arXiv:2305.03295v1 [stat.ML] 5 May 2023

非参数有限先验知识条件下的去中心化扩散式学习

预印本

Paweł Wachel
弗罗茨瓦夫科技大学控制系统与机电一体化系，
波兰弗罗茨瓦夫
pawel.wachel@pwr.edu.pl

Krzysztof Kowalczyk
弗罗茨瓦夫科技大学控制系统与机电一体化系，
波兰弗罗茨瓦夫
krzysztof.kowalczyk@pwr.edu.pl

Cristian R. Rojas瑞典皇家理工学院电气工程
与计算机科学学院，瑞典斯德哥尔摩
crro@kth.se

摘要

我们研究了在噪声环境中，利用本地智能体收集的测量值对非线性现象 m 进行扩散式网络学习的问题。针对去中心化网络以及仅在直接相邻节点之间传播信息的情形，我们提出了一种非参数学习算法，该算法避免了原始数据交换，且仅需要关于m的温和的先验知识。我们推导了所提出方法的非渐近估计误差界。通过仿真实验展示了该方法的潜在应用。

关键词 去中心化学习，分布式估计，非参数学习

1 引言

去中心化和分布式学习领域与现代物联网（IoT）及无线传感器网络（WSN）应用领域密切相关。得益于技术进步以及与鲁棒性和可扩展性相关的功能优势 [20], , 去中心化和分布式技术在工业界日益普及，并成为持续科学研究的主题。在去中心化网络中，学习与推断有各种特定目标，例如线性模块学习 [15], 、分布式经济调度 [6] 或目标跟踪 [11], ；其中值得一提的是 [2] 或 [4]中的估计任务。在这种场景下，传感器（或智能体）散布在给定区域内，收集有关某种未知现象的数据，该现象建模为非线性函数 m: $\mathbb{R} \to \mathbb{R}$。由于通信可能受到限制且缺乏专用融合中心，智能体可能只能依赖本地/私有测量值以及可用的网络信息。遵循 [12] 和 [7]，我们先简要概述几种众所周知的策略。

在增量方法中， [9, 3], 局部参数估计值会通过网络按环路传递，从而显著降低通信成本。然而，确定网络环路（哈密顿路径）已知是 NP困难 [14], , 因此每当网络架构发生变化时，计算开销都会很高，例如由于节点故障或智能体移动所致。不过，在静态、节能且具备故障安全能力的网络中，此类算法的性能可能接近其集中式对应方法 [3]。

基于共识和扩散的方法提供了更通用的解决方案，即在智能体的局部邻域中广播数据样本。然而，其中一些方法在实时问题中对实现施加了限制；例如，参见 [19]中的讨论。在此背景下， [18]中提出了改进；另请参见 [10, 16]。值得指出的是，如 [19], 所示，在网络稳定性和信息传播领域，扩散策略的表现优于共识。文献中也有对广为人知的网络内.....的详细讨论。

学习障碍，例如与异步学习 [17, 23], 灵活的用户自定义测量和通信周期 [8, 22], 或网络拓扑变化 [13]相关的障碍。

上述大多数方法都需要参数化先验信息，从而可以将所考虑的问题重新表述为有限参数集的估计。本文所采用的方法放宽了这些要求，并允许先验知识显著有限且为非参数形式。事实上，我们仅假设所关注的潜在非线性现象是利普希茨连续函数。此外，我们不采用梯度下降程序，并避免相关的网络稳定性问题 [19]。与此同时，所提出的算法通过利用网络传入的信息，力求补偿智能体本地测量值的低质量。

本文的主要贡献如下：

- 我们提出了一种新的去中心化学习算法，用于建模传感器网络中智能体在本地观测到的潜在非线性现象。
- 我们对所提出算法的准确度进行了形式化研究，并通过集中不等式推导出适用于用户指定显著性水平的误差界。
- 在对解释变量数据、测量噪声以及所关注的非线性现象施加弱假设的条件下，我们研究了该算法的形式性质。
- 该算法的理论分析与其实践应用相关，因为误差界对任意有限（非渐近）的观测值数量均成立。
- 该算法及其相应的误差界简单易懂，因此可以在计算资源有限的智能体中实现和评估。
- 该算法能够保护数据隐私：节点不会共享其本地（私有）输出测量值；网络中仅交换原始预估值。
- 该算法能够适应网络节点间测量噪声水平的差异。具体而言，它倾向于利用可获取的在线网络信息，弥补本地观测值质量不佳所带来的影响。

本文结构如下。第2节和第3节正式阐述所研究的网络学习任务，介绍模型学习和模型利用流程，并讨论该方法的形式性质。数值实验的结果见第4节，结论见第5节。

2 问题公式化

我们研究一个由 M 智能体组成的去中心化网络，在对感兴趣的未知非线性现象的先验知识有限的情况下进行推断。在所考虑的问题中，我们将该网络建模为一个连通的无向图$\mathcal{G} = (\mathcal{M},\ \mathcal{E})$，包含 $\mathcal{M} = \{1, 2, \ldots, M\}$ 个节点和一组无权边 $\mathcal{E}$。如果两个节点 $i,\ j \in \mathcal{M}$ 直接相连，则当且仅当它们可以交换信息，即，如果 $\{i, j\} \in \mathcal{E}$。对于任意节点 $i \in \mathcal{M}$，其邻域是集合 $\mathcal{N}_i = \{j :\ \{i, j\} \in \mathcal{E}\}$。
我们考虑一种网络学习场景：在每个时间步 $t \in \mathbb{N}$，每个智能体 $k \in \mathcal{M}$ 都会获取解释变量数据 $\xi_k^{loc}{}_{,t} \in \mathbb{R}$, 并观测到某个未知底层非线性现象 m的结果 $\mathcal{D} \subset \mathbb{R} \to \mathbb{R}$，这些结果以含噪测量 $y_{k\,t}^{loc}$,的形式呈现，其中

$$y_{k,t}^{loc} = m(\xi_{k,t}^{loc}) + \eta_{k,t}^{loc}, \quad k \in \mathcal{M}, \quad t \in \mathbb{N}, \tag{1}$$

以及 $\eta_k^{loc}{}_{,\ t}$ 表示传感器噪声/不准确性。未知映射 m 对于整个网络而言是相同的，尽管各个智能体的测量值可能仅具有局部特征[1],，即对于任意智能体 $k \in \mathcal{M}$，相应的数据集 $\{(\xi_k^{loc}{}_{,\ n},\ y_k^{loc}{}_{,\ n})\ n = 1,\ 2,\ \ldots\ldots,\ t\}$ 仅允许在域 $\mathcal{D}$的某个局部子集 $\mathcal{D}_k$ 内推断出 m；$\xi_{k,\ n} \in \mathcal{D}_k$，参见图 1。

尽管智能体对所研究现象的了解可能有限，但我们的目标是设计一种去中心化网络学习算法，使其能够将自身的局部（私有）推断能力扩展到更广泛的范围$\mathcal{D}_1 \cup \mathcal{D}_2 \cup \cdots \cup \mathcal{D}_M$。然而，由于所有智能体只能与其直接邻居进行通信，因此当前问题具有信息扩散的性质。

