arXiv:2303.04555v1 [cs.LG] 8 Mar 2023

小空间流式核主成分分析算法

Yichuan Deng* Zhao Song† Zifan Wang‡ Han Zhang§

摘要

主成分分析（PCA）是机器学习、数据分析和信号处理领域广泛使用的一种技术。随着数据集规模和复杂性的增加，开发低空间开销的主成分分析算法变得日益重要。近年来，流式主成分分析备受关注，因为它能够高效处理大型数据集。核方法常用于支持向量机（SVM）等学习算法，也已应用于主成分分析算法中。

我们提出了一种基于 Oja 传统方案的核主成分分析问题流式算法。我们的算法旨在降低主成分分析的内存使用量，同时保持其准确性。我们通过研究算法成功运行所需的条件来分析其性能。具体而言，我们证明，当目标协方差矩阵的谱比 $R := \lambda_1/\lambda_2$ 下界为 $C \cdot \log n \cdot \log d$ 时，流式主成分分析可以通过 $O(d)$ 的空间开销求解。

我们提出的算法相比现有方法具有多项优势。首先，它是一种流式算法，能够高效处理大型数据集。其次，它采用核方法，能够捕捉数据点之间复杂的非线性关系。第三，它占用空间少，因此适用于内存有限的应用场景。

*ethandeng02@gmail.com. 中国科学技术大学。
†zsong@adobe.com. Adobe Research。
‡Zifan.wang@stonybrook.edu. 石溪大学。
§micohan@cs.washington.edu. 华盛顿大学。

1 引言

主成分分析（PCA）是一种用于降低数据维度的技术。它是一种线性方法，利用正交变换将一组相关变量转换为一组相关性较低的变量，这些变量称为主成分。在最简单的情况下，我们关注第一个主成分。

核主成分分析（核 PCA）是主成分分析的一种扩展（也是一种推广），结合了核方法。核 PCA 有许多应用，例如基于距离的算法 [Sch00], 在高维特征空间中计算主成分 [SSM97]，人脸识别 [YAK00，Liu04], 谱嵌入 [BDR+04]，新颖性检测 [Hof07], 特征空间中的去噪[MSS+98], 以及非线性过程的故障检测与识别[CLL+05]。

在最简单的主成分分析情形中，给定数据集 $X = \{x_1, x_2, \ldots, x_N\} \subseteq \mathbb{R}^d$，数据集的协方差矩阵为 $C := \frac{1}{N}\sum_{i\in[N]} x_i x_i^\top$。目标是找到与 C 的最大特征值 λ 相对应的特征向量 $v^* \in \mathbb{R}^d$。

为了理解核 PCA 的动机 [LMTS20，XLL+19，MSS+98]，尤其是在聚类方面，请注意，尽管 N 个点通常无法在 $d < N$ 维空间中线性分离，但几乎总能在 $d \geq N$ 维空间中线性分离。也就是说，给定 N 个点，x_i，若将它们映射到一个 N维空间，其中 $\phi(x_i)$，其中 ϕ： $\mathbb{R}^d \to \mathbb{R}^N$，就很容易构造一个超平面，将这些点划分为任意的簇。因此，核 PCA 是一种广泛使用的工具，用于提取非线性特征，而传统的（线性）主成分分析只能检测线性特征。

由于核空间的维度可能非常高（这由核函数 ϕ隐含表示），因此在该空间中计算精确乘积的代价会非常高。因此，自然且合理的做法是使用 Mercer 核 [Gir02、ZG04、XY19], 一个函数 $k(x, y)$： $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ 对于输入数据集 $X = \{x_i\}_{i\in[N]} \subseteq \mathbb{R}^d$，该函数会生成一个正定矩阵 $K \in \mathbb{R}^{n\times n}$，其中$K$ 的每个元素由下式给出

$$K_{i,j} := k(x_i, x_j).$$

通过定义

$$k(x, y) := \phi(x)^\top \phi(y),$$

可以使用 k 将数据点映射到核空间，而无需显式计算内积。请注意，矩阵 K 的每一列 K_i 都是核空间中的一个乘积，即从一个点 x_i 到 X中的所有 N 个点。

由于我们没有显式地在特征空间中进行运算（这可能因维度过高而代价非常大），因此得到的主成分是针对投影数据的。对于数据点 x，它在第 k 个主成分 v_k 上的投影是 $v_k^\top \phi(x)$，而不是线性主成分分析中的原始 $v_k^\top x$ 。

在传统的主成分分析问题中 [CMW13、ZB05、BBZ07、VL12], 需要访问所有数据点 $\{x_i\}_{i\in[n]}$。因此，存储这些数据所需的内存可能非常大。流式主成分分析是一种在数据量过大、无法全部装入内存时执行主成分分析的方法。传统的主成分分析算法要求一次性将所有数据加载到内存中，因此不适用于非常大的数据集。相比之下，流式主成分分析允许分批处理数据，从而减少内存需求，并能够分析非常大的数据集。

在流式场景中，我们需要维护一种数据结构，使其能够接收以数据流方式到来的数据点，并在数据流结束时输出估计的主成分。形式上，该数据结构接收由 x_i 构成的数据流。然后，通过某种

维护操作，它可以输出一个向量 u ，使得 $u \approx x^*$，其中 x^* 是该数据集的第一主成分。

受核主成分分析算法的启发，并结合算法应当运行快速、占用空间少这一自然期望，我们提出以下问题：

Can we solve the kernel PCA in a small space?

在本文中，我们为这个问题给出了肯定的答案。

1.1 相关工作

流式算法。过去几十年来，人们设计了大量流式算法，因为人们担心在某些情况下，数据量过大，无法存储在单台机器上。一些流式算法主要针对图问题[BFK+20]，例如最短路径和直径 [FKM+05, FKM+09]、极大独立集 [ACK19, CDK18]，、最大匹配和最小顶点覆盖 [FKM+05, GKK12, Kap 13],、谱稀疏化 [KNST19, KMPV19]，、最大割 [KKS14]，、核方法和草图技术[SWYZ21, AS23, ACSS20, KP20, KKP18]。除图问题外，流式算法也为其他领域提供了有益见解，例如多臂老虎机问题 [LSPY18]。由于已证明许多问题无法在低于 n 的空间复杂度下求解，其中我们用 n 表示图中的节点数，因此一系列研究 [FKM+05, M+05] 聚焦于半流式模型。在这一设定中，流式算法允许使用 $O(n\text{poly}\log n)$ 空间。

近年来，研究重点已转向多遍流式模型。在这种设定下，模型可以多次查看流式更新。原因是，让模型多遍处理更新能够有效减少所需空间。例如，极大独立集的 O（loglog n）遍算法 [ACK19、CDK18、GGK+18]，以及近似匹配的 O（1）遍算法[GKMS19、GKK12、Kap13、McG05]。

主成分分析。从统计学角度研究主成分分析的工作已有很多，这些研究在特定条件下考察不同算法的性能。其中包括使用数据的生成模型 [CMW13],，并对特征值间隔 [ZB05] 和协方差矩阵谱[BBZ07、VL12]作出假设。尽管这些研究确实针对有限数据量提供了保证，但它们并不适用于实际应用，因为它们要么仅适用于完整数据集，要么需要大量计算资源。因此，实际应用需要一种高效的增量算法。

谈到增量算法，Warmuth 和 Kuzmin 的工作 [WK06] 对最坏情况下的流式主成分分析进行了分析。此前，通用的增量主成分分析算法在有限样本条件下的性能尚未得到分析。 [ACLS12]。最近，研究者尝试通过放宽问题的非凸性 [ACS13] ，或作出生成式假设 [MCJ13]，以解决缺乏有限样本分析的问题。

由于这是一个热门课题（人们自然会想到从以数据流方式到来的数据集中提取主成分），多年来，研究者一直关注流式主成分分析。流式主成分分析有两种传统算法：一种是 Oja 算法 [Oja82] ，另一种是 Krasulina 提出的经典方案 [Kra69]。Balsubramani、Dasgupta 和 Freund[BDF13] 的研究分析了 Krasulina 算法和 Oja 算法的收敛速度。Hardt 和 Price [HP14] 对一种著名的幂方法进行了稳健的收敛性分析。该方法用于计算矩阵的主奇异向量，我们称之为噪声幂方法。随后

Allen-Zhu 和 Li 的研究 [AZL17] 证明了 Oja 算法在具有 $k > 1$ 多个第一主成分时的全局收敛性，并提出了一个运行速度更快的 Oja 算法变体。另一系列研究 [HNWW21、HNWTW22] 表明，即使更新不只是秩为 1，Oja 算法的性能也几乎可以匹配最优离线算法。还有一些研究关注如何对 Oja 算法所得主特征向量的估计误差进行不确定性量化 [LSW21]。一项最新研究 [MP22]给出了正确性保证：在满足谱比特定条件时，Oja 算法可用于传统设定下的流式主成分分析。

1.2 我们的结果

在本节 中，我们给出主要结果：一种用于核主成分分析的流式算法 .

定理 1.1（定理 5.2 的非正式版本）。设 $\phi: \mathbb{R}^d \to \mathbb{R}^m$。设 $\Sigma = \frac{1}{n}\sum_{i=1}^n \phi(x_i)\phi(x_i)^\top \in \mathbb{R}^{m\times m}$。我们定义 $R := \lambda_1(\Sigma)/\lambda_2(\Sigma)$，其中 $\lambda_1(\Sigma)$ 是 Σ 的最大特征值，而 $\lambda_2(\Sigma)$ 是 Σ 的第二大特征值。令 x^* 表示 Σ 的最大特征向量。令 $C > 10^4$ 表示一个足够大的常数。若 $R \geq C \cdot (\log n) \cdot (\log d)$，则存在一种流式算法，只使用 $O(d)$ 空间，并以在线/流式方式接收 $x_1, x_2, \cdots, x_n$，输出一个单位向量 u，使得

$$1 - \langle x^*, u\rangle^2 \leq (\log d)/R$$

以至少 $1 - \exp(-\Omega(\log d))$ 的概率成立。

内容概要。第2节中，我们概述技术概述。随后，第3节介绍所需的预备知识。第4节中，我们给出带有更新规则的流式核主成分分析算法的基本定义及一些性质。第5节中，我们分析流式核主成分分析算法，并得到一个理论结果。第6节中，我们作出结论。

2 技术概述

本节中，我们概述用于设计我们的算法的技术。总体而言，我们的算法结合了 Oja 的流式主成分分析算法[Oja82] 以及对其中应用核函数的一种新分析。

2.1 流式主成分分析

我们的第一种技术基于 Oja 用于流式主成分分析问题的传统方案。该算法基于赫布学习规则，该规则指出，如果两个神经元的活动相关联，它们之间的连接强度就应增强。在主成分分析的语境中，该算法沿当前数据点的方向更新主成分（PC）向量，但学习率会随时间递减。该算法旨在使 PC 向量收敛到数据协方差矩阵的主特征向量。这个特征向量对应于数据变化最显著的方向。利用这种方法，可以识别数据分布随时间发生的任何变化。形式上，当数据结构接收到数据点流 $x_1, \ldots, x_n \in \mathbb{R}^d$ 时，它会迭代更新向量 $v \in \mathbb{R}^n$ （从一个随机高斯向量开始），使得 $v_i = v_{i-1} + \eta \cdot x_i x_i^\top v_{i-1}$，其中$\eta \in \mathbb{R}$ 是学习率。最后，数据结构输出向量 $v_n = \prod_{i=1}^n (I_n + \eta x_i x_i^\top) v_0$，其中 $v_0 \sim \mathcal{N}(0, I_n)$。众所周知，以很高的概率，该输出向量接近第一主成分。

