arXiv:2304.03191v1 [cs.DS] 6 Apr 2023

Krylov 方法对于低秩近似而言几乎是最优的

AineshBakshi ShyamNarayanan
ainesh@mit.edu shyamsn@mit.edu MIT
MIT

摘要

我们考虑矩阵-向量乘积模型下，在各种 Schatten 范数意义下的秩-1 低秩近似（LRA）问题：

$$\min_{\|u\|_2=1} \left\| A(I-uu^\top) \right\|_{\mathcal{S}_p},$$

其中 $\|M\|_{\mathcal{S}_p}$ 表示M的奇异值的 ℓ_p 范数。给定 $\varepsilon > 0$，我们的目标是输出一个单位向量 v，使得

$$\left\| A\left(I-vv^\top\right) \right\|_{\mathcal{S}_p} \leqslant (1+\varepsilon) \min_{\|u\|_2=1} \left\| A\left(I-uu^\top\right) \right\|_{\mathcal{S}_p}.$$

我们的主要结果表明，Krylov 方法（几乎）达到了信息论意义下的最优[1] 矩阵-向量乘积数量，适用于谱范数（$p = \infty$）、Frobenius 范数（$p = 2$）和核范数（$p = 1$）LRA。

特别地，对于谱低秩近似，我们证明任何算法都需要 $\Omega\left(\log(n)/\varepsilon^{1/2}\right)$次矩阵-向量乘积，这与通过 Krylov 方法 [MM15]获得的上界完全匹配。我们的下界回应了 [Woo14], 中的开放问题 1，为谱低秩近似算法缺乏进展提供了证据，并解决了 [BCW22]中的开放问题 1.2。接下来，我们证明对于任意固定常数 p，即 $1 \leqslant p = O(1)$，存在 $O\left(\log(1/\varepsilon)/\varepsilon^{1/3}\right)$ 次矩阵-向量乘积的上界，这意味着复杂度不会随输入规模变化而增长。这改进了 [BCW22], 中近期获得的 $O\left(\log(n/\varepsilon)/\varepsilon^{1/3}\right)$ 上界，并且与其 $\Omega\left(1/\varepsilon^{1/3}\right)$ 下界相匹配，误差仅相差一个 $\log(1/\varepsilon)$ 因子。

[1]对于谱低秩近似，上界和下界在一个固定的普适常数范围内相匹配。对于 Frobenius 范数低秩近似和核范数低秩近似，它们相差一个对数$(1/\varepsilon)$ 因子。

目录

1 引言 11.1 我们的结果.21.2 开放问题.4

2 技术概述
42.1 针对 Krylov 方法的困难实例与下界.52.2 将 Krylov 下界提升为矩阵-向量下界.92.3 更精确的 Krylov 子空间算法.10**3 其他相关工作** 11**4 预备知识** 12**5 针对块 Krylov 方法的下界** 14**6 针对任意自适应算法的下界** 176.1 将任意自适应查询归约为块 Krylov 方法.186.2 一般矩阵-向量下界.18**7 Schatten-p LRA 的上界** 19

引理证明6.3 29
A.1 附加预备知识.30A.2 条件数引理.30A.3 从查询算法到块Krylov算法.33

1 引言

迭代算法是现代优化方法的主力，并广泛应用于科学计算、数值线性代数和机器学习的各个领域。这类算- 法如今已用于各种任务，从训练大型机器学习模- 型 [BBV04， GBC16], 运行大规模流体力学模拟 [ESW96， ES96， Elm96],结构分析 [Rou89， TW04]以及计算化学 [Sch09，ZVY+15]，到量子机器学习 [HHL09， GSLW19]。为了建立迭代算法的一般理论，并系统地比较其性能，我们需要一种计算模型，能够同时涵盖所有这类算法，并允许证明细粒度下界。虽然标准 RAM模型可以轻松涵盖迭代算法，但我们没有任何工具来获得其性能的细粒度下界。

近期受到广泛关注的另一种计算模型是矩阵-向量乘积模型 [WWZ14，SWYZ19，RWZ20，SAR18，BHSW20，MMMW21，BCW22，NSW22]
这里，算法仅通过自适应矩阵-向量查询访问输入矩阵 A 。具体而言，算法选择查询向量 v^1,，获得乘积 $A \cdot v^1$,，选择下一个查询向量 v^2,，它可以是 v^1 和 $A \cdot v^1$, 的任意随机化函数，接收 $A \cdot v^2$,，如此继续。在此模型中，衡量复杂度的基本指标是解决给定问题所需的矩阵-向量乘积的最小数量，我们称之为矩阵-向量复杂度。

这种矩阵-向量乘积模型刻画了自然的迭代算法，并已在科学计算和数值线性代数领域得到广泛研究（例如，参见[KK04], 及其中的参考文献）。此外，在许多现实应用中，矩阵-向量乘积的次数决定了整体运行时间 [MCG04]。最后，可以为各种问题推导出无条件的信息论下界，从而确定其矩阵-向量复杂度，正如 [SAR18、BHSW20], 在计算矩阵最大特征值时所展示的那样。

一类常见的迭代算法基于计算 Krylov 子空间：我们将这类算法统称为 *Krylov* 子空间方法。这里，算法不是丢弃中间的矩阵-向量乘积，而是为中间向量张成的子空间构造一组基，即 $\mathcal{K} = [v$、Av、A^2v、.....$A^tv]$。Krylov 子空间方法的典型例子包括 Krylov 迭代（算法 *Krylov* 子空间方法（算法 2.2））， 用于计算前 k 个特征值和低秩近似 [RST10、HMT11、MM15], 共轭梯度法用于求解线性系统，以及 Lanczos 迭代用于将低次多项式作用于特征值（参见 [Saa81] 及其中的参考文献）。

在这项工作中，我们着重研究低秩近似的矩阵-向量复杂度，特别是目标秩为 1 的情形。具体而言，给定一个 $n \times d$ 矩阵 A 和精度参数 $0 < \varepsilon < 1$，目标是计算一个单位向量 v ，使得

$$\left\| A\left(I - vv^\top\right)\right\|_{\mathcal{S}_p} \leqslant \min_{\|u\|_2=1} (1+\varepsilon) \left\| A\left(I - uu^\top\right)\right\|_{\mathcal{S}_p},$$

其中， $\|M\|_{\mathcal{S}_p}$ 是 M 的 Schatten-p 范数，其定义为 A 的奇异值的 ℓ_p 范数。在 Schatten-p 范数下表述低秩近似，为比较各种算法在经过充分研究的矩阵范数上的表现提供了一种便捷方式：谱范数 ($p = \infty$)、Frobenius 范数 ($p = 2$)和核范数 ($p = 1$)。我们注意到，秩 1 低秩近似的任何下界，都意味着当秩为固定的通用常数时的下界；而我们的上界可自然推广到秩-k近似情形。为便于阐述，我们将重点讨论秩 1 低秩近似问题。

最近，Bakshi、Clarkson 和 Woodruff [BCW22] 研究了 Schatten-p 低秩近似的矩阵-向量乘积模型中的 Krylov 方法。他们得到了一个上界

$O\left(p^{1/6}\log(n/\varepsilon)/\varepsilon^{1/3}\right)$ 矩阵-向量乘积，适用于任意 $p \geqslant 1$，其方法利用了迭代次数与分块大小之间的权衡（即每一步中与 A 相乘的起始向量数量）。他们构造了两个相互独立的 Krylov 子空间，$\mathcal{K}_1 = [g, Ag, A^2g, \ldots]$ 和 $\mathcal{K}_2 = [G, AG, A^2G, \ldots]$,，其中 $\mathcal{K}_2$ 以一个分块矩阵而非单个向量作为起点。[BCW22] 还为任意固定的通用常数 p 获得了一个 $\Omega\left(1/\varepsilon^{1/3}\right)$下界。另一方面，对于谱低秩近似，Musco 和 Musco [MM15] 对 Krylov 迭代的无间隙分析得到了 $O\left(\log(n)/\varepsilon^{1/2}\right)$ 上界。此外，据我们所知，目前尚无关于谱低秩近似的矩阵-向量下界（参见第 1 节，2 [BCW22], 其中明确将此表述为一个开放问题）。