[1]为便于表述，本文使用上标‘*loc*’、‘*acq*’、‘*req*’和‘*shr*’分别作为术语本地、已获取、已请求 和已共享的缩写。

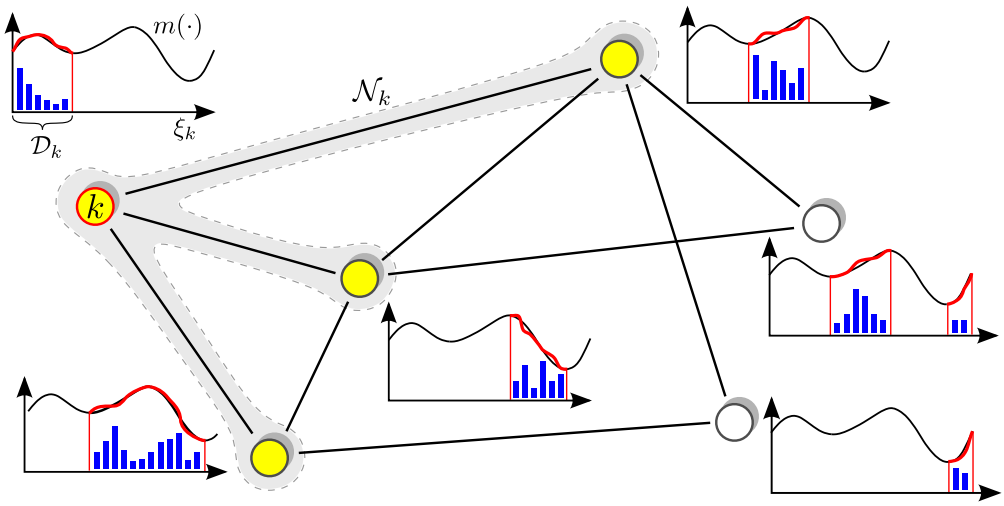


图 1：由智能体组成的去中心化网络，其中包含局部解释变量数据的直方图。智能体 k 的邻域 $\mathcal{N}_k$ 已突出显示。

为确保所提方法的通用性，我们仅假设对潜在现象具有非常宽松的先验知识。实际上，只需满足 m 是利普希茨连续的即可。因此，当前问题是非参数的，而参数估计技术可能会导致不一致结果。

假设1. 感兴趣的潜在现象 m： $\mathcal{D} \subset \mathbb{R} \to \mathbb{R}$ 是一个利普希茨连续映射，即

$$|m(\xi) - m(\xi')| \leq L|\xi - \xi'|, \quad \forall\, \xi, \xi' \in \mathcal{D}, \tag{2}$$

其中 $L < \infty$ 为已知常数。

一般而言，每个智能体 $k \in \mathcal{M}$ 的目标都是利用其本地信息和获取的（网络化）信息来学习 m。我们在对智能体的测量值及其相应的不确定性仅作出较弱假设的情况下，对这一场景进行了详细刻画。

假设 2. 对于网络 $\mathcal{G}$ 中的任意智能体 $k \in \mathcal{M}$，局部解释变量序列 $\{\xi_{k,t}^{loc}\colon\ t \in \mathbb{N}\}$ 是一个任意随机过程$\xi_{k,t}^{loc} \in \mathcal{D}_k \subset \mathcal{D}$。

假设 3. 对于网络 $\mathcal{G}$ 中的任意智能体 $k \in \mathcal{M}$，扰动 $\{\eta_{k,t}^{loc}\colon\ t \in \mathbb{N}\}$ 是一个次高斯随机过程，其含义是存在一个 $\sigma_k > 0$，使得对于任意 $\gamma \in \mathbb{R}$ 和 $t \in \mathbb{N}$,

$$\mathbb{E}\{\exp(\gamma \eta_{k,t}^{loc}) | \eta_{k,1}^{loc}, \ldots, \eta_{k,t-1}^{loc}, \xi_{k,1}^{loc}, \ldots, \xi_{k,t}^{loc}\} \leq \exp\left(\frac{\gamma^2 \sigma_k^2}{2}\right). \tag{3}$$

上述假设需要作一些说明。首先注意，局部解释变量序列的类别非常一般，并未对 $k, l \in \mathcal{M}$ 时 $\xi_{k,t}^{loc}$ 与 $\xi_{l,t}^{loc}$ 之间的函数依赖关系施加任何限制。关于条件 (3)，可以证明它意味着 $E\{\eta_{k,t}^{loc} | \eta_{k,1}^{loc}, \ldots, \eta_{k,t-1}^{loc}, \xi_{k,1}^{loc}, \ldots, \xi_{k,t}^{loc}\} = 0$ 和$E\{(\eta_{k,t}^{loc})^2 | \eta_{k,1}^{loc}, \ldots, \eta_{k,t-1}^{loc}, \xi_{k,1}^{loc}, \ldots, \xi_{k,t}^{loc}\} \leq \sigma_k^2$，参见 [21, 练习 2.5]。此外，网络中不同节点的方差代理量 σ_k 可以各不相同。特别地，与过程 $\{\xi_{k,t}^{loc}\}$ 独立的独立同分布零均值高斯序列满足条件 (3)。最后，若 $\eta_{k,t}^{loc} \in [a, b]$ 几乎处处成立，其中某个 $a, b \in \mathbb{R}$，则根据霍夫丁引理，σ_k^2 可取为 $(b-a)^2/4$（尽管在这种情况下，任何大于 $(b-a)^2/4$ 的 σ_k^2 值也都满足 (3)）。

3 非参数去中心化学习

根据问题陈述，网络中的每个智能体都能够收集其本地（私有）观测值$\{(\xi_{k,n}^{loc}, y_{k,n}^{loc})\ n = 1, 2, \ldots\ldots, t\}$，因此只能对映射 m 进行本地估计。接下来，我们将重点讨论

非参数、数值高效的 Nadaraya–Watson估计量

$$\hat{\mu}_{k,t}(x) := \sum_{n=1}^{t} \frac{K_h(x, \xi^{loc}_{k,n})}{\kappa_{k,t}(x)} y^{loc}_{k,n}, \qquad \kappa_{k,t}(x) := \kappa_{k,t}(x,h) = \sum_{n=1}^{t} K_h(x, \xi^{loc}_{k,n}), \tag{4}$$

其中， k 表示智能体索引， $K_h(x,\xi) := K((x-\xi)/h)$，其中 K 和 h 分别表示核函数和带宽参数。

一般而言，在关于 $m(x)$ 的温和要求下，只要观测值"趋于"集中在 x 的邻域内，$\hat{\mu}_{k,t}(x)$ 的逐点一致性就可以在各种概率意义下得到证明。然而，这类结果通常具有渐近性质，无法直接适用于有限数据集任务。因此，针对所考虑的去中心化非参数学习，我们建立了一个非渐近界。为此，考虑以下假设：

假设 4. 核函数 $K\colon \mathbb{R} \to \mathbb{R}$ 是一个非负有界函数，满足对于所有 $v \in \mathbb{R}$， $0 \le K(v) \le 1$。此外，对于所有 $|v| > 1$， $K(v) = 0$。

