高维数据中表征漂移的Fubini–Study几何

Arturo Tozzi（通讯作者）ASL Napoli 1 Centro, Distretto 27，意大利那不勒斯 Via Comunale del Principe 13/a 80145tozziarturo@libero.it

摘要

高维表征漂移通常使用欧几里得距离或余弦距离来量化，这些度量在比较不同时间、训练或预处理阶段的表征时，预设了固定坐标。尽管在许多情形下有效，这些度量将数据的内在变化与任意参数化所引起的变异混杂在一起。我们提出一种基于Fubini–Study度量的表征漂移射影几何观点，该度量将仅因规范变换（如全局缩放或符号翻转）而不同的表征视为同一。将这一框架应用于实证高维数据集，我们显式构建表征轨迹，并通过累积几何漂移追踪其演化。沿这些轨迹比较欧几里得距离、余弦距离和Fubini–Study距离表明，每当表征带有真实的射影歧义时，传统度量会系统性地高估变化。相比之下，Fubini–Study度量在规范诱导的波动下保持不变，从而分离出内在演化。我们进一步表明，余弦漂移与Fubini–Study漂移之差定义了一个可计算的单调量，直接刻画可归因于规范自由度的表征变动。这种分离提供了一种诊断手段，用以区分有意义的结构演化与参数化伪影，且无需引入特定于模型的假设。总体而言，我们建立了评估高维系统中表征稳定性的几何判据，并阐明了角距离的局限。将表征动力学嵌入射影空间，把数据分析与既有的几何研究计划联系起来，并产生可在实证工作流程中直接检验的可观测量。

关键词：投影　几何；表征稳定性；规范不变性；累积漂移；特征向量歧义 ty.

引言

高维数据分析依赖几何相似性与变化概念，欧几里得距离和余弦距离仍是量化不同时期、预处理阶段或学习迭代之间表征漂移的主要工具（Tang 等，2012；Canessa 等，2020；Feng 等，2022；Li 等，2025）。这些度量对坐标和参数化是否具有意义作出了强假设。它们隐含地将表征视为固定向量空间中的点，即使底层对象仅在归一化、符号或更一般的变换意义下才有定义。表征空间中看似发生的变化，可能反映的是规范诱导的变异，而非数据的内在演化（Green 等，2021；Hu 和 Szymczak，2023；Han 等，2023；Rahnenführer 等，2023）。这一局限在多个领域均有论述，包括降维、潜变量建模和表示学习：等价解在数值实现上可能存在显著差异，却编码着相同的结构信息（Lesort 等，2018；Sinha 等，2021；Zhang 等，2022；Goteti 等，2024；Vo 等，2024；Chiossi 等，2025）。现有方法通常通过临时采用归一化或对齐过程来处理这一问题，而不是重新定义测量表征漂移所依据的几何结构。因此，至今仍缺乏一种按构造即具不变性的判据，能够区分内在变化与参数化诱导的漂移。这使得经验实践与支配高维表征的等价关系之间存在概念上的鸿沟。弥合这一鸿沟，需要发展一种尊重许多数据表征之射影性质的距离概念，以便有意义地比较它们的演化。

我们基于 Fubini–Study度量提出一种理解表征漂移的射影几何视角。该度量为等价类空间而非原始向量空间提供了自然的距离度量（Elvang 2021；Espinosa-Champo 和 Naumis 2023；Hofmann 等，2023）。通过将表征视为射线而非点，并仅依据内在方向来衡量变化，该度量能够明确识别仅因规范变换而不同的状态。我们通过追踪经验表征轨迹上的累积漂移，并比较沿相同路径计算的欧几里得距离、余弦距离和 Fubini–Study 距离，将这一思路付诸实践。由此得到一个可计算的量，用于量化规范诱导的变异对观测到的漂移所作的贡献，而无需进行对齐、选择参考对象或采用依赖模型的校正。我们设计了实验，使用存在真实射影歧义的经验数据集，直接检验这种区别。此外，我们还评估了角度度量与射影度量之间由此产生的差异是否具有系统性和稳健性。综上，这些考量确立了表征漂移几何的概念基础，使其能够区分内在演化与参数化伪影。

我们将按以下方式进行 。首先，我们将描述表征轨迹的构建及其计算 of
欧几里得、余弦 以及 Fubini–Study 累积漂移。随后，我们在免费公开的数据集上展示实证结果， ith