因此，一个自然的问题是：

Krylov 迭代在谱范数、*Frobenius* 范数和核范数低秩近似中，是否能达到最优的矩阵-向量乘积次数？

1.1 我们的结果

参考文献	谱范数($p = \infty$)	Frobenius 范数($p = 2$)	核范数($p = 1$)
同时迭代 [RST10, HMT11, MM15]	$O(\log(n)/\varepsilon)$	$O(\log(n)/\varepsilon)$	N.A.
块 Krylov [MM15]	$O\left(\log(n)/\varepsilon^{1/2}\right)$	$O\left(\log(n)/\varepsilon^{1/2}\right)$	N.A.
修改的块 Krylov [BCW22]	$O\left(\log(n/\varepsilon)/\varepsilon^{1/2}\right)$	$O\left(\log(n/\varepsilon)/\varepsilon^{1/3}\right)$	$O\left(\log(n/\varepsilon)/\varepsilon^{1/3}\right)$
先前的下界 [BCW22]	N.A	$\Omega\left(1/\varepsilon^{1/3}\right)$	$\Omega\left(1/\varepsilon^{1/3}\right)$
我们的结果 定理 1.1 , 定理 1.5	$\Omega\left(\log(n)/\varepsilon^{1/2}\right)$	$O\left(\log(1/\varepsilon)/\varepsilon^{1/3}\right)$	$O\left(\log(1/\varepsilon)/\varepsilon^{1/3}\right)$

表 1：我们将结果与先前工作进行比较，衡量谱范数、Frobenius 范数和核范数低秩近似所需的矩阵-向量乘积次数。我们指出，我们的下界意味着块 Krylov [MM15] 是谱低秩近似的最优算法。我们针对 Frobenius 范数和核范数低秩近似给出的上界，与 [BCW22]的下界相差至多一个 $\log(1/\varepsilon)$ 因子。

我们对上述问题给出了肯定的回答，并证明了使用单个初始向量的 Krylov 迭代（算法 2.2）针对谱范数、Frobenius 范数和核范数低秩近似获得了（近乎）最优的矩阵-向量乘积次数。对于谱低秩近似，矩阵-向量复杂度为$\Theta\left(\log(n)/\sqrt{\varepsilon}\right)$。对于 Frobenius 范数和核范数低秩近似，矩阵-向量复杂度为 $\tilde{\Theta}\left(1/\varepsilon^{1/3}\right)$，其中 $\tilde{\Theta}$ 会抑制一个 $\log(1/\varepsilon)$ 因子（显式的上界和下界见表 1.1 ）。

我们首先陈述谱低秩近似的下界：

定理 1.1（谱低秩近似的下界）。 存在一个对称实数 $n \times n$ 矩阵上的分布 $\mathcal{D}$，使得给定 $A \sim \mathcal{D}$ 和 $0 < \varepsilon < 1$，任何随机化算法都需要$\Omega\left(\log(n)/\varepsilon^{1/2}\right)$ 次矩阵-向量乘积，才能输出向量 v，使其以至少2/3的概率满足

$$\left\|A\left(I - vv^\top\right)\right\|_{op} \leqslant (1+\varepsilon)\min_{\|u\|_2=1}\left\|A\left(I - uu^\top\right)\right\|_{op}.$$

备注 1.2（关于最优性）。 Krylov 迭代 [MM15] 需要$O\left(\log(n)/\varepsilon^{1/2}\right)$ 次矩阵-向量乘积，因此我们确定了秩一谱低秩近似的矩阵-向量复杂度。

备注 1.3（矩阵-向量与RAM）。在RAM 模型中，Krylov迭代可以在 $O\left(\text{nnz(A) log(n)}/\varepsilon^{1/2}\right)$ 时间内实现，而Woodruff的专著 [Woo14] 中的开放问题 1询问是否可以将其改进为 O(nnz(A) + npoly(1/ε))。这一问题也已在近期的几篇论文 [BBK+21、KW21、WY22]中被重新提出。我们的结果表明，这一问题缺乏算法进展。

备注 1.4（与先前工作比较）。据我们所知，目前尚无针对谱低秩近似的已知矩阵-向量下界。Simchowitz、Alaoui 和 Recht [SAR18] 获得了一个用于估计最大k 个特征值的矩阵-向量下界。然而，这并不能转化为针对谱低秩近似的任何下界（详见 [BCW22] 中的附录 A）。Braverman、Hazan、Simchowitz 和 Woodworth [BHSW20] 为估计最大特征值引入了一个不同的困难实例，而 Bakshi、Clarkson 和 Woodruff 将该实例调整为 Schatten-$\Omega\left(1/\varepsilon^{1/3}\right)$ 的下界，用于 Schatten-p 低秩近似，其中 p 是一个固定常数。事实上，他们将获得任何随 $1/\varepsilon$ 增长的、针对谱低秩近似的下界列为一个开放问题（见 [BCW22]中的第 1.2 节）。

接下来，我们陈述 Schatten-p 低秩近似的上界，其中任意 $p \geqslant 1$ 均由一个固定常数所界定。

定理 1.5（Schatten-p LRA 的上界）。 给定一个 $n \times d$ 矩阵 A、$0 < \varepsilon < 1$ 和 $1 \leqslant p$，存在一种算法，需要 $O\left(p\log(1/\varepsilon)/\varepsilon^{1/3}\right)$ 次矩阵-向量乘积，并输出一个单位向量 v，使得其概率至少为99/100，

$$\left\|A\left(I - vv^{\top}\right)\right\|_{\mathcal{S}_p} \leqslant (1+\varepsilon) \min_{\|u\|_2=1} \left\|A\left(I - uu^{\top}\right)\right\|_{\mathcal{S}_p}.$$

注 1.6（关于最优性）。 Bakshi、Clarkson 和 Woodruff [BCW22] 获得了$O\left(\log（n/\varepsilon）\, p^{1/6}/\varepsilon^{1/3}\right)$ 的上界，以及 $\Omega\left(1/\varepsilon^{1/3}\right)$ 的下界，其中p为固定常数。当p 是固定的通用常数时（Frobenius 范数和核范数低秩近似均满足这一条件），我们将上述下界匹配到相差一个（$1/\varepsilon$） 倍的对数因子。

备注 1.7 (Krylov 迭代与块 Krylov)。 Bakshi、Clarkson 和 Woodruff 并行运行 Krylov 方法的两种实例化，一种使用 分块大小 1，另一种使用 分块大小 $O(1/\varepsilon^{1/3})$。我们的算法只需要一种实例化，使用 分块大小1，而已知这在实践中具有更好的数值稳定性 [CLR21]。结合我们的下界，我们的结果表明，对于秩-1 低秩近似，我们从不需要运行分块大小 大于 *1* 的 Krylov 迭代。

我们相信我们的算法和分析同样可以推广到秩-k 低秩近似，但为便于阐述，我们聚焦于秩-1情形。

最后，我们重点介绍一个提升定理。该定理表明，在相当一般的条件下，针对大分块大小 Krylov 迭代的下界，可以转化为矩阵-向量模型中针对自适应查询的信息论下界；这一结果本身或许也值得独立关注。

这一提升结果出自 [CdDL+23]一篇尚未发表的论文，其证明收录于附录A 中，以供参考。[2] 我们在此对这一提升结果作简要说明。

定理 1.8 (提升分块 Krylov 下界，引理 6.3 (非正式表述))。 设 $\mathcal{A}$ 是任意一个自适应算法，它对一个 $n \times\times n$ 对称矩阵 A 进行 k 次矩阵 - 向量查询，其中 A 的特征向量均匀地服从（*Haar*）随机分布。给定 k 个随机高斯向量 $V = [v_1,\ldots,v_k]$，仅凭 *Krylov* 矩阵 $\mathcal{A}AV\mathcal{K} = [, \mathrm{A}^2V, A^3V, \ldots A^kV]$，且不掌握 A 的任何其他信息，就能完美模拟 $\mathcal{A}$的 k 次自适应查询及其响应的分布。