引理1。设假设*1–4*成立。考虑网络 $\mathcal{G}$中的智能体 $k \in \mathcal{M}$，其配备估计量$\hat{\mu}_{k,t}$，并固定带宽 h。设 $x \in \mathcal{D}_k$ 为一个固定的确定性数值，或者一般而言，是$\eta^{loc}_{k\,1,}$、.....、 $\eta^{loc}_{k\,,t-1}$、 $\xi^{loc}_{k\,1,}$、.....、 $\xi^{loc}_{k\,,t}$ 的函数（例如，$x = \xi^{loc}_{k\,t,}$）。那么，对于每个 $0 < \delta < 1$，以至少 $1-\delta$的概率，若$\kappa_{k,t}(x) \neq 0$,

$$|\hat{\mu}_{k,t}(x) - m(x)| \le \beta_{k,t}(x), \tag{5}$$

$$\textit{where} \quad \beta_{k,t}(x) := Lh + 2\sigma_k \frac{\alpha_{k,t}(x,\delta)}{\kappa_{k,t}(x)} \tag{6}$$

and

$$\alpha_{k,t}(x,\delta) := \begin{cases} \sqrt{\log(\sqrt{2}/\delta)}, & \text{for } \kappa_{k,t}(x) \le 1 \\ \sqrt{\kappa_{k,t}(x)\log(\sqrt{1+\kappa_{k,t}(x)}/\delta)}, & \text{for } \kappa_{k,t}(x) > 1. \end{cases}$$

证明。参见附录。 □

上述引理中的估计误差界具有直接的实际应用价值，因为如果已知 σ_k 和 L，或者能够通过已知常数对其进行上界估计，那么对于任意有限数量的测量值，均可计算 $\beta_{k,t}(x)$。值得注意的是，与Nadaraya-Watson估计中常用的假设（尤其是核函数的对称性或单位积分）不同，核函数 K只需满足假设4中的宽松要求。然而，由于置信界 $\beta_{k,t}$高度依赖于$\kappa_{k,t}$，因此，我们可以选择核函数 K 和带宽参数 h，以使 $\beta_{k,t}$最小化。尽管如此，我们将这些问题留待未来研究。为简单起见，在第4节所介绍的实现中，我们采用经典的箱形核。

基于界（6），我们可以进一步介绍所提出的网络学习算法，见算法1。接下来，我们聚焦于网络中一个任意但固定的智能体 $k \in \mathcal{M}$。显然，对于给定的局部解释样本 $\xi^{loc}_{k\,t,}$，可以构造相应的局部元组 $T(\xi^{loc}_{k\,t,}) := (\xi^{loc}_{k\,t,}, \hat{\mu}^{loc}_{k\,t,}, \beta^{loc}_{k\,t,})$，其中 $\hat{\mu}^{loc}_{k,t} = \hat{\mu}_{k,t}(\xi^{loc}_{k\,t,})$ 且 $\beta^{loc}_{k,t} = \beta_{k,t}(\xi^{loc}_{k\,t,})$。我们将所考虑的智能体 k收集的所有此类元组组成的集合记为 $\mathbb{T}^{loc}_k$。类似地，令 $T(\xi^{acq}_{l\,t,}) := (\xi^{acq}_{l\,t,}, \hat{\mu}^{acq}_{l\,t,}, \beta^{acq}_{l\,t,})$ 表示智能体 k 从节点 $l \in \mathcal{N}_k$获取的元组。智能体 k的所有此类元组组成的集合记为 $\mathbb{T}^{acq}_k$ 。

备注 1. 算法*1*和算法*2*给出了所提出的非参数去中心化学习方法。它们基于以下命令构建：

- 获取 – 获取局部解释样本以及与现象结果相对应的含噪测量值$(\xi^{loc}_{k\,t,}, y^{loc}_{k\,t,})$，适用于 $t = 12, \ldots$
- 评估 – 根据公式(4) 和/或引理1，分别评估估计量值和/或其对应的误差界。
- 追加 – 向本地元组集合或已获取元组集合中追加一个新元组。通常，追加 命令可以包含一种更新机制，在获得更准确的数据时，从给定集合中删除准确度较低的元组（第*4*节所用实现中采用了一个简单的基于利普希茨条件的例程来实现这一点）。

- 可用性检查是否已从邻居 获取元组 $(\xi_{l,t-1}^{acq})$ 或请求的样本 $\xi_{l,t-1}^{req}$

- 发送 – 将给定的元组发送给请求节点 $l \in \mathcal{N}_k$。
- 选择 – 选择智能体要请求的参数。一般而言，该参数可以是确定性的，也可以是 $\eta_{k,1}^{loc}, \ldots, \eta_{k,t-1}^{loc}, \xi_{k,1}^{loc}, \ldots, \xi_{k,t}^{loc}$ 在时间 t 的任意（可测）函数。在第*4*节所用的实现中，该参数从 $\mathcal{D}$ 上的均匀分布中抽取。

- 广播 – 向邻居发送请求。

请注意，论文中提供了第4节所用 MATLAB 代码实现的链接。

算法1 基于非参数扩散的**模型学习** ▷ 智能体 k

输入： $\{T(\xi_{l,t}^{acq}) : l \in \mathcal{N}_k\}, \{\xi_{l,t}^{req} : l \in \mathcal{N}_k\}, \delta, L, \sigma_k$
初始化： $\mathbb{T}_k^{loc} = \varnothing, \mathbb{T}_k^{acq} = \varnothing$
for $t = 1, 2, \ldots$ **做**
  获取 $(\xi_{k,t}^{loc}, y_{k,t})$ ▷ 获取局部测量
  评估 $\hat{\mu}_k^{loc}$ 和 β_k^{loc} 如式 (4)–(6) 所示
  追加 $T(\xi_{k,t}^{loc}) := (\xi_{k,t}^{loc}, \hat{\mu}_{k,t}^{loc}, \beta_{k,t}^{loc})$ 到 $\mathbb{T}_k^{loc}$
  对每个 $l \in \mathcal{N}_k$ **执行** ▷ 获取共享元组
    如果 ISAVAILABLE$(T(\xi_{l,t-1}^{acq}))$，**然后**
      追加 $T(\xi_{l,t-1}^{acq})$ 至 $\mathbb{T}_k^{acq}$
    end
    如果 ISAVAILABLE$(\xi_{l,t-1}^{req})$，**然后** ▷ 发送请求的元组
      $\xi_{l,t}^{shr} \leftarrow \operatorname{argmin}_{\xi} |\xi - \xi_{l,t-1}^{req}|$ 满足条件 $T(\xi) \in \{\mathbb{T}_k^{loc} \cup \mathbb{T}_k^{acq}\}$
      发送 $T(\xi_{l,t}^{shr})$ 至节点 l
    end
  end
  选择 $\xi_{k,t}^{req}$
  广播 $\xi_{k,t}^{req}$ ▷ 向所有智能体广播请求 $l \in \mathcal{N}_k$
end