射影歧义。最后，我们讨论这些发现对于理解高维数据中的表征漂移所具有的意义。

方法

我们介绍用于量化不同几何距离概念下表征漂移的数学构造、数据操作和计算程序。我们着重说明可复现的步骤，详细介绍如何构建表征轨迹、如何沿轨迹计算距离，以及如何通过解析和数值方法处理射影歧义。

数据集描述。所有分析均基于 scikit-learn 库提供的手写数字数据集。该数据集包含 1,797 张由人类受试者书写的真实数字灰度图像，每张图像均表示为一个 8×8 像素网格，并展平为 64 维实值特征向量。像素强度对应测得的灰度级。数据集包含用于标识每张图像所描绘数字的类别标签；但构建表征轨迹时不使用这些标签。数据本身未提供内在的时间顺序。我们仅通过人为排序来定义连续的分析窗口，从而能够在可控且可复现的条件下研究表征漂移。选择该数据集是因为它可公开获取、来源记录完善、维度固定，且除了中心化之外无需额外预处理即可用于主成分分析。

表征轨迹的构建。 下一步是将经验数据转换为有序轨迹，以便进行几何分析。令 $X = \{x_i\}_{i=1}^{N} \subset \mathrm{R}^{D}$ 表示一组经验观测，其中每个 x_i都是一个 D维特征向量。为研究表征漂移，我们将数据嵌入为有序序列，以体现某种进程。在我们的分析中，顺序由针对数据集索引执行的滑动窗口程序确定。对于固定的窗口长度 W 和步长

$$s \qquad X_k = x_k\ x_{k+1} \quad x_{k+W-1} \qquad k = 1\,1 + s\ 1 +$$

，提取连续子集 { , , … , }，其中 , , , ... 每个子集都定义了数据的一个局部表征，而这些子集构成的序列则形成一条以 k为索引的离散轨迹。这种构造不假设存在时间因果关系；它只施加足以定义连续表征状态的顺序。随后，根据分析步骤，将每个状态映射为 R^{D} 中的一个向量或一个派生对象。轨迹长度 T 由窗口数量决定，即 $T = \lfloor (N - W)/s \rfloor + 1$。所有后续几何量均沿这条离散轨迹 $\{r_k\}_{k=1}^{T}$计算，其中 r_k 表示从窗口 X_k中提取的表征。这种排序使我们能够定义逐步变化量 $r_{k+1} - r_k$ 以及沿轨迹累积的漂移度量，构成了我们比较不同距离度量的基础。

欧氏几何距离与累积漂移。 对于表征 $r_k \in \mathrm{R}^{D}$，欧几里得距离定义为

$$d_{\mathrm{E}}(r_k, r_{k+1}) = \| r_{k+1} - r_k \|_2 = \sqrt{\sum_{j=1}^{D} (r_{k+1}^{(j)} - r_k^{(j)})^2}.$$

该距离将表征视为固定向量空间中的点，同时对方向和幅值敏感。为刻画沿轨迹的漂移，将逐步欧氏距离累加，构成累积漂移函数

$$\Delta_{\mathrm{E}}(n) = \sum_{k=1}^{n-1} d_{\mathrm{E}}\,(r_k, r_{k+1}), n = 2, \ldots, T.$$

按定义，该量单调递增，反映在 R^{D} 中 traversed 的总路径长度。除非另有明确说明，否则不进行归一化，因此全局尺度的变化会直接计入 Δ_{E}。欧氏累积漂移可作为参照，用于比较尺度不变度量和射影度量。在实践中，欧氏距离通过向量化线性代数例程计算，并利用双精度运算确保数值稳定性。欧氏度量并非旨在捕捉表征之间的等价关系；相反，它用于量化在所选参数化下变化的原始幅值。

余弦距离与角度漂移。 此处我们引入角度比较，作为一种尺度不变的替代方案。为消除对全局缩放的敏感性，将表征归一化为单位范数。对于任意非零向量 $r \in \hat{r} = \mathrm{R}^{D}$，定义 $/\| r \|_2$。相继状态之间的余弦相似度为

$$s_{\mathrm{C}}(r_k, r_{k+1}) = \langle \hat{r}_k, \hat{r}_{k+1} \rangle,$$

其中 $\cdot,\cdot$ 表示标准内积。相应的角距离为

$$d_{\mathrm{C}}(r_k, r_{k+1}) = \arccos\left(\langle \hat{r}_k, \hat{r}_{k+1} \rangle\right),$$

其取值范围为 $[0,\pi]$。该距离度量衡量归一化表征之间的夹角，并且在正标量乘法下保持不变。与欧几里得情形一样，累积余弦漂移定义为

$$\Delta_{\mathrm{C}}(n)=\sum_{k=1}^{n-1} d_{\mathrm{C}}\left(r_k, r_{k+1}\right).$$