2.2 将核函数应用于流式主成分分析

Oja 的原始流式算法仅支持传统的线性主成分分析问题。我们希望将其推广到支持核函数。为此，我们需要克服以下几个障碍：

- 应该在哪里应用核函数？如前所述，我们需要将输入数据点“映射”到某个“核”空间中。但在流式场景下，如何处理数据流（这与离线算法不同）就成了一个问题。

- 流式算法能否结合核方法使用？经典的流式主成分分析算法大多适用于线性主成分分析问题。在这里应用核方法可能会遇到一些意想不到的障碍。

为克服这些障碍，我们提出了基于 Oja 算法推广而来的流式主成分分析算法。具体而言，给定一个核函数 ϕ: $\mathbb{R}^d \to \mathbb{R}^m$，我们的算法接收一个由数据点组成的数据流 x_1，……，$x_n \in \mathbb{R}^d$. 它首先在算法开始时生成一个随机高斯向量 $v_0 \in \mathbb{R}^m$，随后迭代更新一个向量 $v_i = v_{i-1} + \eta \cdot \langle \phi(x_i)$ $v_{i-1} \rangle \cdot \phi(x_i)$，其中 $\eta \in \mathbb{R}$ 是学习率。算法停止时，会输出一个向量 $v_n = \prod_{i=1}^{n}(I_n + \eta \cdot \phi(x_i)^\top \phi(x_i))$ $\cdot v_0$.通过对该算法进行分析，我们将证明，以高概率而言，这个向量 v_n 与定理 1.1 中所需的第一主成分足够接近。

2.3 特征值比率蕴含算法的存在

在传统的（线性）流式主成分分析算法中，研究表明，维护向量趋近主特征向量的速度取决于最大特征值与第二大特征值之间的关系。具体而言，若 λ_1 和 λ_2 是协方差矩阵的前两个特征值，我们定义 $R := \frac{\lambda_1}{\lambda_2}$ 为它们的比率。设 $\epsilon \in (0,0.1)$ 为误差参数，则可以保证，在经过 $O(\log_R(\frac{d}{\epsilon}))$ 次迭代后，$1 - \langle v_n, v^* \rangle = \sin^2(v_n, v^*) \leq \epsilon$。

在我们的核设置中，我们首次对流式主成分分析算法的这一收敛结果进行了分析。我们证明，当 $R \geq C \cdot \log n \cdot \log d$ 时，修改后的 Oja 算法（在其中加入核技巧）可为主成分分析问题提供一个 ϵ-解。

2.4 我们的分析方法概述

我们的分析方法可概括为以下几段。我们的证明概要主要遵循[MP22], ，同时我们在算法和分析的不同阶段应用核函数。

更新规则蕴含的性质。 根据我们算法的更新规则，即 $v_i = v_{i-1} + \eta x_i x_i^\top v_{i-1}$，我们首先证明维护向量具有若干简单但有用的性质（详细陈述和证明见命题 4.6 ），这些性质为后续分析奠定了基础。例如，我们证明，在迭代维护过程中，向量的范数持续增长，即 $\|v_i\|_2^2 \geq \|v_{i-1}\|_2^2$ 对于任意 $i \in [n]$，这一点（下一段将对此进行说明）非常有用，因为误差界中包含最终向量范数的倒数比例项。文献 [MP22] 证明了在传统设定下（不使用核函数），范数的增长有一个下界。我们沿用他们的方法并证明了一个核版本，即证明

$$\log(\|v_b\|_2^2 / \|v_a\|_2^2) \geq \eta \sum_{i=a+1}^{b} \langle \phi(x_i), \widehat{v}_{i-1} \rangle^2.$$

这些性质对于正确性证明至关重要，相关内容将在后续段落中介绍。

永不远离性质。如前所述，我们的算法会迭代地维护一个向量 v_i，使其收敛到协方差矩阵的最大特征向量 v^*（即第一主成分）。我们关注算法的收敛性和鲁棒性：当数据流以对抗方式到来时，例如将多个数据点置于某些特殊方向上，我们的算法是否仍能保证收敛？基于这一点，[MP22] 提出了一种方法，表明无论维护过程从何处开始，只要维护向量曾经接近目标 v^*，它就永远不会离目标太远。我们进一步进行了详细分析，证明即使采用核函数，这一结论仍然成立。具体来说，我们定义 $P := I - v^* v^{*\top} \in \mathbb{R}^{d\times d}$，那么对于任意 v_0 和 i，都有如下结果：

$$\|P\widehat{v}_i\|_2 \le \sqrt{\alpha} + \|Pv_0\|_2/\|v_i\|_2,$$

对于某个常数 α。由于我们的数据结构具有零记忆能力，即在某个时刻 i，其未来输出只取决于当前状态 v_i，而与过去无关v_j（对于 $j < i$），这意味着：一旦它接近目标，就永远不会离目标太远。我们称之为“永不远离”性质。这一结果还意味着，随着维护向量 ℓ_2 的范数 $\|v_i\|_2$增大，最终输出会更好。该性质已在引理C.1中得到形式化证明。

序列界。 根据引理 C.1，我们证明，如果某个时刻距离 v^*足够近，那么之后它与 v^*的距离永远不会超过 $\sqrt{\alpha}$ 。基于此，我们进一步证明，即使移动 $\sqrt{\alpha}$，也会增加 v的范数；也就是说，我们在引理 C.2 中证明，如果 $v_0 = v^*$，那么对任意两个步骤$0 \le a \le b \le n$，都有

$$\|P\widehat{v}_b - P\widehat{v}_a\|_2^2 \le 50 \cdot \alpha \cdot \log(\|v_b\|_2/\|v_a\|_2).$$

根据上述分析，我们得到如下结论：若要使最终输出接近目标，就需要让 $\|v_i\|_2$ 足够大。我们首先注意到，当 v_i 偏离我们期望的方向时，它会导致 $\|v_i\|_2$ 减小，即：

$$\|v_i\|_2 \ge \exp(\sum_{j\in[i]} \eta\langle\phi(x_j), \widehat{v}_{j-1}\rangle^2).$$

我们希望确保每一项 $\eta\langle\phi(x_j), \widehat{v}_{j-1}\rangle$ 对 $\|v_i\|$ 的影响都足够小，从而使 v_N 的最终范数足够大。因此，我们给出如下分解

$$\begin{aligned}&\langle\phi(x_j), \widehat{v}_{j-1}\rangle^2\\ \ge\ &\frac{1-\alpha}{2}\langle\phi(x_j), v^*\rangle^2 - \langle\phi(x_j), P\widehat{v}_{j-1}\rangle^2.\end{aligned}$$

因此，只需证明第二项足够小，不会破坏范数的增长。形式上，我们需要证明，如果 $v_0 = v^*$，那么对所有 $i \in [N]$，均有

$$\eta\sum_{i=1}^{n}\langle\phi(x_i), P\widehat{v}_{i-1}\rangle^2 \le 100 \cdot \alpha^2 \cdot \log^2 n \cdot \log\|v_n\|_2.$$

如前面的分析所示，这意味着，如果所维护的向量曾接近目标特征向量，那么这些乘积之和将有界，因此范数会持续增长。形式化表述见引理 C.5。

下界。 在 [MP22], 中，他们给出了输出向量范数的下界。我们通过在此处应用核函数，将他们的方法加以推广。证明的下一步是求最终输出范数的下界。我们的方法如下。首先证明，Claim 4.6 中的性质推出 $\|v_n\|_2$的下界结果。我们在引理 C.7 中证明:

$$\|v_n\|_2 \geq \sqrt{\eta} \cdot (\sum_{i\in[n]} \langle \phi(x_i), v_{i-1}\rangle^2)^{1/2}.$$

将这些结果与引理 C.5结合起来，我们得到

$$\log(\|v_n\|_2) \geq \frac{\eta \sum_{i\in[n]} \langle v^*, \phi(x_i)\rangle^2}{8 + 8 \cdot C \cdot \alpha^2 \log^2 n},$$

它们给出了输出向量范数的下界。形式化证明见附录中的引理C.6 。

3 预备知识

在第 3.1 节中，我们给出用于概率计算的马尔可夫不等式。在第 3.2 节中，我们介绍一些有用的代数工具。在第 3.3 节中，我们展示高斯分布的性质。

记号。对于矩阵 A，我们用 $A^\top$ 表示其转置。对于方阵 A，我们用 $\mathrm{tr}[A]$ 表示其迹。对于向量 $x \in \mathbb{R}^n$，我们用 $\|x\|_2$ 表示其 ℓ_2 范数，即$\|x\|_2 := (\sum_{i=1}^n x_i^2)^{1/2}$。

如果方阵 $P \in \mathbb{R}^{d\times d}$满足 $P^2 = P$，我们称其为投影矩阵。

对于两个函数 f、g，我们用简写 $f \lesssim g$ （分别地， $\gtrsim$）表示：对于某个绝对常数 C， $f \leq Cg$ （分别地， $\geq$）。我们用 $f \approx g$ 表示对于常数 $c > 0$ 和 $C > 0$， $cf \leq g \leq Cf$。

对于定义域为 X的函数 $h(j)$，我们用 $\arg\max_{j\in X} h(j)$ 表示函数 $h(j)$ 取得最大输出值时对应的索引 j。

我们用 $\mathbb{E}[\cdot]$ 表示期望，用 $\Pr[\cdot]$ 表示概率。

对于一个分布 D 和一个随机变量 x，我们用 $x \sim D$ 表示从分布 D中抽取一个随机变量。

我们用 $\mathcal{N}(\mu, \sigma^2)$ 表示均值为 μ、方差为 σ^2的高斯分布。

定义 3.1. 设 $\phi : \mathbb{R}^d \to \mathbb{R}^m$ 表示一个核函数。我们定义 $\Sigma := \frac{1}{n}\sum_{i=1}^n \phi(x_i)\phi(x_i)^\top$.

