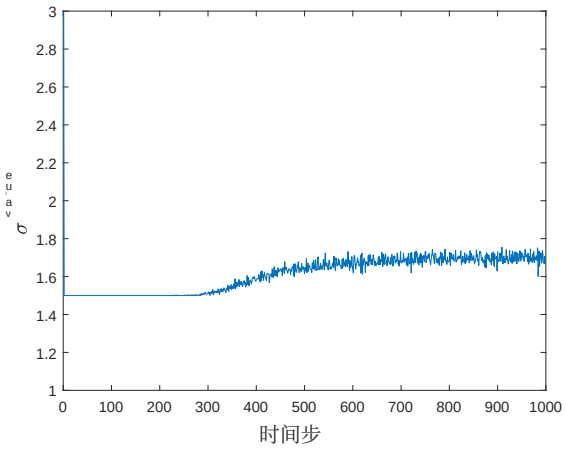

图2. 每个时间步的核带宽 (σ) 值

其中

$$F = \begin{bmatrix} 1 & T & \frac{1}{2}T^2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix},$$

$$\Delta F = \begin{bmatrix} 0 & 0 & \delta_1 T^2 \\ 0 & 0 & \delta_2 T \\ 0 & 0 & 0 \end{bmatrix}$$

$H = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}$，采样时间 T 取为 0.1 分钟，且 $q_k = \begin{bmatrix} q_{1k} & q_{2k} & q_{3k} \end{bmatrix}^T$ 和 $|\delta_1| \leq 0.005$ & $|\delta_2| \leq 0.05$。项 $\delta_1 T^2 a_k$ 表示位置不确定性，它被建模为加速度

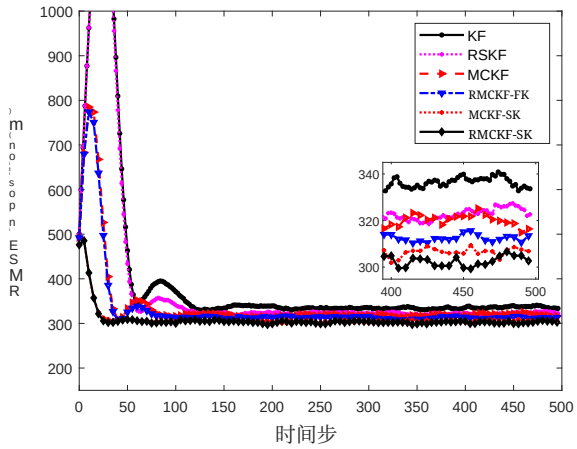

图 3. 位置均方根误差

（a_k）和采样时间（T）的函数。此外，项 $\delta_2 T a_k$ 定义了速度不确定性，它是加速度（a_k）和采样时间（T）的函数。我们将不相关噪声视为高斯混合形式，其均值为零，分布为 $q_{1k} = q_{2k} = q_{3k} = 0.9\mathcal{N}(0, 0.0005) + 0.1\mathcal{N}(0, 0.05)$ 和 $r_k = 0.8\mathcal{N}(0, 0.005) + 0.2\mathcal{N}(0, 50)$。初始状态 $x_0 = \begin{bmatrix} 50 & 4 & 1 \end{bmatrix}^T$ 和初始误差协方差矩阵取为 $P_0 = \begin{bmatrix} 0 & 5^2 & 0 \\ .0 & 0 & .0 \\ 0 & 0 & .1^2 \end{bmatrix}$。风险敏感参数 $\mu_{1,k-1}$ 和 $\mu_{2,k-1}$

按照问题1中的说明选取。

卡尔曼滤波、鲁棒平方根卡尔曼滤波、最大相关熵卡尔曼滤波、RMCKF-FK、MCKF-SK 和 RMCKF-SK 的位置和速度均方根误差（RMSE）分别如图3和图4所示。可以观察到，RMCKF-SK