备注 2. 要应用估计量(4)，必须选择与核的选择性相对应的带宽 $h > 0$ K。请注意，在所考虑的方法中，合理的做法是选择 h，使其最小化(6)中的上界。对于固定的 $x \in \mathcal{D}$，可以通过数值方法找到合适的带宽，例如使用黄金分割搜索算法，因为 $\beta_{k,t}(x)$ 中的两项关于 h 均为单调函数（特别是，对于任意一组解释变量数据 $\{\xi_{k,n}^{loc}\}$，$\kappa_{k,t}(x)$ 会随 h 增大而增大）。

在所考虑的问题设置中，智能体根据网络拓扑$\mathcal{G}$与其直接邻居交换元组。元组扩散协议是所提出的非参数学习方法的一部分，该方法在算法1中进行了描述。对于给定的智能体 k，在时间步 t，该过程有两个输入：一组已获取的元组 $\{T(\xi_{l,t}^{acq}) : l \in \mathcal{N}_k\}$，以及一组请求的参数 $\{\xi_{l,t}^{req} : l \in \mathcal{N}_k\}$。

注释 3。 算法*1* 的关键部分之一是元组请求机制“SELECT $\xi_{k,t}^{req}$”。如定理 *1* 所示，最终估计值 $\hat{m}_{k,t}$ 在任意给定点 x 处的准确度，直接取决于从智能体的邻居处获取的元组。因此，$\xi_{k,t}^{req}$ 的选择应取决于那些对网络中相应智能体尤其重要的参数 $x \in \mathcal{D}$。

关于所研究现象的实际推断由智能体 k 根据算法 2 执行。对于任何请求的 $x \in \mathcal{D}$（不一定来自 $\mathcal{D}_k$），均使用来自算法1的数据集 $\mathbb{T}_k^{loc}, \mathbb{T}_k^{acq}$，以评估估计值 $\hat{m}_{k,t}(x)$ 及其对应的误差界 $\beta_{k,t}(x)$。对于给定的 $\delta \in (0,1)$，该算法会将智能体 k 的本地预测与获取的网络信息进行比较，并返回 $m(x)$ 的非参数估计值，该估计值具有可用的最紧误差上界，置信水平为 $1 - \delta$。

算法 2 非参数扩散式**模型利用** ▷ 智能体 k

1: **输入：** $x \in \mathcal{D}, \delta, L, \sigma_k$2: $\mathbb{T}_k^{loc} = \{(\xi_i^{loc}, \hat{}\mu_i^{loc}, \beta_i^{loc}) : i = 1, 2, \dots\}$ ▷ 本地数据集

3: $\mathbb{T}_k^{acq} = \{(\xi_i^{acq}, \hat{}\mu_i^{acq}, \beta_i^{acq}) : i = 1, 2, \dots\}$ ▷ 已获取数据集

4: 输出： $\{\hat{}m_{k,t}(x)\beta_{k,t}(x)\}$**5: 评估** $\beta_{k,t}(x)$**，如式 (6) 所示 6:** $\mathbb{I}_k \leftarrow \{i: L|x - \xi_i^{acq}| < \beta_{k,t}(x) - \beta_i^{acq}\}$**7: 如果** $\mathbb{I}_k = \varnothing$**，则 8: 评估** $\hat{}\mu_{k,t}(x)$**，如式 (4) 所示 9:** $\hat{}m_{k,t}(x) \leftarrow \hat{}\mu_{k,t}(x)$ **10: 否则 11:** $j \leftarrow \textbf{argmin}_{i \in \mathbb{I}_k}\{L|x - \xi_i^{acq}| + \beta_i^{acq}\}$**12:** $\hat{}m_{k,t}(x) \leftarrow \hat{}\mu_j^{acq}(x)$ **13:** $\beta_{k,t}(x) \leftarrow \beta_j^{acq}(x)$ **14: 结束**

在模型利用算法（算法 2）的第 6 行，根据 m 的利普希茨连续性以及定理 1 中提出的思路，计算已获取元组索引的集合 $\mathbb{I}_k$；对于给定的 x，该集合提供的置信界比局部估计更紧。此外，在第 11 行，如果 $\mathbb{I}_k$ 非空，则通过计算 $\text{argmin}_{i \in \mathbb{I}_k}\{L|x - \xi_i^{acq}| + \beta_i^{acq}\}$ 找出可用置信界最紧的元组。

So所提出算法的非渐近性质将在以下定理中给出 em.

定理 1. 设假设*1–4*均成立。考虑网络 $\mathcal{G}$ 中的任意智能体 $k \in \mathcal{M}$，其使用算法*1*的元组交换协议，并按照算法 *2*估计 m。那么，对于任意 $0 < \delta < 1$、任意 $t \in \mathbb{N}$ 以及任意固定的$x \in \mathcal{D}$（可以是确定性的，也可以是一般意义上关于$\eta_{k\,1}^{loc}$、…，$\eta_{k\,t-1}^{loc}$、$\xi_{k\,1}^{loc}$、…，$\xi_{k\,t}^{loc}$的函数），在给定$\eta_{k\,1}^{loc}$、…，$\eta_{k\,t-1}^{loc}$、$\xi_{k\,1}^{loc}$、…，$\xi_{k\,t}^{loc}$的条件下，以至少$1-\delta$的概率，

$$|\hat{m}_{k,t}(x) - m(x)| \leq \min\{\beta_{k,t}(x), B^{acq}(x)\}, \tag{7}$$

其中 $B^{acq}(x) := \min_i\{B_i^{acq}(x)\}$ 且 $B_i^{acq}(x) = L|x - \xi_i^{acq}| + \beta_i^{acq}$。

证明。根据算法 2，在 $\beta_{k,t}(x) \leq B^{acq}(x)$ 的情况下，$\mathbb{I} = \varnothing$，且估计量 $\hat{}m_{k,t}(x)$ 等于 $\hat{}\mu_{k,t}(x)$。因此，由引理1可知，不等式 (7) 至少以概率 $1-\delta$ 成立。反过来，如果 $\beta_{k,t}(x) > B^{acq}(x)$，那么 $\hat{}m_{k,t}(x) = \hat{}\mu_j^{acq}$，其中 $j = \text{argmin}_{i \in \mathbb{I}}\{L|x - \xi_i^{acq}| + \beta_i^{acq}\}$；这对应于获得的元组 $(\xi_j^{acq}, \hat{}\mu_j^{acq}, \beta_j^{acq})$，对于该元组，$|\hat{}\mu_j^{acq}(\xi_j^{acq}) - m(\xi_j^{acq})| \leq \beta_j^{acq}$ 以概率 $1-\delta$ 成立。因此，根据假设1，

$$\begin{aligned} |\hat{\mu}_j^{acq} - m(x)| &= |\hat{\mu}_j^{acq}(\xi_j^{acq}) - m(\xi_j^{acq}) + m(\xi_j^{acq}) - m(x)| && (8)\\ &\leq |\hat{\mu}_j^{acq}(\xi_j^{acq}) - m(\xi_j^{acq})| + |m(\xi_j^{acq}) - m(x)| && (9)\\ &\leq \beta_j^{acq} + L|\xi_j^{acq} - x| = B_j^{acq}(x) = B^{acq}(x), && (10) \end{aligned}$$