3.1 基础概率工具

引理 3.2（马尔可夫不等式）。如果 X 是一个非负随机变量，且 $a > 0$，那么

$$\Pr[X \geq a] \leq \mathbb{E}[X]/a.$$

3.2 基础代数工具

命题 3.3 (命题的重述 B.1)。 令 $P = (I - v^*(v^*)^\top)$ 其中 $P \in \mathbb{R}^{d\times d}$。令 $u \in \mathbb{R}^d$表示任意单位向量 $\|u\|_2 = 1$，若 $\|Pu\|_2 \leq \epsilon$，则有

$$1 - \langle u, v^*\rangle^2 \leq \epsilon^2.$$

事实 3.4 (事实的重述 B.2)。 对于任意整数 A，以及整数 k，定义 $f_k := \lfloor A/2^k \rfloor$以及 $f_{k+1} := 2 \cdot \lfloor A/2^{k+1} \rfloor$。于是有

$$|f_k - f_{k+1}| \leq 1.$$

断言 3.5 (断言 B.4的重述)。 对于任意 $x \in \mathbb{R}$， $y \in \mathbb{R}$，我们有

$$(x+y)^2 \geq \frac{1}{2}x^2 - y^2.$$

3.3 随机高斯分布的基本性质

断言 3.6 (断言 B.6的重述)。 令 $a \sim \mathcal{N}(0,1)$。

对于任意两个向量 $u \in \mathbb{R}^d$ 和 $v \in \mathbb{R}^d$，我们有

$$\Pr_{a \sim \mathcal{N}(0,1)}[\|au + v\|_2 \geq \delta \|u\|_2] \geq 1 - \delta.$$

4 流式核主成分分析算法和更新规则的基本定义与性质

本节中 n，我们给出用于证明主要定理 1.1的论述。

在第 4.1节中，我们定义用于核主成分分析的样本向量。在第 4.2节中，我们为流式算法提供更新规则。在第 4.3节中，我们进一步介绍由更新规则所蕴含的一些性质。

4.1 向量的定义

我们将 α、 $\eta > 0$ 、 $v^* \in \mathbb{R}^d$ 和 $\beta > 0$ 正式定义如下:

定义 4.1. 令 β 和 α 表示两个 $\beta \geq \alpha > 0$ 参数。

对于每个 $i \in [n]$，我们用 $x_i \in \mathbb{R}^d$ 表示该样本。令 $\eta \in$ (0, 0.1) 为学习率。我们将向量 $v^* \in \mathbb{R}^d$ 定义如下:

- $\|v^*\|_2 = 1$,
- $\eta \sum_{i=1}^n \langle v^*, \phi(x_i) \rangle^2 = \beta$,
- f或者，对于所有满足 w with $\|w\|_2 \leq 1$ and $\langle w, v^* \rangle = 0$ 的向量，我们有 $\eta \sum_{i=1}^n \langle w, \phi(x_i) \rangle^2 \leq \alpha$.

不失一般性，在整个算法分析过程中，我们固定 $\|v^*\|_2 = 1$ 。我们基于 v^*定义投影算子。

定义 4.2。我们将 $P = I - v^*(v^*)^\top$ 定义为移除 v^* 分量的投影矩阵。

我们有如下结论。

断言 4.3。由于 $P = I - v^*(v^*)^\top$ 且 $\|v^*\|_2 = 1$，因此我们有 $Pv^* = 0$。

证明。我们有

$$\begin{aligned} P \cdot v^* = (I - v^*(v^*)^\top)v^* = & \ v^* - v^* \cdot \|v^*\|_2^2 \\ = v^* - v^* = & \ 0 \end{aligned}$$

其中第二步根据向量的 ℓ_2 范数的定义得出。

至此，我们完成了证明。 □

4.2 更新规则

定义 4.4（　　更新规则）。令 η 表示某个参数。我们将更新规则定义如下：

$$v_i := v_{i-1} + \eta\langle\phi(x_i), v_{i-1}\rangle\phi(x_i).$$

然后，我们可以将其改写为

$$v_i = (I + \eta\phi(x_i)\phi(x_i)^\top)v_{i-1}.$$

为了保证稳定性，实现时只会跟踪归一化向量 $\widehat{v}_i = v_i/\|v_i\|_2$。出于分析目的，我们通常会考虑未归一化向量 v_i。

定义 4.5。设 v_i 表示所有 $i \in [n]$ 对应的未归一化向量。我们定义 $\widehat{v}_i$ 如下

$$\widehat{v}_i := v_i/\|v_i\|_2.$$

4.3 更新规则蕴含的性质

命题 4.6。 对于任意参数 $\eta > 0$。根据 v_i 与 v_{i-1} (见定义 4.4) 之间的关系，我们有

- 性质 1。

$$\begin{aligned}\|v_i\|_2^2 = &\ \|v_{i-1}\|_2^2 \\ &\cdot (1 + (2\eta + \eta^2\|\phi(x_i)\|_2^2) \cdot \langle\phi(x_i), \widehat{v}_{i-1}\rangle^2)\end{aligned}$$

- 性质 2。

$$\|v_i\|_2^2 \geq \|v_{i-1}\|_2^2, \forall i \in [n]$$

- 性质 3。若进一步假设 $\eta \leq 0.1/\max_{i\in[n]} \|\phi(x_i)\|_2^2$,

$$\log(\|v_i\|_2^2/\|v_{i-1}\|_2^2) \geq \eta\langle\phi(x_i), \widehat{v}_{i-1}\rangle^2.$$

- 性质 4。

$$\log(\|v_b\|_2^2/\|v_a\|_2^2) \geq \sum_{i=a+1}^{b} \eta\langle\phi(x_i), \widehat{v}_{i-1}\rangle^2$$

- 性质 5。对于任意整数 $b > a$

$$v_b - v_a = \sum_{i=a+1}^{b} \eta\phi(x_i)\phi(x_i)^\top v_{i-1}$$

证明。Pr 性质 1 的证明。 将定义 4.4 两边平方，得到

$$\begin{aligned}v_i^2 = &\ v_{i-1}^2 \\ &+ 2\eta v_i\langle\phi(x_i), v_{i-1}\rangle\phi(x_i) + \eta^2\langle\phi(x_i), v_{i-1}\rangle^2\phi^2(x_i).\end{aligned}$$

由于 $v^2 = \|v\|_2^2$，我们将其改写为

$$\begin{aligned}
&\|v_i\|^2 \\
=\ &\|v_{i-1}\|_2^2 \\
+\ &2\eta v_{i-1}\langle\phi(x_i), v_{i-1}\rangle\phi(x_i) + \eta^2\langle\phi(x_i), v_{i-1}\rangle^2\|\phi(x_i)\|_2^2 \\
=\ &\|v_{i-1}\|_2^2 \\
+\ &2\eta\langle\phi(x_i), v_{i-1}\rangle^2 + \eta^2\langle\phi(x_i), v_{i-1}\rangle^2\|\phi(x_i)\|_2^2 \\
=\ &\|v_{i-1}\|_2^2 \cdot (1 + (2\eta + \eta^2\|\phi(x_i)\|_2^2) \cdot \langle\phi(x_i), \widehat{v}_{i-1}\rangle^2)
\end{aligned}$$

其中最后一步由定义 4.5得出。

性质 2 的证明。该命题的证明将在某一步将性质 1 作为黑箱使用。我们先考虑项 $(2\eta + \eta^2\|\phi(x_i)\|_2^2)$ 和 $\langle\phi(x_i)\ \widehat{v}_{i-1}\rangle^2$。

对于 $(2\eta + \eta^2\|\phi(x_i)\|_2^2)$，我们有 $\|\phi(x_i)\|_2^2 \geq 0$。

根据定义 4.1，我们得到 $2\eta > 0$ 和 $\eta^2 > 0$。因此，

$$2\eta + \eta^2\|\phi(x_i)\|_2^2 > 0.$$

对于 $\langle\phi(x_i)\widehat{v}_{i-1}\rangle^2$，显然该项大于或等于 0。因此，我们有 ave

$$\langle\phi(x_i), \widehat{v}_{i-1}\rangle^2 \geq 0.$$

因此，我们得出

$$\begin{aligned}
\|v_i\|^2 &= \|v_{i-1}\|_2^2 \cdot (1 + (2\eta + \eta^2\|\phi(x_i)\|_2^2) \cdot \langle\phi(x_i), \widehat{v}_{i-1}\rangle^2) \\
&\geq \|v_{i-1}\|_2^2 \cdot (1 + 0) \\
&= \|v_{i-1}\|_2^2,
\end{aligned}$$

其中，第二步由不等式关系和 $i \in [n]$ 得出。

性质 3 的证明。由性质 1 可得

$$\frac{\|v_i\|_2^2}{\|v_{i-1}\|_2^2} = 1 + (2\eta + \eta^2\|\phi(x_i)\|_2^2) \cdot \langle\phi(x_i), \widehat{v}_{i-1}\rangle^2.$$

两边取对数，得

$$\log(\frac{\|v_i\|_2^2}{\|v_{i-1}\|_2^2}) = \log(1 + (2\eta + \eta^2\|\phi(x_i)\|_2^2) \cdot \langle\phi(x_i), \widehat{v}_{i-1}\rangle^2).$$

我们定义 $u = (2\eta + \eta^2\|\phi(x_i)\|_2^2) \cdot \langle\phi(x_i)\ \widehat{v}_{i-1}\rangle^2$。我们需要证明 $u \in [0, 1.5]$。

对于下界情形，显然 $u \geq 0$，因为 $\eta \geq 0$。

接下来，我们证明上界情形，

$$\begin{aligned}
&u \\
=\ &(2\eta + \eta^2\|\phi(x_i)\|_2^2) \cdot \langle\phi(x_i), \widehat{v}_{i-1}\rangle^2 \\
=\ &(2\eta + \eta^2\|\phi(x_i)\|_2^2) \cdot \|\phi(x_i)\|_2^2 \cdot \langle\phi(x_i)/\|\phi(x_i)\|_2, \widehat{v}_{i-1}\rangle^2 \\
\leq\ &(2\eta + \eta^2\|\phi(x_i)\|_2^2) \cdot \|\phi(x_i)\|_2^2, \\
\leq\ &2 \cdot 0.1 + 0.1^2 \\
\leq\ &0.3
\end{aligned}$$

其中，第三步由 $\langle a,b\rangle^2 \le 1$ 得出，适用于任意 $\|a\|_2 = \|b\|_2 = 1$；第四步由$\eta \le 0.1/\|\phi(x_i)\|_2^2$ 得出。

不难看出，对于任意 $u \in [0, 1.5]$

$$\log(1+u) \ge 0.25 \cdot u.$$

因此，

$$\begin{aligned}\log(1+u) &\ge 0.25 \cdot (2\eta + \eta^2 \|\phi(x_i)\|_2^2) \cdot \langle \phi(x_i), \widehat{v}_{i-1}\rangle^2 \\ &\ge 0.5\eta \langle \phi(x_i), \widehat{v}_{i-1}\rangle^2.\end{aligned}$$

性质 4 的证明。根据性质 3，我们有

$$\log(\|v_i\|_2^2/\|v_{i-1}\|_2^2) \ge \eta \langle \phi(x_i), \widehat{v}_{i-1}\rangle^2.$$

$\forall a$、 $b \in [n]$ 和 $a < b$，我们有

$$\begin{aligned}&\log(\|v_b\|_2^2/\|v_a\|_2^2) \\ =\ &\log(\frac{\|v_b\|_2^2}{\|v_{b-1}\|_2^2} \cdot ... \cdot \frac{\|v_{a+1}\|_2^2}{\|v_a\|_2^2}) \\ =\ &\log(\frac{\|v_b\|_2^2}{\|v_{b-1}\|_2^2}) + ... + \log(\frac{\|v_{a+1}\|_2^2}{\|v_a\|_2^2}) \\ \ge\ &\eta \langle \phi(x_b), \widehat{v}_{b-1}\rangle^2 + ... + \eta \langle \phi(x_{a+1}), \widehat{v}_a\rangle^2 \\ =\ &\sum_{i=a+1}^{b} \eta \langle \phi(x_i), \widehat{v}_{i-1}\rangle^2\end{aligned}$$

其中，第二步由 $\log(ab) = \log(a) + \log(b)$ 得出，第三步由性质 3 得出。

P性质 5 的证明。根据定义。4.4，我们有 $v_i = (I + \eta\phi(x_i)\phi(x_i)^\top)v_{i-1}$。
我们将其改写为

$$\begin{aligned}v_i - v_{i-1} &= (I + \eta\phi(x_i)\phi(x_i)^\top)v_{i-1} - v_{i-1} \\ &= \eta\phi(x_i)\phi(x_i)^\top v_{i-1}, \end{aligned} \tag{1}$$