从高层次来看，该结果表明，如果存在任何矩阵-向量算法能够通过 k 次自适应查询解决给定的特征值或特征向量问题，那么使用分块大小为 k 的 Krylov 迭代可以通过 k 次迭代解决同一问题。一般而言，该模拟器需要 k^2 次矩阵-向量乘积，而自适应算法使用 k 次矩阵-向量乘积，因此产生二次开销。然而，在谱低秩近似的场景下，我们证明从大于 1 的分块大小开始不会带来任何额外价值（见定理 5.2），而提升技术能够得到最优下界。

注1.9。 我们注意到 [CdDL+23] 证明了定理 1.8的一个稍微简单的版本，因为他们只要求提升来解决一个依赖于特征值的问题（具体来说，是估计 $\mathrm{Tr}(A^{-1})$，目的是从 $\mathcal{N}(0, A^{-1})$中生成一个样本）。然而，我们关于低秩近似的下界基于识别良好的 特征向量，因此定理 1.8事实上是他们结果的一个略微推广。

1.2 开放问题

我们强调以下源于我们工作的开放问题：

开放问题 1.10 (更大的目标秩，细化了 [Bak22]中的开放问题 34）。 虽然我们关注秩为 1 的低秩近似，但当目标秩为k 时，谱低秩近似的上界为$O\left(k\log(n)/\sqrt{\varepsilon}\right)$ 次矩阵-向量乘积 [MM15]。我们的下界表明，对于任何固定的通用常数目标秩，这都是最优的。关于k、$\log(n)$ 和 $1/\varepsilon$ 同时变化时，矩阵-向量乘积的复杂度应当是多少?

开放问题 1.11 (大 p 的相变)。 或许令人惊讶的是，当 p 是一个固定的通用常数时，矩阵-向量复杂度不会随输入规模增长，甚至不会以对数速度增长。然而，对于 $\log(n)$ 而言，这是必要的 $p = \infty$ (见定理 1.1)。能否在这两种情形之间平滑过渡，同时得到关于p、$\log(n)$和 $1/\varepsilon$ 的正确依赖关系?

2 技术概述

在本节中，我们首先介绍困难实例，并给出针对谱低秩近似中 Krylov 迭代（算法 2.2）下界的完整证明。随后，我们概述如何建立针对块 Krylov（算法 5.1）的下界，并解释如何将针对块 Krylov 的下界推广为一般的矩阵-向量乘积下界。最后，我们将讨论为在 p 为固定常数时获得更好的 Schatten-p 低秩近似上界所需的新思路。

[2]我们已获得 [CdDL+23] 的作者明确许可，可以转载该证明。

2.1 针对 Krylov 方法的困难实例与下界

作为热身，我们给出一个针对 Krylov 迭代（算法 2.2）的下界。从高层次来看，我们的证明通过构造一个困难实例A 来进行，使得 Krylov 子空间K = [g, Ag, A^2g, . . .,A^qg], 对于 q = c log(n)/$\sqrt{\varepsilon}$，对于足够小的常数c，不张成任何与A的最大特征向量具有至少 $\sqrt{\varepsilon}/100$ 相关性的向量（见引理2.6）。然后我们证明，任何 $(1+\varepsilon)$ 相对误差谱低秩近似都必须与A的最大特征向量具有至少 $\sqrt{\varepsilon}/100$ 的相关性（见引理 2.7）。

形式化地说，我们得到以下定理:

定理 2.1 （谱低秩近似对 Krylov 方法而言是困难的）。 给定 $\varepsilon > 0$，令 $n = \Omega(1/\varepsilon^2)$。那么，在 $n \times n$ 个矩阵 A 上存在一个分布，使得算法 2.2 需要 $q = \Omega(\log n/\sqrt{\varepsilon})$次矩阵-向量乘积，才能输出一个单位向量 v，使得其满足以下条件，且其发生的概率至少为99/100,

$$\left\|A - Avv^\top\right\|_{\text{op}} \leqslant (1+\varepsilon) \min_{\|u\|_2=1} \left\|A - Auu^\top\right\|_{\text{op}}.$$

算法 2.2（Krylov 迭代， [MM15]）。

输入： 一个 $n \times n$ 矩阵 A，迭代次数 q。

1. 令g 为从 $\mathcal{N}(0, I)$ 中抽取的向量。令 $\mathcal{K} = \left[Ag;\ A^2g;\ A^3g;\ \ldots\ldots;\ A^qg\right]$ 为通过拼接向量Ag、……、A^qg得到的$n \times q$ Krylov 矩阵。

2.计算 Q 作为 $\mathcal{K}$ 的列空间的一个正交归一基。令 $M = Q^\top A^2 Q$。

3.计算M 的顶级左奇异向量，并将其记为 y_1。

输出： $v = Qy_1$。

切比雪夫多项式背景。 我们首先定义我们的主角：切比雪夫多项式。

定义 2.3 （第一类切比雪夫多项式。）。 对于任意 $d \in \mathbb{N}$，第 d 个切比雪夫多项式定义如下:

$$T_d(x) = \begin{cases} \cos(d \arccos(x)) & \text{if } |x| \leqslant 1 \\ \frac{1}{2}\left(\left(x - \sqrt{x^2-1}\right)^d + \left(x + \sqrt{x^2-1}\right)^d\right) & \text{if } |x| \geqslant 1 \end{cases}$$

根据上述定义，很容易看出，当$|x| \leqslant 1$ 时，切比雪夫多项式是有界的，并在 -1 和 1 之间振荡。因此，我们可以定义它们取得极值的位置:

定义 2.4 （切比雪夫多项式的极值）。 对于所有 $i \in [d]$, $T_d(x_i) \in \{-1, 1\}$ 当且仅当$x_i = \cos(i \cdot \pi/d)$。

谱低秩近似的困难实例。 回想一下，我们的目标是构造一个困难实例，使得当 $q = c\log(n)/\sqrt{\varepsilon}$ 时，对于足够小的常数 c，一个t维 Krylov 子空间不张成任何与最大特征向量$(\sqrt{\varepsilon}/10)$相关的向量。直观地说，其中的任意向量都

$$A = U \begin{bmatrix} \lambda_1 & & & & & & \\ & \lambda_2 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_2 & & & \\ & & & & \ddots & & \\ & & & & & \lambda_{q+2} & \\ & & & & & & \ddots \\ & & & & & & & \lambda_{q+2} \end{bmatrix} U^\top$$

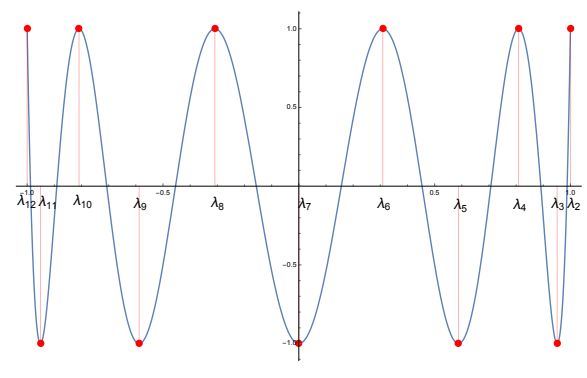


图 1：令 $n = \Omega(1/\varepsilon^2)$，令 $t = \Theta(n\sqrt{\varepsilon}/\log(n))$，并令 $q = \Theta(\log(n)/\sqrt{\varepsilon})$。令 $\lambda_2, \lambda_3, \lambda_4, \ldots, \lambda_{q+2}$ 为达到次数为 q 的切比雪夫多项式极值的点（参见定义 2.4）。我们构造一个困难实例，其中设 $\lambda_1 = 1 + 2\varepsilon$，并将其余特征值设为 λ_i，每个 $i \in [2, q+2]$ 对应的代数重数为 t。最后，设 $A = U\Lambda U^\top$，其中 U 是均匀随机正交矩阵。