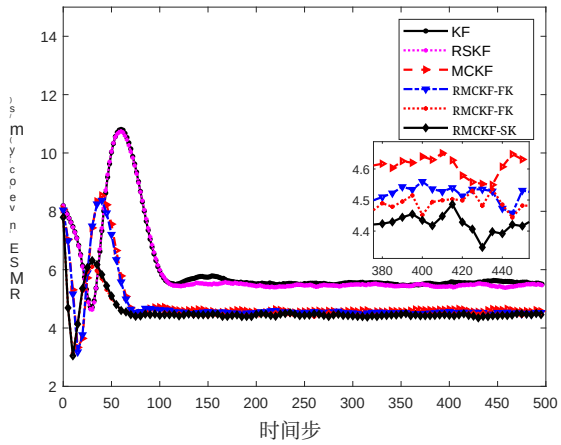

图 4. 速度的均方根误差

与其他滤波器相比，表现更好。所选核带宽的值如图5所示。表2分别展示了随着系统不确定性参数 δ_2 变化，位置均方根误差和速度均方根误差的变化情况。从表中可以得出，RMCKF-SK 的表现优于所有其他所述滤波器。可能有人会质疑为什么要改变 δ_2? 根据系统描述，可以认为 δ_1 只影响位置，而 δ_2 直接影响速度，因此位置也会受到影响。所以，改变 δ_2 意味着同时改变位置和速度的不确定性。这就是我们选择改变 δ_2 的原因。

VII. 结论

我们开发了一种新的滤波算法，用于处理存在非高斯噪声时的不确定系统模型，而名义鲁棒卡尔曼滤波器在这种情况下会失效。我们提出了一种代价函数，采用

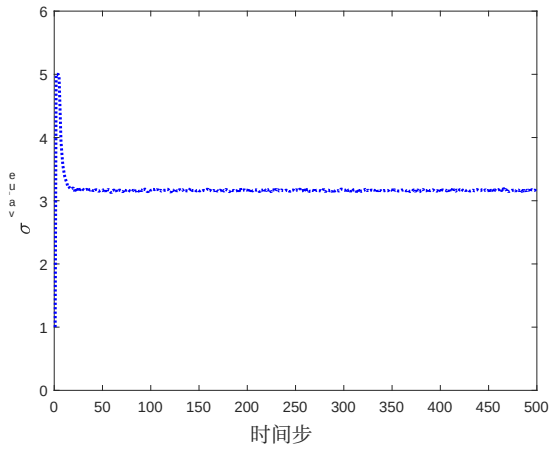

图5. 每个时间步的核带宽（σ）值

最大相关熵准则，并通过将其最大化建立了所提出的滤波递推方程。我们还提出了一种新的数值方法，用于在每个时间步选择核带宽，以获得更好的性能。文中给出了稳定性条件、滤波器收敛条件，以及用于计算后验状态的不动点迭代的收敛条件

参考文献

[1] B. Chen、L. Dang、Y. Gu、N. Zheng 和 J. C. Pr´ıncipe，“最小误差熵卡尔曼滤波器”，*IEEE*《系统、人类与控制论：系统汇刊》，第 51 卷，第 9 期，第 5819–5829 页，2019 年。[2] Y. Zhang、K. Yang 和 Z. Lei，“基于非高斯环境中贝叶斯推断的多径幅度估计”，载于 *2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)*。IEEE，2018 年，第 1–4 页。[3] B. Chen、X. Liu、H. Zhao 和 J. C. Principe，“最大相关熵卡尔曼滤波器”，*Automatica*，第 76 卷，第 70–77 页，2017 年。