其中第一步由定义得出。 4.4。
然后，对于 $\forall a$、 $b \in [n]$ 和 $a < b$，我们有

$$\begin{aligned}&v_b - v_a \\ =\ &v_b - v_{b-1} + ... + v_{a+1} + v_a \\ =\ &\eta\phi(x_b)\phi(x_b)^\top v_{b-1} + ... + \eta\phi(x_{a+1})\phi(x_{a+1})^\top v_a \\ =\ &\sum_{i=a+1}^{b} \eta\phi(x_i)\phi(x_i)^\top v_{i-1}\end{aligned}$$

其中第二步由等式 (1) 得出。 □

5 我们的核主成分分析结果

5.1 最终输出的保证

定理 5.1。设 $C \geq 1000$ 为一个足够大的常数。假设 $\alpha \in (0, C^{-1} \cdot \log n)$ 且 $\beta \geq C \log d$。我们的算法输出一个向量 $\widehat{v}_n \in \mathbb{R}^d$，使得

$$\Pr[\|P\widehat{v}_n\|_2 \leq \sqrt{\alpha} + \exp(-\beta/200)] \geq 1 - \exp(-\beta/200)$$

证明。我们的算法从一个均匀随机方向 $\widehat{v}_0$ 开始，而 $\widehat{v}_i$ 的序列不依赖于 $\|v_0\|_2$，因此我们可以假设 $v_0 \sim \mathcal{N}(0, I_d)$。

根据这一假设，我们知道对于每个 $i \in [d]$，$v_0 \sim \mathcal{N}(0,1)$。因此，我们对序列 $\widehat{v}_i$ 的所有初始向量 v_0 求和，得到

$$\mathbb{E}[\|v_0\|_2^2] = \sum_{i=1}^{d} \mathbb{E}[\|(v_0)_i\|_2^2] = \sum_{i=1}^{d} 1 = d$$

其中，第一步根据我们的证明假设得出，第二步则根据高斯分布的定义得出。

我们定义向量 $v_0 \in \mathbb{R}^d$

$$v_0 := a \cdot v^* + u$$

对于 $u \perp v^*$ 和 $a \sim \mathcal{N}(0,1)$。

我们定义矩阵 $B \in \mathbb{R}^{d \times d}$

$$B := \prod_{i=1}^{n} (1 + \eta \cdot \phi(x_i) \cdot \phi(x_i)^\top),$$

因此，根据定义 4.4 (更新规则)，

$$v_n = B v_0.$$

以 $1 - \delta$ 的概率，我们得到

$$\begin{aligned} \|v_n\|_2 &= \|Bv_0\|_2 \\ &= \|aBv^* + Bu\|_2 \\ &\geq \delta \cdot \|Bv^*\|_2 \\ &\geq \delta \cdot \exp(\beta/20) \end{aligned} \tag{2}$$

其中，第一步由 $v_n = Bv_0$ 得出，第二步由 $v_0 = av^* + u$ 得出，第三步由断言 3.6 得出，最后一步由引理C.6 得出。

我们可以计算期望，

$$\begin{aligned} \mathbb{E}[\|u\|_2^2] = &\ \mathbb{E}[\|v_0\|_2^2 - \|av^*\|_2^2 - 2\langle av^*, u\rangle] \\ = &\ \mathbb{E}[\|v_0\|_2^2] - \mathbb{E}[\|av^*\|_2^2] - \mathbb{E}[2\langle av^*, u\rangle] \\ = &\ d - \mathbb{E}[\|av^*\|_2^2] - \mathbb{E}[2\langle av^*, u\rangle] \\ = &\ d - 1 - \mathbb{E}[2\langle av^*, u\rangle] \\ = &\ d - 1 \end{aligned}$$

其中，第一步根据我们对证明的定义得出： $v_0 := a \cdot v^* + u$；第二步根据简单代数得出；第三步根据高斯分布的定义得出；第四步根据$\mathbb{E}[a^2] = 1$ 和 $\|v^*\|_2^2 = 1$得出；最后一步根据 $\langle u^*、\ u\rangle = 0$得出。

然后应用引理3.2，可得

$$\begin{aligned}\Pr[\|u\|_2^2 \geq d/\delta] \leq\ & \mathbb{E}[\|u\|_2^2]/(d/\delta) \\ =\ & (d-1)\frac{\delta}{d} \\ \leq\ & \delta \end{aligned} \tag{3}$$

最后一步由 $(d-1)/d \leq 1$ 得出。

上述方程意味着

$$\Pr[\|u\|_2 \leq \sqrt{d/\delta}] \geq 1 - \delta.$$

以 $1 - 3\delta$ 的概率，有

$$\begin{aligned}\|P\widehat{v}_n\|_2 &\leq \sqrt{\alpha} + \frac{\|u\|_2}{\|v_n\|_2} \\ &\leq \sqrt{\alpha} + \frac{\sqrt{d/\delta}}{\|v_n\|_2} \\ &\leq \sqrt{\alpha} + \frac{\sqrt{d/\delta}}{\delta \cdot \exp(\beta/20)} \\ &\leq \sqrt{\alpha} + 8 \cdot \sqrt{d} \cdot \exp(-\beta/30) \\ &\leq \sqrt{\alpha} + \exp(-\beta/40) \\ &\leq \sqrt{\alpha} + \exp(-\beta/200)\end{aligned}$$

其中，第一步由引理 C.1 得出，第二步由方程(3)得出，第三步由方程(2)得出，第四步由选择 $\delta = \exp(-\beta/200)/4$ 得出，第五步由 $\beta \geq C$ 对数 d 得出，其中 $C \geq 500$。

□

5.2 主要结果

定理 5.2（定理 1.1 的形式化版本）。设 $\phi: \mathbb{R}^d \to \mathbb{R}^m$。设 $\Sigma = \frac{1}{n}\sum_{i=1}^n \phi(x_i)\phi(x_i)^\top \in \mathbb{R}^{m\times m}$。我们定义 $R := \lambda_1(\Sigma)/\lambda_2(\Sigma)$，其中 $\lambda_1(\Sigma)$ 是 Σ 的最大特征值，而 $\lambda_2(\Sigma)$ 是 Σ的第二大特征值。令 x^* 表示 Σ的最大特征向量。令 $C > 10^4$ 表示一个足够大的常数。如果 $R \geq C \cdot (\log n) \cdot (\log d)$，则存在一个流式算法（算法 1），仅使用 $O(d)$ 空间，并以在线/流式方式接收 $x_1, x_2, \cdots, x_n$，输出一个单位向量 u，使得

$$1 - \langle x^*, u\rangle^2 \leq (\log d)/R$$

以至少 $1 - \exp(-\Omega(\log d))$ 的概率成立。

证明。令 $C \geq 1000$为一个充分大的常数。假设 $\alpha \in (0,\ C^{-1} \text{对数}\ n)$ 且 $\beta \geq C$~~对数~~ d。由定理5.1，我们有

$$\|Pu\|_2 \leq \epsilon$$

其中 $\epsilon = \sqrt{\alpha} + \exp(-\beta/200)$。

根据命题 3.3，我们知道

$$1 - \langle u, x^* \rangle^2 \leq \epsilon^2.$$

根据我们的证明假设，可得

$$R \geq C \cdot (\log n) \cdot (\log d) \geq \frac{1}{4} C \cdot (\log n) \cdot (\log d) \tag{4}$$

其中，第二步由 $C \cdot (\text{对数 } n) \cdot (\text{对数 } d) \geq 0$ 得出。

重写公式（4），得到

$$\frac{1}{4}(\log d)/R \leq \frac{1}{C \log n}.$$

因此，我们可以选择

$$\alpha = \frac{1}{4}(\log d)/R \tag{5}$$

根据其定义域 $\alpha \in (0, C \frac{1}{\text{对数 } n})$。

式（5）等价地可得

$$\sqrt{\alpha} = \frac{1}{2}\sqrt{(\log d)/R}.$$

由于根据定义有 $R \geq 1$，且我们选择

$$\beta \geq C \log(R/(\log d)),$$

then

$$\begin{aligned} \exp(-\beta/200) &\leq \exp(-(C/200)\log(R/\log d)) \\ &\leq ((\log d)/R)^2 \\ &\leq \frac{1}{2}\sqrt{(\log d)/R}. \end{aligned}$$

其中，第二步由 $C/200 \geq 4$ 得出，最后一步由 $R \geq 4\text{对数 } d$ 得出。因此，我们有

$$\begin{aligned} \epsilon &\leq \frac{1}{2}\sqrt{(\log d)/R} + \frac{1}{2}\sqrt{(\log d)/R} \\ &= \sqrt{(\log d)/R}, \end{aligned}$$

其中，第一步由 $\epsilon = \sqrt{\alpha} + \exp(-\beta/200)$ 得出。

两边平方后，可得

$$\epsilon^2 \leq (\log d)/R.$$

因此，对于总体条件，我们选择 β 为

$$\beta \geq C \cdot (\log d + \log(R/\log d)).$$

由公式（5）可知， R 必须满足

$$R \geq (C/4) \log n \cdot \log d.$$

失败概率至多为

$$\begin{aligned} & \exp(-\beta/200) \\ \leq\ & \exp(-(C/200)\log(d) - \log((C/4)\log n)) \\ \leq\ & \exp(-\Omega(\log d)). \end{aligned}$$

因此，我们得出结论：满足条件 $1 - \langle x^*, u \rangle^2 \leq$(对数 d)$/R$ 的概率至少为 $1 -$ exp(−Ω(对数 d)), 正如预期。 □

6 结论

总之，我们的工作提出了一种适用于核主成分分析的流式算法，该算法仅需 $O(m)$ 的空间，其中 m 是核空间的维度。我们的工作扩展（推广）了 Oja 提出的传统方案，并给出了算法能够良好运行时顶部特征向量之间的比率条件。我们相信，我们的工作还有望影响学习算法和在线算法中的其他基础性问题。

附录

内容概要。 在第 A 节中，我们将提供更多符号说明。在第 B 节中，我们将为理论证明提供代数工具。在第 C 节中，我们将提供核主成分分析算法的更多证明和分析。

更多记号

对于任意函数 $f(x) \in \mathbb{R}$和 $g(x) \in \mathbb{R}$，若 $\exists M \in \mathbb{R}^+$和 $x_0 \in \mathbb{R}$，使得对所有 $x > x_0$，$|f(x)| \leq M \cdot g(x)$。我们记作 $f(x) = O(g(x))$。

对于任意函数 $f(x) \in \mathbb{R}$和 $g(x) \in \mathbb{R}$，若 $\exists k \in \mathbb{R}^+$和 $x_1 \in \mathbb{R}$，使得对所有 $x > x_1$，$|f(x)| \geq k \cdot g(x)$。我们记作 $f(x) = \Omega(g(x))$。

对于任意函数 $f(x) \in \mathbb{R}$和 $g(x) \in \mathbb{R}$，若 $f(x) = O(g(x))$ 且 $f(x) = \Omega(g(x))$，我们记作 $f(x) = \Theta(g(x))$。

B 更多代数工具

命题 B.1 (命题 3.3的重述)。 设 $P = (I - v^*(v^*)^\top)$ 其中 $P \in \mathbb{R}^{d \times d}$。设 $u \in \mathbb{R}^d$表示任意单位向量 $\|u\|_2 = 1$，若 $\|Pu\|_2 \leq \epsilon$，则有

$$1 - \langle u, v^* \rangle^2 \leq \epsilon^2.$$

证明。我们有

$$\begin{aligned}
\epsilon^2 &\geq \|Pu\|_2^2 \\
&= u^\top P P u \\
&= u^\top P u \\
&= u^\top u - u^\top v^* (v^*)^\top u \\
&= 1 - \langle u, v^* \rangle^2
\end{aligned}$$