Krylov 子空间可以写成将次数为 q 的多项式作用于输入矩阵的特征值后得到的随机线性组合。因此，我们构造一个实例，其中

1. 最大特征向量沿着一个均匀随机方向，且最大特征值与第二大特征值（按绝对值计）之间的间隔为 ε。 2. 不同特征值（不包括绝对值最大的特征值）的位置满足：任何次数有界的 t 多项式，都必须在至少一个不同特征值处取得较大的值（绝对值接近 1）。 3. 每个特征值（不包括最大特征值）的代数重数都足够大，以至于任何在此类特征值处赋予非平凡权重的多项式，都意味着 Krylov 子空间中的对应向量与最大特征向量几乎不相关。

如前所述，我们将不同特征值的位置设在次数为 $(\log(n)/\sqrt{\varepsilon})$ 切比雪夫多项式取得极值的位置（见图 1）。随后，我们将每个特征值重复足够多次，并将特征向量设为 Haar 随机正交矩阵。与以往的困难实例 [SAR18、BHSW20、BCW22], 不同，我们直接设计困难实例的特征值（而不是从 Wishart 系综或变形 Wigner 系综中随机选取一个矩阵）。此外，在我们的分析中，除最大特征值外，每个特征值都具有较高的代数重数，这一点至关重要；而在以往的困难实例中，每个特征值的代数重数恰好为 1。

现在，我们可以正式定义这个困难实例了。

定义 2.5（困难分布）。给定 $\varepsilon > 0$，令 $n \geqslant 1/\varepsilon^2$。令 $A = U\Lambda U^\top$ 为特征分解，其中 U 是一个列向量正交的 $n \times n$ 矩阵，且从中均匀随机选取（见定义 4.12）； Λ 是由特征值构成的对角矩阵。为定义 Λ，我们首先选择最大特征值 $\lambda_1 := 1 + 2\varepsilon$，令其重数为 1。接下来，加入 $q+1$ 个额外的不同特征值 λ_2、……、λ_{q+2}，其中 $q = c\log(n)/\sqrt{\varepsilon}$，使得 $\lambda_{i+2} = \cos\left(\frac{i}{q} \cdot \pi\right)$，对 $0 \leqslant i \leqslant q$ 均成立。（注意：此处的 q 与我们证明 Krylov 迭代下界时所针对的 q 相同。）这些特征值的重数均为 t，其中 $t = \Theta(n \cdot \sqrt{\varepsilon}/\log(n))$。

我们建立的关键引理表明，Krylov 子空间中的任何向量都无法与A的最大特征向量产生非平凡相关，除非 Krylov 子空间的维数为 $\Omega(\log(n)/\varepsilon^{1/2})$。

引理 *2.6*（*Krylov*子空间中向量的对齐）。给定 $g \sim \mathcal{N}(0,I)$和 $\varepsilon > 0$，令A从定义*2.5* 中的困难分布采样，令 $\mathcal{K} = [g; Ag; A^2g, \ldots, A^qg]$ 为当 $q = c\,log(n)/\sqrt{\varepsilon}$时的 *Krylov*子空间，其中 c为充分小的常数。进一步地，令$u_1, u_2, \ldots, u_n$ 为 A 的特征向量，其中 u_1 是最大特征向量，$u_2, \ldots, u_{t+1}$ 对应于第二个特征值 λ_2，依此类推。那么，以至少 $1 - 1/n$ 的概率，对于 $\mathcal{K}$ 的列空间中的任意向量 w，都有$\langle w, u_1\rangle^2 \leqslant \varepsilon/100$。

证明。我们首先回顾一下，由于u_i's张成整个空间，我们可以将$g = \sum_{i\in[n]} a_i u_i$ 重写为，其中$a_i = \langle u_i$为标量，$g\rangle$。此外，根据高斯分布的旋转不变性，我们知道每个a_i i.i.d.$\sim \mathcal{N}(0,1)$。现在，我们可以将与每个重数为t 的特征值相对应的所有特征向量归为一组，其中 t 在定义2.5 中设定。形式化地，对于 $j \in [2,q+2]$，令

$$\hat{u}_j = \sum_{\ell\in[t]} \frac{a_{(j-2)\cdot t+1+\ell} u_{(j-2)\cdot t+1+\ell}}{\sqrt{\sum_{\ell\in[t]} a^2_{(j-2)\cdot t+1+\ell}}},$$

为对应于第j个不同特征值的平均特征向量。此外，设

$$\hat{a}_j = \sqrt{\sum_{\ell\in[t]} a^2_{(j-2)*t+1+\ell}}$$

为g 在特征基展开中的对应系数。于是，我们可以将g重写如下：

$$g = a_1u_1 + \sum_{j\in[2,q+2]} \hat{a}_j\hat{u}_j.$$

回顾一下记号 $\lambda_1, \lambda_2, \ldots, \lambda_{q+2}$，它们表示 A 的不同特征值。注意，在算法 第 r 次迭代中2.2，我们得到向量

$$\begin{aligned} A^rg &= U\Lambda^rU^\top \left(a_1u_1 + \sum_{j\in[2,q+2]} \hat{a}_j\hat{u}_j \right) \\ &= \lambda_1^r a_1u_1 + \sum_{j\in[2,q+2]} \lambda_j^r\hat{a}_j\hat{u}_j. \end{aligned} \tag{1}$$

经过$q = c\log(n)/\sqrt{\varepsilon}$ 次迭代后，Krylov 子空间中的任意向量都可以表示为这些列的线性组合，因此具有以下形式：

$$\begin{aligned} w &= \sum_{r=1}^{q} c_r \left(\lambda_1^r a_1u_1 + \sum_{j\in[2,q+2]} \lambda_j^r\hat{a}_j\hat{u}_j \right) \\ &= a_1p(\lambda_1)u_1 + \sum_{j\in[2,q+2]} p(\lambda_j)\hat{a}_j\hat{u}_j, \end{aligned} \tag{2}$$

其中 $p(x) = \sum^q{}_{r=1} c_r x^r$，其中标量 c_r 任意。此外，由于特征向量具有正交归一性，

$$\|w\|_2^2 = a_1^2p(\lambda_1)^2 + \sum_{j\in[2,q+2]} p(\lambda_j)^2\hat{a}_j^2.$$

注意，p 是一个次数为 q 的多项式，因此至多有 q 个根。根据构造，A 有$\gg q$ 个不同的特征值，因此，p 在A的特征值中，除去一个较小的常数比例外，均必须取非零值。我们在此证明的主要结论是

$$a_1^2 p(\lambda_1)^2 \leqslant \frac{\varepsilon}{100} \max_{j\in[2,q+2]} \hat{a}_j^2 p(\lambda_j)^2, \tag{3}$$

其中，λ_j表示 A 的互不相同的特征值。我们首先证明如何利用等式（3）完成证明。考虑 w 与最大特征向量的内积：

$$\begin{aligned}
\langle w, u_1\rangle^2 &= \left\langle a_1 p(\lambda_1) u_1 + \sum_{j\in[2,q+2]} p(\lambda_j)\hat{a}_j \hat{u}_j, u_1 \right\rangle^2 \\
&= a_1^2 p(\lambda_1)^2 \|u_1\|^2 \\
&\leqslant \frac{\varepsilon}{100} \cdot \max_{j\in[2,q+2]} \hat{a}_j^2 p(\lambda_i)^2 \cdot \|u_1\|^2 \\
&\leqslant \frac{\varepsilon}{100} \|w\|_2^2,
\end{aligned} \tag{4}$$

其中，最后一个不等式由等式（2）得出。

因此，剩下的任务是证明等式（3）。首先，考虑 $\max_{i\in[2:q+2]} |p(\lambda_i)| = 0$ 的情形。在这种情况下，p 有 q 个 $+1$ 根，但其次数至多为 q，因此等式（3）显然成立。或者，由于等式（3）关于 p 是尺度不变的，不失一般性，我们可以通过适当缩放 p，假设 $\max_{i\in[2:q+2]} |p(\lambda_i)| = 1$。根据次数为 q 的切比雪夫多项式 T_q 具有极值性，即在区间 $[-1,1]$ 之外，其增长速度快于任何其他在 $\cos(i\,q\cdot\pi)$ 处有界的 q 次多项式（见事实 4.5），

$$|p(\lambda_1)| = |p(1+2\varepsilon)| \leqslant T_q(1+2\varepsilon) \leqslant e^{C\min\left(\sqrt{\varepsilon}q, \varepsilon q^2\right)} \leqslant n^{0.01} \max_{i\in[2,q+2]} |p(\lambda_i)|, \tag{5}$$