以至少为 $1-\delta$ 的概率，从而完成证明。 □

4 数值实验

本节针对图 2 所示、包含 50 个节点的随机拓扑网络，对模型学习 和 模型利用 算法进行了数值研究。潜在非线性为 $m(x) = \sin(x)\exp(-0.2x) + 3$。在实验中[2]，，假设网络中的所有智能体都有局部解释变量测量值，$\xi_{k\,t,}^{loc}$，这些测量值独立采样自正态分布 $\mathcal{N}(\mu_k^{inp}, \sigma_k^{inp})$，其均值 μ_k^{inp}和离散度 σ_k^{inp}均为随机选取。同样，输出噪声序列 $\eta_{k\,t,}^{loc}$ 由均值为零、离散度随机选取的相互独立正态分布生成；每个智能体的离散度分别固定，并从区间 (0，0.7) 中均匀采样。假设已知的利普希茨常数 L 等于 1，总感兴趣区域 $\mathcal{D}$为区间 $[0, 10]$, 和 $\delta = 0.01$，即 对于每个固定的 $x \in \mathcal{D}$，所有所得误差界都应以 0.99 的概率成立。

[2]用于获得数值结果的 MATLAB 代码可在 https://github.com/githuban10/ Decentralized-diffusion-based-learning 获取。

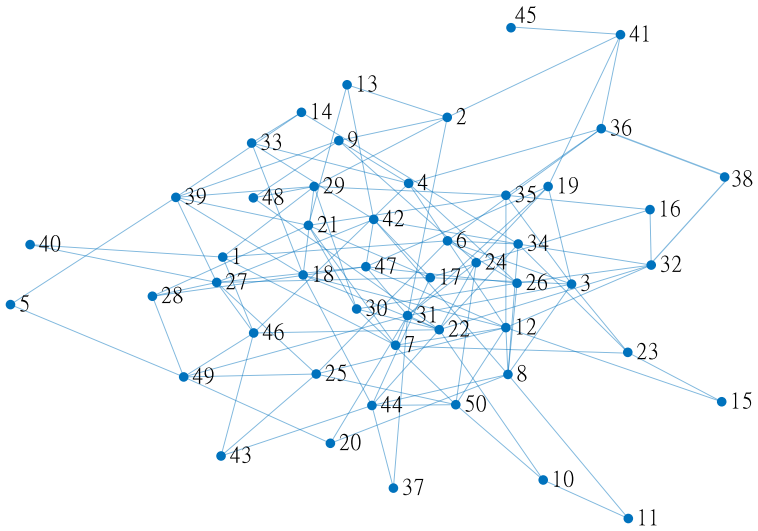


图2：含50个节点的随机拓扑网络。

图 3（下）展示了节点18在时间范围 $t = 1000$ 内收集的本地数据直方图示例。该图还包含节点通过元组扩散过程获取的元组直方图。在实验中，我们假设网络中的所有智能体都从整个感兴趣区域 $\mathcal{D} = [0, 10]$上的均匀分布中，完全随机地生成元组请求 $\xi_{k\,t}^{req}$ ，。

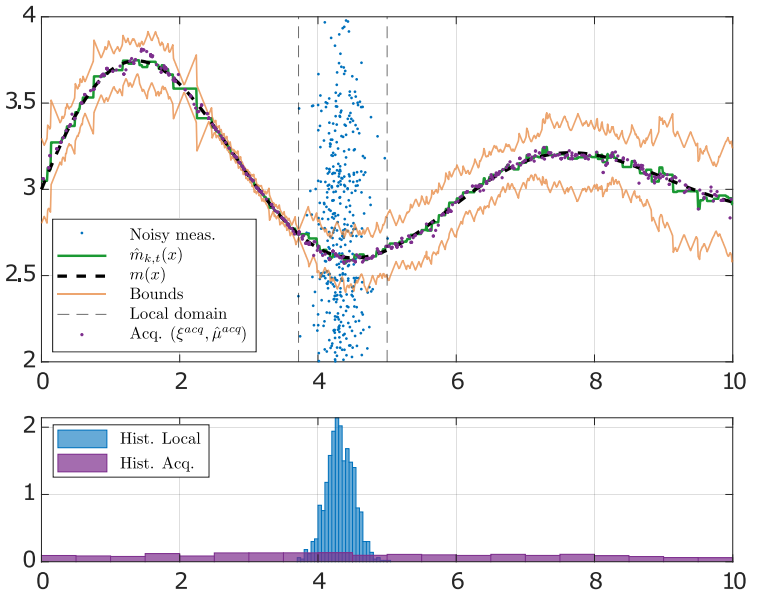


图3：节点 $k = 18$ 在时间 $t = 1000$的学习结果。上图：误差界（橙色）包含潜在现象 m （黑色），但智能体的局部测量值（蓝色）远远超出了由此得到的误差界。下图：局部测量值（蓝色）和从邻居 $\mathcal{N}_k$ 处获取的元组（紫色）的直方图。

为了研究基于扩散的模型学习（算法1）带来的准确度提升，我们将所得模型的结果 $\hat{m}_{k,\,t}$ 与仅基于本地数据的推断结果进行比较。得益于信息扩散，单个智能体能够在一个远大于其自身本地测量值支撑集的范围内，恰当地推断所研究的现象。这在图4中十分明显：我们将所得的（基于扩散的）模型 $\hat{m}_{k,\,t}$ 与基于本地数据的估计结果进行了比较。

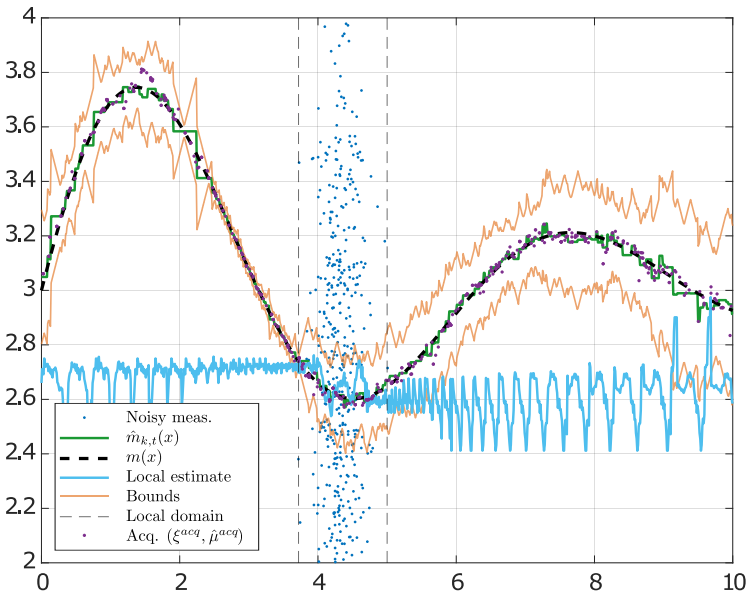
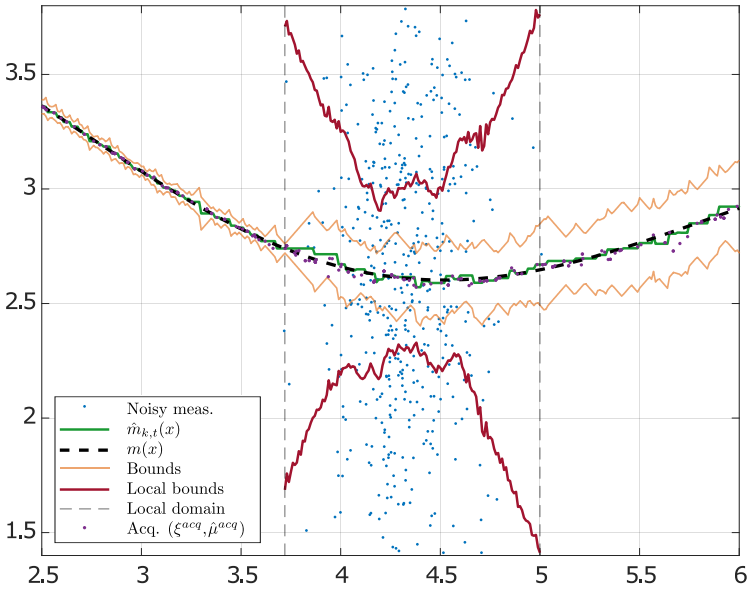