其中，第一步由证明所需的假设得出，第二步由范数的性质得出，第三步由投影矩阵 $P^2 = P$ 的定义得出，第四步由我们在证明中对 $P = (I - v^*(v^*)^\top)$ 的定义得出，最后一步由 $a^\top b = \langle a, b \rangle$ 得出。

□

事实 B.2 (事实 3.4的重述)。 对于任意整数 A，以及整数 k，定义 $f_k := \lfloor A/2^k \rfloor$以及 $f_{k+1} := 2 \cdot \lfloor A/2^{k+1} \rfloor$。那么，我们有

$$|f_k - f_{k+1}| \leq 1$$

证明。我们总可以写成 A

$$A = B \cdot 2^{k+1} + C \cdot 2^k + D$$

其中 $B \geq 0$、 $C \in \{0, 1\}$ 和 $D \in [0, 2^k - 1]$。

我们有

$$|f_k - f_{k+1}| = |(2B + C) - 2B| = C \leq 1$$

至此，证明完成。

□

命题 B.3. 令 $0 \leq a_1, a_2, ..., a_n$. 对于每个 $i \in \{0, 1, \cdots, n\}$，我们定义

$$b_i := \exp(\sum_{j=0}^{i} a_j)$$

对于 $i \in \{0, 1, ..., n\}$。

那么：

$$\sum_{i=1}^{n} a_i b_{i-1} \leq b_n.$$

证明。由对 n 的归纳法可得。 $n = 0$ 是显然的，接下来对于 $k \in \{0, 1, ..., n\}$ 和 $k < n$，我们针对 $k + 1 \in \{0, 1, ..., n\}$ 有以下情形。

$$\begin{aligned}
\sum_{i=1}^{k+1} a_i b_{i-1} &\leq b_k + a_{k+1} b_k \\
&= (1 + a_{k+1}) b_k \\
&\leq e^{a_{k+1}} b_k \\
&\leq b_k,
\end{aligned}$$

其中，第一步由归纳法得出，第二步由乘法分配律得出，第三步由指数函数的麦克劳林级数得出，最后一步则根据我们的证明定义得出。

□

命题 B.4 (命题的重述 3.5)。 对于任意 $x \in \mathbb{R}$， $y \in \mathbb{R}$，我们有

$$(x + y)^2 \geq \frac{1}{2} x^2 - y^2.$$

证明。这等价于

$$x^2 + 2xy + y^2 \geq \frac{1}{2} x^2 - y^2,$$

这等价于

$$\frac{1}{2} x^2 + 2xy + 2y^2 \geq 0,$$

这进一步等价于

$$\frac{1}{2} (x + 2y)^2 \geq 0.$$

至此，我们完成了证明。 □

引理 B.5 (高斯分布反集中性，示例见 [SY19] 中的引理 A.4)。 令 $X \sim \mathcal{N}(0, \sigma^2)$，即 X 的概率密度函数为 $\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-x^2/(2\sigma^2))$。则

$$\frac{2}{3} t/\sigma \leq \Pr[|X| \leq t] \leq \frac{4}{5} t/\sigma.$$

命题 B.6 (命题 3.6 的形式化陈述)。 令 $a \sim \mathcal{N}(0,1)$。

对于任意两个向量 $u \in \mathbb{R}^d$ 和 $v \in \mathbb{R}^d$，我们有

$$\Pr_{a \sim \mathcal{N}(0,1)}[\|au+v\|_2 \geq \delta \|u\|_2] \geq 1-\delta.$$

证明。定义

$$x := \|au+v\|_2^2.$$

情形 1。存在某个标量 $b \in \mathbb{R}$，使得 $v = b \cdot u$。于是我们有

$$x = (a+b)^2 \|u\|_2^2.$$

回顾一下，本引理的目标是证明

$$\Pr_{a \sim \mathcal{N}(0,1)}[\sqrt{x} \geq \delta \|u\|_2] \geq 1-\delta.$$

这等价于

$$\Pr_{a \sim \mathcal{N}(0,1)}[x \geq \delta^2 \|u\|_2^2] \geq 1-\delta.$$

利用 $x = (a+b)^2\|u\|_2^2$ 的方程，该陈述等价于

$$\Pr_{a \sim \mathcal{N}(0,1)}[(a+b)^2 \|u\|_2^2 \geq \delta^2 \|u\|_2^2] \geq 1-\delta,$$

这等价于

$$\Pr_{a \sim \mathcal{N}(0,1)}[(a+b)^2 \geq \delta^2] \geq 1-\delta.$$

根据高斯分布的性质，我们知道

$$\Pr_{a \sim \mathcal{N}(0,1)}[(a+b)^2 \geq \delta^2] \geq \Pr_{a \sim \mathcal{N}(0,1)}[(a+0)^2 \geq \delta^2].$$

因此，我们只需证明

$$\Pr_{a \sim \mathcal{N}(0,1)}[a^2 \geq \delta^2] \geq 1-\delta.$$

上述方程直接由引理 B.5。

情形 2。存在某个标量 b 和向量 w，使得 $\langle u、 w \rangle = 0$ 以及

$$v = b \cdot u + w.$$

在这种情形下，

$$\begin{aligned} x &= \|(a+b)u+w\|_2^2 \\ &= (a+b)^2\|u\|_2^2 + \|w\|_2^2 \\ &> (a+b)^2\|u\|_2^2. \end{aligned}$$

其余证明与情形 1 相同，因为 x 现在正在变大 . □

我们的核主成分分析算法的 C 分析

章节 C.1，我们给出所定义向量的性质：增长蕴含正确性。
章节 C.2中，我们给出投影算子，并展示所定义向量的范数递增性质。

在 C.3 节中，我们给出序列界。
在 C.4 节中，我们给出内积求和的上界。
在 C.5 节中，我们给出流式算法最终输出的范数的对数下界。

在第 C.6 节中，我们给出了 ℓ_2 我们算法生成的最终向量的范数下界。

算法 1 我们的流式核主成分分析算法

过程 KernelPCA(n, d, m, ϕ) ▷ 定理1.1
  $v_0 \sim \mathcal{N}(0, I_m)$
  对于 $i = 1 \to n$，执行
    接收 x_i
    $v_i \leftarrow v_{i-1} + \eta \cdot \langle \phi(x_i), v_{i-1} \rangle \cdot \phi(x_i)$
  结束循环
  $u \leftarrow v_n$
  返回 u
结束过程

C.1 增长蕴含正确性

引理 C.1（增长蕴含正确性）。对于任意 v_0 和所有 $i \in [n]$，我们有

$$\|P\widehat{v}_i\|_2 \leq \sqrt{\alpha} + \|Pv_0\|_2 / \|v_i\|_2.$$

此外，如果 $v_0 = v^*$，那么我们有

$$\|P\widehat{v}_i\|_2 \leq \sqrt{\alpha}.$$

证明。我们将针对最终索引 $i = n$ 证明这一点。
不失一般性，在整个证明过程中，我们可以假设 $\|v_0\|_2 = 1$ 。
那么，对于任意单位向量 $w \perp v^*$，

$$\begin{aligned}
& \langle v_n - v_0, w \rangle \\
= & \eta \sum_{i=1}^n \langle \phi(x_i), v_{i-1} \rangle \langle \phi(x_i), w \rangle \\
\leq & \eta (\sum_{i=1}^n \langle \phi(x_i), v_{i-1} \rangle^2)^{1/2} \cdot (\sum_{i=1}^n \langle \phi(x_i), w \rangle^2)^{1/2} \\
\leq & \|v_n\|_2 \cdot \sqrt{\eta} \cdot (\sum_{i=1}^n \langle \phi(x_i), w \rangle^2)^{1/2} \\
\leq & \|v_n\|_2 \cdot \sqrt{\alpha} \qquad (6)
\end{aligned}$$

其中，第一步由命题4.6的性质 5 得出，第二步由柯西-施瓦茨不等式得出，第三步由引理C.7得出，最后一步由定义4.1得出。因此

$$\begin{aligned}\langle \widehat{v}_n, w\rangle &\le \frac{1}{\|v\|_2}\langle v_n, w\rangle \\ &= \frac{1}{\|v_n\|_2}(\langle v_n - v_0, w\rangle + \langle v_0, w\rangle) \\ &\le \sqrt{\alpha} + \frac{\langle v_0, w\rangle}{\|v_n\|_2}. \end{aligned} \tag{7}$$

其中，第一步根据 $\widehat{v}_n$的定义得出，第二步通过减去并加上同一项得出，第三步则由式（6）得出。

令 $w = P\widehat{v}_n/\|P\widehat{v}_n\|_2$，我们有

$$\begin{aligned}\langle \widehat{v}_n, w\rangle &= \langle \widehat{v}_n, P\widehat{v}_n/\|P\widehat{v}_n\|_2\rangle \\ &= \langle \widehat{v}_n, P^2\widehat{v}_n/\|P\widehat{v}_n\|_2\rangle \\ &= \widehat{v}_n^\top P^2\widehat{v}_n/\|P\widehat{v}_n\|_2 \\ &= \|P\widehat{v}_n\|_2 \end{aligned} \tag{8}$$

其中，第二步根据 P 是投影矩阵（这意味着 $P^2 = P$）得出，第三步根据欧几里得向量空间内积的性质得出，最后一步则根据对于任意矩阵 B 和向量 a， $a^\top B^2 a = \|Ba\|_2^2$ 得出。

我们还知道

$$\begin{aligned}\langle v_0, w\rangle &= \langle v_0, P\widehat{v}_n/\|P\widehat{v}_n\|_2\rangle \\ &= \langle Pv_0, P\widehat{v}_n/\|P\widehat{v}_n\|_2\rangle \\ &\le \|Pv_0\|_2 \cdot \|P\widehat{v}_n\|_2/\|P\widehat{v}_n\|_2 \\ &\le \|Pv_0\|_2, \end{aligned} \tag{9}$$

其中，第二步根据 P 是投影矩阵（这意味着 $P^2 = P$）得出，第三步根据 $\langle a、\ b\rangle \le \|a\|_2 \cdot \|b\|_2$ 得出。

现在，我们可以得出结论

$$\begin{aligned}\|P\widehat{v}_n\|_2 &= \langle \widehat{v}_n, w\rangle \\ &\le \sqrt{\alpha} + \frac{\langle v_0, w\rangle}{\|v_n\|_2} \\ &\le \sqrt{\alpha} + \|Pv_0\|_2/\|v_n\|_2 \end{aligned}$$

其中第一步由式（8）得出，第二步由式（7）得出，最后一步由式（9）得出。

对于 $v_0 = v^*$ 的情况，由于 $Pv^* = 0$，我们得到 $\|P\widehat{v}_i\|_2 \le \sqrt{\alpha}$，如所需。
因此，我们完成了证明。 □

C.2 投影算子

利用引理 C.1，我们证明，如果从 v^* 开始，移动距离就不会超过 $\sqrt{\alpha}$ 。现在我们证明，即使移动 $\sqrt{\alpha}$ ，也无法避免增大 v 的范数。