其中，第二个不等式由切比雪夫多项式的标准界得到（见事实 4.6），最后一个不等式则是根据我们的假设 $\max_{i\in[2:q+2]} |p(\lambda_i)| = 1$，以及 $q = c\log n/\sqrt{\varepsilon}$ 得到的。

接下来，由于 $a_1 \sim \mathcal{N}(0,1)$，根据事实 4.9可知，以至少 $1 - 1/n^2$，$|a_1| \leqslant O(\sqrt{\log(n)})$的概率。进一步地，对于每个$i \in [2,q+2]$，$\hat{a}_i$ 是一个 t 维高斯向量的欧几里得范数。利用标准集中界（见事实 4.10），我们知道，以至少 $1 - 1/2n$ 的概率，对于所有 $j \in [2,q+2]$，

$$\hat{a}_i = \Theta\left(\sqrt{t}\right) = \Theta\left(\sqrt{n\sqrt{\varepsilon}/\log(n)}\right) \geqslant \frac{c' n^{1/2}\varepsilon^{1/4}}{\log(n)} |a_1|,$$

对于固定常数 c'。对所有 $i \in [n]$，应用并界，以至少 $1 - 1/n$ 的概率，对所有$j \in [2,q+2]$，

$$|a_1| \leqslant \frac{c'\log(n)}{n^{1/2}\varepsilon^{1/4}} \hat{a}_j \tag{6}$$

结合（5）和（6），我们有

$$|a_1 p(\lambda_1)| \leqslant \frac{c'\log(n) n^{0.01}}{n^{1/2}\varepsilon^{1/4}} \max_{i\in[2,q+2]} \hat{a}_i |p(\lambda_i)| \leqslant \frac{\sqrt{\varepsilon}}{10} \max_{i\in[2,q+2]} \hat{a}_i |p(\lambda_i)|$$

其中，最后一个不等式源于回忆起 $n \geqslant 1/\varepsilon^2$，从而完成了方程（3）。

□

接下来，我们证明，如果任意单位向量与最大特征向量的相关性很小（内积平方小于 ε），则该向量不可能是A 的良好谱低秩近似。

引理 2.7。 *[与谱低秩近似对齐]* 给定 $\varepsilon > 0$，设 A 是一个矩阵，使得 u_1 是 A 的最大特征向量。此外，设 $\lambda_1 = 1 + 2\varepsilon$ 和 $\lambda_2 = 1$ 是 A 的两个最大特征值，并假设$|\lambda_i| \leqslant 1$ 对所有 i 均成立 $\geqslant 2$。设 w 是一个向量，使得 $\langle u_1, w\rangle^2$，$\langle u_1, Aw\rangle^2 \leqslant \varepsilon/2$。那么，

$$\left\|A\left(I - ww^\top\right)\right\|_{\text{op}} > (1+\varepsilon) \min_{\|u\|_2=1} \left\|A\left(I - uu^\top\right)\right\|_{\text{op}}.$$

证明。 根据算子范数的定义，对于任意单位向量 v，我们有

$$\left\|A\left(I - ww^\top\right)\right\|_{\text{op}} \geqslant v^\top A\left(I - ww^\top\right)v.$$

特别地，对于 u_1,，我们有

$$u_1^\top A\left(I - ww^\top\right)u_1 = u_1^\top Au_1 - \langle u_1, Aw\rangle \cdot \langle u_1, w\rangle \geqslant 1 + 2\varepsilon - \varepsilon/2 > 1 + \varepsilon.$$

此外，$\min_{\|u\|_2=1} \left\|A\left(I - uu^\top\right)\right\|_{\text{op}} = 1$。因此，

$$\left\|A\left(I - ww^\top\right)\right\|_{\text{op}} \geqslant u_1^\top A(1 - ww^\top)u_1 > 1 + \varepsilon = (1+\varepsilon) \min_{\|u\|_2=1} \left\|A\left(I - uu^\top\right)\right\|_{\text{op}},$$

由此证明得证。 □

现在我们已经准备好完成定理的证明 2.1。

定理的证明2.1。 注意，引理 2.6 意味着，以至少 99/100 的概率，任何 w 若属于 Krylov 子空间（特别地，由算法 2.2输出的向量 $\langle w, u_1\rangle^2 \leqslant \varepsilon/100$。此外，$\langle u_1, Aw\rangle^2 = (1+2\varepsilon)^2\langle u_1, w\rangle^2 \leqslant \varepsilon/10$。我们调用引理 2.7，

$$\left\|A\left(I - ww^\top\right)\right\|_{\text{op}} > (1+\varepsilon) \min_{\|u\|_2=1} \left\|A\left(I - uu^\top\right)\right\|_{\text{op}},$$

从而证明完毕。 □

2.2 将 Krylov 下界提升为矩阵-向量下界

在第 2.1 节中，我们说明了为什么对于足够小的常数 $c\log n/\sqrt{\varepsilon}$，Krylov 子空间经过这些迭代后，仍无法足够准确地定位最大特征向量，以执行谱低秩近似。或许出人意料的是，事实证明，在 Krylov 迭代中以一个矩阵而非单个向量作为起点并无帮助。具体而言，我们证明，即使有 $c\log n/\sqrt{\varepsilon}$ 个起始向量，并对每个起始向量运行 $c\log n/\sqrt{\varepsilon}$ 次 Krylov 迭代（算法 5.1），所得的 Krylov 子空间中仍不包含任何与输入矩阵的最大特征向量相关性达到 ε 的向量。

在单起始向量 Krylov 迭代的下界证明中，我们表明，对于 Krylov 子空间中的任意单位向量 w，$|\langle u_1, w\rangle|$ 都非常小。直观地说，这意味着对于分块大小S，块 Krylov 子空间中的单位向量的范数不应超过$S \cdot |\langle u_1, w\rangle|$。确实，我们可以写成$w = w_1 + w_2 + \cdots + w_S$，其中每个$w_i$都来自由第$i$个起始向量生成的 Krylov 子空间，因此 $|\langle u_1, w\rangle| \leqslant \sum_{i=1}^{S} |\langle u_1, w_i\rangle|$。

然而，执行这种方法的障碍在于，向量w_i可能存在反相关并相互抵消。因此，某些w_i的范数可能远大于 1。我们克服了这一障碍，并在构造中证明，以高概率，任意选取的潜在w_i与w_j之间的反相关都非常弱。这一洞见使得我们的单向量 Krylov 迭代下界可以推广到块 Krylov 迭代。

最后，我们可以应用提升结果来证明，任何进行$\log n/\sqrt{\varepsilon}$次查询的自适应算法都可以由块 Krylov算法进行模拟。直观地说，对于特征向量构成均匀Haar随机矩阵的输入实例，自适应算法所能做的最好选择，就是在迄今为止已探索的子空间的补空间中探索一个均匀随机方向。我们通过证明自适应查询序列可以建模为$\{v_i\}_{i\in[k]}$，其中$v_i = v^{\|}{}_i + v_i^{\perp}$，满足$v^{\|}{}_i$是$v_i$在先前查询的张成空间中的分量，而$v_i^{\perp}$与该张成空间正交。此外，$v_i^{\perp}$的分布在剩余子空间上是均匀随机的。然后我们证明，具有较大分块大小的 Krylov 迭代可以模拟此类自适应查询及其响应（详见第 6 节）。

2.3 更优的 Krylov子空间算法

最后，我们证明，在p有界于某个固定常数时，可以改进 Bakshi、Clarkson 和 Woodruff 得到的 Schatten- [BCW22], p 低秩近似上界。这包括 Frobenius 范数和核范数低秩近似这两个重要特例。从高层次来看，他们的算法构造了两个 Krylov 子空间：一个以单个向量为起始向量，迭代$O\left(\log(n/\varepsilon)/\varepsilon^{1/3}\right)$次；另一个子空间的起始分块大小为$O\left(1/\varepsilon^{1/3}\right)$，迭代$O(\log(n/\varepsilon))$次。他们的分析关键在于利用较大的起始分块大小来发挥奇异值间隙的作用。