图4：在时间 $t = 1000$ 时节点 $k = 18$ 的学习结果，与仅基于局部严重扰动的测量值进行估计的结果（浅蓝色）相比。

所提出方法的另一个有趣特性与智能体执行的噪声消除有关。考虑一个位于环境中噪声极强区域的单个智能体，或者一个配备低质量传感器（信噪比较低）的智能体。显然，在这种情况下，其局部测量值的质量较低，从而导致局部推断的准确度较低（见图 5），且本地误差界缺乏信息价值。尽管如此，所获取的邻近元组仍能在置信水平保持不变的情况下，获得准确度高得多的估计结果 δ。



图五月 图 5：节点 $k = 18$ 和 $t = 1000$ 的学习结果与本地智能体的界限进行比较（续） d).
丁香

在最后一个实验中，我们研究了估计值 $\hat{m}_{k,t}$ 的误差上界随时间的变化（随着局部测量数量和获取的元组数量增加）。图 6 展示了智能体 k =18（六个连接）和 $k = 45$ （一个连接）的误差界在三种不同置信水平 $\delta \in \{0.01; 0.001; 0.0001\}$ 下的收敛情况。可以观察到，δ 对误差界的影响相对较小，因此可以使用较低的 δ 值运行算法，*i.e.* 即对所得误差界具有较高的置信度。

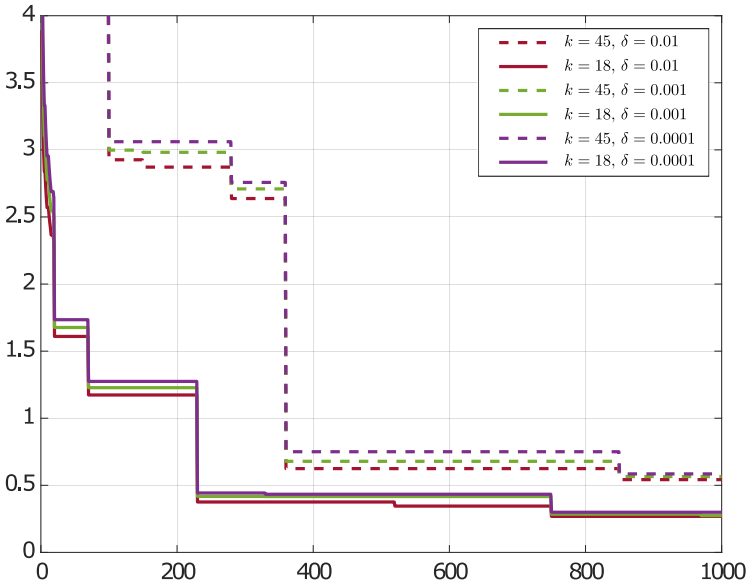


图6：节点 $k = 18,\ \ k = 45$ 的误差上界随时间 t 的演化，适用于 $\delta \in \{10^{-2}, 10^{-3}, 10^{-4}\}$。

5 结论

本文提出了一种针对利普希茨连续非线性现象的新型非参数扩散式去中心化学习方法。该算法由两个协同过程组成：模型学习 和模型利用，并在对测量信号和加性扰动的要求较为宽松的条件下进行了形式化研究。所得估计误差上界具有非渐近性质，并适用于用户指定的置信水平。

我们通过数值仿真展示了该方法的适用性。根据实验结果，所提出的算法能够确保智能体之间的协作，并促进网络中的平稳数据扩散。因此，单个智能体能够克服高噪声水平或测量域狭窄等局部障碍，并以已知准确度为所研究的现象提供高质量模型。此外，置信水平 δ 对所得的（非渐近）误差界影响相对较小，因此可以提高推断置信度，而不会显著损失准确度。

附录

证明(引理*1*的证明)。根据式（1），我们首先观察到，参见 [5],

$$\left|\sum_{n=1}^{t} \frac{K_h(x, \xi_{k,n}^{loc})}{\kappa_{k,t}(x)} y_{k,n}^{loc} - m(x)\right| \leq \sum_{n=1}^{t} \theta_n |m(\xi_{k,n}^{loc}) - m(x)| + \left|\sum_{n=1}^{t} \theta_n \eta_{k,n}^{loc}\right|, \tag{11}$$

其中 $\theta_n := K_h(x,\ \xi_{k,\ n}^{loc})/\kappa_{k,\ t}(x)$。注意，$\Sigma_{n=1}^{t}\theta_n = 1$。根据假设4，若 $K_h(x,\ \xi_{k,\ n}^{loc}) > 0$，则 $|x - \xi_{k,\ n}^{loc}|/h \leq 1$。因此，（参见 假设1）

$$K_h(x, \xi_{k,n}^{loc}) > 0 \quad \Longrightarrow \quad |m(\xi_{k,n}^{loc}) - m(x)| \leq L|x - \xi_{k,n}^{loc}| \leq Lh,$$

由于权重 θ_n 之和为 1,

$$\sum_{n=1}^{t} \theta_n |m(\xi_{k,n}^{loc}) - m(x)| \leq Lh.$$

对于式 (11) 中的最后一项，注意到

$$\left|\sum\nolimits_{n=1}^{t} \theta_n \eta_{k,n}^{loc}\right| = \frac{1}{\kappa_{k,t}(x)} \left|\sum\nolimits_{n=1}^{t} K_h(x, \xi_{k,n}^{loc}) \eta_{k,n}^{loc}\right|. \tag{12}$$

Acc根据引理 2，式 (12) 的右侧以概率 $1-\delta$ 为上界 by

$$\frac{1}{\kappa_{k,t}(x)}\sigma_k\sqrt{2\log\left(\delta^{-1}\sqrt{1+\sum_{n=1}^{t}K_h^2(x,\xi_{k,n}^{loc})}\right)\left(1+\sum_{n=1}^{t}K_h^2(x,\xi_{k,n}^{loc})\right)}.$$