引理 C.2。设 $v_0 = v^*$。对于任意两个时间步 $0 \le a < b \le n$,

$$\|P\widehat{v}_b - P\widehat{v}_a\|_2^2 \le 50 \cdot \alpha \log(\|v_b\|_2/\|v_a\|_2).$$

证明。我们有

$$\begin{aligned}\|P\widehat{v}_a\|_2 &\le \sqrt{\alpha} + \|Pv_0\|_2/\|v_n\|_2 \\ &\le \sqrt{\alpha} + \|Pv^*\|_2/\|v_n\|_2 \\ &\le \sqrt{\alpha}\end{aligned}$$

其中第一步由引理 C.1得出，第二步由 $v_0 = v^*$ 得出，最后一步由 P 的定义得出（参见定义 4.2，这意味着 $Pv^* = 0$，参见命题 4.3）。

我们可以证明

$$\begin{aligned}\|P\widehat{v}_b - P\widehat{v}_a\|_2^2 &\le (\|P\widehat{v}_b\|_2 + \|P\widehat{v}_a\|_2)^2 \\ &\le (2\sqrt{\alpha})^2 \\ &\le 4\alpha.\end{aligned}$$

其中第二步由 $\|P\widehat{v}_b\|_2 \le \sqrt{\alpha}$ 和 $\|P\widehat{v}_a\|_2 \le \sqrt{\alpha}$得出。

现在，我们可以考虑两种情况。

情形 1。若 $\log(\|v_b\|_2/\|v_a\|_2) \ge 1$，则证明已经完成。

Ca情况 2：如果 $\log(\|v_b\|_2/\|v_a\|_2) < 1$。我们将在下一段中证明这种情况。

我们将 w 定义为沿 $P(\widehat{v}_b - \widehat{v}_a)$ 方向的单位向量，即，

$$w = P(\widehat{v}_b - \widehat{v}_a)/\|P(\widehat{v}_b - \widehat{v}_a)\|_2.$$

利用引理C.1，我们可以证明如下结论，

$$\begin{aligned}&\langle v_b - v_a, w\rangle^2 \\ =& (\sum_{i=a+1}^{b} \eta\langle\phi(x_i), v_{i-1}\rangle\langle\phi(x_i), w\rangle)^2 \\ \le& (\sum_{i=a+1}^{b} \eta\langle\phi(x_i), v_{i-1}\rangle^2)(\eta\sum_{i=a+1}^{b}\langle\phi(x_i), w\rangle^2) \\ \le& (\sum_{i=a+1}^{b} \|v_i\|_2^2 \cdot \eta\langle\phi(x_i), \widehat{v}_{i-1}\rangle^2)(\eta\sum_{i=1}^{n}\langle\phi(x_i), w\rangle^2) \\ \le& (\|v_b\|_2^2 \cdot \sum_{i=a+1}^{b} \eta\langle\phi(x_i), \widehat{v}_{i-1}\rangle^2)(\eta\sum_{i=1}^{n}\langle\phi(x_i), w\rangle^2) \\ \le& (\|v_b\|_2^2 \cdot \sum_{i=a+1}^{b} \eta\langle\phi(x_i), \widehat{v}_{i-1}\rangle^2) \cdot \alpha \\ \le& \|v_b\|_2^2 \cdot \log(\|v_b\|_2^2/\|v_a\|_2^2) \cdot \alpha. \end{aligned} \tag{10}$$

其中第一步由断言 4.6 的性质 5 得出，第二步由柯西-施瓦茨不等式得出，第三步由定义 4.5 得出，第四步由 $\|v_i\|_2 \le \|v_b\|_2$ 对所有 $i \le b$ (参见断言 4.6 的性质 2)，第五步由 α 的定义得出，并且

最后一步由 $\log(\|v_b\|_2^2/\|v_a\|_2^2) \geq \sum_{i=a+1}^{b} \eta\langle x_i$ 得出，$\widehat{v}_{i-1}\rangle^2$ 对所有 $a < b$ (参见断言 4.6 的性质 4)。

因此，我们可以按以下方式给出 $\|P\widehat{v}_b - P\widehat{v}_a\|_2^2$ 的上界，

$$\begin{aligned}
&\|P\widehat{v}_b - P\widehat{v}_a\|_2^2 \\
&= \langle \widehat{v}_b - \widehat{v}_a, w\rangle^2 \\
&= \langle \widehat{v}_b - \frac{\|v_a\|_2}{\|v_b\|_2}\widehat{v}_a + \frac{\|v_a\|_2}{\|v_b\|_2}\widehat{v}_a - \widehat{v}_a, w\rangle^2 \\
&\leq 2\langle \widehat{v}_b - \frac{\|v_a\|_2}{\|v_b\|_2}\widehat{v}_a, w\rangle^2 + 2\langle \frac{\|v_a\|_2}{\|v_b\|_2}\widehat{v}_a - \widehat{v}_a, w\rangle^2
\end{aligned} \tag{11}$$

其中，第一步根据 w 的定义得出，第二步是通过加上一个项再减去同一项得出，最后一步根据 $\langle a+b、\ c\rangle^2 \leq 2\langle a、\ c\rangle^2 + 2\langle b、\ c\rangle^2$ 得出（这只是三角不等式，并将其应用于向量的每个坐标）。

对于第一 项，如上方方程 Eq. (11)（忽略常数因子 2），我们有

$$\begin{aligned}
\langle \widehat{v}_b - \frac{\|v_a\|_2}{\|v_b\|_2}\widehat{v}_a, w\rangle^2 &= \langle \frac{v_b}{\|v_b\|_2} - \frac{\|v_a\|_2}{\|v_b\|_2}\widehat{v}_a, w\rangle^2 \\
&= \langle \frac{v_b}{\|v_b\|_2} - \frac{v_a}{\|v_b\|_2}, w\rangle^2 \\
&= \frac{1}{\|v_b\|_2^2} \cdot \langle v_b - v_a, w\rangle^2 \\
&\leq \alpha \cdot \log(\|v_b\|_2^2/\|v_a\|_2^2) \\
&= 2\alpha \cdot \log(\|v_b\|_2/\|v_a\|_2)
\end{aligned} \tag{12}$$

其中，第一步根据 $\widehat{v}_b$ 的定义得出，第二步根据 $\widehat{v}_a$ 的定义得出（参见定义 4.5），第四步根据方程 Eq. (10) 得出。

对于　该方程 Eq. (11) 的第二项（忽略常数因子 2），我们有

$$\begin{aligned}
\langle \frac{\|v_a\|_2}{\|v_b\|_2}\widehat{v}_a - \widehat{v}_a, w\rangle^2 &= (\frac{\|v_a\|_2}{\|v_b\|_2} - 1)^2 \cdot \langle \widehat{v}_a, w\rangle^2 \\
&= (\frac{\|v_a\|_2}{\|v_b\|_2} - 1)^2 \cdot \langle \widehat{v}_a, P(\widehat{v}_b - \widehat{v}_a)\rangle^2 \\
&= (\frac{\|v_a\|_2}{\|v_b\|_2} - 1)^2 \cdot \langle P\widehat{v}_a, P(\widehat{v}_b - \widehat{v}_a)\rangle^2 \\
&\leq (\frac{\|v_a\|_2}{\|v_b\|_2} - 1)^2 \cdot 4\|P\widehat{v}_a\|_2^2 \\
&\leq (\frac{\|v_a\|_2}{\|v_b\|_2} - 1)^2 \cdot 4\alpha \\
&\leq 4\log(\frac{\|v_b\|_2}{\|v_a\|_2}) \cdot 4\alpha
\end{aligned} \tag{13}$$

其中第二步由 w的定义得出，第三步由 $P = P^2$ （于是$\langle\langle a,\ P^2 b\rangle = a^\top PPb = \langle Pa,\ Pb\rangle$）得出，第四步由 $\widehat{v}_a$和 $\widehat{v}_b$均为单位向量得出，第五步由$\|P\widehat{v}_a\|_2 \leq \sqrt{\alpha}$得出，最后一步由 $(\frac{1}{x}-1)^2 \leq 4$对数 x对所有 $x \in [1,2]$成立得出（注意，此处我们将 $x = \|v_b\|_2/\|v_a\|_2$视为 $x \geq 1$。我们可以假设4.6的原因是命题 $x \leq 2$，这是因为在这种情况下我们将 $\log(x)$ 限制为 ≤ 1，这意味着 $x \leq 2$。）

因此，

$$\begin{aligned}
&\|P\widehat{v}_b - P\widehat{v}_a\|_2^2 \\
\leq\ & 2\langle \widehat{v}_b - \frac{\|v_a\|_2}{\|v_b\|_2}\widehat{v}_a, w\rangle^2 + 2\langle \frac{\|v_a\|_2}{\|v_b\|_2}\widehat{v}_a - \widehat{v}_a, w\rangle^2 \\
\leq\ & 2\cdot 2\alpha \log(\|v_b\|_2/\|v_a\|_2) + 2\cdot 16\alpha \log(\|v_b\|_2/\|v_a\|_2) \\
\leq\ & 50\alpha \log(\|v_b\|_2/\|v_a\|_2).
\end{aligned}$$

其中，第一步由公式（11）得出，第二步由公式（12）和公式（13）得出。现在，我们完成证明。□

C.3 关于序列的结果

命题 C.3. 设 $a \in \mathbb{R}^n$，并假设 $a_1 = 0$。对于每个 $j \in [n]$ 和 $k \in [\text{对数 } n]$，我们定义

$$b_{j,k} := a_{1+2^k \cdot j}$$

注意，若 $1 + 2^k \cdot j > n$，则我们假设 $b_{j,\ k} = 0$。

于是，我们有

$$\max_{j\in[n]} a_j^2 \leq (\log n) \sum_{k=0}^{(\log n)-1} \sum_{j=1}^{n} (b_{j,k} - b_{j-1,k})^2.$$

证明。我们定义 $j^* := \arg\max_{j\in[n]} a_j^2$。

我们定义 $j_k := 1 + 2^k \lfloor \frac{j^*-1}{2^k} \rfloor$。

根据 j_k 的定义，我们有

$$j_0 = 1 + 2^0 \lfloor \frac{j^* - 1}{2^0} \rfloor = j^*$$

and

$$j_{\log n} = 1 + 2^{\log n} \lfloor \frac{j^* - 1}{2^{\log n}} \rfloor = 1.$$

因此，

$$\begin{aligned}
a_{j^*} &= a_{j^*} - a_1 \\
&= a_{j_0} - a_{j_{\log n}} \\
&= \sum_{k=0}^{(\log n)-1} (a_{j_k} - a_{j_{k+1}}) \qquad (14)
\end{aligned}$$

其中，第一步由 $a_1 = 0$ 的定义得出。

设 $j_k = 1 + 2^k y$ 且 $j_{k+1} = 1 + 2^{k+1} z$。显然，$2z \geq y \geq z$。根据事实 3.4，我们知道 $|2z - y| \leq 1$。

现在，我们考虑两种情况。

情形 1. $j_k = j_{k+1}$。在这种情况下，我们有

$$a_{j_k} - a_{j_{k+1}} = 0.$$

情形 2. $j_k \neq j_{k+1}$。
那么我们有

$$\begin{aligned} a_{j_k} - a_{j_{k+1}} &= b_{y,k} - b_{2z,k} \\ &= (b_{y,k} - b_{y+1,k}). \end{aligned}$$

