面向流形流数据高效降维的稳定滤波

Nicholas P. Bertrand,[1+]Eva Yezerets,[2+]Han Lun Yap,[3]Adam S. Charles,[2†]Christopher J. Rozell[1†*]

[1]佐治亚理工学院电气与计算机工程系，North Ave，亚特兰大，GA 30332，美国[2]约翰斯·霍普金斯大学生物医学工程系，3400 N. Charles Street，巴尔的摩，MD 21218，美国[3]新加坡国防科技研究院，20 Science Park Dr，新加坡，118230

+共同第一作者，†共同资深作者
*通讯请联系：crozell@gatech.edu。

arXiv:2601.08685v1 [eess.SP] 13 Jan 2026

如今，科学与工程领域的许多方向都可以利用相关技术，快速采集海量数据。尽管这些数据集对于理解从物理系统到生物系统和社会系统的各种现象至关重要，但数据量之大使得即便是简单的存储、传输和基本处理也面临巨大挑战。为高效、准确地执行这些数据处理任务，我们需要新的降维工具，它们应当：1）无需耗费大量资源和时间进行训练；2）保留数据的内在几何结构，其中包含理解被测系统所需的信息。具体而言，需要保留的几何结构源于这样一个事实：在许多应用中，流式高维数据沿着低维吸引子流形演化。重要的是，我们可能事先并不知道该流形的确切结构先验地。为应对这些挑战，我们提出随机滤波（RF），它利用随机降维的一种特定实现，经证明能够在嵌入空间中保留非线性流形结构，同时不依赖于数据且计算高效。在这项工作中，我们以随机降维丰富的理论前景为基础，将 RF 发展为一种切实可行的方法。我们介绍了新的方法、分析和实验验证，以阐明 RF 在各种科学应用中的实用性，其中包括若干模拟数据和真实数据示例，展示 RF 带来的实际益处。

1 引言

过去几十年的技术发展极大地加快了众多应用领域的数据采集速度。从新型光学显微镜 [35, 49, 55] 和相机 [51] 到转录组学 [12]和社交媒体 [45], , 科学家和分析人员如今面临着数据洪流：将原始观测转化为知识的瓶颈，已经从数据采集转移到了数据的遥测、存储和处理。例如，为了深入理解人脑，神经科学家正投入大量资源开发空间分辨率和时间分辨率不断提高的新型神经记录技术 [12, 27, 35, 49, 53, 55]。如果此类技术最终能够同时记录人脑中每个神经元的活动，那么仅以每个神经元每秒记录一比特这一最低标准计算，所产生的数据速率也将达到约100 Gbps。事实上，利用现有的材料和制造技术 [30]，实现这一目标或许已并非遥不可及。将这些数据从系统中提取出来并转换为数字格式所需的带宽需求将极其庞大，因为即使在当前，高维探针也已将数据读出的能力推向极限 [27]。以灵活且高效的方式降低此类数据的维度，对于推动这类重要的数据密集型应用至关重要。此外，为实现无损组织成像，此类记录硬件必须受到严格的功耗限制，从而进一步压缩了机载数据压缩的选择空间。

压缩高吞吐量流数据需要采用高效（在功率受限的环境中计算速度快）、与数据无关（不需要训练数据）且通用（能够适应不断变化的数据统计特性和多样化任务）的技术。从经典的主成分分析 [23] 到现代的变分自编码器 [29, 43], , 大多数压缩方法都依赖数据，因此存在若干缺点。首先，必须先借助训练数据计算降维映射，才能压缩新数据。这限制了这些方法在缺乏此类训练数据的流式和在线应用中的适用性。随着数据规模扩大，学习映射也变得愈发困难，因此人们持续关注重新训练、微调和基础模型。其次，当底层统计特性偏离训练数据集时，这类方法不适用于非平稳数据。最后，数据采集往往与发现过程紧密交织，实践者并不预先知道哪些数据特征会很重要，因此难以确定应保留哪些特征。尽管在线主成分分析 [8] 可以弥补其中一些不足，但它会在在线

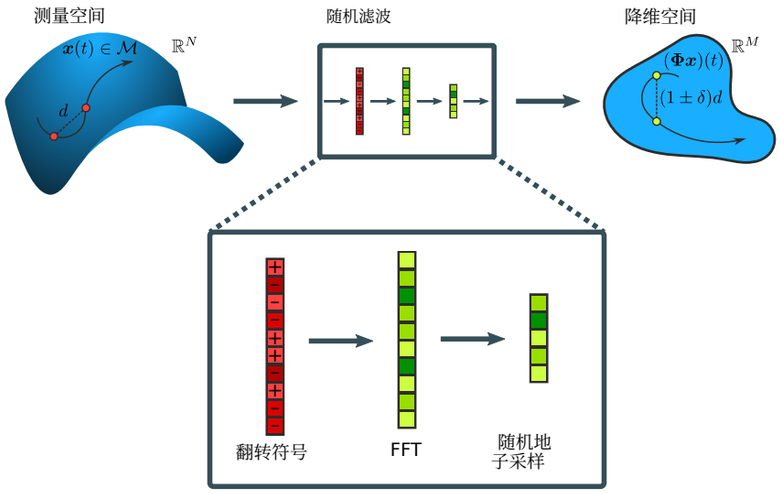

图 1：随机滤波（RF）的示意图，包含三个步骤：1. 将输入向量的符号随机化；2. 对结果计算快速傅里叶变换；3. 随机子采样傅里叶系数。RF 将位于 $\mathbb{R}^N$ 中的 aD 维流形上的点映射到降维空间 $\mathbb{R}^M$，其中 $M < N$。我们的理论保证，当 M 足够大时，该映射是稳定嵌入，即成对距离近似保持不变。由于主要计算步骤涉及快速傅里叶变换，该算法的复杂度为 $\mathcal{O}(N \log\ N)$。

训练过程中产生时间延迟，其时间复杂度相对于精度参数呈立方增长，而且仅适用于线性子空间模型。此外，强调最大化解释方差，可能与分类或回归等其他分析目标不一致；这些分析可能依赖于数据中较小但显著的变化 [26]。因此，在压缩之后还需要用于分析的数据时，仅保留方差最大的维度往往并不足够。

我们并非旨在保留方差解释度最高的维度，而是希望保留许多高维数据集中存在的底层低维结构。这一侧重点源于一个事实：底层非线性几何往往体现了科学家和分析人员希望研究的数据属性。这类结构可能以多种方式产生，例如数据源自某个底层动力系统，或由数据的空间属性/组成属性所导致的复杂物理约束和相关性 [14, 18, 21, 32]。因此，正如图像压缩需要保留感知特性，科学数据压缩也需要保留这些几何特性。

为提供这样一种方法，我们提出随机滤波（RF），这是一种基于随机投影的在线数据压缩替代方案 [9]。RF 具有计算效率高、通用且与数据无关等特点，因此能够有效地用于流式高维数据的降维。此外，压缩感知这一丰富领域及其后续的随机化线性代数研究表明，随机投影能够超越主成分分析式压缩，保留数据几何 [9, 54]。具体而言，经过适当设计的随机投影能够保留多种结构中的数据几何，包括受限集合、子空间并集、流形和动力系统吸引子。传统方法研究的是稀疏点集（即大部分元素为零），而更一般的低维流形结构也同样能够得到保留 [54]。然而，过去的研究主要停留在理论层面，所采用的投影在实际系统中并不可行（例如随机高斯投影），或依赖于可能不适用于现实应用的参数和常数（例如关于流形性质的假设）。我们证明，RF 能够克服这些挑战，超越理论基础，成为一种实用工具，可用于压缩来自多种科学应用领域的动力系统测量值。我们首先介绍 RF 的实用算法实现，该算法可在线运行且复杂度低；随后表明，许多真实数据集都满足相应条件，能够使用 RF 进行精确压缩；最后给出若干更为详尽的案例研究，涉及流体动力学和神经元成像等领域。这些示例展示了 RF 作为通用降维工具的价值，满足了科学数据集持续增长所带来的迫切需求。

2 结果

RF 嵌入保持数据几何

在许多应用中，我们会观测到动力系统沿着低维吸引子流形 $\boldsymbol{x}_t \in \mathcal{M} \subset \mathbb{R}^N$演化时产生的流式测量 $\boldsymbol{x}_t \in \mathbb{R}^N$for $t = 1,\ 2,\ \ldots\ldots$。将随机滤波应用于 $\boldsymbol{x}_t$ 包括三个步骤（图 1）：随机翻转每个测量值 $x_{it}^{\pm} = x_{it}$的符号，计算符号翻转后结果 $\boldsymbol{x}_t^{\pm}$的离散傅里叶变换（DFT），最后对傅里叶系数 [1]进行随机下采样。由于 DFT 是算法核心，快速傅里叶变换（FFT）可实现快速高效的计算，仅需 $\mathcal{O}(N \log\ N)$ 次运算（如果数据也是稀疏的，速度甚至更快 [22]）。此外，滤波器的参数只有二进制符号翻转和一组采样的傅里叶系数。因此，滤波器可以独立于数据生成，并存储在 $\mathcal{O}(N)$ 内存中；与主成分分析中的完整线性映射相比，它更加紧凑，而且效率高得多

[1]读者可能会发现，这些计算步骤与快速 Johnson-Lindenstrauss（JL）变换 [1] 中的步骤相同。值得注意的关键区别是，随机滤波的理论保证基于一种强大得多的流形数据模型，该模型适用于流式测量；而快速 JL 变换使用的是有限点云模型。

于高亮　自编码器模型中参数过多的编码器　（自编码器）。

随机滤波的优势源于这样一个事实：许多观测数据采集自一个 D 维流形，该流形位于 N 维环境空间中，其中 $N \gg D$。直观而言，流形可视为二维曲面从三维空间向更高维空间中嵌入的超曲面的一种推广。事实证明，基于流形的建模能够有效描述状态取决于少数参数且具有非线性关系的系统，例如不同角度下物体的图像 [17]或神经元群体活动 [21]。随机滤波的核心前提是：如果所关注的动力系统在一个维度为 $D \ll N$ 的吸引子流形 $\mathcal{M}$ 上演化，那么系统状态可以直接压缩至一个随 D 而非随 N 增长的维度。

随机滤波的理论基础是稳定嵌入的概念；也就是说，这类功能能够保留输入空间中点对之间的距离。保留距离是保留数据内在特征的关键条件。回归、分类等关键分析都依赖于对点之间相对距离的分析，而任何无法保留这些距离的映射都会造成失真，进而影响后续分析。一般而言，完全等距——即将高 N 维空间映射到低 M 维空间时保留任意两点之间的距离——是不可能的。然而，鉴于流形模型，该映射无需保留所有距离，只需保留感兴趣点之间的距离；也就是说，只需保留高维空间中受限 于低维流形上的点之间的距离。因此，更有针对性的条件是受限等距性质 （RIP）。

F通常，我们称映射 $\mathbf{\Phi}$ 是 $\mathcal{M}$ 的稳定嵌入，如果对于每一对点 $\boldsymbol{x}$，$\boldsymbol{y} \in \mathcal{M}$，我们都有 t
满足以下
的 RIP 条件

$$1 - \delta \leq \frac{\|\mathbf{\Phi}\boldsymbol{x} - \mathbf{\Phi}\boldsymbol{y}\|_2^2}{\|\boldsymbol{x} - \boldsymbol{y}\|_2^2} \leq 1 + \delta. \tag{1}$$