表II 不同滤波器的位置和速度（m/s）（m）平均均方根误差比较

δ_2	位置						速度					
	KF	RSKF	MCKF	RMCKF-FK	MCKF-SK	RMCKF-SK	KF	鲁棒平方根卡尔曼滤波	最大相关熵卡尔曼滤波	RMCKF-FK	最大相关熵卡尔曼滤波-SK	RMCKF-SK
0.000	407.10	409.86	341.67	343.23	305.76	312.56	5.88	5.91	4.75	4.79	4.42	4.57
0.005	406.47	409.50	338.90	339.63	306.56	315.00	5.88	5.91	4.76	4.75	4.47	4.61
0.01	409.08	407.68	340.44	336.87	307.38	306.67	5.99	5.93	4.79	4.70	4.44	4.33
0.015	405.33	405.09	341.10	337.07	307.33	300.37	5.95	5.91	4.79	4.75	4.45	4.36
0.02	405.01	403.26	343.62	333.84	307.89	304.54	5.99	5.90	4.80	4.73	4.50	4.38
0.025	405.67	401.03	342.56	333.11	306.56	302.28	5.99	5.90	4.80	4.60	4.48	4.43
0.03	407.21	402.73	341.89	333.57	308.13	303.05	6.00	5.95	4.85	4.75	4.52	4.42
0.035	405.60	403.01	342.49	327.74	307.90	304.79	6.01	5.94	4.82	4.73	4.52	4.47
0.04	404.34	401.90	342.55	337.27	307.81	300.17	6.01	5.89	4.85	4.80	4.52	4.43
0.045	404.75	401.61	343.53	332.85	307.64	304.70	6.07	5.96	4.89	4.74	4.55	4.50
0.05	399.94	397.95	343.57	330.89	308.66	304.95	6.04	6.00	4.87	4.74	4.53	4.44

[4] W. Liu、P. P. Pokharel 和 J. C. Principe，“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。

[5] C. Hu、G. Wang、K. Ho 和 J. Liang，《基于拉普拉斯核最大相关熵准则的鲁棒椭圆拟合》，*IEEE*图像处理汇刊，第 30 卷，第 3127–3141 页，2021 年。

[6] X. Wang、Q. Sun、L. Chen、D. Mu 和 R. Liu，《用于鲁棒荷电状态估计的混合最大相关熵准则无迹卡尔曼滤波器》，载于*2022*年*IEEE*第五届电子信息与通信技术国际会议（*ICEICT*）。IEEE，2022 年，第 670–676 页。

[7] S. Bhaumik、S. Sadhu 和 T. K. Ghoshal，《建模不准确系统的风险敏感估计器》，载于*2005*年*IEEE*印度年度会议（*Indicon*）。IEEE，2005 年，第 86–91 页。[8] S. Bhaumik、S. Sadhu 和 T. Ghoshal，《无迹卡尔曼滤波器的风险敏感表述》，*IET* 控制理论 & 应用，第 3 卷，第 4 期，第 375–382 页，2009 年。[9] S. Bhaumik，《航空航天问题的改进滤波与估计方法》，博士学位论文，博士论文，工程系，Jadavpur 大学，2008 年。

[10] G. T. Cinar 和 J. C. Principe，“使用带有不动点更新和自适应核大小的相关熵滤波器进行隐状态估计”，载于 *2012* 年国际神经网络联合会议 *(IJCNN)*。IEEE，2012 年，第 1–6 页。

[11] B. Hou, Z. He, X. Zhou, H. Zhou, D. Li, and J. Wang, “非高斯噪声下用于 α-加加速度跟踪模型的最大相关熵准则卡尔曼滤波器，” 熵，第 19 卷，第 12 期，第 648 页，2017 年。[12] S. Fakoorian, R. Izanloo, A. Shamshirgaran, and D. Simon, “具有自适应核大小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*）。IEEE，2019 年，第 581–584 页。[13] R. K. Tiwari and S. Bhaumik, “具有随机延迟测量值的风险敏感滤波，”*Automatica*，第 142 卷，110409，2022 年，第。[14] K. S. Riedel, “一种适用于秩增广矩阵并用于中心化的 Sherman–Morrison–Woodbury恒等式，”*SIAM*矩阵分析与应用杂志，第 13 卷，第 2 期，第 659–662 页，1992 年。[15] A. Singh and J. C. Principe, “信息论代价函数中的核宽度自适应，”载于*2010*年*IEEE*声学、语音与信号处理国际会议。IEEE，2010 年，第 2062–2065 页。[16] J. Havil and F. Dyson，2010 年。[17] A. H. Jazwinski，随机过程与滤波理论。Courier Corporation，2007 年。[18] H. V. Henderson and S. R. Searle，“关于矩阵和的逆的推导，”*SIAM*评论，第 23 卷，第 1 期，第 53–60 页，1981 年。[19] L. Guo，“通过基于卡尔曼滤波器的算法估计时变参数：稳定性与收敛性，”*IEEE*自动控制汇刊，第 35 卷，第 2 期，第 141–147 页，1990 年。[20] R. P. Agarwal, M. Meehan, and D. O’regan，不动点理论及其应用。剑桥大学出版社，2001 年，第 141 卷。