虽然余弦距离消除了对尺度的依赖，但它仍会区分仅符号不同的向量，并为对跖点赋予最大距离 π 。当表征仅在符号或更一般的规范变换意义下确定时，这一性质便尤为重要。在计算时，通过将内积裁剪至$[-1,1]$ 来计算余弦角，以避免浮点误差导致的数值伪影。

射影几何与 Fubini–Study 度量。 接下来，我们定义用于识别规范等价表征的射影距离。射影几何将相差一个非零标量的向量视为等价。在实数情形下，相关空间为实射影空间 RP^{D-1}，它通过将$r \sim \lambda r$ 对所有$\lambda \neq 0$视为等同而得到。Fubini–Study度量在该空间上诱导出一种自然的距离。对于归一化代表元 $\hat{r}_k$、 $\hat{r}_{k+1}$，实数情形下的 Fubini–Study 距离定义为

$$d_{\mathrm{FS}}(r_k, r_{k+1})=\arccos\ \left(\left|\left\langle \hat{r}_k, \hat{r}_{k+1}\right\rangle\right|\right),$$

其取值范围为$\left[0,\frac{\pi}{2}\right]$。绝对值确保将对跖点视为等同，因此仅符号不同的向量之间的距离为零。与前文一样，累积射影漂移定义为

$$\Delta_{\mathrm{FS}}(n)=\sum_{k=1}^{n-1} d_{\mathrm{FS}}\left(r_k, r_{k+1}\right).$$

该构造度量的是等价类之间而非原始向量之间的变化。与余弦距离不同，Fubini–Study度量在重新缩放和符号反转下均保持不变。数值实现遵循与余弦距离相同的归一化和截断程序，并在计算反余弦之前对内积额外取绝对值。

主成分分析与诱导的射影歧义. 为确保存在非平凡的射影等价，表征r_k 被导出为每个数据窗口的主成分方向。对于窗口$X_k \in \mathrm{R}^{W\times D}$，通过减去均值向量$\mu_k=\frac{1}{}\sum_{i=1}^{W} x_i$对数据进行中心化。随后应用奇异值分解，

$$X_k-\mathbf{1}\mu_k^{\top}=U_k\overset{W}{\Sigma}_k V_k^{\top},$$

其中$V_k \in \mathrm{R}^{D\times D}$ 包含标准正交的右奇异向量。第一主成分方向为$r_k=v_{k,1}$，即V_k的第一列。根据构造，$v_{k,1}$仅在符号意义下被定义，因为$v_{k,1}$与$-v_{k,1}$均对应于同一方差最大化方向。这种符号不确定性将真正的射影歧义引入表征轨迹。各窗口之间不进行符号对齐，并保留分解的原始输出。这一选择使余弦距离与Fubini–Study距离之间的比较能够直接反映对对跖等价的处理方式。

沿轨迹检测符号翻转和规范诱导漂移。。给定相邻的主成分向量 r_k 和 r_{k+1}，我们会监测内积 r_k, r_{k+1}。负值表示相邻窗口之间发生了符号反转。这些事件不会被校正，而是作为轨迹的一部分予以记录。在余弦距离下，每次符号翻转都会使累积漂移跃增约 π 。在 Fubini–Study度量下，同一事件的贡献为零，这反映了规范等价性。因此，累积差异

$$\Delta_{\mathrm{C}}(n)-\Delta_{\mathrm{FS}}(n)$$

由此提供了一个定量指标，用于衡量完全由符号歧义导致的漂移。该差异沿轨迹非负且单调不减。计算该差异无需引入任何额外参数，只需使用余弦距离和 Fubini–Study距离中已有的参数。

欧几里得漂移与射影漂移的对数比较。 为对比原始的基于幅值的漂移与射影漂移，本文考察欧几里得累积测度与 Fubini–Study累积测度之比。为保证数值稳定性，在这两个量上都加上一个很小的常数 $\varepsilon>0$ ，并将对数比定义为

$$R(n)=\log_{10}\left(\frac{\Delta_{\mathrm{E}}(n)+\varepsilon}{\Delta_{\mathrm{FS}}(n)+\varepsilon}\right).$$

该量突出显示欧几里得变化与射影变化之间的数量级差异，并以轨迹索引为横轴绘制。欧几里得漂移的增长速度远快于射影漂移时，对数变换有助于可视化。ε 的取值固定，不影响定性趋势。