在此条件中，RIP常数 δ是一个介于 0 和 1 之间的值，在确定映射 $\mathbf{\Phi}$对点之间距离的保真程度方面起着关键作用。当 $\delta \to 0$时，映射会发生塌缩：相距较远的点（$\|\boldsymbol{x} - \boldsymbol{y}\|_2^2$ 较大）可能被映射到彼此任意接近的位置，从而变得几乎无法区分。反之，当 $\delta \to 1$时，关注点之间的距离会得到精确保留，这意味着高维空间中的任何功能都可以在低维空间中精确复现。请注意，此定义与 RIP 的典型用法略有不同；在典型用法中，流形是所有稀疏信号的集合（例如 [10, 13]），而此处描述的是任意低维流形这一更一般的情况。

这类稳定嵌入能够有效抵御测量噪声；在压缩过程中，测量噪声可能会严重破坏测量值。通过界定由广义受限等距性质衡量的失真，可以保证只要测量值的数量 M 满足

$$M \geq \frac{K_1}{\delta^2}\left(D\log\left(\frac{K_2 N}{\delta}\right) + \log(K_3)\right)\log^4(N), \tag{2}$$

其中，K_i 是取决于流形性质和失效概率[2][3] [54]的常数。请注意，所需测量值的数量随 D 线性增长（直观上可将其理解为流形所编码的“信息”量），而仅随环境维数 N对数增长。再次注意其对 δ的关键依赖：映射越忠实，分母就越接近 1，M 的下界也随之降低。

将随机投影用于降维，源自 Johnson 和 Lindenstrauss [25], 的开创性工作；他们证明了存在一种随机投影算子，能够稳定嵌入任意点云。后续研究将这一理论扩展到可高效应用的随机嵌入 [2] ，并证明了适用于更一般流形的嵌入结果 [5]。我们的理论统一了这些结果，提供了一种基于随机投影的高效嵌入方法，并将其推广至流形数据模型。对于由潜在低维动力系统支配的高维数据所构成的流式科学测量，将流形模型的推广与高效计算相结合，对于实现无模型在线压缩至关重要。

尽管我们的理论结果保证，当满足条件 (2) 时，RF 是一种稳定嵌入，但我们通常无法获得该表达式中涉及的流形参数。因此，实际应用中需要多少测量值可能并不明确。给定代表性数据，我们可以实证地探究等距常数[4] δ如何随压缩比 N/M变化。值得注意的是，尽管没有利用关于各类数据具体细节的先验知识，RF 在各种合成数据集和真实数据集上都能得到较低的 δ 估计值（图 2 和补充文本）。作为比较，低通滤波（LPF）是一种降维方法，采用了我们的

[2]我们注意到，通过在滤波器输出端应用适当的后处理步骤，所需的测量值数量还可以进一步减少，低于定理 1 给出的界 [3,40]。不过，为了使表述简洁，我们采用 [54] 中描述的原始流程。[3]**定理 1** 设 $\mathcal{M}$ 是 $\mathbb{R}^N$ 中一个紧致的 D 维黎曼子流形，其测地正则性为 R，体积为 V，条件数为 τ^{-1}。设 $F \in \mathbb{C}^{M\times N}$ 为一个子采样傅里叶矩阵，其各行从 $N \times N$ *DFT* 矩阵中均匀选取；设 D_ξ 为一个对角拉德马赫矩阵，即 D_ξ 的对角元素以概率 $1/2$ 取值为 ± 1。如果

$$M \geq \frac{C_1}{\delta_{\mathcal{M}}^2}\left(D\log\left(\frac{RN}{\tau\delta_{\mathcal{M}}}\right) + \log\left(\frac{V}{\rho}\right)\right)\log^4(N)\log\left(\rho^{-1}\right), \tag{3}$$

那么，以大于 $1 - C_2\rho$ 的概率，$\Phi = FD_\xi$ 稳定嵌入 $\mathcal{M}$，其等距常数为 $\delta_{\mathcal{M}}$。定理中不随任何其他量变化的普适常数用 C_1 和 C_2 表示。

[4]对于给定的数据集，会在输入空间和输出空间中穷举计算成对距离，并使用失真最严重的距离来计算等距常数 δ。δ 的值会在随机化滤波参数的不同实例上取平均值。

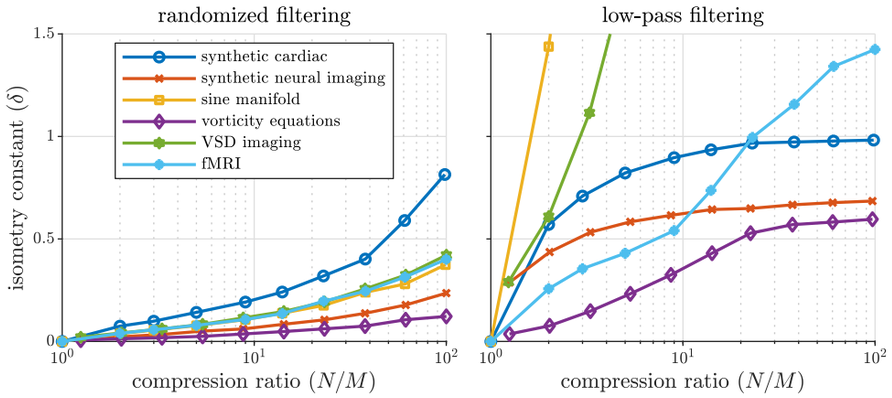

图2：等距常数 δ 量化嵌入对输入空间几何结构的保持效果。 $\delta = 0$ 对应于完美嵌入，此时所有输入点对之间的距离都等于它们在降维空间中的对应距离。图中展示了若干合成数据集（心脏模型 [20], 图3中使用的神经成像数据、正弦流形 [19], 以及图5所示的涡度方程解）和真实数据集（来自啮齿动物实验的电压敏感染料成像 [57]以及功能性磁共振成像 [41]）的等距常数估计值。与缺乏类似稳定性保证的低通滤波（右）相比，随机滤波（左）只需少得多的测量值即可达到更低的等距常数。

所需的高效性、通用性和与数据无关性——具有更高的等距常数，因此无法保留后续分析所必需的数据几何。

为进一步评估随机滤波在实践中的适用性，我们探究其几何保持特性如何应用于真实世界的数据处理任务：从钙成像推断神经活动，以及根据涡量测量进行流动分类。

RF 嵌入保留神经元钙成像中的神经活动

在不久的将来，快速在线压缩对于遏制数据洪流至关重要，而一个迅速发展的应用领域是高密度神经记录，例如光学功能显微镜 [16, 48, 53] 和高密度电极设计 [27]。尽管捕捉神经活动的诸多细微变化离不开这些技术，但其极高的分辨率（也因此具有极高的比特率）限制了对这些数据集进行高效处理的能力，也使得从接受慢性植入、自由活动的受试对象身上无线传输数据变得困难。在此，我们探讨其中一种记录技术——钙成像。这既是一种常用的神经影像技术，也正在迅速扩展至更大的神经体积，并实现更高的保真度 [31, 49, 55]。

在钙成像中（图 3a），荧光蛋白 [4, 56] 会被导入目标神经元（或其他细胞，例如星形胶质细胞 [28]）中。蛋白质在细胞内表达后，在与钙离子结合之前会保持惰性。结合钙离子后，可采用光学显微镜技术（例如多光子、光片或宽场技术）对这些蛋白质进行成像。由此得到一系列荧光图像（或体积），这些图像必须经过处理，才能提取单个荧光单元（例如神经元）[6]的活动时间轨迹。尽管这些技术的总体比特率可能极高，但人们真正关注的底层活动实际上维度要低得多。事实上，行秩等简单模型在事后去噪 [42], 和源提取 [39], 方面已十分有效，甚至已被纳入成像流程，以实现快速图像采集 [53]。在此，我们转而探讨能否随着数据到来，使用随机滤波（RF）进行在线压缩，以及能否利用随机滤波（RF）的几何保持特性，直接从压缩数据中提取功能单元的活动。

为测试 RF 对此类数据集的适用性，我们生成合成钙成像数据，旨在直接从降维测量中恢复神经活动 [5]（图 3b）。模拟数据能够完全控制数据特征，从而可以在特定参数范围内评估 RF。记 $\boldsymbol{A} \in \mathbb{R}^{N \times N_c}$为一个矩阵，其列对应于矢量化的细胞轮廓，即每一列中的非零值是空间中对应于一个细胞的像素。同样，令 $\boldsymbol{S} \in \mathbb{R}^{N_t \times N_c}$ 为一个矩阵，其列包含每个细胞的时间轮廓，即由钙浓度变化引起的荧光变化。根据这些定义，我们可以将数据矩阵建模为

$$\boldsymbol{D} = \boldsymbol{A}\boldsymbol{S}^T + \sigma\boldsymbol{\epsilon}, \tag{4}$$

其中， $\boldsymbol{D}$ 的第 t^{th} 列表示时间 t 测得的荧光帧，而 $\sigma\boldsymbol{\epsilon}$ 表示标准差为 σ 的加性白高斯噪声。因此，压缩数据可以表示为对每一帧独立应用 RF 算子$\boldsymbol{\Phi}$，这是一项快速操作，可以在获取帧的同时完成。因此，降维数据为 $\widetilde{\boldsymbol{D}} = \boldsymbol{\Phi}\boldsymbol{D} = \boldsymbol{\Phi}\boldsymbol{A}\boldsymbol{S}^T = \widetilde{\boldsymbol{A}}\boldsymbol{S}^T$。

在没有噪声和细胞重叠的情况下，细胞 i （记作 $\boldsymbol{t}_i$）的时间序列可以通过数据矩阵与其空间细胞轮廓的内

[5]数据的生成过程首先是在 $W \times H$ 像素场景中放置 N_c 个圆形神经元细胞轮廓。每个神经元都有固定概率与之前生成的神经元重叠。时间动态通过在 N_t 时间采样点上对泊松尖峰过程进行卷积生成，随后再与高斯核进行卷积，以近似钙成像中观测到的时间轮廓。