[21] H. Zhao, B. Tian, and B. Chen, “基于最大相关熵准则的鲁棒稳定迭代无迹卡尔曼滤波器，”*Automatica*，第 142 卷，第 110410 页，2022 年。

[22] B. Chen、J. Wang、H. Zhao、N. Zheng 和 J. C. Principe，“最大相关熵准则下不动点算法的收敛性”，*IEEE Signal Processing Letters*，第22卷，第10期，第1723–1727页，2015年。

附录

A. 引理*3*的证明

˜根据引理2，可以得出$\mathbf{E}[\|\theta_{\tilde{n}}\|] \to 0$当$\mathbf{E}[V_{\tilde{n}}] \to 0$时成立。因此，我们的任务归结为证明，对于 ˉ $n \to \infty$，有$\mathbf{E}[V_{\tilde{n}}] \to 0$。考虑式（40）中的（$I - K_k H_k$）$= G_k$，并利用式（25），可得

$$P_{k|k} = G_k F_{k-1}(P_{k-1|k-1}^{-1} - 2\mu_1 I)^{-1} F_{k-1}^T G_k^T + K_k R_k K_k^T + \tilde{Q}_{k-1}, \tag{49}$$

其中，$\tilde{Q}_{k-1} = G_k Q_{k-1} G^T{}_k$。误差 $\tilde{\theta}_k$ 可构造如下。

$$\begin{aligned} \tilde{\theta}_k &= (\mathcal{X}_k - \hat{\mathcal{X}}_{k|k}) \\ &= \mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1} - K_k(\mathcal{Y}_k - H_k\hat{\mathcal{X}}_{k|k-1}) \\ &= (\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}) - K_k(H_k\mathcal{X}_k + r_k - H_k\hat{\mathcal{X}}_{k|k-1}), \end{aligned} \tag{50}$$

ˆ 将式（1）中的 $\mathcal{X}_k$ 和式（24）中的 $\mathcal{X}_{k|k-1}$ 代入并整理各项，得 t

$$\begin{aligned} \tilde{\theta}_k &= G_k F_{k-1}\tilde{\theta}_{k-1} + G_k q_{k-1} - K_k r_k \\ &= G_k F_{k-1}\tilde{\theta}_{k-1} + z_k, \end{aligned} \tag{51}$$

whe关于 $z_k = -K_k r_k + G_k q_{k-1}$。利用所定义的李雅普诺夫函数，可将其写为 a s

$$\begin{aligned} V_k &= [G_k F_{k-1}\tilde{\theta}_{k-1} + z_k]^T P_{k|k}^{-1}[G_k F_{k-1}\tilde{\theta}_{k-1} + z_k] \\ &= \tilde{\theta}_{k-1}^T F_{k-1}^T G_k^T P_{k|k}^{-1} G_k F_{k-1}\tilde{\theta}_{k-1} + \tilde{\theta}_{k-1}^T F_{k-1}^T G_k^T P_{k|k}^{-1} z_k + z_k^T P_{k|k}^{-1} G_k F_{k-1}\tilde{\theta}_{k-1} + z_k^T P_{k|k}^{-1} z_k. \end{aligned} \tag{52}$$

Now,

$$\begin{aligned} F_{k-1}^T G_k^T P_{k|k}^{-1} G_k F_{k-1} &= F_{k-1}^T G_k^T [G_k F_{k-1}(P_{k-1|k-1}^{-1} - 2\mu_1 I)^{-1} F_{k-1}^T G_k^T + K_k R_k K_k^T + \tilde{Q}_{k-1}]^{-1} G_k F_{k-1} \\ &\le F_{k-1}^T G_k^T [G_k F_{k-1} P_{k-1|k-1} F_{k-1}^T G_k^T + K_k R_k K_k^T + \tilde{Q}_{k-1}]^{-1} G_k F_{k-1} \\ &\le [P_{k-1|k-1} + F_{k-1}^{-1} G_k^{-1}(K_k R_k K_k + \tilde{Q}_{k-1})(F_{k-1}^{-1} G_k^{-1})^T]^{-1} \end{aligned} \tag{53}$$