相反，我们证明，从单个起始向量出发，运行 Krylov 迭代（算法 2.2），经过$t = O\left(\log(1/\varepsilon)/\varepsilon^{1/3}\right)$次迭代后，会收敛到一个满足以下条件的单位向量$v$：

$$\left\|A\left(I - vv^{\top}\right)\right\|_{\mathcal{S}_p} \leqslant (1+\varepsilon)\min_{\|u\|_2=1}\left\|A\left(I - uu^{\top}\right)\right\|_{\mathcal{S}_p}.$$

为简便起见，我们讨论 Frobenius 范数低秩近似的情形（$p = 2$）；一般情形只需将所有使用勾股定理的地方替换为 [BCW22] 中得到的适当推广即可（详见引理 7.5）。此外，为便于概述，我们假设$\sigma_1 = 1$（最终的证明实际上并不需要知道谱范数）。

我们进行的分类分析与 [BCW22]中类似。首先，考虑前t个奇异值较大且不会形成间隙的情形，即$\sum_{i\in[t]}\sigma_i^2 \geqslant 1/\varepsilon^{1/3}$。在这种情况下，我们注意到最优解本身的代价很大：

$$\min_{\|u\|_2=1}\left\|A\left(I - uu^{\top}\right)\right\|_F^2 = \sum_{i=2}^{n}\sigma_i^2 \geqslant \varepsilon^{-1/3} - 1.$$

对上述代价的$(1+\varepsilon)$相对误差解对应于加性$\varepsilon^{2/3}$误差。Krylov 迭代 [MM15], 的标准分析指出，经过$q = \log(n)/\sqrt{\zeta}$次迭代后，对于任意$0 < \zeta < 1$，算法都会输出一个向量v，使得$v^{\top}A^{\top}Av \geqslant \|A\|^2_{\mathrm{op}} - \zeta\sigma_2^2$。根据勾股定理

定理,

$$\left\|A\left(I-vv^{\top}\right)\right\|_F^2=\|A\|_{\mathrm{op}}^2-\|Av\|_2^2\leqslant\min_{\|u\|_2=1}\left\|A\left(I-uu^{\top}\right)\right\|+\sigma_2^2\zeta.$$

由于 $\sigma_2^2\leqslant 1$，只需设定 $\zeta=\varepsilon^{2/3}$，因此 $\mathrm{q}=\mathrm{O}(\log(\mathrm{n})/\varepsilon^{1/3})$ 次迭代即可。我们通过证明一个维数显著更低的 Krylov子空间（对应于$\mathrm{O}\left(\log(1/\varepsilon)/\varepsilon^{1/3}\right)$ 次迭代）张成了一个向量v，使得 $\left\|\mathrm{A}\left(\mathrm{I}-\mathrm{vv}^{\top}\right)\right\|^2\mathrm{F}\leqslant\min_{\|u\|_2=1}\left\|\mathrm{A}\left(\mathrm{I}-\mathrm{uu}^{\top}\right)\right\|+\varepsilon^{2/3}$，从而强化了这一分析。我们通过显式分析切比雪夫多项式来实现这一点（而不是分析 [MM15] 中阈值函数的多项式近似），并证明算法 2.2 的输出至少与输出上述向量 v 一样好（详见引理 7.7）。

在互补情形下，我们与此前对 Krylov 迭代的任何分析（包括 [MM15、BCW22]都有显著不同。在这里，我们知道，范围$[1/2,1-\varepsilon]$ 内的奇异值数量至多为 $O(\varepsilon^{-1/3})$。因此，我们构造了一个完全不同的多项式，它不再基于切比雪夫多项式。该多项式旨在将范围 $[1/2,1-\varepsilon]$内的所有奇异值显式置零。我们注意到，该多项式的次数为$p_0,$，仅为$O(\varepsilon^{-1/3})$，这使我们能够消除所有中等大小奇异值的贡献，类似于从更大的分块大小开始。然而，对于某些位于区间 $[1/2,1-\varepsilon]$之外的 σ_j，仍有可能出现$p_0(\sigma_1)$显著小于$p_0\left(\sigma_j\right)$的情况。

为了解决这个问题，我们考虑多项式 $\mathrm{p}_1(\mathrm{x})=\mathrm{x}^q\mathrm{p}_0(\mathrm{x})$，其中 $\mathrm{q}=\mathrm{O}\left(\log(1/\varepsilon)/\varepsilon^{1/3}\right)$。我们证明，由于 p_1 会显著放大最大的奇异值，因此对于任意$\sigma_j<1/2$，$\mathrm{p}_1(\sigma_1)\gg\mathrm{p}_1\left(\sigma_j\right)$。接着我们证明，向量 $\mathrm{v}=\mathrm{p}_1(\mathrm{A})\mathrm{g}/\|\mathrm{p}_1(\mathrm{A})\mathrm{g}\|$ 可以得到一个 $(1+\varepsilon)$ 相对误差低秩近似；而经过 $\mathrm{O}\left(\log(1/\varepsilon)/\varepsilon^{1/3}\right)$ 次迭代后，Krylov 迭代输出的向量至少能达到同样好的效果。

3 其他相关工作

近年来，矩阵-向量乘积模型在理论计算机科学界受到了广泛关注，因为该模型已针对若干问题进行了形式化 [SWYZ19、RWZ20]。与此同时， [SAR18、BHSW20] 得到了估计最大特征向量和特征值的近乎紧致界。接下来，对于估计半正定矩阵迹的问题， [MMMW21] 得到了紧致界（另见 [DM23] 关于对角线估计的工作）。对于从随机图中恢复隐藏团的问题， [RWYZ21]得到了上界和下界。最后， [BCW22] 研究了低秩近似问题， [NSW22] 则在矩阵-向量模型中研究了检验矩阵是否为半正定的问题。

与矩阵-向量模型密切相关的一种设定是以非自适应方式访问输入，即在向$v^1,$、...、v^q量A进行任何查询之前，就已选定。这一模型等价于草图模型，该模型本身已得到充分研究（例如，参见[Nel11、Woo14]），也在数据流背景下得到研究 [Mut05、LNW14b]。在该模型下，基于 Schatten 范数的低秩近似已得到广泛研究 [CW13、LW20]。

迭代方法（例如基于 Krylov 子空间的方法）可由矩阵-向量乘积框架刻画，而线性草图方法允许选择一个矩阵 $S\in\mathbb{R}^{t\times n}$，其中$t$ 是“查询”的次数，然后观察乘积 $S\cdot A$，依此类推（参见 [Woo14] 及其中的参考文献）。该模型在流式和分布式算法中有重要应用，近期的多项研究聚焦于估计谱范数和最大的奇异值

值 [AN13，LNW14a，LW16b，BBK+21]，估计 Schatten 范数和 Ky-Fan 范数 [LW16b，LW17，LW16a，BKKS19] 以及低秩近似 [CW13，MM13，NN13，BDN15，Coh16]。

最后，矩阵-向量乘积模型也与次线性时间/查询算法和量子启发算法密切相关。在对输入 [MW17、BW18、IVWW19、SW19、BCW20]以及量子启发模型 [KP16、CLW18、Tan19、GLT18、GSLW19、CCHW20、BT23]下研究次线性低秩近似的工作大量涌现。