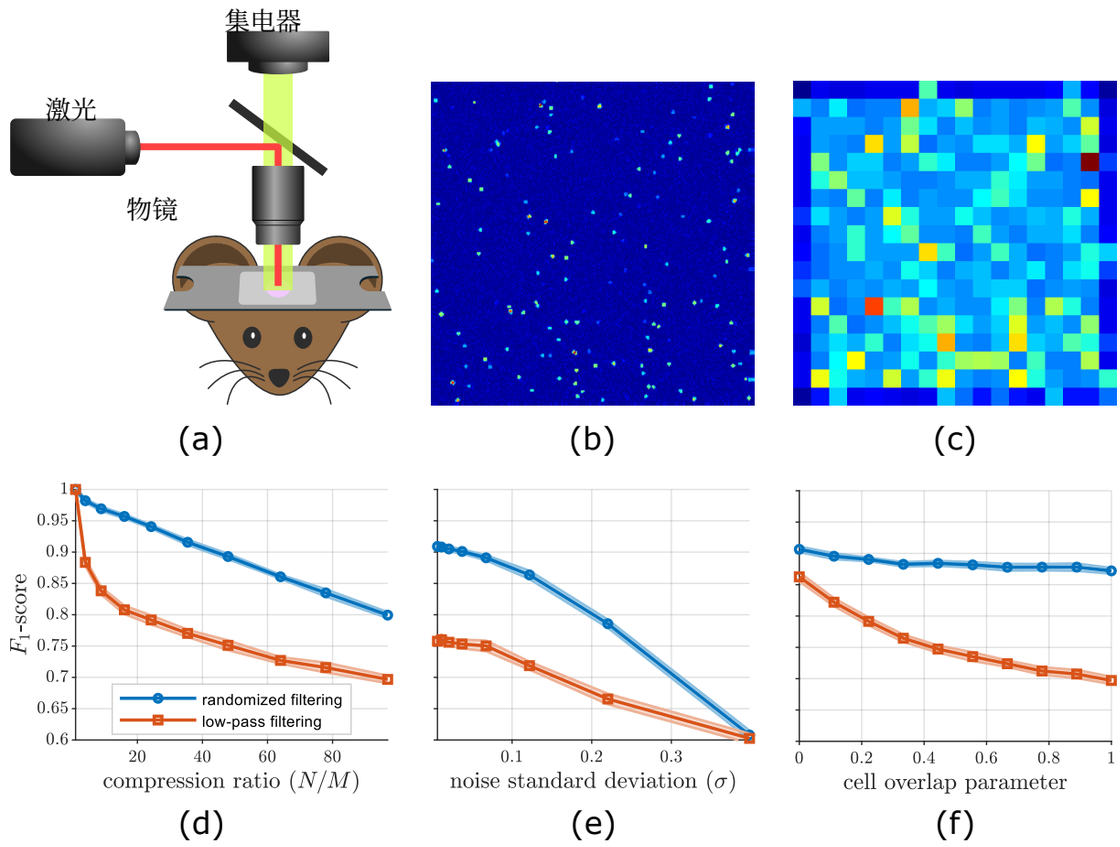

图3：**(a)** 钙成像利用荧光钙指示剂测量大脑活动。**(b)** 模拟钙成像数据的示例帧。**(c)** 在低通滤波（LPF）中，每一帧都会在每个空间维度上模糊并下采样 15 倍，以此作为一种粗略的降维方式。图像的精细细节会丢失，尤其是在包含重叠细胞的区域。**(d-f)** 合成钙成像数据中事件检测的性能比较。通过对方程（6）的输出进行阈值处理来估计事件，并以 F_1-分数作为性能指标（越高越好）。**(d)** 即使经过大幅压缩，随机滤波（RF）仍能取得良好效果。**(e)** 在噪声水平适中时，随机滤波（RF）的表现优于低通滤波（LPF）。**(f)** 随着重叠细胞数量增加，低通滤波（LPF）无法区分相邻或重叠细胞之间的活动。相比之下，随机滤波（RF）能够分离重叠程度很高的细胞活动。

积恢复：[6]

$$\widehat{\boldsymbol{s}}_i = \boldsymbol{D}^T \boldsymbol{a}_i / \|\boldsymbol{a}_i\|_2^2, \tag{5}$$

其中，$\boldsymbol{a}_i$ 是 i^{th} 细胞的空间轮廓。尽管我们无法为原始数据的完全恢复提供理论保证，但仍可在降维空间 [15]中估计感兴趣的量。具体而言，在温和的假设条件下，稳定嵌入也能保留内积，因此我们可以近似得到未压缩的估计值：

$$\widehat{\boldsymbol{s}}_i = \Re\left\{(\boldsymbol{\Phi}\boldsymbol{D})^T \boldsymbol{\Phi}\boldsymbol{a}_i\right\} / \|\boldsymbol{\Phi}\boldsymbol{a}_i\|_2^2 = \Re\left\{\widetilde{\boldsymbol{D}}^T \widetilde{\boldsymbol{a}}_i\right\} / \|\widetilde{\boldsymbol{a}}_i\|_2^2, \tag{6}$$

其中，$\boldsymbol{\Phi}$ 是 RF 算子（补充文本，[7]）。降维空间中的这一估计量与原始空间中的估计量类似，都是内积，但前者将降维算子同时应用于数据和细胞轮廓。使用公式(6)中的 RF，我们恢复了每个模拟源在压缩前后的时间轨迹。作为比较，我们采用了同样的流程，使用目前主要的替代方法之一：在线、参数化且无需学习的低通滤波（LPF）（图 3c）。

为了检验恢复轨迹的质量，我们比较了恢复时间轨迹中事件的能力。这些事件也称为“瞬态”，表现为荧光的急剧上升，随后缓慢衰减回基线，代表构成神经放电的动作电位或动作电位爆发。在 RF压缩的情况下，我们仅使用一部分原始测量值，就能够准确恢复这些事件（图 3）。为对此进行量化，我们计算了F1分数，其定义为

$$F_1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}, \tag{7}$$

其中，精确率 表示真阳性数占所有检测结果的比例，而召回率 表示真阳性数占所有实际事件的比例。检测结果的 F_1-分数始终高于 0.9

[6]在高斯噪声存在的情况下，可以证明这是最大似然估计量。细胞重叠也可以视为一种额外的噪声来源。[7]由于测量值是复值的，我们需要取所示内积的实部。该估计量的完整推导见补充材料。或者，也可以考虑一种高效的实值RIP矩阵，并采用 [54] 中的方法，得到与定理 1 类似的保证。例如，基于子采样离散余弦变换的算子是实值、高效且RIP最优的，不过其构造比本文所介绍的方法稍复杂 [3]。

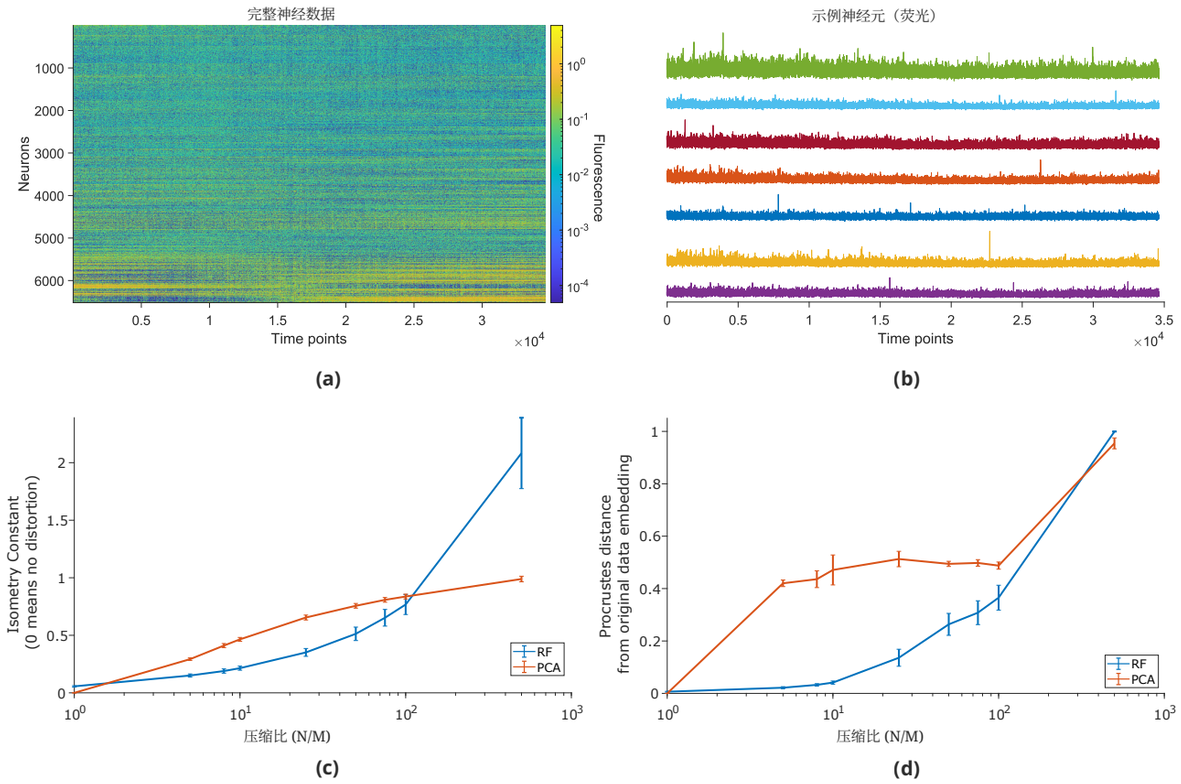

图4：**(a)** 小鼠神经荧光数据：来自 6519 个神经元的记录，共 34600 个时间点，采样频率为 9.6 赫兹 [35, 36]。**(b)**从数据中选取的示例神经元。**(c)** 根据等距常数 δ 的测量结果，在一系列压缩比下，随机滤波比主成分分析更能保留小鼠脑数据的成对距离。主成分分析结果取 100 个样本的平均值（无放回抽样，使用 80% 的数据；误差线：标准差）。随机滤波结果取 10 个样本的平均值，每个样本均通过 10 个随机种子运行随机滤波。**(d)** 与主成分分析压缩相比，RF压缩能更好地保留数据整体的嵌入流形形状。我们将原始数据（6519 维）中的 34600 个时间点，以及 RF 和主成分分析压缩后的数据，分别使用局部线性嵌入将每个数据矩阵嵌入到 4 个嵌入维度中。我们使用数据嵌入与 RF 和主成分分析压缩数据的 4 维嵌入之间的普罗克鲁斯特斯距离，比较 4 维空间中的流形形状（横轴为压缩比）。普罗克鲁斯特斯距离为 0 表示两个嵌入的形状没有差异。在压缩比约达 100 倍（65 维）时，随机滤波比主成分分析更能保留整体的 4 维流形形状。

在压缩比最高达 40X 时（图 3d）。低通滤波的表现差得多，在压缩比 < 5X 时，其 F_1 很快降至 0.9 以下（图 3d）。类似地，随着噪声水平升高（图 3e）或空间重叠神经元数量增加（图 3f），检测问题变得更加复杂，而随机滤波在检测结果的 F_1 分数方面表现要好得多。

RF 嵌入保留了神经动力学的几何结构

尽管神经成像是神经科学中一种重要的大数据形式，但能够记录数万个神经元的尖端技术，即使在提取神经活动轨迹后，也会产生庞大的数据集。这些数据集本身具有复杂的内部几何结构，通常称为“神经流形”。虽然人们已经创建了许多模型和《方法》，试图根据数据估计此类潜在流形 [44],，但我们认为，流形的存在意味着数据本质上是可压缩的，而且仅凭压缩数据，就能像凭神经元空间数据一样出色地存储和恢复该流形。

为了验证 RF 保持流形结构的能力，我们考察了来自小鼠脑的真实神经钙成像数据（图 4a、b） [36] 。该数据由一段约一小时的记录组成（34,600 个时间点，采样频率为 9.6 赫兹），涵盖 6,519 个独立神经元（图 4a、b）。将原始数据 $\boldsymbol{S}$（神经元×时间）记为……，我们观察到，使用 RF 压缩~数据 $\boldsymbol{S}_{RF} = \boldsymbol{\Phi}\boldsymbol{S}$ 后，等距常数 δ 更低，即使与数据驱动的主成分分析压缩相比也是如此。具体而言，在压缩比为 5 至 100 时，可以观察到显著改善（图 4c）。压缩超过 100 倍后，两种方法的等距常数都接近 1。