因此，

$$\begin{aligned} a_{j^*}^2 &= (\sum_{k=0}^{(\log n)-1} (a_{j_k} - a_{j_{k+1}}))^2 \\ &\leq (\log n) \cdot \sum_{k=0}^{(\log n)-1} (a_{j_k} - a_{j_{k+1}})^2 \\ &\leq (\log n) \cdot \sum_{k=0}^{(\log n)-1} \cdot \sum_{j=1}^{n} (b_{j,k} - b_{j-1,k})^2 \end{aligned}$$

其中，第一步由式（14）得出，第二步则由我们对 j_k 的定义及证明得出。

□

引理 C.4. 设 $A \in \mathbb{R}^{d \times n}$ 的第一列全为零，即对所有 $i \in [d]$, $A_{i,\ 1} = 0$。对于每个$j \in [n]$ 和 $k \in [$对数 $n]$，定义 $b_{j,\ k}$ 为 A的第 $1 + 2^k \cdot j$ 列。若 $1 + 2^k \cdot j > n$，则我们假设$b_{j,\ k}$ 为零列。

- 性质 1. 对于每个 $i \in [d]$，有

$$\max_{j \in [n]} A_{i,j}^2 \leq (\log n) \sum_{k=0}^{\log n} \sum_{j=2}^{n+1} (b_{j,k} - b_{j-1,k})_i^2$$

- 性质 2. 则：

$$\sum_{i=1}^{d} \max_{j \in [n]} A_{i,j}^2 \leq (\log n) \sum_{k=0}^{\log n} \sum_{j=2}^{n+1} \|b_{j,k} - b_{j-1,k}\|_2^2$$

证明。 利用命题 C.3，我们可以证明性质 1。
应用命题 C.3对于 d 不同的行，我们有

$$\sum_{i=1}^{d} \max_{j \in [n]} A_{i,j}^2 \leq (\log n) \sum_{k=0}^{\log n} \sum_{j=2}^{n+1} \|b_{j,k} - b_{j-1,k}\|_2^2.$$

因此，我们证明了性质 2。

□

C.4 内积求和的上界

我们回到流式主成分分析的设定。本节的目标是证明，如果 $v_0 = v^*$，那么$\|v_n\|_2$ 很大。

引理 C.5. 如果 $v_0 = v^*$，那么对于 $i \in [n]$，我们有

$$\eta \sum_{i=1}^{n} \langle \phi(x_i), P\widehat{v}_{i-1} \rangle^2 \leq 100 \cdot \alpha^2 \cdot \log^2 n \cdot \log \|v_n\|_2.$$

证明。对于 $i \in [n]$，我们定义 $u_i := P\widehat{v}_i$。这也意味着 $\|u_i\|_2 \leq 1$。

由于 u_i 属于 P 的张成空间，且根据命题 4.3 ， $Pv^* = 0$，我们知道 $u_i \perp v^*$。因此，我们有

$$\langle u_i, v^* \rangle = 0.$$

对于每个 $i \in [d]$ 和每个 $j \in [n]$，我们定义一个矩阵 $A_{i,\ j} \in \mathbb{R}^{d \times n}$ 如下

$$A_{i,j} := \langle \phi(x_i), u_{j-1} \rangle.$$

我们可以证明

$$\begin{aligned}
&\sum_{i=1}^{d} \max_{j \in [n]} \langle \phi(x_i), u_j \rangle^2 \\
\leq\ & (\log n) \sum_{k=0}^{\log n} \sum_{j=2}^{n+1} \|b_{j,k} - b_{j-1,k}\|_2^2 \\
=\ & (\log n) \sum_{k=0}^{\log n} \sum_{j=2}^{n+1} ((b_{j,k})_i - (b_{j-1,k})_i)^2 \\
=\ & (\log n) \sum_{k=0}^{\log n} \sum_{j=2}^{n+1} \sum_{i=1}^{d} (\langle \phi(x_i), u_{2^k j} \rangle - \langle \phi(x_i), u_{2^k (j-1)} \rangle)^2. \qquad (15)
\end{aligned}$$

其中，第一步由引理 C.4 得出，第二步根据 ℓ_2 范数的定义得出，第三步由 $(b_{j,\ k})_i = A_{i,\ 1+2^k \cdot j} = \langle \phi(x_i), u_{1+2^k \cdot j-1} \rangle = \langle \phi(x_i), u_{2^k \cdot j} \rangle$ 得出。

对于上述方程中的每个 (k, j) 项，我们有

$$\begin{aligned}
&\sum_{i=1}^{d} (\langle \phi(x_i), u_{2^k j} \rangle - \langle \phi(x_i), u_{2^k (j-1)} \rangle)^2 \\
=\ & \sum_{i=1}^{d} (\langle \phi(x_i), u_{2^k j} - u_{2^k (j-1)} \rangle)^2 \\
\leq\ & \frac{\alpha}{\eta} \cdot \|u_{2^k j} - u_{2^k (j-1)}\|_2^2. \qquad (16)
\end{aligned}$$

其中，第一步由简单代数得出，第二步由 $\langle u_i, v^* \rangle = 0$ 以及 $\|u_i\|_2$对所有 $i \in [n]$ 和定义 4.1中的性质 3 得出。

那么，对于每个 $k \in [\log n]$，我们有

$$\begin{aligned}
\sum_{j=2}^{n+1} \|u_{2^k j} - u_{2^k (j-1)}\|_2^2 &\leq 50\alpha \log \frac{\|v_n\|_2}{\|v_0\|_2} \\
&= 50\alpha \log \|v_n\|_2 \qquad (17)
\end{aligned}$$

其中第一步是对 $j \in [2, n+1]$ 应用引理 C.2，对每个 j求和得到的；第二步则由 $v_0 = v^*$（参见引理 C.5陈述中的假设）以及 $\|v^*\|_2 = 1$得出。因此，

$$\begin{aligned}
&\eta \sum_{i=1}^{d} \langle \phi(x_i), P\widehat{v}_{i-1} \rangle^2 \\
\leq\ &\eta \sum_{i=1}^{d} \max_{j \in [n]} \langle \phi(x_i), u_j \rangle^2 \\
=\ &\eta (\log n) \sum_{k=0}^{\log n} \sum_{j=2}^{n+1} \sum_{i=1}^{d} (\langle \phi(x_i), u_{2^k j} \rangle - \langle \phi(x_i), u_{2^k (j-1)} \rangle)^2 \\
=\ &\alpha (\log n) \sum_{k=0}^{\log n} \sum_{j=2}^{n+1} \| u_{2^k j} - u_{2^k (j-1)} \rangle \|_2^2 \\
\leq\ &(\log n) \sum_{k=0}^{\log n} 50 \alpha^2 \log \|v_n\|_2 \\
\leq\ &100 \cdot \alpha^2 \cdot \log^2 n \cdot \log \|v_n\|_2
\end{aligned}$$

其中，第一步由本证明中的定义得出，第二步由式 (15）得出，第三步由式 (16）得出，第四步由式 (17）得出，最后一步则由简单代数运算得出。

因此，证明得证。 □

C.5 范数对数的下界

引理 C.6（右方向增长）。设 $\alpha \in (0, 0.1)$。设 $C_1 \geq 200$ 为某个固定常数。则如果 $v_0 = v^*$，我们有

$$\log(\|v_n\|_2) \geq \frac{\beta/8}{1 + C_1 \cdot \alpha^2 \log^2 n}.$$

此外，如果 $\alpha \in (0, 1/(10 C_1 \text{对数 } n))$，我们有

$$\|v_n\|_2 \geq \exp(\beta/20).$$

证明。我们将 $\widehat{v}_i = a_i \cdot v^* + u_i$ 改写为 $u_i \perp v^*$。

于是，我们有

$$\begin{aligned}
&\langle \phi(x_i), \widehat{v}_{i-1} \rangle^2 \\
=\ &\langle \phi(x_i), a_{i-1} \cdot v^* + u_{i-1} \rangle^2 \\
\geq\ &\frac{a_{i-1}^2}{2} \langle \phi(x_i), v^* \rangle^2 - \langle \phi(x_i), u_{i-1} \rangle^2. \qquad (18)
\end{aligned}$$

其中第二步由命题 3.5 得出。

应用引理C.1，结合 $v_0 = v^*$，我们得到

$$\|P\widehat{v}_i\|_2^2 \leq \alpha. \qquad (19)$$

注意

$$\begin{aligned}\|P\widehat{v}_i\|_2^2 &= \|P(a_i v^* + u_i)\|_2^2 \\ &= \|Pu_i\|_2^2 \\ &= \|u_i\|_2^2 \\ &\geq \frac{1}{2}\|\widehat{v}_i\|_2^2 - \|a_i v^*\|_2^2 \\ &= \frac{1}{2} - a_i^2 \end{aligned} \tag{20}$$

其中，第一步根据我们的定义 $\widehat{v}_i = a_i \cdot v^* + u_i$得出，第二步根据$Pv^* = 0$(见断言 4.3) 得出，第三步根据 P的定义得出，第四步根据断言3.5得出，最后一步则由简单代数运算得出。

因此，我们有

$$\begin{aligned}a_i &\geq (\frac{1}{2} - \alpha)^{1/2} \\ &\geq \frac{1}{2} - \alpha \end{aligned} \tag{21}$$

其中，第一步由合并公式 (19) 和公式 (20) 得出，最后一步根据$\alpha \in (0,\ 0。1)$ 得出。

现在，对 $i \in [n]$ 求和可得

$$\begin{aligned}&\eta \sum_{i=1}^{n} \langle \phi(x_i), \widehat{v}_{i-1} \rangle^2 \\ \geq\ &\eta \sum_{i=1}^{n} (\frac{a_{i-1}^2}{2} \langle \phi(x_i), v^* \rangle^2 - \langle \phi(x_i), u_{i-1} \rangle^2) \\ \geq\ &\frac{1}{4}\beta - \eta \sum_{i=1}^{n} \langle \phi(x_i), u_{i-1} \rangle^2.\end{aligned}$$

其中，第一步是对每个 i，将公式 (18) 对 $i \in [n]$ 求和得出；第二步根据公式 (21) 得出。

我们可以如下给出 $\log(\|v_n\|_2)$ 的下界：

$$\begin{aligned}\log \|v_n\|_2 &\geq \frac{1}{2}\eta \sum_{i=1}^{n} \langle \phi(x_i), \widehat{v}_{i-1} \rangle^2 \\ &\geq \frac{1}{8}\beta - C_1 \cdot \alpha^2 \log^2 n \log \|v_n\|_2,\end{aligned}$$

其中，第一步由引理 C.7 得出，第二步由引理 C.5 得出，其中 $C_1 \geq 200$是一个足够大的常数。

上述方程意味着以下结论

$$\log \|v_n\|_2 \geq \frac{\beta/8}{1 + C_1 \cdot \alpha^2 \log^2 n}.$$

□

C.6 $\|v_n\|_2$下界

引理 C.7。我们有

$$\|v_n\|_2 \geq \sqrt{\eta} \cdot (\sum_{i=1}^{n} \langle \phi(x_i), v_{i-1} \rangle^2)^{1/2}$$

证明。我们定义

$$B_i := \|v_i\|_2^2,$$

我们还定义

$$A_i := \log \frac{B_i}{B_{i-1}}$$

然后利用命题 4.6 的性质 3，可以很容易看出

$$A_i \geq \eta \langle \phi(x_i), \widehat{v}_{i-1} \rangle^2.$$

因此，

$$\begin{aligned} A_i \cdot B_{i-1} &\geq \eta \langle \phi(x_i), \widehat{v}_{i-1} \rangle^2 \cdot B_{i-1} \\ &\geq \eta \langle \phi(x_i), \widehat{v}_{i-1} \rangle^2 \cdot \|v_{i-1}\|_2^2 \\ &= \eta \langle \phi(x_i), v_{i-1} \rangle^2 \end{aligned}$$