利用 [18] in (53) 的式（16）中给出的矩阵求逆公式，可得

$$\begin{aligned} F_{k-1}^T G_k^T P_{k|k}^{-1} G_k F_{k-1} &\le P_{k-1|k-1}^{-1} - P_{k-1|k-1}^{-1} F_{k-1}^{-1} G_k^{-1}[(K_k R_k K_k^T + \tilde{Q}_{k-1})^{-1} + (G_k^{-1})^T (F_{k-1}^{-1})^T P_{k-1|k-1}^{-1} F_{k-1}^{-1} G_k^{-1}]^{-1}(G_k^{-1})^T (F_{k-1}^{-1}) \\ &\le P_{k-1|k-1}^{-1} - [P_{k-1|k-1} + P_{k-1|k-1} F_{k-1}^T G_k^T (K_k R_k K_k^T + \tilde{Q}_{k-1})^{-1} G_k F_{k-1} P_{k-1|k-1}]^{-1} \\ &\le S_{k-1|k-1}^{-1T}(I - [I + S_{k-1|k-1}^T F_{k-1}^T G_k^T (K_k R_k K_k^T + \tilde{Q}_{k-1})^{-1} G_k F_{k-1} S_{k-1|k-1}]^{-1}) S_{k-1|k-1}^{-1} \end{aligned} \tag{54}$$

现在，利用性质 $CAB \le C\|A\|B \ \forall A > 0, B > 0, C > 0$，我们得到

$$F_{k-1}^T G_k^T P_{k|k}^{-1} G_k F_{k-1} \le S_{k-1|k-1}^{-1T}(1 - [1 + \frac{\|S_{k-1|k-1}^T F_{k-1}^T G_k^T G_k F_{k-1} S_{k-1|k-1}\|}{\|(K_k R_k K_k^T + \tilde{Q}_{k-1})\|}]^{-1}) S_{k-1|k-1}^{-1} \tag{55}$$

利用矩阵范数性质 $||A^TA|| = ||AA^T|| \ \forall A$，我们可以写成

$$F_{k-1}^T G_k^T P_{k|k}^{-1} G_k F_{k-1} \leq [1 - (1 + ||(K_k R_k K_k^T + \tilde{Q}_{k-1})^{-1} G_K F_{k-1} P_{k-1|k-1} F_{k-1}^T G_k^T||)^{-1}] P_{k-1|k-1}^{-1} \tag{56}$$

将（49）代入（56），可得

$$F_{k-1}^T G_k^T P_{k|k}^{-1} G_k F_{k-1} \leq [1 - (1 + ||(K_k R_k K_k^T + \tilde{Q}_{k-1})^{-1} P_{k|k}||)^{-1}] P_{k-1|k-1}^{-1} \tag{57}$$

˜现在，$P_{k-1|k-1} + Q_{k-1} \geq P_{k|k}$。因此，（57）为

$$\begin{aligned} F_{k-1}^T G_k^T P_{k|k}^{-1} G_k F_{k-1} &\leq [1 - (1 + ||\tilde{Q}_{k-1}^{-1}(P_{k-1|k-1} + \tilde{Q}_{k-1})||)^{-1}] P_{k-1|k-1}^{-1} \\ &\leq [1 - (2 + ||\tilde{Q}_{k-1}^{-1}||||P_{k-1|k-1}||)^{-1}] P_{k-1|k-1}^{-1} \\ &= P_{k-1|k-1}^{-1} - \frac{P_{k-1|k-1}^{-1}}{2 + ||\tilde{Q}_{k-1}^{-1}||||P_{k-1|k-1}||} \end{aligned} \tag{58}$$