虽然等距常数提供了一种衡量距离保持程度的方法，但我们旨在通过后续流形学习方法的视角，明确检验 RF 保持流形的能力：局部线性嵌入（LLE）。LLE 是一种经典方法，用于根据高维数据识别低维流形坐标。我们希望确保通过 LLE 学到的参数化在 RF 压缩后仍得以保留，因此分别对原始数据、RF 压缩数据和 PCA 压缩数据运行了 LLE（图 4d）。对于这些数据中的每一种，LLE 都为每个时间点生成了一个 4D 表示。为了衡量各个压缩数据版本的 LLE 输出与原始数据的 LLE 输出之间的接近程度，我们计算了普罗克鲁斯特斯距离。该距离衡量的是两个点云在经过仿射变换（旋转和缩放）后之间的差异。从数学上讲，如果 $\boldsymbol{L}$、$\boldsymbol{L}_{RF}$ 和 $\boldsymbol{L}_{PCA}$ 分别是原始数据时间轨迹 $\boldsymbol{S}$、RF

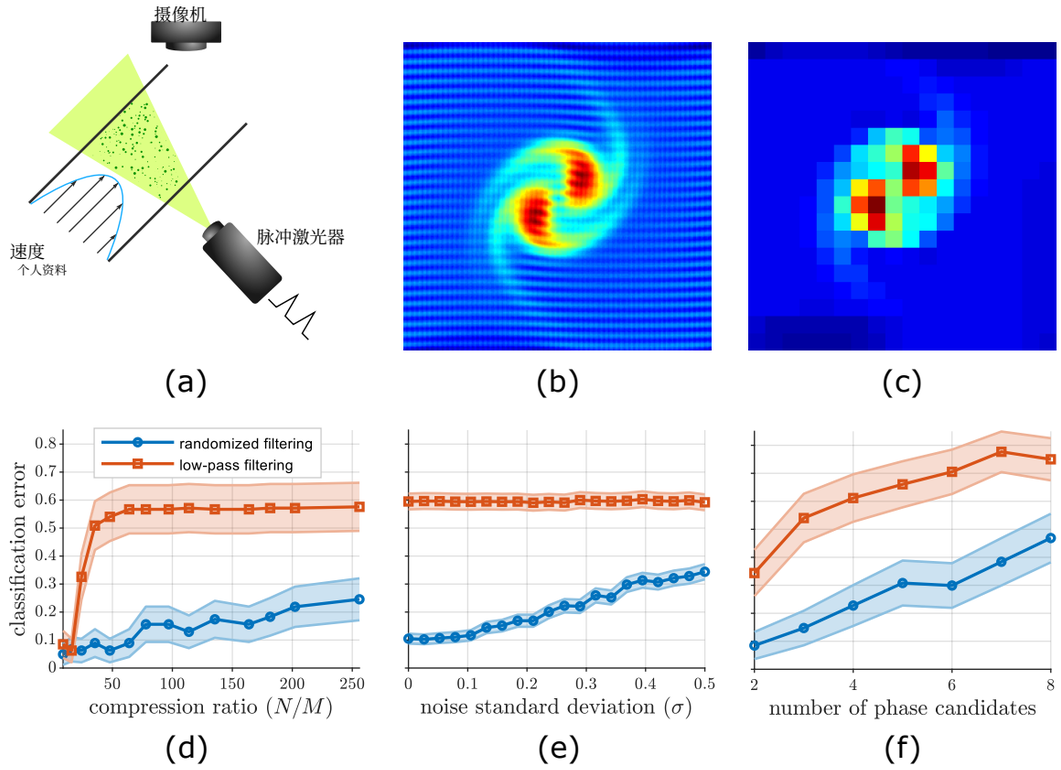

图5：**(a)** 涡量是描述流体局部旋转趋势的量，可通过粒子图像测速等实验方法测得，如图所示。**(b)** 涡度方程（9）解的示例帧。**(c)** 通过低通滤波和降采样，将每个维度的维度降低 15 倍的示例。正如预期，高频成分已消失。**(d-f)** 涡度方程解的相位分类任务表现。该任务是根据解的压缩测量值，判断生成该解时所使用的强迫函数相位。**(d)** 即使压缩达到 2 到 3 个数量级，随机滤波仍能实现较低的分类错误率。**(e)** 在有噪声的条件下，随机滤波的相位分类表现更优；而低通滤波会完全丢失分类所需的信息。**(f)** 即使通过从更大的候选集合中抽取相位使问题更难，随机滤波仍比低通滤波更常实现正确分类。

~ ~压缩数据 $\boldsymbol{S}_{RF}$ 和 PCA 压缩数据 $\boldsymbol{S}_{PCA}$ 的 LLE 嵌入，则普罗克鲁斯特斯距离为

$$d_{proc}(\boldsymbol{L}, \boldsymbol{L}_{PCA\ or\ RF}) = \min_{\boldsymbol{G}} \|\boldsymbol{L} - \boldsymbol{G}\boldsymbol{L}_{PCA\ or\ RF}\|_F^2, \tag{8}$$

其中，$\boldsymbol{G}$ 是定义仿射变换的矩阵。压缩前后流形之间较小的普罗克鲁斯特斯距离，表明数据几何的一致性更高。事实上，我们发现，在压缩比为 5 到 100 时，与主成分分析相比，随机滤波得到的结果与原始数据嵌入之间的普罗克鲁斯特斯距离始终更小（图 4d）。此外，主成分分析与随机滤波在局部线性嵌入距离上的差距要大得多，这表明，即使等距常数仅有细微提升，在更复杂的分析中也会对数据几何的保留产生深远影响。总而言之，等距常数和普罗克鲁斯特斯距离均表明，与主成分分析相比，随机滤波能更好地保留复杂的流形结构。

可从 RF 嵌入中解码涡旋动力学

第二个高维复杂动力学数据的例子，来自对非线性流体流动行为的研究，例如湍流——这是流体力学中的一种基本现象，且因难以建模而闻名——该研究通过对涡量场进行光学测量来开展。涡量量化了流体局部发生旋转运动的趋势；对于二维不可压缩流，可将其描述为满足以下偏微分方程组的标量 ω 场[8] [34]

$$\begin{aligned} \partial_t \omega + \left(\nabla^\perp \psi\right) \cdot \nabla \omega &= \gamma \Delta \omega + f \\ \Delta \psi &= \omega. \end{aligned} \tag{9}$$

受 [46]，中所展示的模拟启发，我们考虑两个涡旋斑块在高频驱动函数 f 的作用下相互作用，该函数定义为

$$f(x,y) = 0.075 \frac{\sin(64x+\phi) + \sin(32y+\phi)}{1 + 0.25(\cos(128x+\phi) + \cos(64y+\phi))}. \tag{10}$$

[8]T标量涡量场为函数 $\omega(x,\ y,\ t)$，流函数为 $\psi(x,\ y,\ t)$，γ 为运动黏度，$f(x,\ y)$ 为强迫函数 g 其中，$\nabla^\perp = (-\partial_y,\ \partial_x)^T$ 是垂直梯度算子，$\Delta = \partial_{xx} + \partial_{yy}$ 是拉普拉斯算子。

我们使用由方程（10）描述的强迫函数生成了方程（9）的解，其中未知相位ϕ从一组相位候选值$\{\phi_1, \ldots, \phi_P\}$中均匀随机选取。随后，我们进行了分类，以确定哪个相位生成了观测到的涡量场。该分类过程仅选择这样一个相位：它使给定时刻的涡量场与使用每个可能相位生成的强迫函数之间的距离最小，

$$\widehat{\phi} = \underset{\phi_i}{\operatorname{argmin}} \|\mathbf{\Phi}\omega - \mathbf{\Phi} f_{\phi_i}\|_2^2, \tag{11}$$

其中，$f_{\phi_i}(x,\ y)$ 是相位 ϕ_i 产生的强迫函数（参见补充文本）。RF 能够捕获强迫项的高频成分（图 5b），因此与使用 LPF 进行同等降维相比，RF 能获得更准确的相位估计（图 5c）。具体而言，当压缩比高于 $\approx$ 10X 时，RF 始终表现良好，分类误差为 < 0.2；而 LPF 的分类误差则迅速升高，在压缩比达到 40X 时就已达到 > 0.5。与钙成像示例（图 3）类似，RF 对噪声水平（图 5e）和任务复杂度（图 5f）的变化均保持稳定。

3 讨论

本文提出了随机滤波（RF）：一种通用、计算高效且理论可靠的降维方法。我们着重探讨如何在处理大规模流数据的最初阶段快速降低其维度，同时保留有关数据流的关键信息。我们在多个应用中展示了 RF 的优势和性能，包括神经影像学、高维神经元群体活动和涡量估计。综合来看，这些应用展现了 RF 通用降维的强大能力。尽管数据集差异极大，所采用的降维过程却完全相同。该过程无需了解数据特征或压缩后任务，因此在尚不清楚目标特征为何的探索性实验中尤其有用。

一项关键的概念性进展，是将重点放在保留低维数据流形上。压缩始终旨在保留特定的“重要”特性。例如，JPEG 等图像压缩技术 [24]和 JPEG2000 [37] 旨在保留经人类视觉感知评估的图像质量。然而，对于许多其他数据集，究竟应保留哪些关键因素，事先 并不明确。流形假设提供了一种通用的质量标准，它已有充分确立且可进行定量评估；正如我们通过等距性计算以及分类、事件检测和流形估计中的补充分析所展示的那样。

使用任何降维技术时，一个关键参数是嵌入维度的大小，也就是压缩比。我们注意到，在特定应用中首次使用随机滤波时，选择压缩比这一单一参数所需的数据，远少于数据依赖方法所需的数据；后者需要经过完整训练才能刻画内在结构。此外，这种初始调参只需进行一次；此后，只要适用相同的底层流形模型，所得数值便可用于各种实验条件。随着随机滤波的应用日益广泛，压缩比的选择可能会越来越多地依赖经验。

我们通过评估 RF 在实际滤波任务中的表现，探究了它的潜力，但它的适用范围绝不仅限于此。由于 RF 保留了观测数据所依托的流形几何，因此它与依赖这种结构的其他技术兼容，例如回归、分类和流形学习方法 [50], , 包括局部线性嵌入 [44], ; 我们在神经活动分析示例中展示了这一点。事实上，在更广泛的矩阵分解和时间轨迹分析领域 [11, 38], , 随机投影已被用于加速计算 [7], ; 随着这些数据集的规模不断增长，RF 将成为进一步加速数据分析的关键助力。在成像应用中，甚至可以通过透镜以光学方式实现 RF 来计算 DFT，从而彻底消除计算成本。

开发新的降维工具对于确保海量数据成为福音而非负担至关重要。本文展示了一种简单而强大的降维工具，其适用范围广泛。从确定性方案转向随机化方案，代表着我们对降维方式的一种根本性转变；但随着数据量持续激增，这或许是我们保持竞争力的唯一途径。

4 《方法》

内积的保持

本节介绍一项关于内积保持的结果，该结果为神经影像学模拟中未压缩估计量的近似提供了理论依据。具体而言，在假设 $\mathbf{\Phi}$ 是 $\mathcal{M} \cup -\mathcal{M}$ 的稳定嵌入（即 $\mathbf{\Phi}$ 也能稳定嵌入该流形关于原点的镜像）的前提下，我们推导了一个界，用于约束在降维空间与原始空间中计算的内积之间的偏差。我们注意到，由于 [15, Theorem 4] 隐含地假设 $\mathbf{\Phi}$ 也是 $\mathcal{M}$ 上每个点方向上的单位向量的稳定嵌入，因此该定理中的界不能直接应用。为此，我们提出以下定理。

定理 2. 假设 ℓ、$s \in \mathcal{M}$，且 Φ 是 $(\mathcal{M} \cup -\mathcal{M})$ 的 δ-稳定嵌入，则

$$|\langle \Phi\ell, \Phi s\rangle - \langle \ell, s\rangle| \leq \frac{\delta}{2}(\|\ell\|_2^2 + \|s\|_2^2).$$