计算工具与实现. 所有计算均使用标准科学计算库完成。线性代数运算，包括范数、内积和奇异值分解，均通过优化的数值例程实现。主成分提取依靠奇异值分解来确保结果具有确定性。向量归一化、累积求和和三角函数计算均采用双精度。轨迹和累积量的可视化使用标准绘图库，并明确控制坐标轴和缩放。除窗口大小和步长的选择外，不涉及随机优化、学习过程或参数调优；这两个参数在分析前即已固定。

总之，我们在此说明了用于分析欧几里得、余弦和 Fubini–Study度量下表征漂移的数据构造、几何距离及计算步骤。通过详细给出明确的数学定义和可复现的流程，这些内容使人们能够比较高维表征轨迹中变化的原始、角度和射影概念。

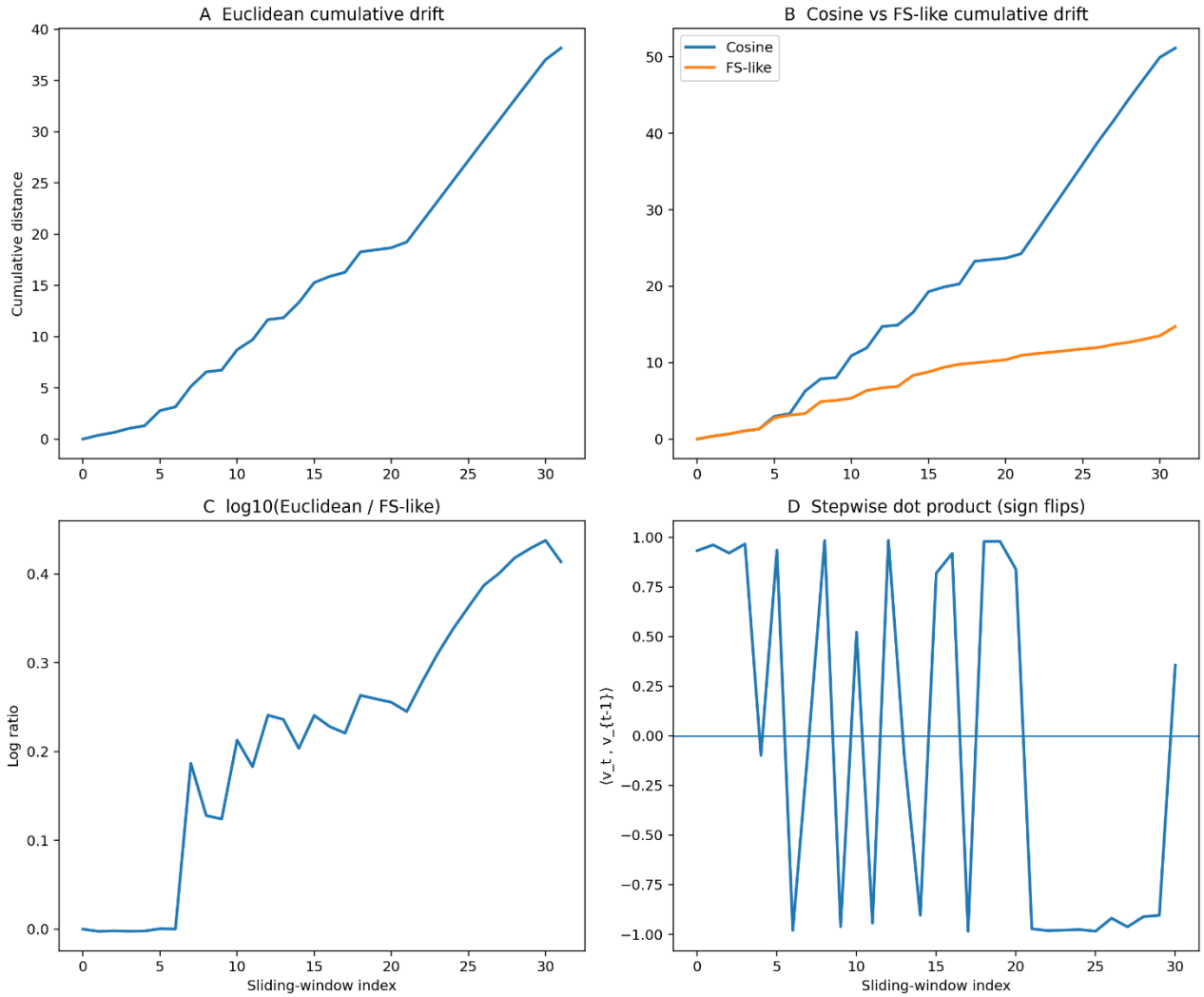
结果

我们报告了将欧几里得距离、余弦距离和 Fubini–Study距离应用于从真实高维数据中提取的表征轨迹后获得的经验观测。我们重点关注真实射影歧义存在时这些度量之间的定量差异，以及累积漂移如何沿相同的经验路径分解为内在分量和规范诱导分量。