将（58）代入（52），可得

$$\begin{aligned} V_k &\leq V_{k-1} - \frac{V_{k-1}}{2 + ||\tilde{Q}_{k-1}^{-1}||||P_{k-1|k-1}||} + \tilde{\theta}_{k-1}^T F_{k-1}^T G_k^T P_{k|k}^{-1} z_k + z_k^T P_{k|k}^{-1} G_k F_{k-1} \tilde{\theta}_{k-1} + z_k^T P_{k|k}^{-1} z_k \\ &\leq V_{k-1} - \frac{V_{k-1}}{2 + ||\tilde{Q}_{k-1}^{-1}||||P_{k-1|k-1}||} + 2||z_k^T P_{k|k}^{-1} G_k F_{k-1} \tilde{\theta}_{k-1}|| + ||z_k^T P_{k|k}^{-1} z_k|| \end{aligned} \tag{59}$$

利用初等不等式 $2|ab| \leq a^2 + b^2$, ，可得

$$\begin{aligned} 2||z_k^T P_{k|k}^{-1} G_k F_{k-1} \tilde{\theta}_{k-1}|| &\leq 2||z_k^T S_{k|k}^{-1}||||S_{k|k}^{-1} G_k F_{k-1} \tilde{\theta}_{k-1}|| \\ &\leq ||z_k^T P_{k|k}^{-1} z_k|| + ||\tilde{\theta}_{k-1}^T F_{k-1}^T G_k^T P_{k|k}^{-1} G_k F_{k-1} \tilde{\theta}_{k-1}|| \end{aligned} \tag{60}$$

利用（58），可得 $\tilde{\theta}^T$k₋₁F^{T}k₋₁G^{T}k P^{-1} k|kG$_k$F$_{k-1}\tilde{\theta}_{k-1} \leq$ V$_{k-1}$，代入 z$_k$ = –K$_k$r$_k$ + G$_k$q$_{k-1}$后，可得 ||z^{T}k P^{-1} k|kz$_k$|| ≤ ||S $^{-1}$ k|k(–K$_k$r$_k$ + G$_k$q$_{k-1}$)||^2 ≤ O(||P^{-1} k|k||(||r$_k$||^2 + ||q$_{k-1}$||^2)）。因此，由（59）可得

$$V_k \leq 2[V_{k-1} - \frac{V_{k-1}}{4 + 2||\tilde{Q}_{k-1}^{-1}||||P_{k-1|k-1}||}] + O(||P_{k|k}^{-1}||(||r_k||^2 + ||q_{k-1}||^2)). \tag{61}$$

现在，考虑一个函数 ϕ(˜n,k)，使得

$$\phi(\bar{n}, k) = (1 - \frac{1}{4 + 2||\tilde{Q}_{\bar{n}-1}^{-1}||||P_{\bar{n}-1|\bar{n}-1}||})\phi(\bar{n} - 1, k), \tag{62}$$

$\forall$˜n ≥k ≥ 0。因此我们可以写成

$$V_{\bar{n}} \leq \phi(\bar{n} - 1, 0) V_0. \tag{63}$$

在引理5的条件下 [19],

$$\mathbf{E}[\phi(\bar{n}, k)] \leq M\gamma^{n-k}, \tag{64}$$

$\forall$˜n ≥k ≥ 0, 0 < γ < 1, M < ∞。现在，对于 ˜n → ∞和 k = 0, $\gamma \to 0$即 ϕ(˜n, k)) → 0。因此，

$$\begin{aligned} \mathbf{E}[V_{\bar{n}}] &\leq O(\mathbf{E}[\phi(\bar{n}, 0)||\tilde{\theta}_0||^2]) \\ &\leq O(\mathbf{E}[\phi(\bar{n}, 0))]\mathbf{E}[||\tilde{\theta}_0||^2]) \to 0 \end{aligned} \tag{65}$$

■

B. 引理*4*的证明

$\hat{f}(\hat{X}_{k|k})$ 的不动点迭代在 [3] 的第3节中推导。为了证明不动点算法的收敛性，压缩映射定理（也称为巴拿赫不动点定理） [20]是一个非常重要的工具。利用该定理，可以证明所提出的 RMCKF 的收敛性。从 [21], 的第4节中，我们可以写出

$$\begin{aligned}\hat{f}(\hat{X}_{k|k}) &= (W_k^T \Pi_k W_k)^{-1}(W_k \Pi_k D_k) \\ &= [\hat{R}_{WW}^G]^{-1}[\hat{P}_{dW}^G].\end{aligned} \tag{66}$$