4 预备知识

给定一个秩为 r 的n ×d 矩阵 A，以及 n ⩾ d，我们可以计算其奇异值分解，记为 SVD(A) = $U\Sigma V^\top$，其中 U 是一个具有正交归一列的n ×r 矩阵，$V^\top$ 是一个具有正交归一行的r × d 矩阵，而 Σ 是一个r ×r对角矩阵。对角线上的元素是A的奇异值，记为 $\sigma_1, \sigma_2 \ldots \sigma_r$。给定一个整数k ⩽ r，我们定义A的截断奇异值分解，将A中除前k个奇异值以外的所有奇异值置零，即 $A_k = U\Sigma_k V^\top$，其中 Σ_k 的对角线上只有k 个非零元素。众所周知，在任意酉不变范数下，截断 SVD 都能计算出 A的最佳秩 k 近似；特别地，对于任意 Schatten-p 范数（定义如下），我们有$A_k = \min_{\mathrm{rank}(X)=k} \|A - X\|_{\mathcal{S}_p}$。更一般地，对于任意矩阵 M，我们使用记号 M_k和$M^{\backslash k}$ 分别表示前 k 个分量和除前k 个分量之外的所有分量。我们分别使用 $M_{i,*}$ 和$M_{*,j}$ 表示 M 的第 i^{th} 行和第 j^{th} 列。

我们使用记号 I_k 表示截断单位矩阵，即一个方阵，其前k个对角元素为 1，其余元素均为 0。I_k 的维度将由上下文确定。

Schatten 范数。 我们回顾 Schatten-p 范数的一些基本事实。我们还需要以下迹不等式和算子不等式。

定义 4.1 （Schatten-p 范数）。 给定一个矩阵 $A \in \mathbb{R}^{n\times d}$，设 $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_d$ 为A的奇异值。那么，对于任意$p \in [0,\infty)$，A 的 Schatten-p 范数定义为

$$\|A\|_{\mathcal{S}_p} = \mathrm{tr}\left(\left(A^\top A\right)^{p/2}\right)^{1/p} = \left(\sum_{i\in[d]} \sigma_i^p(A)\right)^{1/p}.$$

事实 4.2 （Schatten-p 范数具有酉不变性）。 给定一个 $n \times d$ 矩阵 M，对于任意一个列正交的 $m \times n$ 矩阵 U，若范数 $\|\cdot\|_X$ 满足 $\|UM\|_X = \|M\|_X$，则称其具有酉不变性。对于所有 $p \geqslant 1$，*Schatten-p* 范数都具有酉不变性。

对于 Schatten-p 范数下的低秩近似问题，存在一个闭式表达式：

事实 4.3（Schatten-p 低秩近似）。 给定一个矩阵 $A \in \mathbb{R}^{n\times d}$ 和一个整数 $k \in \mathbb{N}$，

$$A_k = \arg\min_{\mathrm{rank}(X)\leqslant k} \|A - X\|_{\mathcal{S}_p},$$

其中 A_k 是 A 的截断 *SVD*。

Cheby 舍夫多项式。接下来，我们回顾一些关于切比雪夫多项式的基本知识。 s.

定义4.4（切比雪夫多项式的极值 [MH02]）。d 阶切比雪夫多项式的局部极值，T_d（x），位于范围 $[-1,1]$ 内，具体为

$$x_i = \cos\left(\frac{i}{d}\cdot\pi\right),$$

对于所有 $i \in [d-1]$。此外，对于每个极值点 x_i，$T_d(x_i)\in\{-1,1\}$。最后，$T_d(x)\in\{-1,1\}$ 的解集为，

$$x_i = \cos\left(\frac{i}{d}\cdot\pi\right),$$

对所有 $i \in \{0,1,\ldots,d\}$.

事实 4.5 (切比雪夫多项式是极值的 [MH02])。 设 $q(x)$ 为任意一个次数为 d 的多项式，满足对所有 $x_i = \cos\left(i\,d\ \cdot\pi\right) for\text{-}i \in \{0,1,\ldots,d\}$, $|q(x_i)| \leqslant 1$。那么，对于任意 $x > 1$, $|q(x)| \leqslant T_d(x)$。

事实 4.6（切比雪夫多项式的增长 [MH02]）。 对于任意 $0 < \varepsilon < 0.5$,

$$e^{c\cdot\min(\sqrt{\varepsilon}\cdot d,\varepsilon\cdot d^2)} \leqslant T_d(1+\varepsilon) \leqslant e^{C\cdot\min(\sqrt{\varepsilon}\cdot d,\varepsilon\cdot d^2)},$$

对于某些固定常数 $C > c > 0$。

以下是事实 4.6 的一个著名推论

推论 4.7。 存在常数 $c > 0$，使得对于任意 $\varepsilon < 0.5$，存在一个次数为 d 的多项式 $T_{d,\,\varepsilon}$ 使得 $|T_{d,\,\varepsilon}(x)| \leqslant e^{-c\cdot\min(\sqrt{\varepsilon}\cdot d,\varepsilon\cdot d^2)}$ 对所有 $0 \leqslant \bar{x} \leqslant 1-\varepsilon$ 均成立，且 $T_{d,\,\varepsilon}(1) = 1$。

证明。令 T_d 为次数为 d 的切比雪夫多项式。定义 $T_{d,\varepsilon}(x) := T_d(x+\varepsilon)/T_d(1+\varepsilon)$。于是，$T_{d,\varepsilon}(1) = 1$。此外，对于任意 $x \in [0,1-\varepsilon]$, $|T_{d,\varepsilon}(x)| \leqslant |T_d(x+\varepsilon)|/|T_d(1+\varepsilon)|$。然而，根据切比雪夫多项式的定义（当 $0 \leqslant x \leqslant 1-\varepsilon$ 时）可得 $|T_d(x+\varepsilon)| \leqslant 1$，并且 $|T_d(1+\varepsilon)| \geqslant e^{c\cdot\min(\sqrt{\varepsilon}\cdot d,\varepsilon\cdot d^2)}$，其中 c 与事实 4.6 中的常数相同。因此，对所有 $0 \leqslant x \leqslant 1-\varepsilon$，均有 $|T_{d,\varepsilon}(x)| \leqslant e^{-c\cdot\min(\sqrt{\varepsilon}\cdot d,\varepsilon\cdot d^2)}$。□

此外，作为事实 4.5 和4.6的直接推论，我们有如下结论。

推论 4.8。对于任意 $0 < \varepsilon < 0.5$，以及任意次数不超过 d 的多项式 P,

$$P(1+\varepsilon) \leqslant e^{C\cdot\min(\sqrt{\varepsilon}\cdot d,\varepsilon\cdot d^2)}\cdot\max_{i\in\{0,1,\ldots,d\}}\left|P\left(\cos\left(\frac{i}{d}\cdot\pi\right)\right)\right|.$$

证明.令 $x_i := \cos\left(i\,d\ \cdot\pi\right)$. 首先，如果 $P(x_i) = 0$ 对所有 $i \in \{0,1,\ldots,d\}$, 那么 P至少有d 个根，因此 $P \equiv 0$. 或者，如果 $\max_{i\in\{0,1,\ldots,d\}}|P(x_i)| = \kappa > 0$，那么 $q(x) = P(x)\,\kappa$ 满足

$|\quad_i\,| \leqslant 1$ 4 5 $|\ 1+\varepsilon\ | \leqslant\ _d\ 1+\varepsilon$ $|\ 1+\varepsilon\ | \leqslant \kappa\cdot\ _d\ 1+q(x)$

$\varepsilon \leqslant \kappa\cdot\ ^{C\cdot\min(\sqrt{\varepsilon}\cdot d,\varepsilon\cdot d^2}$ 4 6

对所有 i，这意味着根据事实. , $q(\,)\,T\,(\,)$。所以，$P(\,)\,T\,(\,)\,e\,)$，使用事实 . . □

概率背景。 我们还需要了解以下关于概率分布的基本事实：

事实 4.9 （高斯分布的偏差）。 设 $g \sim \mathcal{N}(0,1)$。那么，对于任意 $\delta > 0$，以至少$1 - 1/\delta$的概率，$|g| \leqslant \sqrt{\log(1/\delta)}$。