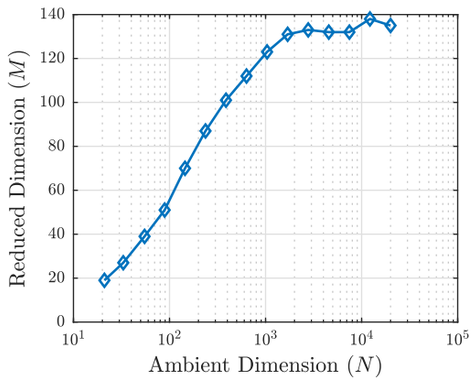

图6：在固定等距常数 $\delta = 1/3$ 下，RF 与正弦流形的缩放行为。

证明. 由于 Φ 是 $(\mathcal{M} \cup -\mathcal{M})$ 的稳定嵌入，因此我们有

$$1 - \delta \leq \frac{\|\Phi\ell \pm \Phi s\|^2}{\|\ell \pm s\|^2} \leq 1 + \delta,$$

或者等价地（通过展开分母中的范数）

$$1 - \delta \leq \frac{\|\Phi\ell \pm \Phi s\|^2}{\|\ell\|^2 + \|s\|^2 \pm 2\langle \ell, s\rangle} \leq 1 + \delta.$$

根据平行四边形法则，

$$\begin{aligned}
\langle \Phi\ell, \Phi s\rangle &= \frac{\|\Phi\ell + \Phi s\|^2 - \|\Phi\ell - \Phi s\|^2}{4} \\
&\leq \frac{(1+\delta)(\|\ell\|^2 + \|s\|^2 + 2\langle \ell, s\rangle) - (1-\delta)(\|\ell\|^2 + \|s\|^2 - 2\langle \ell, s\rangle)}{4} \\
&= \langle \ell, s\rangle + \frac{\delta}{2}(\|\ell\|^2 + \|s\|^2),
\end{aligned}$$

类似地，我们可以证明

$$\langle \Phi\ell, \Phi s\rangle \geq \langle \ell, s\rangle - \frac{\delta}{2}(\|\ell\|^2 + \|s\|^2).$$

综合这两个不等式，即可得到所需结果

$$-\frac{\delta}{2}(\|\ell\|^2 + \|s\|^2) \leq \langle \Phi\ell, \Phi s\rangle - \langle \ell, s\rangle \leq \frac{\delta}{2}(\|\ell\|^2 + \|s\|^2).$$

□

注。 如果我们进一步假设 $x/\|x\|_2 \in \mathcal{M}\ \forall x \in \mathcal{M}$，那么就可以得到 *[15, 定理 4]* 中的结果，该定理主要讨论稀疏信号模型，而不是本文所考虑的流形模型。

通过正弦流形验证缩放行为

当 M 大于一个随环境维数 N 呈对数增长的量时，RF 是输入流形的稳定嵌入。我们通过计算在达到固定等距常数 δ 时所需的 M 值来验证这一事实。作为参考，我们设定 $\delta = 1/3$，这是相关流形恢复结果 [47] 中使用的阈值。此模拟的输入数据是由公式 (18) 描述的正弦流形。对于 $N < 1000$ 的取值，缩放行为如预期般呈对数增长。对于更大的 N，降维空间所需的维数增长得更为缓慢。

复值测量下的估计：实值噪声

令 $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{s}, \sigma^2 \boldsymbol{I}_N)$，且 $\boldsymbol{Z} = \boldsymbol{\Phi}\boldsymbol{X}$，其中 $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$。假设我们希望描述与 $\boldsymbol{Z}$ 相关的密度函数，以便计算其中一个模型参数的最大似然估计（MLE）。由于 $\boldsymbol{Z}$ 的均值和协方差矩阵分别为

$$\mathbb{E}[\boldsymbol{Z}] = \mathbb{E}[\boldsymbol{\Phi}\boldsymbol{X}] = \boldsymbol{\Phi}\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\Phi}\boldsymbol{s},$$

and

$$\begin{aligned}
\mathrm{Cov}[\boldsymbol{Z}, \boldsymbol{Z}] &= \mathbb{E}[(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])^*] \\
&= \mathbb{E}[(\boldsymbol{\Phi}\boldsymbol{X} - \mathbb{E}[\boldsymbol{\Phi}\boldsymbol{X}])(\boldsymbol{\Phi}\boldsymbol{X} - \mathbb{E}[\boldsymbol{\Phi}\boldsymbol{X}])^*] \\
&= \boldsymbol{\Phi}\mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^*]\boldsymbol{\Phi}^* \\
&= \boldsymbol{\Phi}\mathrm{Cov}[\boldsymbol{X}, \boldsymbol{X}]\boldsymbol{\Phi}^* \\
&= \boldsymbol{\Phi}\boldsymbol{\Phi}^*,
\end{aligned} \tag{12}$$

因此，使用多元高斯分布的著名公式来表示该分布或许很诱人，即

$$f_{\boldsymbol{Z}}(\boldsymbol{z} \mid \boldsymbol{s}) = (2\pi)^{-n/2} \det(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{z}-\boldsymbol{\Phi}\boldsymbol{s})^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}(\boldsymbol{z}-\boldsymbol{\Phi}\boldsymbol{s})\right\}, \tag{13}$$

其中， $\boldsymbol{\Phi}^*$ 表示矩阵 $\boldsymbol{\Phi}$的共轭转置。不幸的是，仅凭均值和协方差矩阵，无法充分确定 $\boldsymbol{Z}$的分布。为说明这一点，我们不妨考虑通过典范同构从 $\mathbb{C}^m \to \mathbb{R}^{2m}$ 得到的实值随机向量，其定义如下：

$$\boldsymbol{a} + i\boldsymbol{b} \mapsto \begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{bmatrix}.$$

在我们的问题中，可以通过测量矩阵生成这个向量：

$$\boldsymbol{A} \coloneqq \begin{bmatrix}\boldsymbol{\Phi}_R\\ \boldsymbol{\Phi}_I\end{bmatrix},$$

其中， $\boldsymbol{\Phi}_R = \Re\{\boldsymbol{\Phi}\}$ 和 $\boldsymbol{\Phi}_I = \Im\{\boldsymbol{\Phi}\}$ 分别表示 $\boldsymbol{\Phi}$的实部和虚部。将得到的测量值记为 $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$，根据对实高斯向量的分析，我们知道 $\boldsymbol{Y}$的分布由均值 $\boldsymbol{A}\boldsymbol{s}$ 和协方差矩阵 $\boldsymbol{A}\boldsymbol{A}^T$确定。因此，似然函数为[9]

$$f_{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{s}) = (2\pi)^{-n/2} \det(\boldsymbol{A}\boldsymbol{A}^T)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{s})^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{y}_t-\boldsymbol{A}\boldsymbol{s})\right\}.$$

为理解为何不能使用似然函数（13），我们将说明均值和协方差矩阵不足以描述 $\boldsymbol{Y}$的分布。首先，我们注意到协方差矩阵 $\boldsymbol{A}\boldsymbol{A}^T$可以写成分块形式：

$$\boldsymbol{A}\boldsymbol{A}^T = \begin{bmatrix}\boldsymbol{\Phi}_R\boldsymbol{\Phi}_R^T & \boldsymbol{\Phi}_R\boldsymbol{\Phi}_I^T\\ \boldsymbol{\Phi}_I\boldsymbol{\Phi}_R^T & \boldsymbol{\Phi}_I\boldsymbol{\Phi}_I^T.\end{bmatrix}$$

相比之下，与 $\boldsymbol{y}$ 相关的协方差矩阵为：

$$\boldsymbol{\Phi}\boldsymbol{\Phi}^* = (\boldsymbol{\Phi}_R\boldsymbol{\Phi}_R^T + \boldsymbol{\Phi}_I\boldsymbol{\Phi}_I^T) + i(\boldsymbol{\Phi}_I\boldsymbol{\Phi}_R^T - \boldsymbol{\Phi}_R\boldsymbol{\Phi}_I^T).$$

由于 $\boldsymbol{\Phi}\boldsymbol{\Phi}^*$ 只包含 $\boldsymbol{A}\boldsymbol{A}^T$各分块的和，因此无法仅根据 $\boldsymbol{Z}$ 的协方差矩阵推断出 $\boldsymbol{Y}$的协方差矩阵，进而也无法推断出 $\boldsymbol{Z}$的分布。

要确定 $\boldsymbol{Z}$的分布，我们还需要一个称为伪协方差矩阵的量，其定义为 [52]

$$\mathbb{E}\left[\boldsymbol{Z}\boldsymbol{Z}^T\right] = \boldsymbol{\Phi}\boldsymbol{\Phi}^T = (\boldsymbol{\Phi}_R\boldsymbol{\Phi}_R^T - \boldsymbol{\Phi}_I\boldsymbol{\Phi}_I^T) + i(\boldsymbol{\Phi}_I\boldsymbol{\Phi}_R^T + \boldsymbol{\Phi}_R\boldsymbol{\Phi}_I^T). \tag{14}$$

The 利用 $\boldsymbol{Z}$ 的协方差矩阵和伪协方差矩阵，可以计算出 $\boldsymbol{Y}$的完整协方差矩阵。因此， e mea协方差和伪协方差可以完全确定 $\boldsymbol{Z}$ 的分布。

如果噪声的伪协方差矩阵为零，则它服从一种特殊类型的复高斯分布，称为圆对称高斯分布 [33]。此时，式 (14) 得出

$$\boldsymbol{\Phi}_R\boldsymbol{\Phi}_R^T = \boldsymbol{\Phi}_I\boldsymbol{\Phi}_I^T, \quad \boldsymbol{\Phi}_I\boldsymbol{\Phi}_R^T = -\boldsymbol{\Phi}_R\boldsymbol{\Phi}_I^T.$$

$\boldsymbol{Y}$的协方差矩阵随之变为

$$\boldsymbol{A}\boldsymbol{A}^T = \begin{bmatrix}\boldsymbol{\Phi}_R\boldsymbol{\Phi}_R^T & \boldsymbol{\Phi}_R\boldsymbol{\Phi}_I^T\\ -\boldsymbol{\Phi}_R\boldsymbol{\Phi}_I^T & \boldsymbol{\Phi}_R\boldsymbol{\Phi}_R^T\end{bmatrix},$$

且每个块都可以利用 $\boldsymbol{Z}$ 的协方差矩阵计算，该矩阵为

$$\boldsymbol{\Phi}\boldsymbol{\Phi}^* = 2\boldsymbol{\Phi}_R\boldsymbol{\Phi}_R^T - 2i\boldsymbol{\Phi}_R\boldsymbol{\Phi}_I^T.$$

换言之，均值和协方差矩阵足以确定圆对称高斯向量的分布。此外，可以证明，相应的密度函数具有式 (13) 的形式。在本文所讨论的情形中，如正文所述，噪声并非圆对称，因此我们使用 $\boldsymbol{Y}$ 进行分析。不过，假设圆对称复噪声模型是对实噪声模型的近似，可以简化分析，并建立与未压缩空间中估计量的联系，因此我们也会在后续章节中考虑这种噪声模型。