其中第三步由定义 4.5 得出。
因此，我们可以证明以下几点，

$$\begin{aligned} \eta \sum_{i=1}^{n} \langle \phi(x_i), v_{i-1} \rangle^2 &\leq \sum_{i=1}^{n} A_i B_{i-1} \\ &\leq B_n \\ &= \|v_n\|_2^2 \end{aligned}$$

其中，第一步由 $\eta \langle \phi(x_i)\ v_{i-1} \rangle^2 \leq A_i B_{i-1}$得出，第二步由断言 B.3得出，第三步则依据我们对证明的定义得出。 □

参考文献

[ACK19] Sepehr Assadi、Yu Chen 和 Sanjeev Khanna。用于 ($\delta+$ 1) 顶点着色的次线性算法。载于《第三十届 ACM-SIAM 离散算法年度研讨会论文集》，第 767–786 页。SIAM，2019年。[ACLS12] Raman Arora、Andrew Cotter、Karen Livescu 和 Nathan Srebro。用于主成分分析和偏最小二乘的随机优化。载于《2012 年第 50 届 Allerton 通信、控制与计算年会》（Allerton），第 861–868 页。IEEE，2012。[ACS13] Raman Arora、Andy Cotter 和 Nati Srebro。采用截断 MSG 的主成分分析随机优化。《神经信息处理系统进展》，第 26 卷，2013年。

[ACSS20] Josh Alman、Timothy Chu、Aaron Schild 和 Zhao Song。几何图上线性代数的算法与难度。载于《2020 年 IEEE 第 61 届计算机科学基础年会》（FOCS），第 541–552 页。IEEE，2020年。[AS23] Josh Alman 和 Zhao Song。快速注意力要求有界条目。arXiv 预印本 arXiv:2302.13214，2023年。[AZL17] Zeyuan Allen-Zhu 和 Yuanzhi Li。流式 k-PCA 的首个高效收敛算法：全局、无间隙且近最优的速率。载于《2017 年 IEEE 第 58 届计算机科学基础年会》（FOCS），第 487–492 页。IEEE，2017年。[BBZ07] Gilles Blanchard、Olivier Bousquet 和 Laurent Zwald。核主成分分析的统计性质。《Machine Learning》，66(2):259–294，2007年。[BDF13] Akshay Balsubramani、Sanjoy Dasgupta 和 Yoav Freund。增量 PCA 的快速收敛。《神经信息处理系统进展》，第 26 卷，2013年。[BDR+04] Yoshua Bengio、Olivier Delalleau、Nicolas Le Roux、Jean-Fran¸cois Paiement、Pascal Vincent 和 Marie Ouimet。学习特征函数将谱嵌入与核主成分分析联系起来。《Neural Computation》，16(10):2197–2219，2004年。[BFK+20] Maciej Besta、Marc Fischer、Vasiliki Kalavri、Michael Kapralov 和 Torsten Hoefler。流式图和动态图实践：概念、模型、系统与并行性。arXiv，第 1912–12740 页，2020年。[CDK18] Graham Cormode、Jacques Dark 和 Christian Konrad。顶点到达数据流中的独立集。arXiv 预印本 arXiv:1807.08331，2018年。[CLL+05] Sang Wook Choi、Changkyu Lee、Jong-Min Lee、Jin Hyun Park 和 In-Beum Lee。基于核主成分分析的非线性过程故障检测与识别。《Chemometrics and Intelligent Laboratory Systems》，75(1):55–67，2005年。[CMW13] T Tony Cai、Zongming Ma 和 Yihong Wu。稀疏主成分分析：最优速率与自适应估计。《The Annals of Statistics》，41(6):3074–3110，2013年。[FKM+05] Joan Feigenbaum、Sampath Kannan、Andrew McGregor、Siddharth Suri 和 Jian Zhang。半流式模型中的图问题。《Theoretical Computer Science》，348(2-3):207–216，2005年。[FKM+09] Joan Feigenbaum、Sampath Kannan、Andrew McGregor、Siddharth Suri 和 Jian Zhang。数据流模型中的图距离。《SIAM Journal on Computing》，38(5):1709–1727，2009年。[GGK+18] Mohsen Ghaffari、Themis Gouleakis、Christian Konrad、Slobodan Mitrovi´c 和 Ronitt Rubinfeld。面向极大独立集、匹配和顶点覆盖的改进型大规模并行计算算法。载于《2018 年 ACM 分布式计算原理研讨会论文集》，第 129–138 页，2018年。[Gir02] Mark Girolami。特征空间中基于 Mercer 核的聚类。《IEEE Transactions on Neural Networks》，13(3):780–784，2002年。

[GKK12] Ashish Goel、Michael Kapralov 和 Sanjeev Khanna。最大二分图匹配的通信复杂度与流式复杂度。载于第二十三届 ACM-SIAM离散算法研讨会论文集，第 468–485 页。SIAM，2012。[GKMS19] Buddhima Gamlath、Sagar Kale、Slobodan Mitrovic 和 Ola Svensson。通过无权增广实现加权匹配。载于 2019 年 ACM 分布式计算原理研讨会论文集，第 491–500 页，2019。[HNWTW22] De Huang、Jonathan Niles-Weed、Joel A Tropp 和 Rachel Ward。矩阵乘积的集中性。计算数学基础，22(6):1767–1799，2022。[HNWW21] De Huang、Jonathan Niles-Weed 和 Rachel Ward。流式 k-PCA：Oja 算法的高效保证，超越秩一更新。载于学习理论会议，第 2463–2498 页。PMLR，2021。[Hof07] Heiko Hoffmann。用于新颖性检测的核 PCA。《模式识别》，40(3):863–874，2007。[HP14] Moritz Hardt 和 Eric Price。噪声幂方法：一种元算法及其应用。神经信息处理系统进展，27，2014。[Kap13] Michael Kapralov。流式模型中匹配问题的更优界。载于第二十四届 ACM-SIAM离散算法研讨会论文集，第 1679–1697 页。SIAM，2013。[KKP18] John Kallaugher、Michael Kapralov 和 Eric Price。图与超图计数的草图复杂度。载于 2018 年 IEEE 第 59 届计算机科学基础年会（FOCS），第 556–567 页。IEEE，2018。[KKS14] Michael Kapralov、Sanjeev Khanna 和 Madhu Sudan。近似最大割的流式下界。载于第二十六届 ACM-SIAM离散算法研讨会论文集，第 1263–1282 页。SIAM，2014。[KMPV19] John Kallaugher、Andrew McGregor、Eric Price 和 Sofya Vorotnikova。在邻接表流式模型中计数环的复杂度。载于第 38 届 ACM SIGMOD-SIGACT-SIGAI数据库系统原理研讨会论文集，第 119–133 页，2019。[KNST19] Michael Kapralov、Navid Nouri、Aaron Sidford 和 Jakab Tardos。近线性时间与空间的动态流式谱稀疏化。arXiv 预印本 arXiv:1903.12150，2019。[KP20] John Kallaugher 和 Eric Price。转闸流式处理与线性草图之间的分离与等价关系。载于第 52 届 ACM SIGACT计算理论研讨会论文集，第 1223–1236 页，2020。[Kra69] Tatiana Pavlovna Krasulina。一种用于确定对称矩阵最小特征值的随机逼近方法。《计算数学与数学物理杂志》，9(6):1383–1387，1969。

[Liu04] 刘成军。用于人脸识别的基于 Gabor 的核 PCA 与分数幂多项式模型。IEEE 模式分析与机器智能汇刊，26(5):572–581，2004。[LMTS20] 李宇宰、Gamini P Mendis、Matthew J Triebe 和 John W Sutherland。利用核主成分分析和核密度估计监测加工过程。《智能制造杂志》，31(5):1175–1189，2020。[LSPY18] David Liau、Zhao Song、Eric Price 和 Ger Yang。常数空间中的随机多臂老虎机。载于国际人工智能与统计会议，第 386–394 页。PMLR，2018。[LSW21] Robert Lunde、Purnamrita Sarkar 和 Rachel Ward。对 Oja 算法的误差进行自举。神经信息处理系统进展，34:6240–6252，2021。[M+05] Shanmugavelayutham Muthukrishnan 等。数据流：算法与应用。计算机科学理论前沿，1(2):117–236，2005。[McG05] Andrew McGregor。在数据流中寻找图匹配。载于近似、随机化与组合优化：算法与技术，第 170–181 页。施普林格，2005。[MCJ13] Ioannis Mitliagkas、Constantine Caramanis 和 Prateek Jain。内存受限的流式 PCA。神经信息处理系统进展，26，2013。[MP22] Nikos Mouzakis 和 Eric Price。对抗式流式 PCA 的谱保证。2022。[MSS+98] Sebastian Mika、Bernhard Sch¨olkopf、Alex Smola、Klaus-Robert M¨uller、Matthias Scholz 和 Gunnar R¨atsch。特征空间中的核 PCA 与去噪。神经信息处理系统进展，11，1998。[Oja82] Erkki Oja。简化的神经元模型作为主成分分析器。《数学生物学杂志》，15(3):267–273，1982。[Sch00] Bernhard Sch¨olkopf。用于距离计算的核技巧。神经信息处理系统进展，13，2000。[SSM97] Bernhard Sch¨olkopf、Alexander Smola 和 Klaus-Robert M¨uller。核主成分分析。载于国际人工神经网络会议，第 583–588 页。施普林格，1997。[SWYZ21] Zhao Song、David Woodruff、Zheng Yu 和 Lichen Zhang。多项式次数核的快速草图化。载于国际机器学习会议，第 9812–9823 页。PMLR，2021。[SY19] Zhao Song 和 Xin Yang。借助矩阵 Chernoff 界，二次规模足以实现过参数化。arXiv 预印本 arXiv:1906.03593，2019。[VL12] Vincent Vu 和 Jing Lei。高维稀疏 PCA 的极小极大估计速率。载于人工智能与统计会议，第 1278–1286 页。PMLR，2012。

[WK06] Manfred KK Warmuth 和 Dima Kuzmin。带有维度对数级遗憾界的随机化主成分分析算法。神经信息处理系统进展，19，2006年。[XLL+19] Zhou Xu、Jin Liu、Xiapu Luo、Zijiang Yang、Yifeng Zhang、Peipei Yuan、Yutian Tang 和 Tao Zhang。基于核主成分分析和加权极限学习机的软件缺陷预测。信息与软件技术，106：182–200，2019年。[XY19] Yuesheng Xu 和 Qi Ye。广义 Mercer 核与再生核巴拿赫空间，第258卷。美国数学学会，2019年。[YAK00] M-H Yang、Narendra Ahuja 和 David Kriegman。使用核特征脸进行人脸识别。载于《2000年国际图像处理会议论文集》（目录号：00CH37101），第1卷，第37–40页。IEEE，2000年。[ZB05] Laurent Zwald 和 Gilles Blanchard。核主成分分析中特征子空间的收敛性。神经信息处理系统进展，18，2005年。[ZG04] Shangming Zhou 和 John Q Gan。Mercer 核、模糊 C 均值算法与聚类原型。载于《智能数据工程与自动化学习国际会议》，第613–618页。施普林格，2004年。