取范数值，可得

$$\|\hat{f}(\hat{X}_{k|k})\|_1 \le \|[\hat{R}_{WW}^G]^{-1}\|_1 \|[\hat{P}_{dW}^G]\|_1. \tag{67}$$

现在，利用 [3],中的式(20, 23)，可得

$$[\hat{R}_{WW}^G] = \frac{1}{L}\sum_{i=1}^{L}[G_\sigma(e_{k,i}) w_{k,i} w_{k,i}^T], \tag{68}$$

and

$$[\hat{P}_{dW}^G] = \frac{1}{L}\sum_{i=1}^{L}[G_\sigma(e_{k,i}) d_{k,i} w_{k,i}]. \tag{69}$$

根据 [22],中的定理 1，可写为

$$\|[\hat{R}_{WW}^G]^{-1}\|_1 \le \sqrt{L}\|[\hat{R}_{WW}^G]^{-1}\|_2 \le \sqrt{L}\lambda_{max}[[\hat{R}_{WW}^G]^{-1}], \tag{70}$$

其中 $\lambda_{max}[\circ]$ 表示给定矩阵的最大特征值。由(70)，可写为 e

$$\begin{aligned}\lambda_{max}[[\hat{R}_{WW}^G]^{-1}] &= \frac{1}{\lambda_{min}[[\hat{R}_{WW}^G]^{-1}]} \\ &= \frac{L}{\lambda_{min}[\sum_{i=1}^{L}[G_\sigma(e_{k,i}) w_{k,i} w_{k,i}^T]]} \\ &\le \frac{L}{\lambda_{min}[\sum_{i=1}^{L}[G_\sigma(\beta\|w_{k,i}\|_1 + |d_{k,i}|) w_{k,i} w_{k,i}^T]]},\end{aligned} \tag{71}$$

其中， $\lambda_{min}[.]$ 表示给定矩阵的最小特征值，且 $\|e_{k,\,i}\|_1 = \|d_{k,\,i} - w_{k,\,i}x_{k,\,i}\|_1 \le |d_{k,\,i}| + \beta\|w_{k,\,i}\|_1$。现在，

$$\begin{aligned}\|[\hat{P}_{dW}^G]\|_1 &= \|\frac{1}{L}\sum_{i=1}^{L}[G_\sigma(e_{k,i}) d_{k,i} w_{k,i}]\|_1 \\ &\le \frac{1}{L}\sum_{i=1}^{L}\|[G_\sigma(e_{k,i}) d_{k,i} w_{k,i}]\|_1 \\ &\le \frac{1}{L}\sum_{i=1}^{L}|d_{k,i}|\|w_{k,i}\|_1,\end{aligned} \tag{72}$$

对于任意i、k，均有 $\|G_\sigma(e_{k,\,i})\| \le 1$。

利用（67）、（70）、（71）和（72），我们可以得到

$$\|\hat{f}(\hat{X}_{k|k})\|_1 \le \phi(\sigma) = \frac{\sqrt{L}[\frac{1}{L}\sum_{i=1}^{L}|d_{k,i}|\|w_{k,i}\|_1] \times L}{\lambda_{min}[\sum_{i=1}^{L}[G_\sigma(\beta\|w_{k,i}\|_1 + |d_{k,i}|) w_{k,i} w_{k,i}^T]]} \tag{73}$$