复值测量下的估计：复值噪声

现在我们来考虑上一节问题的一种变体，其中假设噪声为复值噪声。我们用 $\mathcal{N}_{\mathbb{C}}(\boldsymbol{0},\ \boldsymbol{K})$ 表示协方差矩阵为 $\boldsymbol{K}$ 的圆对称高斯随机向量的分布。于是，我们的信号模型可表示为 $\boldsymbol{X} \sim \boldsymbol{s} + \boldsymbol{\epsilon}$，其中 $\boldsymbol{\epsilon} \sim \mathcal{N}_{\mathbb{C}}\left(0,\ \sigma^2\boldsymbol{I}_N\right)$。与之前一样，我们通过复测量算子 $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ 观测数据，即 $\boldsymbol{Z} = \boldsymbol{\Phi}\boldsymbol{X}$，并希望估计模型的一个参数。与上一节一样， $\boldsymbol{Z}$ 的协方差矩阵为 $\boldsymbol{\Phi}\boldsymbol{\Phi}^*$，但在这种情况下，由于噪声是圆对称的， $\boldsymbol{s}$ 和协方差矩阵足以完全确定密度函数：

$$f_{\boldsymbol{Z}}(\boldsymbol{z} \mid \boldsymbol{s}) = \frac{1}{\pi^n\sigma^2 \det(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)} \exp\left\{-\frac{1}{\sigma^2}(\boldsymbol{z}-\boldsymbol{\Phi}\boldsymbol{s})^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}(\boldsymbol{z}_t-\boldsymbol{\Phi}\boldsymbol{s})\right\}. \tag{15}$$

我们将在后续章节中看到，在这种噪声模型和随机滤波矩阵 $\boldsymbol{\Phi}$ 下，估计量尤其容易化简。

[9]我们在此假设 $\boldsymbol{A}\boldsymbol{A}^T$ 是非奇异的。在奇异情形下，可以考虑 $\boldsymbol{Y}$中的一个元素子集，使相应的协方差矩阵为非奇异矩阵。或者，我们可以分别用伪行列式和伪逆替换行列式和逆。

应用：基于降维测量的时间序列估计

现在我们推导式 (5) 中给出的估计量。假设 $\boldsymbol{s}$ 是由向量化二值掩码表示的空间细胞轮廓（即，如果像素 i 包含感兴趣的细胞，则 $\boldsymbol{s}[i]=1$ ；否则为 0 ）。令 $\boldsymbol{x}_t=\mu_t\boldsymbol{s}+\sigma\boldsymbol{\epsilon}$ 表示时间 t 的含噪图像，其均值为由我们希望估计的时变参数 μ_t 调制的细胞模板。在原始空间中， $\boldsymbol{y}_t$ 的分布函数因此为

$$f_{\boldsymbol{X}}(x_t\mid\mu_t,\boldsymbol{s})=(2\pi)^{-n/2}\prod_i\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{x}_t-\mu_t\boldsymbol{s})^T(\boldsymbol{x}_t-\mu_t\boldsymbol{s})\right\},$$

令对数似然关于 μ_t 的导数等于零，可得

$$\widehat{\mu}_t=\frac{\boldsymbol{x}^T\boldsymbol{s}}{\|\boldsymbol{s}\|^2},$$

这就是估计量 (5)。

现在，假设我们已知降维空间中的测量值 $\boldsymbol{z}_t=\boldsymbol{\Phi}\boldsymbol{x}_t$。定义等效测量矩阵 $\boldsymbol{A}=[\boldsymbol{A}_R;\ \boldsymbol{A}_I]$，并令 $\boldsymbol{y}_t=\boldsymbol{A}\boldsymbol{x}_t$, 则分布函数为

$$f_{\boldsymbol{Y}}(\boldsymbol{y}_t\mid\mu_t,\boldsymbol{s})=C\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}_t-\mu_t\boldsymbol{A}\boldsymbol{s})^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{y}_t-\mu_t\boldsymbol{A}\boldsymbol{s})\right\},$$

where $C=(2\pi)^{-n/2}\det(\boldsymbol{A}\boldsymbol{A}^T)^{-1/2}$。令对数似然的导数等于零，可得以下 g esti估量:

$$\widehat{\mu}_t=\frac{\boldsymbol{y}_t^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\boldsymbol{s}}{(\boldsymbol{A}\boldsymbol{s})^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\boldsymbol{s}}. \tag{16}$$

因此，该估计量是在白化矩阵 $(\boldsymbol{A}\boldsymbol{A}^T)^{-1}$ 后，降维空间中的测量值与细胞模板之间的内积。

虽然式（16）是针对 μ_t 在技术上正确的最大似然估计，但考虑圆对称复噪声模型下的估计量也很有启发意义。在这种情况下，压缩测量值 $\boldsymbol{z}$ 的分布为

$$\boldsymbol{z}_t\sim\mathcal{N}_{\mathbb{C}}\left(\mu_t\boldsymbol{\Phi}\boldsymbol{s},\sigma^2\boldsymbol{\Phi}\boldsymbol{\Phi}^*\right),$$

根据式（15），我们得到分布函数

$$f_{\boldsymbol{Z}}(\boldsymbol{z}_t\mid\mu_t,\boldsymbol{s})=\frac{1}{\pi^n\det(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)}\exp\left\{-\frac{1}{\sigma^2}(\boldsymbol{z}_t-\mu_t\boldsymbol{\Phi}\boldsymbol{s})^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}(\boldsymbol{z}_t-\mu_t\boldsymbol{\Phi}\boldsymbol{s})\right\}.$$

最大化似然等价于最小化指数项中的二次型，我们通过计算其导数来实现:

$$\begin{aligned}
&\frac{\partial Q}{\partial\mu_t}=\frac{\partial}{\partial\mu_t}\left((\boldsymbol{z}_t-\mu_t\boldsymbol{\Phi}\boldsymbol{s})^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}(\boldsymbol{z}_t-\mu_t\boldsymbol{\Phi}\boldsymbol{s})\right)\\
=&\frac{\partial}{\partial\mu_t}\left(\boldsymbol{z}_t^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{z}_t-\mu_t\boldsymbol{z}_t^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s}-\mu_t(\boldsymbol{\Phi}\boldsymbol{s})^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{z}_t+\mu_t^2(\boldsymbol{\Phi}\boldsymbol{s})^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s}\right)\\
=&\frac{\partial}{\partial\mu_t}\left(\boldsymbol{z}_t^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{z}_t-\mu_t\boldsymbol{z}_t^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s}-\mu_t(\boldsymbol{z}_t^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s})^*+\mu_t^2(\boldsymbol{\Phi}\boldsymbol{s})^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s}\right)\\
=&\frac{\partial}{\partial\mu_t}\left(\boldsymbol{z}_t^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{z}_t-2\mu_t\Re\left\{\boldsymbol{z}_t^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s}\right\}+\mu_t^2(\boldsymbol{\Phi}\boldsymbol{s})^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s}\right)\\
=&-2\Re\left\{\boldsymbol{z}_t^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s}\right\}+2\mu_t(\boldsymbol{\Phi}\boldsymbol{s})^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s}
\end{aligned}\tag{17}$$

与之前一样，将对数似然关于 μ_t 的导数设为零，可得一般感知矩阵 $\boldsymbol{\Phi}$ 对应的最大似然估计:

$$\widehat{\mu}_t=\frac{\Re\left\{\boldsymbol{z}_t^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s}\right\}}{(\boldsymbol{\Phi}\boldsymbol{s})^*(\boldsymbol{\Phi}\boldsymbol{\Phi}^*)^{-1}\boldsymbol{\Phi}\boldsymbol{s}}.$$

在射频场景中，估计量可大幅简化。注意，随机滤波中 $\boldsymbol{\Phi}$ 的构造可表示为 $\boldsymbol{\Phi}=\boldsymbol{S}\boldsymbol{F}\boldsymbol{D}$，其中 $\boldsymbol{S}$ 是用于从其相乘的矩阵中选取部分行的矩阵，$\boldsymbol{F}$ 是离散傅里叶变换矩阵，而 $\boldsymbol{D}$ 的对角线上包含一个拉德马赫序列。于是有

$$\boldsymbol{\Phi}\boldsymbol{\Phi}^*=\boldsymbol{S}\boldsymbol{F}\boldsymbol{D}\boldsymbol{D}\boldsymbol{F}^*\boldsymbol{S}^T=\boldsymbol{I}_M,$$

其中， $\boldsymbol{I}_M$ 是 $M\times M$ 单位矩阵。因此，估计量可简化为

$$\widehat{\mu}_t=\frac{\Re\left\{\boldsymbol{z}_t^*\boldsymbol{\Phi}\boldsymbol{s}\right\}}{\|\boldsymbol{\Phi}\boldsymbol{s}\|^2},$$

这就是估计量（6）。或者，根据定理 2，我们可以将压缩空间中的估计量视为原始空间中估计量的近似。

应用：基于降维测量的分类

假设我们测量了一幅图像的含噪观测，其均值等于已知候选集合 S 中的一个未知元素 $\boldsymbol{s}^\star$：

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{s}^\star, \sigma^2 \boldsymbol{I}_N), \quad \boldsymbol{s}^\star \in S.$$

密度函数为

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{s}_i) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{s}_i)^T(\boldsymbol{x}-\boldsymbol{s}_i)\right\},$$

我们对 i 进行最大化，得到分类器：

$$\widehat{\boldsymbol{s}} = \underset{\boldsymbol{s}_i \in S}{\operatorname{argmin}} \left\|\boldsymbol{x}-\boldsymbol{s}_i\right\|^2.$$

现在，假设我们通过 RF 算子 $\boldsymbol{\Phi}$观测到降维测量。与之前一样，我们考虑通过相应算子 $\boldsymbol{A}$ 得到的等效测量，其似然函数为

$$f_{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{s}_i) = (2\pi)^{-n/2} \det(\boldsymbol{A}\boldsymbol{A}^T) \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{s}_i)^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{s}_i)\right\},$$

因此，相应的分类器为

$$\widehat{\boldsymbol{s}} = \underset{\boldsymbol{s}_i \in S}{\operatorname{argmin}} (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{s}_i)^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{s}_i).$$

最后，如果我们将噪声近似为圆对称，并注意到 $\boldsymbol{\Phi}\boldsymbol{\Phi}^* = \boldsymbol{I}_M$，则似然函数可表示为

$$f_{\boldsymbol{Z}}(\boldsymbol{z} \mid \boldsymbol{s}_i) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{z}-\boldsymbol{\Phi}\boldsymbol{s}_i)^*(\boldsymbol{z}-\boldsymbol{\Phi}\boldsymbol{s}_i)\right\},$$

从而得到分类器

$$\widehat{\boldsymbol{s}} = \underset{\boldsymbol{s}_i \in S}{\operatorname{argmin}} \left\|\boldsymbol{z}-\boldsymbol{\Phi}\boldsymbol{s}_i\right\|^2.$$

同样，我们可以将其视为原始空间中分类器的一种近似，其误差由等距常数 δ 界定，这是因为 $\boldsymbol{\Phi}$是一个稳定嵌入。

等距常数估计

这里我们概述图2中等距常数 δ 的估计过程。对于给定的数据集 $\boldsymbol{D} \in \mathbb{R}^{N\times K}$，其中包含 K 个维度为 N的样本，我们计算原始空间中的所有成对距离：

$$d_{\mathrm{o}}(i,j) = \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|_2, \quad i < j \le K,$$

以及降维空间中的距离

$$d_{\mathrm{r}}(i,j) = \left\|\boldsymbol{y}_i - \boldsymbol{y}_j\right\|_2, \quad i < j \le K,$$