余弦漂移与射影漂移的经验性区分。对滑动窗口主成分方向所导出的表征轨迹进行分析发现，当存在符号歧义时，余弦累积漂移与 Fubini–Study 累积漂移之间会出现系统性偏离（图 A–B）。在包含 32 个窗口的完整轨迹中，余弦累积漂移达到 51.15 rad，而 Fubini–Study 累积漂移达到 14.72 rad，两者相差 36.43 rad，这一差异完全归因于符号反转。对相邻主方向之间逐步计算内积后发现了 17 次符号翻转事件，每次均对应负点积（图 D ）。采用余弦距离时，每次事件都会产生接近 π 的增量；而采用 Fubini–Study 度量时，同一事件产生的增量为零。除这些符号翻转事件外，余弦距离与 Fubini–Study距离测得的增量角度变化在数值上完全一致。因此，对逐步角度增量进行配对比较后，只剩下 17 个非零差异，且全部为正（精确双侧符号检验：$p<0.001$）。这些结果表明，沿着一条固定的经验轨迹，余弦距离累积了内在方向变化和规范诱导的变异，而 Fubini–Study距离仅累积前者。因此，两者之间的累积差异可作为一种能够通过统计检验区分的规范诱导漂移度量，且无需进行对齐或事后校正。

原始幅值漂移与射影变化之间的关系。与欧氏累积漂移的比较进一步阐明了表征变化的结构（图 A–C）。沿整个轨迹，相邻主方向之间的累积欧几里得漂移单调且显著地增加，反映出其同时对幅值和方向的敏感性。与 Fubini–Study 漂移相比，对数比 $\log_{10}[(\Delta_E + \varepsilon)/(\Delta_{FS} + \varepsilon)]$ 沿轨迹持续增加（图 C），表明基于原始幅值的变化增长速度显著快于内在射影变化。即使在未发生符号翻转的区间内，这种分化仍然存在，说明欧几里得距离混淆了平滑的方向变异与尺度及参数化效应，而归一化和射影识别会消除后者。对数比的单调增加证实，这种分离反映的是累积结构，而非孤立事件。结合具有统计显著性的余弦–Fubini–Study 分离，我们的分析将观测到的漂移分解为在同一数据上测量的三个组成部分：原始幅值变化、对符号敏感的角度变化，以及对规范不敏感的内在射影变化。

总体而言，我们的结果表明，余弦距离与Fubini–Study距离一致，但在经验检测到的符号翻转事件处例外；在这些事件中，余弦漂移会累积较大的增量，而这些增量在Fubini–Study度量下并不存在。这些增量具有统计显著性，并占总角度漂移的相当大一部分。与此同时，欧几里得漂移的增长速度远快于射影漂移，使原始度量与内在变化度量之间的对数分离持续扩大。



图。实证高维数据中的射影效应。

（A）在手写数字图像的滑动窗口上计算得到的相邻第一主成分（第一主成分）之间的累积欧几里得漂移。欧几里得距离稳步增长，反映出其对参数化变化的敏感性。

（B）比较余弦距离与射影（类Fubini–Study）距离的累积角度漂移。由于PCA特征向量反复发生符号反转，余弦漂移会累积较大的跳变，而类Fubini–Study度量对这些规范变换保持不敏感。

（C）欧几里得累积漂移与类Fubini–Study累积漂移之间的对数比，突出显示原始距离增长与内在射影变化之间日益扩大的差异。

（D）相邻第一主成分向量之间的逐步点积；负值表示特征向量符号翻转（$v \equiv -v$），这是产生真实射影歧义的来源。

总体而言，各面板表明，当表征仅在符号意义下确定时，欧几里得度量和余弦度量会高估变化，而类Fubini–Study度量则能分离出等价类的内在演化。

结论

我们发现，高维数据中的表征漂移可分解为不同的几何分量，而常用的距离度量会将它们混为一谈。通过使用欧几里得距离、余弦距离和 Fubini–Study距离考察相同的实证轨迹，我们表明，基于原始幅值的漂移会快速且单调地增长；角度漂移则同时累积了内在方向变化和规范诱导的变异，而射影漂移只反映内在分量。从定量结果来看，余弦累积漂移与 Fubini–Study累积漂移之间的差异，完全可归因于实证观测到的主成分方向符号反转，而其余增量则完全一致。由此产生的差异占总角度漂移的相当大比例，并且可通过精确检验在统计上加以区分，这表明这种分离并非数值伪影。与此同时，欧几里得漂移与射影漂移之间不断扩大的对数间隔表明，即使没有发生离散的规范事件，原始距离也会系统性地高估变化。综上，这些观察描绘出一种实证图景，其中不同的