此外，由于 $K_h(x,\ \xi_{k,\ n})\leq 1$ (参见 假设4)，我们得到

$$\frac{1}{\kappa_{k,t}(x)}\left|\sum_{n=1}^{t}K_h(x,\xi_{k,n}^{loc})\eta_{k,n}^{loc}\right|\leq\sigma_k\sqrt{2\log(\delta^{-1}\sqrt{1+\kappa_{k,t}(x)})}\frac{\sqrt{1+\kappa_{k,t}(x)}}{\kappa_{k,t}(x)}.$$

接下来注意，若$\kappa_{k,\ t}(x)>1$，则

$$\frac{\sqrt{1+\kappa_{k,t}(x)}}{\kappa_{k,t}(x)}<\frac{\sqrt{2\kappa_{k,t}(x)}}{\kappa_{k,t}(x)}=\frac{\sqrt{2}}{\sqrt{\kappa_{k,t}(x)}}.$$

因此，以概率 $1-\delta$，对于 $\kappa_{k,\ t}(x)>1$

$$\frac{1}{\kappa_{k,t}(x)}\left|\sum_{n=1}^{t}K_h(x,\xi_{k,n}^{loc})\eta_{k,n}^{loc}\right|\leq\frac{2\sigma_k}{\kappa_{k,t}(x)}\sqrt{\kappa_{k,t}(x)\log\left(\delta^{-1}\sqrt{1+\kappa_{k,t}(x)}\right)},$$

而对于$0<\kappa_{k,\ t}\leq 1$,

$$\begin{aligned}\frac{1}{\kappa_{k,t}(x)}\left|\sum_{n=1}^{t}K_h(x,\xi_{k,n}^{loc})\eta_{k,n}^{loc}\right|&\leq\frac{\sigma}{\kappa_{k,t}(x)}\sqrt{2\log\left(\delta^{-1}\sqrt{1+\kappa_{k,t}(x)}\right)}\sqrt{1+\kappa_{k,t}(x)}\\&\leq\frac{2\sigma}{\kappa_{k,t}(x)}\sqrt{\log(\sqrt{2}/\delta)},\end{aligned}$$

由此完成证明。 □

以下两个引理分别是文献 [1], 中定理 3 和引理1 的一个小幅变体 .

引理 2. 设 $\{v_t\colon\ t\in\mathbb{N}\}$ 是一个有界随机过程，$\{\eta_t\colon\ t\in\mathbb{N}\}$ 是一个随机过程，并且满足次高斯条件，即存在一个 $\sigma>0$，使得对任意 $\gamma\in\mathbb{R}$ 和 $t\in\mathbb{N}$,

$$\mathbb{E}\{\exp(\gamma\eta_t)|\eta_1,\ldots,\eta_{t-1},v_1,\ldots,v_t\}\leq\exp\left(\frac{\gamma^2\sigma^2}{2}\right). \tag{13}$$

另外，设

$$S_t:=\sum_{n=1}^{t}v_n\eta_n\quad\text{and}\quad V_t:=\sum_{n=1}^{t}v_n^2.$$

那么，对于任意 $t\in\mathbb{N}$ 和 $0<\delta<1$，以概率 $1-\delta$,

$$|S_t|\leq\sqrt{2\sigma^2\log\left(\delta^{-1}\sqrt{1+V_t}\right)(1+V_t)}.$$

证明。 不失一般性，设 $\sigma=1$。对于任意 $\lambda\in\mathbb{R}$，令

$$w_t(\lambda):=\exp\left(\lambda S_t-\frac{1}{2}\lambda^2V_t\right).$$

由引理 3，我们注意到，对于任意 $\lambda\in\mathbb{R}$，都有 $\mathbb{E}\{w_t(\lambda)\}\leq 1$。现在令 Λ 为一个 $(0,1)$ 随机变量，且与所有其他变量相互独立。显然，$\mathbb{E}\{w_t(\Lambda)|\Lambda\}\leq 1$。定义

$$w_t:=\mathbb{E}\{w_t(\Lambda)|v_n,\eta_n\colon\ n\in\mathbb{N}\}.$$

于是，由于 $\mathbb{E}\{w_t\}=\mathbb{E}\{\mathbb{E}\{w_t(\Lambda)|v_n,\eta_n\colon\ \ n\in\mathbb{N}\}\}=\mathbb{E}\{w_t(\Lambda)\}=\mathbb{E}\{\mathbb{E}\{w_t(\Lambda)|\Lambda\}\}\leq 1$，我们也可以将 w_t 直接表示为

$$\begin{aligned}w_t&=\frac{1}{\sqrt{2\pi}}\int\exp\left(\lambda S_t-\frac{1}{2}\lambda^2V_t\right)\exp\left(\frac{-\lambda^2}{2}\right)d\lambda\\&=\frac{1}{\sqrt{2\pi}}\int\exp\left(-\frac{1}{2}(V_t+1)\lambda^2+S_t\lambda\right)d\lambda,\end{aligned}$$

进一步可得

$$\begin{aligned} w_t &= \frac{1}{\sqrt{2\pi}} \int \exp\left(-\frac{1}{2}\frac{(\lambda - \frac{1}{1+V_t}S_t)^2}{(1+V_t)^{-1}}\right) \exp\left(\frac{1}{2}\frac{S_t^2}{1+V_t}\right) d\lambda \\ &= \frac{1}{(1+V_t)^{1/2}} \exp\left(\frac{1}{2}\frac{S_t^2}{1+V_t}\right) \int \frac{1}{(1+V_t)^{-1/2}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\lambda - \frac{1}{1+V_t}S_t}{(1+V_t)^{-1/2}}\right)^2\right] d\lambda \\ &= \frac{1}{(1+V_t)^{1/2}} \exp\left(\frac{1}{2}\frac{S_t^2}{1+V_t}\right). \end{aligned}$$

因此， $\mathbb{P}\{\delta\, w_t \geq 1\}$ 等于

$$\begin{aligned} \mathbb{P}\left\{\frac{\delta}{(1+V_t)^{1/2}} \exp\left(\frac{1}{2}\frac{S_t^2}{1+V_t}\right) \geq 1\right\} &= \mathbb{P}\left\{\exp\left(\frac{1}{2}\frac{S_t^2}{1+V_t}\right) \geq (1+V_t)^{1/2}\frac{1}{\delta}\right\} \\ &= \mathbb{P}\left\{\frac{S_t^2}{1+V_t} \geq 2\log\left[(1+V_t)^{1/2}\frac{1}{\delta}\right]\right\} \\ &= \mathbb{P}\left\{S_t^2 \geq 2\log\left[\frac{\sqrt{1+V_t}}{\delta}\right](1+V_t)\right\}. \end{aligned}$$

现在回想一下， $\mathbb{E}\{w_t\} \leq 1$。因此，根据马尔可夫不等式，

$$\mathbb{P}\{\delta\, w_t \geq 1\} \leq \delta\, \mathbb{E}\{w_t\} \leq \delta,$$

至此，证明完成。 □

引理 3. 设 $\{v_t\colon\ t \in \mathbb{N}\}$ 和 $\{\eta_t\colon\ t \in \mathbb{N}\}$ 如引理 *2* 所述。对于任意 $\lambda \in \mathbb{R}$，定义

$$w_t(\lambda) := \exp\left(\sum_{n=1}^{t} \frac{\lambda \eta_n v_n}{\sigma} - \frac{1}{2}\lambda^2 v_n^2\right). \tag{14}$$