其中 $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{D}$。接下来，令 $Q(i,\ j)$ 表示降维空间中的成对距离与原始空间中成对距离的比值，即

$$Q(i,j) = \frac{d_{\mathrm{o}}(i,j)}{d_{\mathrm{r}}(i,j)},$$

并定义

$$Q_{\min} = \min_{ij} Q(i,j), \quad Q_{\max} = \max_{ij} Q(i,j), \quad Q_{\mathrm{mean}} = \frac{2}{K(K-1)} \sum_{i<j} Q(i,j).$$

我们将下界和上界等距常数估计为

$$\delta_{\mathrm{lower}} = 1 - \frac{Q_{\min}}{Q_{\mathrm{mean}}}, \quad \delta_{\mathrm{upper}} = \frac{Q_{\max}}{Q_{\mathrm{mean}}} - 1,$$

其中，我们通过平均比率进行归一化，以消除由 $\boldsymbol{\Phi}$ 导致的任何恒定缩放。最后，我们将 δ 取为上述两个值中较差的一个：

$$\delta = \max(\delta_{\mathrm{lower}}, \delta_{\mathrm{upper}}).$$

数据集

正弦流形

The 正弦流形，也称为复指数曲线 [19], , 是 $\mathbb{C}^N$ 中由 b 定义的一维流形 y

$$\boldsymbol{y}(t) = \begin{bmatrix} e^{-i2\pi f_C t} \\ e^{-i2\pi (f_C-1) t} \\ \vdots \\ e^{i2\pi (f_C-1) t} \\ e^{i2\pi f_C t} \end{bmatrix}, \tag{18}$$

其中， f_C 是一个整数，通过 $N = 2f_C + 1$ 定义环境空间的维数。由于正弦流形可以轻松嵌入任意环境维数的空间，且其流形性质（例如可达半径）存在界限，因此正弦流形是分析流形嵌入的一个有用玩具示例 [19]。图 2 展示的是一个 $f_C = 5$，000 的正弦流形，其环境维数为 $N = 10$，001。

心脏模拟

心脏数学建模在开发心血管疾病患者的治疗方案中发挥着关键作用。图2展示了将随机滤波应用于改进版 O'Hara-Virág-Varró-Rudy（OVVR）模型数据后得到的等距常数 [20]。我们使用的模型纳入了真实临床数据，并能忠实再现心力衰竭患者动作电位的关键特征（例如持续时间、振幅和静息膜电位）。此处使用的数据表示三维网格上的电压，该网格包含 $N = 4$，129 个点，覆盖 $K = 12$，500 个时间采样点。

钙成像（模拟）

钙成像是现代神经科学中常用的一种神经记录技术。我们生成的数据模拟了宽视场钙成像记录的一些关键特征，同时还能全面控制感兴趣的参数。我们的数据生成过程由三个部分组成：生成细胞空间轮廓、生成时间事件，以及合成最终数据集。细胞空间轮廓按顺序生成：首先，在随机均匀选取的位置放置一个半径为 r 的二维高斯分布。半径 r 服从均值为 1 像素、标准差为 0.1 的高斯分布。下一个细胞以相同方式生成，但以概率 p_o 选择的位置会使该细胞与前一个细胞重叠。以概率 $1 - p_o$，细胞位置则再次从均匀分布中随机选取。重复这一过程，直至放置 N_c 个细胞，然后将生成的轮廓向量化，并作为矩阵 $\boldsymbol{S}$ 的列存储在 $N \times N_c$ 矩阵中。图3中的结果使用了分布在 256×256 像素视野内的 $N_c = 512$ 个细胞。

为生成时间事件轮廓，我们对每个细胞执行以下步骤：首先，生成一个脉冲序列，使脉冲之间的时间间隔服从指数分布（即泊松过程），但额外限制脉冲之间的最短时间间隔。对标准泊松过程的这一修改模拟了不应期，即刚刚放电的细胞在指定的持续时间内无法再次放电。每个细胞的速率参数均为 1，不应期为 0.3。接下来，我们将生成的脉冲序列与高斯凸起进行卷积，以模拟钙成像记录中常见的弥散时间轮廓。每个细胞生成的时间轮廓都存储在 $N_t \times N_c$ 矩阵 $\boldsymbol{T}$ 中。

最后，通过用相应的时间轮廓调制每个空间轮廓，并通过 $\boldsymbol{D} = \boldsymbol{S}\boldsymbol{T}^T + \sigma\boldsymbol{\epsilon}$ 添加噪声，构建数据矩阵。

小鼠钙成像

Manley 等人 [35, 36] 利用光珠钙成像生成了一组开源的大规模神经数据集。在本文中，我们分析了对一只小鼠右侧大脑皮层半球进行的单区域记录，记录持续了 1 小时（6519 个神经元、34600 个时间点、9.6 赫兹体积速率）。Vaziri 实验室的这项工作以其规模著称：最多可同时记录一百万个神经元。此外，Manley等人 观察到，其中大量数据需要高维表示。在使用 RF 和主成分分析分析这批数据中一个较小的单区域子集时，我们发现，尽管主成分分析即使在较弱的压缩下也会导致显著畸变（就等距常数和普罗克鲁斯特斯距离而言），因而需要高维表示，但 RF 即使在更强的压缩水平下也能避免流形畸变。未来的研究或许可以确定这些 RF 发现能否扩展到数十万个神经元，或者如文中所述，更大的数据集是否本质上具有高维性。

涡度方程

涡量用于量化流体局部旋转的趋势。描述涡量的方程可以由质量守恒和动量守恒定律推导得到，也可以由纳维-斯托克斯方程推导得到。这里我们考虑浅层流体流动，将问题设置在二维空间而非三维空间中即可对其进行很好的近似。对于图5中的结果，我们指定周期边界条件，并设定一个由两个相邻涡旋斑块构成的初始条件，其定义为

$$\omega_0(x, y) = e^{-17.5(x-0.45)^2-0.7y^2} + e^{-17.5(x+0.45)^2-0.7y^2}.$$

我们通过谱方法求解涡度方程（9）。对于定义在具有周期边界条件的矩形区域上的问题，谱方法尤其适用，能够快速获得精度超越所有代数阶的解。

首先，我们将（9）中的顶部方程改写为

$$\partial_t \omega = \gamma \Delta \omega + f - (\nabla^\perp \psi) \cdot \nabla \omega.$$

对两边先关于 x，再关于 y进行傅里叶变换，可得

$$\begin{aligned} \mathcal{F}\{\partial_t \omega\} = & \ \mathcal{F}\left\{\gamma \Delta \omega + f - (\nabla^\perp \psi) \cdot \nabla \omega\right\} \\ = & \ -\gamma \mathcal{F}\{\omega\}(k_x^2 + k_y^2) + \mathcal{F}\left\{f(\nabla^\perp \psi) \cdot \nabla \omega\right\}, \end{aligned} \tag{19}$$

wher其中，k_x 和 k_y 分别是 x 和 y 上的波数。类似地，对于（9）中的底部方程，我们有

$$\mathcal{F}\{\omega\} = \frac{\mathcal{F}\{\psi\}}{k_x^2 + k_y^2}.$$

We t然后使用 MATLAB 的 ode45 常微分方程求解器计算 $\mathcal{F}\{\omega\}$，并通过逆变换恢复 ω e 换 形式。

功能性磁共振成像

功能性磁共振成像通过将血流作为神经活动的代理指标来测量大脑。图2显示了公开提供的增强版纳撒尼尔·克莱因研究所－罗克兰样本数据集 [41] 的等距常数。我们使用第30–33切片，这些切片在时间上包含900个样本，共计24,964个体素。

电压敏感染料

电压敏感染料（VSD）是一种成像技术，可用于记录多个皮层柱的激活，以支持科学应用。图2展示了文献 [57] 中所报告数据集的等距常数。该研究使用VSD记录麻醉大鼠的初级感觉皮层，以研究胡须通路对刺激的适应。该数据集包含多个实验试次的记录，共有3,000个图像样本，每幅图像包含22,632个像素。

致谢

EY和ASC部分获得了美国国家科学基金会职业奖（编号2340338）的资助。NPB和CJR获得了美国国家科学基金会资助（编号CCF-1409422）。HLY和CJR获得了美国国家科学基金会资助（编号CCF-0830456）。CJR获得了詹姆斯·S·麦克唐纳基金会资助（编号220020399）。我们感谢Linwei Wang博士的实验室分享心脏模拟数据，感谢Garrett Stanley博士的实验室分享电压敏感染料数据 [57], ，并感谢Alipasha Vaziri博士的实验室分享开源小鼠数据 [35]。

数据与代码可用性

代码将在发表后于 GitHub 上提供；数据集可按照附录“数据集”部分所引用出版物中的说明获取。

参考文献与注释

[1] Nir Ailon 和 Bernard Chazelle。近似最近邻与快速 Johnson-Lindenstrauss 变换。载于第三十八届 *ACM*计算理论研讨会论文集，STOC ’06，第 557–563 页，美国纽约州纽约市，2006 年。ACM。[2] Nir Ailon 和 Edo Liberty。近乎最优的无约束快速 Johnson-Lindenstrauss 变换。*ACM*算法汇刊，9(3):21:1–21:12，2013 年 6 月。[3] Nir Ailon 和 Holger Rauhut。快速且 RIP最优的变换。离散与计算几何，52(4):780–798，2014 年 12 月。[4] Jasper Akerboom、Tsai-Wen Chen、Trevor J Wardill、Lin Tian、Jonathan S Marvin、Sevinç Mutlu、Nicole Carreras Calderón、Federico Esposti、Bart G Borghuis、Xiaonan Richard Sun 等。用于神经活动成像的 GCaMP钙指示剂优化。《神经科学杂志》，32(40):13819–13840，2012 年。[5] Richard G. Baraniuk 和 Michael B. Wakin。光滑流形的随机投影。《计算数学基础》，9(1):51–77，2009 年 2 月。[6] Hadas Benisty、Alexander Song、Gal Mishne 和 Adam S Charles。神经科学功能性光学显微镜数据处理综述。《神经光子学》，9(4):041402–041402，2022 年。[7] Alejandro Estrada Berlanga、Gabrielle Kang、Amanda Kwok、Thomas Broggini、Jennifer Lawlor、Kishore Kuchibhotla、David Kleinfeld、Gal Mishne 和 Adam S Charles。快速且易于使用的无形态学功能荧光成像分析。*bioRxiv*，第 2025–04 页，2025 年。[8] Christos Boutsidis、Dan Garber、Zohar Karnin 和 Edo Liberty。在线主成分分析。载于第二十六届 *ACM-SIAM*离散算法研讨会论文集，SODA ’15，第 887–901 页，美国宾夕法尼亚州费城，2015 年。工业与应用数学学会。[9] E. J. Candes 和 T. Tao。通过随机投影实现近乎最优的信号恢复：通用编码策略？《*IEEE*信息论汇刊》，52(12):5406–5425，2006 年 12 月。

[10] Emmanuel J Candès 和 Michael B Wakin。《压缩采样导论》。*IEEE*信号处理杂志，25(2):21–30，2008。

[11] Adam S Charles、Nathan Cermak、Rifqi O Affan、Benjamin B Scott、Jackie Schiller 和 Gal Mishne。《Graft：用于功能性神经成像的图滤波时序字典学习》。*IEEE*图像处理汇刊，31:3509–3524，2022。

[12] Adam S Charles, Benjamin Falk, Nicholas Turner, Talmo D Pereira, Daniel Tward, Benjamin D Pedigo, Jaewon Chung, Randal Burns, Satrajit S Ghosh, Justus M Kebschull 等。《迈向社区驱动的大规模开放脑科学：面向结构、功能和遗传学的开放大数据与工具》。《神经科学年度评论》，43(1):441–464，2020年。