度量会响应表征变异性的不同方面，而射影几何则提供了一种方法，可在保留对真实演化的敏感性的同时，消除等价性导致的波动。这一对定量结果的综合分析，为将异质表征中的表征漂移理解为一种复合现象奠定了清晰的基础。

我们将表征漂移重新表述为一个定义在等价类上、而非固定坐标向量上的几何问题。现有技术通过在距离计算之前应用的预处理、归一化或对齐步骤来处理不变性，并假定单一校正后的表征已足够。相比之下，使用Fubini–Study度量将不变性直接纳入距离本身，从而确保规范等价表征在构造上即被等同。这一区分阐明了为何余弦距离虽具尺度不变性，却仍对符号及相关歧义敏感，以及为何欧几里得距离对尺度和参数化均敏感。我们并不引入新的相似度分数，而是对同一经验轨迹上漂移的内在贡献与规范诱导贡献给出明确分离。与信息论度量或基于分布的度量相比，我们的射影度量直接作用于表征，且不需要概率建模（Vasudevan et al. 2018; Sechidis et al. 2018; Liu et al. 2022）。与基于对齐的方法相比，它避免了参考选择和事后校正（Galpert et al. 2018; Kobus et al. 2020; Expósito et al. 2022; He et al. 2024）。我们的方法既不是对标准距离的替代，也不是临时校正，而是一种在表征被定义到等价关系时生效的几何细化。在此意义上，我们的方法可归类为表征分析的射影几何方法，与欧几里得方法、角度方法和统计方法互补，并自然地位于原始向量空间技术与更为抽象的基于流形的分析之间。

需要承认本研究存在若干局限。首先，我们的实证展示依赖于一个本质上不具时间属性的数据集；轨迹是通过滑动窗口构建的，观察到的漂移反映了这种人为设定的顺序。虽然这一过程在数学上定义明确，但无法捕捉自然演化系统中存在的所有变异来源。其次，主成分分析引入了射影歧义，而符号反转的频率取决于窗口大小和重叠程度。第三，统计检验仅限于对单条轨迹中检测到的符号翻转事件进行精确检验，无法确立总体层面的普遍性。第四，分析聚焦于与符号等价相关的实射影几何；本研究未通过实证探讨更一般的规范自由度，例如潜在空间中的旋转。最后，图示展示的是一个有代表性的案例，而非对不同数据集或表征进行的系统调查。

我们的结果指向若干具体的应用方向和后续研究方向。漂移的射影分解可应用于已知存在规范自由度的表征，例如协方差或连接矩阵的特征模态、降维中的潜变量，或通过独立训练运行获得的嵌入。未来研究可以检验：余弦漂移与 Fubini–Study 漂移之间的累积差异，是否会随着表征不可辨识程度的增加而系统性扩大，从而为规范自由度提供一种定量表征。另一项可检验的假设是，在学习流程或预处理流程中，内在射影漂移是否比原始漂移或角度漂移更早趋于稳定，这将表明即使参数化仍在变化，表征内容也已趋于收敛。在方法层面，当表征并非唯一确定时，应同时报告射影漂移和传统距离，并明确检验观察到的变化在射影识别后是否仍然存在。

总之，我们探讨了如何在高维数据中区分内在表征变化与不可辨识性所导致的变异。通过比较同一组实证轨迹上的欧几里得距离、余弦距离和Fubini–Study距离，我们表明，射影几何可以提供一种可计算的方法，明确区分二者。表征漂移并非单一量，而是由原始、角度和射影分量构成的复合量，每个分量都刻画了变化的不同方面。识别并分离这些分量，有助于更准确地解释表征如何演变，并提供一种具有数学依据且可通过实证检验的几何判据。