然后， $\mathbb{E}\{w_t(\lambda)\} \leq 1$。

证明。令

$$D_n := \exp\left(\frac{\lambda \eta_n v_n}{\sigma} - \frac{1}{2}\lambda^2 v_n^2\right).$$

显然， $w_t(\lambda) = D_1 D_2 \ldots D_t$。注意， $\mathbb{E}\{D_n|\eta_1, \ldots, \eta_{n-1}, v_1, \ldots, v_n\}$ 等于

$$\begin{aligned} &\mathbb{E}\left\{\exp\left(\frac{\lambda \eta_n v_n}{\sigma}\right) / \exp\left(\frac{1}{2}v_n^2\lambda^2\right) \middle| \eta_1, \ldots, \eta_{n-1}, v_1, \ldots, v_n\right\} \\ &= \mathbb{E}\left\{\exp\left(\frac{\lambda \eta_n v_n}{\sigma}\right) \middle| \eta_1, \ldots, \eta_{n-1}, v_1, \ldots, v_n\right\} \Big/ \exp\left(\frac{1}{2}v_n^2\lambda^2\right) \end{aligned}$$

因此，由于（13），

$$\mathbb{E}\{D_n|\eta_1, \ldots, \eta_{n-1}, v_1, \ldots, v_n\} \leq \exp\left(\frac{(\lambda v_n/\sigma)^2\sigma^2}{2}\right) \Big/ \exp\left(\frac{1}{2}v_n^2\lambda^2\right) = 1.$$

接下来，对于每个 $t \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}\{w_t(\lambda)|\eta_1, \ldots, \eta_{t-1}, v_1, \ldots, v_t\} &= \mathbb{E}\{D_1 \cdots D_{t-1} D_t|\eta_1, \ldots, \eta_{t-1}, v_1, \ldots, v_t\} \\ &= D_1 \cdots D_{t-1}\mathbb{E}\{D_t|\eta_1, \ldots \eta_{t-1}, v_1, \ldots v_t\} \leq w_{t-1}(\lambda). \end{aligned}$$

因此， $\mathbb{E}\{w_t(\lambda)\} = \mathbb{E}\{\mathbb{E}\{w_t(\lambda)|\eta_1, \ldots \eta_{t-1}, v_1, \ldots v_t\}\}$ 可由 $\mathbb{E}\{w_{t-1}(\lambda)\} \leq \cdots \leq \mathbb{E}\{w_1(\lambda)\} = \mathbb{E}\{\mathbb{E}\{D_1|v_1\}\} \leq 1$ 上界，从而完成证明。 □

参考文献

[1] Yasin Abbasi-Yadkori、D´avid P´al 和 Csaba Szepesv´ari。基于自归一化过程的在线最小二乘估计：多臂老虎机问题中的一个应用。arXiv 预印本 arXiv:1102.2670，2011 年。[2] A. Bertrand 和 M. Moonen。自组织无线传感器网络中基于一致性的分布式总体最小二乘估计。《IEEE 信号处理汇刊》，59(5):2320–2330，2011 年。[3] N. Bogdanovi´c、J. Plata-Chaves 和 K. Berberidis。用于节点特定自适应参数估计的分布式增量式最小均方。《IEEE 信号处理汇刊》，62(20):5382–5397，2014 年。[4] F.S. Cattivelli 和 A.H. Sayed。用于分布式估计的扩散式最小均方策略。《IEEE 信号处理汇刊》，58(3):1035–1048，2009 年。[5] S. Dean、B. 和 Recht。基于感知的确定性等价控制。载于《动力学与控制学习》，第 399–411 页，2021 年。[6] F. Guo 、 C. Wen 、 J. Mao 和 Y.-D. Song。随机风电智能电网的分布式经济调度。《IEEE 智能电网汇刊》，7(3):1572–1583，2015 年。[7] Shaoming He、Hyo-Sang Shin、Shuoyuan Xu 和 Antonios Tsourdos。低成本传感器网络上的分布式估计：最新进展综述。《信息融合》，54:21–43，2020 年。[8] B. Jiang、B. Ravindran 和 H. Cho。传感器网络中用于节能目标跟踪的基于概率的预测与休眠调度。《IEEE 移动计算汇刊》，12(4):735–747，2012 年。[9] C.G. Lopes 和 A.H. Sayed。分布式网络上的增量自适应策略。《IEEE 信号处理汇刊》，55(8):4064–4077，2007 年。[10] C.G. Lopes 和 A.H. Sayed。自适应网络上的扩散式最小均方：公式化与性能分析。《IEEE 信号处理汇刊》，56(7):3122–3136，2008 年。˘[11] L. Martinovi´c、Z. Ze˘cevi´c 和 B. Krstaji´c。多个领导者单积分器多智能体系统的协同跟踪控制。《欧洲控制杂志》，63:232–239，2022 年。[12] S. Modalavalasa、U.K. Sahoo、A.K. Sahoo 和 S. Baraha。无线传感器网络中鲁棒分布式估计策略综述。《信号处理》，188:108150，2021 年。[13] L. Paradis 和 Q. Han。无线传感器网络故障管理综述。《网络与系统管理杂志》，15(2):171–190，2007 年。[14] M.G. Rabbat 和 R.D. Nowak。用于分布式优化的量化增量算法。《IEEE 通信选定领域杂志》，23(4):798–808，2005 年。[15] K.R. Ramaswamy、G. Bottegal 和 P.M.J. Van den Hof。利用正则化核方法学习动态网络中的线性模块。《自动化》，129:109591，2021 年。[16] A.H. Sayed 和 C.G. Lopes。自适应网络上的分布式处理。载于《第九届信号处理及其应用国际研讨会》，第 1–3 页，2007 年。[17] A.H. Sayed 和 X. Zhao。异步自适应网络。载于《协同与图信号处理》，第 3–68 页。学术出版社，2018 年。[18] I.D. Schizas、G. Mateos 和 G.B. Giannakis。用于基于一致性的网络内自适应处理的分布式最小均方。《IEEE 信号处理汇刊》，57(6):2365–2382，2009 年。[19] S.-Y. Tu 和 A.H. Sayed。对于自适应网络上的分布式估计，扩散策略优于一致性策略。《IEEE 信号处理汇刊》，60(12):6217–6234，2012 年。[20] M.A.M. Vieira、C.N. Coelho、D.C. da Silva 和 J.M. da Mata。无线传感器网络设备综述。载于《2003 年 IEEE 新兴技术与工厂自动化会议》（EFTA 2003），第 1 卷，第 537–544 页，2003 年。[21] M.J. Wainwright。《高维统计：非渐近视角》，第 48 卷。剑桥大学出版社，2019 年。[22] W. Yang、G. Chen、X. Wang 和 L. Shi。传感器网络上分布式状态估计的随机传感器激活。《自动化》，50(8):2070–2076，2014 年。[23] W.-A. Zhang、S. Liu 和 L. Yu。具有非均匀估计速率的传感器网络融合估计。《IEEE 电路与系统汇刊 I：正规论文》，61(5):1485–1498，2014 年。