由（73）可知，$\lim_{\sigma\to 0^+}\phi(\sigma)=\infty$ 且 $\lim_{\sigma\to\infty}\phi(\sigma)=\frac{\sqrt{L}\ [\sum_{i=1}^{L}\|\cdot k\|\|\cdot k\|]\ d,\ w,\ w,\ w,\ i]]}{\lambda_{min}[\sum_{i=1}^{L}[\ ki\ k}=\epsilon_1$。现在假设 ^ ^ $\phi(\sigma^*)=\beta$。因此，对于任意 $\sigma\geq\sigma^*$，$\phi(\sigma)\leq\beta$ 都成立，这意味着 $\|f(\mathcal{X}_{k|k})\|_1\leq\beta$。再次由（66），我们可以写出

$$\begin{aligned}\frac{\partial\hat{f}(\hat{\mathcal{X}}_{k|k})}{\partial\mathcal{X}_k}&=\frac{\partial}{\partial\mathcal{X}_k}([\hat{R}_{WW}^G]^{-1}[\hat{P}_{dW}^G])\\&=-[\hat{R}_{WW}^G]^{-1}(\frac{\partial}{\partial\mathcal{X}_k}\hat{R}_{WW}^G)[\hat{R}_{WW}^G]^{-1}\\&\quad\times[\hat{P}_{dW}^G]+[\hat{R}_{WW}^G]^{-1}(\frac{\partial}{\partial\mathcal{X}_k}\hat{P}_{dW}^G)\\&=-[\hat{R}_{WW}^G]^{-1}(\frac{1}{L\sigma^2}\sum_{i=1}^{L}\mu_2 e_{k,i}w_{k,i}G_\sigma(e_{k,i})w_{k,i}w_{k,i}^T)\hat{f}(\mathcal{X}_k)\\&\quad+[\hat{R}_{WW}^G]^{-1}(\frac{1}{L\sigma^2}\sum_{i=1}^{L}\mu_2 e_{k,i}w_{k,i}G_\sigma(e_{k,i})d_{k,i}w_{k,i})\\&=Z_1+Z_2\end{aligned}\tag{74}$$

^ ^ 根据不等式 $\|f(\mathcal{X}_{k|k})\|_1\leq\beta$ 和 $\|G_\sigma(e_{k,i})\|_1\leq 1\ \forall i,k$，可得

$$Z_1\leq\frac{\beta}{L\sigma^2}\|[\hat{R}_{WW}^G]^{-1}\|_1\sum_{i=1}^{L}\mu_2(\beta\|w_{k,i}\|_1+|d_{k,i}|)\|w_{k,i}\|_1\|w_{k,i}w_{k,i}^T\|_1\tag{75}$$

and

$$Z_2\leq\frac{1}{L\sigma^2}\|[\hat{R}_{WW}^G]^{-1}\|_1\sum_{i=1}^{L}\mu_2(\beta\|w_{k,i}\|_1+|d_{k,i}|)\|w_{k,i}\|_1|d_{k,i}|\|w_{k,i}\|_1\tag{76}$$

利用 (70)、(71)、(74)、(75) 和 (76)，可写出以下不等式

$$\begin{aligned}\|\frac{\partial\hat{f}(\hat{\mathcal{X}}_{k|k})}{\partial\mathcal{X}_k}\|_1&\leq\frac{\sqrt{L}\sum_{i=1}^{L}\mu_2(\beta\|w_{k,i}\|_1+|d_{k,i}|)\|w_{k,i}\|_1}{\lambda_{min}[\sum_{i=1}^{L}G_\sigma(\beta\|w_{k,i}\|_1+|d_{k,i}|)w_{k,i}w_{k,i}^T]}\\&\quad\times\frac{(\beta\|w_{k,i}w_{k,i}^T\|_1+|d_{k,i}|\|x_{k,i}\|_1)}{\sigma^2}\\&=\psi(\sigma)\end{aligned}\tag{77}$$

由 (77) 可得 $\lim_{\sigma\to 0^+}\psi(\sigma)=\infty$ 和 $\lim_{\sigma\to\infty}\psi(\sigma)=0$。现在考虑 $\psi(\sigma^+)=\alpha$，其中 $0\leq\alpha\leq 1$。^ ^因此，对于任意 $\sigma\geq\sigma^+$，$\psi(\sigma)\leq\alpha$。由此证明了 $\|^{\partial}{}_{f(\mathcal{X}_{k|k})\partial\mathcal{X}_k}\ \underline{\|_1\leq\alpha}\leq 1$。

■