声明

伦理批准与参与同意。 本研究不涉及作者开展的任何人类参与者或动物研究。

发表同意。 若本文发表，作者将转让全部版权所有权。署名作者保证，本文为原创作品，不侵犯任何第三方的版权或其他专有权，未同时提交至其他期刊审议，且此前未曾发表。

数据和材料的可用性。 本研究期间生成或分析的所有数据和材料均已收录于稿件中。作者可全面访问本研究中的所有数据，并对数据的完整性和数据分析的准确性负责。

利益冲突。 作者不存在任何已知或潜在的利益冲突，包括在提交稿件前三年内与其他个人或组织存在的任何财务、个人或其他关系，而这些关系可能不当影响或被认为会影响其工作。

经费。 本研究未获得来自公共、商业或非营利部门任何资助机构的专项拨款。

致谢：无。

作者贡献。 作者负责：研究构思与设计、数据获取、数据分析与解释、稿件撰写、对稿件重要学术内容进行严格修订、统计分析、获取经费、提供行政、技术和材料支持，以及研究监督。

在写作过程中使用生成式人工智能和 AI辅助技术的声明。 在准备本研究期间，作者使用 ChatGPT 5.2 辅助进行数据分析和稿件撰写，并改进拼写、语法和整体编辑。使用该工具后，作者根据需要审阅并编辑了内容，并对出版物内容承担全部责任。

参考文献

1) Canessa, Enrique, Sergio E. Chaigneau, Sebastián Moreno, 和 Rodrigo Lagos。2020。《基于高维概念属性规范数据计算的余弦及其他相似度的信息内容》。《认知加工》21 5 2 4 601 614 10 1007 10339 020 00985 2025 ( ): – 。https://doi.org/ . /s - - -) Chiossi, H. S. C., Matteo Nardin, Gašper Tkačik, 和 Jozsef Csicsvari。 .《学习重塑海马表征层级》。《美国国家科学院院刊》122 2417025122 3 11 2417025122 10 1073 2021( ): e 。https://doi.org/ . /pnas.) Elvang, Henriette。 .《自举法与振幅：量子场论领域的一次探索》。《物理学进展报告》84 (7)。https://doi.org/10.1088/1361-6633/abf97e 4) Espinosa-Champo, Alejandro J., 和 Gerardo G. Naumis。2023。《 Fubini-Study 度量与平带电子态的拓扑性质：具有 s 轨道的原子链案例》。《物理学杂志：凝聚态物质》36 1 5 1 10 1088 1361 648 ( )。https://doi.org/ . / - X/acfbd) Expósito, Rubén R., Manuel Martínez-Sánchez, 和 Juan Touriño。2022。《SparkEC：加速基于对齐的DNA错误校正工具》。《BMC生物信息学》23 (1): 464。https://doi.org/10.1186/s12859-022- 05013-1 6) Feng, Xiaoyu, Fang Fang, Hong Long, Rong Zeng, 和 Yao Yao。2022。《使用图自编码器和图注意力网络进行单细胞RNA测序数据分析》。《遗传学前沿》13: 1003711。https://doi.org/10 1003711 7 3389 2022 . /fgene. .) Galpert, Daniel, Antonio Fernández, Francisco Herrera, António Antunes, Rafael Molina-Ruiz, 和 Gisselle Agüero-Chapin。2018。《使用监督式大数据分类器调查相关酵母蛋白质组中的无比对特征以进行直系同源检测》。《BMC生物信息学》19 (1): 166。https://doi.org/10.1186/s12859- 018-2148-8 8) Goteti, Kiran, Ricardo Garcia, William R. Gillespie, Jeffrey French, Lena Klopp-Schulze, Ying Li, Carlos V. Mateo, Saptarshi Roy, Oliver Guenther, Luca Benincosa, 和 Karthik Venkatakrishnan。2024。《使用潜变量建模进行基于模型的荟萃分析，为系统性红斑狼疮新疗法设定基准》。《CPT：定量药理学与系统药理学》13 (2): 281–295。https://doi.org/10 13083 9 1002 4. /psp .) Green, Benjamin, Heng Lian, Yi Yu, 和 Tingting Zu。2021。《超高维半参数纵向数据分析》。《生物统计学》77 13348 10 3 903 913 10 1111 2023 ( ): – 。https://doi.org/ . /biom.) Han, Qiang, Yifan Wang, Ning Sun, Jie Chu, Wei Hu, 和 Yan Shen。 .《高通量数据中介分析方法综述》。《遗传学与分子生物学中的统计应用》22 (1)。https://doi.org/10 0031 11 1515 2023 . /sagmb- -) He, Kai, Jian Zhu, Lei Li, Feng Gou, 和 Jie Wu。2024。《FASNet：利用数字病理图像辅助诊断医疗系统的特征对齐方法》。《 Heliyon》10 (22): e40350。https://doi.org/10 40350 12 1016 2024 . /j.heliyon. .e) Hofmann, Johannes S., Erez Berg, 和 Debanjan Chowdhury。2023。《可调谐量子度规平带中的超导、电荷密度波与超固态》。《物理评论快报》130 (22): 226001。https://doi.org/10 226001 13 1103 130 2023 . /PhysRevLett. .) Hu, Jing, 和 Sebastian Szymczak。 .《随机森林纵向数据分析综述》。《生物信息学简报》24 002 14 2 002 10 1093 ( ): bbad 。https://doi.org/ . /bib/bbad) Kobus, Ralf, José María Abuín, Andreas Müller, Stefan L. Hellmann, José C. Pichel, Tomás F. Pena, Andreas Hildebrandt, Thomas Hankeln, 和 Bertil Schmidt。2020。《面向所有食品测序的宏基因组学大数据方法》。《BMC生物信息学》21 (1): 102。https://doi.org/10.1186/s12859-020-3429-6