[13] Adam S Charles、Dong Yin 和 Christopher J Rozell。《循环网络中多维输入的分布式序列记忆》。《机器学习研究杂志》，18(7):1–37，2017年。[14] John P. Cunningham 和 Byron M. Yu。《大规模神经记录的降维》。《自然·神经科学》，17(11):1500–1509，2014年11月。[15] M. A. Davenport、P. T. Boufounos、M. B. Wakin 和 R. G. Baraniuk。《利用压缩测量进行信号处理》。《IEEE《信号处理精选主题期刊》》，4(2):445–460，2010年4月。[16] Winfried Denk、James H Strickler 和 Watt W Webb。《双光子激光扫描荧光显微镜》。《科学》，248(4951):73–76，1990年。[17] David L. Donoho 和 Carrie Grimes。《与欧几里得空间等距的图像流形》。《数学成像与视觉杂志》，23(1):5–24，2005年7月。[18] Lea Duncker 和 Maneesh Sahani。《流形上的动力学：从神经群体记录中识别计算动力活动》，2021年10月。[19] Armin Eftekhari 和 Michael B. Wakin。《流形嵌入与压缩测量信号恢复的新分析》。《应用与计算调和分析》，39(1):67–109，2015年7月。[20] Mohamed M. Elshrif 和 Elizabeth M. Cherry。《人类心室心脏电生理模型在组织中的行为定量比较》。《PLOS ONE》，9(1):e84401，2014年1月。[21] Juan A. Gallego、Matthew G. Perich、Stephanie N. Naufel、Christian Ethier、Sara A. Solla 和 Lee E. Miller。《保留的神经流形中的皮层群体活动支撑多种运动行为》。《自然·通讯》，9(1):4233，2018年10月。[22] Haitham Hassanieh、Piotr Indyk、Dina Katabi 和 Eric Price。《近乎最优的稀疏傅里叶变换》。载于：第四十四届ACM计算理论研讨会论文集，第563–578页，2012年。[23] H. Hotelling。《将一组统计变量分解为主成分的分析》。《教育心理学杂志》，24(6):417–441，1933年9月。[24] Graham Hudson、Alain Léger、Birger Niss、István Sebestyén 和 Jørgen Vaaben。《JPEG-1标准25年：成功的过去、现在与未来原因》。《电子成像杂志》，27(4):040901–040901，2018年。[25] William B Johnson 和 Joram Lindenstrauss。《向希尔伯特空间扩展Lipschitz映射》。《当代数学》，26(189-206):1，1984年。[26] Ian T Jolliffe。《关于回归中主成分使用的说明》。《皇家统计学会会刊C辑：应用统计》，31(3):300–303，1982年。[27] James J. Jun、Nicholas A. Steinmetz、Joshua H. Siegle、Daniel J. Denman、Marius Bauza、Brian Barbarits、Albert K. Lee、Costas A. Anastassiou、Alexandru Andrei、Çağatay Aydın、Mladen Barbic、Timothy J. Blanche、Vincent Bonin、João Couto、Barundeb Dutta、Sergey L. Gratiy、Diego A. Gutnisky、Michael Häusser、Bill Karsh、Peter Ledochowitsch、Carolina Mora Lopez、Catalin Mitelut、Silke Musa、Michael Okun、Marius Pachitariu、Jan Putzeys、P. Dylan Rich、Cyrille Rossant、Wei-lung Sun、Karel Svoboda、Matteo Carandini、Kenneth D. Harris、Christof Koch、John O’Keefe 和 Timothy D. Harris。《用于高密度神经活动记录的全集成硅探针》。《自然》，551(7679):232–236，2017年11月。[28] Baljit S Khakh 和 Ken D McCarthy。《星形胶质细胞钙信号：从观察到功能及其面临的挑战》。《冷泉港生物学展望》，7(4):a020404，2015年。[29] Diederik P. Kingma 和 Max Welling。《自编码变分贝叶斯》。载于：国际学习表征会议（ICLR）2014，2013年12月。[30] David Kleinfeld、Lan Luan、Partha P. Mitra、Jacob T. Robinson、Rahul Sarpeshkar、Kenneth Shepard、Chong Xie 和 Timothy D. Harris。《能否同时记录哺乳动物大脑中每个神经元的电脉冲？》。《神经元》，103(6):1005–1015，2019年9月。[31] Lingjie Kong、Jianyong Tang、Justin P Little、Yang Yu、Tim Lämmermann、Charles P Lin、Ronald N Germain 和 Meng Cui。《通过光学锁相超声透镜进行连续体积成像》。《自然·方法》，12(8):759，2015年。[32] Christopher Langdon、Mikhail Genkin 和 Tatiana A. Engel。《关于神经流形与认知回路的统一视角》。《自然综述：神经科学》，24:363–377，2023年6月。[33] Amos Lapidoth。《数字通信基础》。剑桥大学出版社，2009年。[34] Andrew J Majda 和 Andrea L Bertozzi。《涡量与不可压缩流》，第27卷。剑桥大学出版社，2002年。

[35] Jason Manley、Sihao Lu、Kevin Barber、Jeffrey Demas、Hyewon Kim、David Meyer、Francisca Martínez Traub 和 Alipasha Vaziri。利用光珠显微成像记录小鼠背侧皮层中多达一百万个神经元的活动在自发行为期间。2024年。[36] Jason Manley、Sihao Lu、Kevin Barber、Jeffrey Demas、Hyewon Kim、David Meyer、Francisca Martínez Traub 和 Alipasha Vaziri。同步记录多达100万个神经元的全皮层动态揭示了维度随神经元数量无限扩展。《神经元》，2024年3月。[37] Michael W Marcellin、Michael J Gormish、Ali Bilgin 和 Martin P Boliek。JPEG-2000概述。载于*DCC 2000*论文集。数据压缩会议，第523–541页。IEEE，2000年。[38] Gal Mishne 和 Adam S Charles。学习用于钙成像的空间相关时间字典。载于*ICASSP 2019-2019 IEEE*国际声学、语音与信号处理会议（*ICASSP*），第1065–1069页。IEEE，2019年。[39] Eran A Mukamel、Axel Nimmerjahn 和 Mark J Schnitzer。大规模钙成像数据中细胞信号的自动分析。《神经元》，63(6):747–760，2009年。[40] Jelani Nelson、Eric Price 和 Mary Wootters。具有快速乘法和更少行数的新型RIP矩阵构造。载于第二十五届*ACM-SIAM*离散算法年度研讨会论文集，SODA '14，第1515–1528页，美国宾夕法尼亚州费城，2014年。工业与应用数学学会。[41] Kate Brody Nooner、Stanley Colcombe、Russell Tobe、Maarten Mennes、Melissa Benedict、Alexis Moreno、Laura Panek、Shaquanna Brown、Stephen Zavitz、Qingyang Li、Sharad Sikka、David Gutman、Saroja Bangaru、Rochelle Tziona Schlachter、Stephanie Kamiel、Ayesha Anwar、Caitlin Hinz、Michelle Kaplan、Anna Rachlin、Samantha Adelsberg、Brian Cheung、Ranjit Khanuja、Chaogan Yan、Cameron Craddock、Vincent Calhoun、William Courtney、Margaret King、Dylan Wood、Christine Cox、Clare Kelly、Adriana DiMartino、Eva Petkova、Philip Reiss、Nancy Duan、Dawn Thompsen、Bharat Biswal、Barbara Coffey、Matthew Hoptman、Daniel C. Javitt、Nunzio Pomara、John Sidtis、Harold Koplewicz、Francisco Xavier Castellanos、Bennett Leventhal 和 Michael Milham。NKI-Rockland样本：加速精神病学发现科学进程的模型。《神经科学前沿》，6，2012年。[42] Eftychios A Pnevmatikakis、Daniel Soudry、Yuanjun Gao、Timothy A Machado、Josh Merel、David Pfau、Thomas Reardon、Yu Mu、Clay Lacefield、Weijian Yang 等。钙成像数据的同步去噪、反卷积和解混。《神经元》，89(2):285–299，2016年。[43] Danilo Jimenez Rezende、Shakir Mohamed 和 Daan Wierstra。深度生成模型中的随机反向传播与近似推断。载于 Eric P. Xing 和 Tony Jebara（编），第*31*届国际机器学习会议论文集，机器学习研究论文集第32卷，第1278–1286页。PMLR，2014年6月。[44] Sam T. Roweis 和 Lawrence K. Saul。通过局部线性嵌入实现非线性降维。《科学》，290(5500):2323，2000年12月。[45] Derek Ruths 和 Jürgen Pfeffer。用于大规模行为研究的社交媒体。《科学》，346(6213):1063–1064，2014年。[46] Hayden Schaeffer、Russel Caflisch、Cory D. Hauck 和 Stanley Osher。偏微分方程的稀疏动力学。《美国国家科学院院刊》，110(17):6634–6639，2013年4月。[47] P. Shah 和 V. Chandrasekaran。流形上用于信号识别的迭代投影：全局恢复保证。载于*2011*年第*49*届*Allerton*通信、控制与计算年度会议（*Allerton*），第760–767页，2011年9月。[48] Diana Smetters、Ania Majewska 和 Rafael Yuste。利用钙成像检测神经元群体中的动作电位。《方法》，18(2):215–221，1999年6月。[49] Alexander Song、Adam S Charles、Sue Ann Koay、Jeff L Gauthier、Stephan Y Thiberge、Jonathan W Pillow 和 David W Tank。利用立体成像对神经元进行体积式双光子成像（VTwins）。《自然·方法》，14(4):420，2017年。[50] Joshua B. Tenenbaum、Vin de Silva 和 John C. Langford。非线性降维的全局几何框架。《科学》，290(5500):2319–2323，2000年12月。[51] Eric E Thomson、Mark Harfouche、Kanghyun Kim、Pavan C Konda、Catherine W Seitz、Colin Cooke、Shiqi Xu、Whitney S Jacobs、Robin Blazing、Yang Chen 等。利用新型多相机阵列显微镜进行千兆像素成像。《*eLife*》，11:e74988，2022年。[52] David Tse 和 Pramod Viswanath。无线通信基础。剑桥大学出版社，2005年5月。[53] Esther M Whang、Skyler Thomas、Ji Yi 和 Adam S Charles。通过采用长方形随机采集（NORA）的神经影像学实现快速双光子显微镜。*arXiv*，第arXiv–2503页，2025年。

[54] H. L. Yap、M. B. Wakin 和 C. J. Rozell。利用结构化随机矩阵实现稳定流形嵌入。*IEEE*《信号处理精选主题期刊》，7(4):720–730，2013年8月。[55] Hang Yu、Venkatakaushik Voleti、Kripa Patel、Wenze Li、Mohammed A Shaik、Elizabeth M Hillman 等。双光子扫描共聚焦对齐平面激发显微镜（2p-scape）。载于显微镜新技术，第NW4C–3页。美国光学学会，2017年。[56] Yan Zhang、Márton Rózsa、Yajie Liang、Daniel Bushey、Ziqiang Wei、Jihong Zheng、Daniel Reep、Gerard Joey Broussard、Arthur Tsang、Getahun Tsegaye 等。用于神经元群体成像的快速高灵敏度GCaMP钙指示剂。《自然》，615(7954):884–891，2023年。[57] He J. V. Zheng、Qi Wang 和 Garrett B. Stanley。胡须通路中皮层反应选择性的自适应塑造。《神经生理学杂志》，113(10):3850–3865，2015年3月。