15) Lesort, Timothée, Natalia Díaz-Rodríguez, Jean-Isaac Goudou, 和 David Filliat。2018。《控制的状态表示学习：综述》。《神经网络》108: 379–392。https://doi.org/10 006 16 1016 2018 07 . /j.neunet. . .) Li, Xinyu, Hannah Belcher, 和 Hui Wang。2025。《鲁棒球面拉普拉斯嵌入》。《IEEE 神经网络与学习系统汇刊》36 3401010 17 4 7594 7605 10 1109 2024 2022 ( ): – 。https://doi.org/ . /TNNLS. .) Liu, Haibo, Chen Peng, Yu Tian, Shichao Long, Fei Tian, 和 Zhenyu Wu。 .《GDP与LDP：信息论信道视角下的综述》。《熵》24 24030430 18 3 430 10 3390 ( ): 。https://doi.org/ . /e) Rahnenführer, Jörg, Riccardo De Bin, Axel Benner, Federico Ambrogi, Lara Lusa, Anne-Laure Boulesteix, Enrica Migliavacca, Harald Binder, Stefan Michiels, Willi Sauerbrei, 和 Lisa McShane，代表STRATOS 倡议“高维数据”主题组（TG9）。2023。《高维生物医学数据的统计分析：分析目标、常见方法与挑战简介》。《BMC 医学》21 (1): 182。https://doi.org/10.1186/s12916-023-02858-y 19) Sechidis, Konstantinos, Konstantinos Papangelou, Peter D. Metcalfe, Daniel Svensson, James Weatherall, 和 Gavin Brown。2018。《区分预后性与预测性生物标志物：一种信息论方法》。《生物信息学》34 357 20 19 3365 3376 10 1093 2021 ( ): – 。https://doi.org/ . /bioinformatics/bty) Sinha, Pratik, Carolyn S. Calfee, 和 Katherine L. Delucchi。 .《潜在类别分析实务指南：方法学考量与常见陷阱》。《重症医学》49 (1): e63–e79。https://doi.org/10 0000000000004710 21 1097. /CCM.) Tang, Haibin, Simon M. Chu, Mark Hasegawa-Johnson, 和 Thomas S. Huang。2012。《部分监督说话人聚类》。《IEEE 模式分析与机器智能汇刊》34 (5): 959–971。https://doi.org/10 174 22 1109 2011 . /TPAMI. .) Vasudevan, Sridhar, Efrat Flashner-Abramson, Françoise Remacle, Raphael D. Levine, 和 Natalia Kravchenko-Balasha。2018。《通过信息论压缩癌症大数据获得个性化疾病特征》。《美国国家科学院院刊》115 (30): 7694–7699。https://doi.org/10 1804214115 23 1073. /pnas.) Vo, Khanh, Qiang J. Sun, Michael D. Nunez, Joachim Vandekerckhove, 和 Ramesh Srinivasan。2024。《神经信号与人类行为的深度潜变量联合认知建模》。《神经影像》291: 120559。https://doi.org/10 120559 24 1016 2024 . /j.neuroimage. .) Zhang, Xin, Wei Wang, Chao-Xing Ren, 和 Dao-Qing Dai。2022。《通过鲁棒图正则化整合方法学习多个生物网络的表示》。《生物信息学简报》23 (1): bbab409。https://doi.org/10.1093/bib/bbab409