事实 4.10 （Hanson-Wright）。 设 $g \sim \mathcal{N}(0, I_d)$。那么，

$$\mathbf{Pr}\left[\left|\|g\|^2 - d\right| > t\right] \leqslant 2\exp(-ct).$$

事实 4.11 （高斯分布矩阵的奇异值 [Ver10]）。 设 $G \in \mathbb{R}^{n\times d}$ 满足对所有 $i \in [n], j \in [d]$，$G_{i,j} \sim \mathcal{N}(0,1)$。设 $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_d$ 为 G 的奇异值。那么，以至少 $1 - 2\exp(-d/2)$的概率，

$$\sqrt{n} - 2\sqrt{d} \leqslant \sigma_d \leqslant \sigma_1 \leqslant \sqrt{n} + 2\sqrt{d}.$$

我们还需要以下定义：

定义 4.12 （Haar 随机矩阵）。 设 $\mathbb{O}(n)$ 为作用于 $n \times n$ 矩阵的正交群。存在 μ 于 $\mathbb{O}(n)$上的唯一旋转不变概率测度（Haar 测度）。Haar 随机矩阵U 是从 μ中抽取的（矩阵值）样本。

5 针对块 Krylov 方法的下界

本节将展示，在基于 Krylov 方法的算法中增加分块大小，并不能帮助我们解决第 2 节中的困难实例。特别地，我们将展示，从分块大小为 $c\log(n)/\sqrt{\varepsilon}$ 开始，其中 c 是一个较小的固定常数，并运行 $c\log(n)/\sqrt{\varepsilon}$ 次迭代，并不足以获得 $(1+\varepsilon)$-谱范数低秩近似。

算法 5.1(块 Krylov“算法”， [MM15] 的推广**)**。

输入： 一个 $n \times n$ 矩阵 A，迭代次数 r，分块大小 s。

1. 设 $g_1, \ldots, g_s$ 是从 $\mathcal{N}(0, I)$ 中独立同分布抽取的向量。令 $\mathcal{K} = \{A^t g_j\}_{0\leqslant t\leqslant r, 1\leqslant j\leqslant s}$ 为这些向量构成的集合。 2. 选择一个单位向量 $v \in$ 张成空间$(\mathcal{K})$，使 $\left\| A(I - vv^\top)\right\|_{\mathrm{op}}$ 最小。

$Avv^\top$.

定理 5.2 $\Omega(1/\varepsilon^{2.01})$ $0 < \varepsilon < \frac{1}{2}$ $\geqslant$。(谱低秩近似对块 Krylov 方法而言是困难的)。 给定 $\times$，令 n /。那么，存在一个定义在 $n\,n$ 矩阵 A 上的分布和某个较小的绝对常数$1 > c > 0$，使得以至少 9/10 的概率，对于由 $s = c\log n/\sqrt{\varepsilon}$个随机单位向量生成的 *Krylov* 子空间中的每个向量 v，当迭代次数 $r = c\log n/\sqrt{\varepsilon}$时（见算法 5.1），

$$\left\|A\left(I - vv^\top\right)\right\|_{\mathrm{op}} \geqslant (1+\varepsilon)\min_{\|u\|_2=1}\left\|A\left(I - uu^\top\right)\right\|_{\mathrm{op}}.$$

我们的困难分布本质上与之前相同，但为了完整起见，我们在此重新陈述。我们构造一个矩阵 $A = UDU^\top$，其中 U 是一个均匀随机的 $n \times n$ 正交矩阵，

且D 是由特征值构成的固定对角矩阵。D 有$O(\log(n)/\sqrt{\varepsilon})$ 个不同的特征值，其中最大特征值 $\lambda_1 = 1+2\varepsilon$ 的重数为 1，其余不同的特征值 λ_2、λ_3、.....、λ_{r+2} 的重数为$O(n\sqrt{\varepsilon}/\log(n))$。我们将其余特征值设为次数为$r$ 的切比雪夫多项式T_r 等于 1 或 -1 的不同位置。

定义 5.3（困难分布）。给定 $\varepsilon > 0$，令 $\mathrm{n} \in \mathbb{N}$ 满足 $\mathrm{n} = \Omega(1/\varepsilon^{2.01})$，并令 $\mathrm{r} = \mathrm{c}\log(\mathrm{n})/\sqrt{\varepsilon}$，其中 c 为某个固定的小常数。此外，假设 $\mathrm{k} = \mathrm{n}_{-1}\mathrm{r}_{+1}$ 为整数。对于 $\mathrm{t} \in [2, \mathrm{r}+2]$，令 $\lambda_t := \cos\left(\frac{t-2}{\mathrm{R}} \cdot \pi\right)$。然后，我们将 D 定义为对角矩阵，其中 $\mathrm{D}_{1,1} = \lambda_1 = 1+2\varepsilon$，并且对于每个 $2 \leqslant \mathrm{i} \leqslant \mathrm{n}$，$\mathrm{D}_{i,i} = \lambda_{1+\lceil \frac{i-1}{k} \rceil}$，从而每个 λ_t for $2 \leqslant \mathrm{t} \leqslant \mathrm{r}+2$ 恰好重复 k 次。最后，我们定义 $\mathrm{A} = \mathrm{UDU}^\top$，其中 U 是从均匀分布中随机抽取的 $\mathrm{n} \times \mathrm{n}$ 正交矩阵，而 D 是上文定义的固定对角矩阵。

注意，λ_2、. . .、λ_r 都属于值域 $[-1, 1]$，事实上 $\lambda_2 = 1$ 和 $\lambda_{r+2} = 1$。

由埃克哈特-扬定理以及 $\max_{2\leqslant i\leqslant r+2} |\lambda_i| = 1$ 这一事实，以下结论显而易见。

事实 5.4. 我们有 $\min_{\|u\|_2=1} \left\|A - Auu^\top\right\|_{\mathrm{op}} = 1$。

我们还回顾以下结果。

引理 2.7. *[与谱低秩近似的对齐]* 给定 $\varepsilon > 0$，设 A 为一个矩阵，且 u_1 是 A 的最大特征向量。此外，设 $\lambda_1 = 1+2\varepsilon$ 和 $\lambda_2 = 1$ 是 A 的两个最大特征值，并假设对所有 $i \geqslant 2$，均有$|\lambda_i| \leqslant 1$。设 w 为一个向量，使得 $\langle u_1, w\rangle^2$，$\langle u_1, Aw\rangle^2 \leqslant \varepsilon/2$。那么，

$$\left\|A\left(I - ww^\top\right)\right\|_{\mathrm{op}} > (1+\varepsilon) \min_{\|u\|_2=1} \left\|A\left(I - uu^\top\right)\right\|_{\mathrm{op}}.$$

我们首先给出以下引理:

引理 *5.5*（集中性质）。设 $A = V\Sigma V^\top$ 是定义 *5.3* 中困难实例的一个样本，满足 $r = s = c\,log(n)/\sqrt{\varepsilon}$，且 $k = (n-1)/(r+1)$。设 $g_1, g_2, \ldots, g_s \sim \mathcal{N}(0, I)$，使得对每个 $j \in [s]$，g^j 在 A 的特征基下具有如下分解：$g^j = a_{j,1}v_1 + \sum_{t\in[2,r+2]} a_{j,t}v_t$，其中 $v_1 = V_{*,1}$，且对任意 $t \in [2, r+2]$，$v_t = \sum_{i\in[k]} V_{*,(t-2)k+i+1}$。假设 $k \geqslant 100s$，则以下各项均以至少 *0.9* 的概率成立。

1. 对所有 $j \in [s]$, $a_{j,1} \leqslant 5\sqrt{\log n}$。 *2.* 对所有 $j \in [s]$ 以及所有 $2 \leqslant t \leqslant r+2$，$0.5\sqrt{k} \leqslant a_{j,t} \leqslant 2\sqrt{k}$。

*3.*对于所有 $2 \leqslant t \leqslant r+2$，定义 $V^{(t)} \in \mathbb{R}^{n\times s}$ 为行 j 等于$v_{j,t}$的矩阵。然后，对于所有 $t \leqslant r+2$，$V^{(t)}$ 的所有奇异值都介于 1 4 和 4 之间。

证明。 由于最大特征向量的重数为 1，$a_{j,1}$ 是标准高斯分布 $\mathcal{N}(0,1)$的绝对值。因此，通过对 $s \leqslant n$个j 选项应用并合界，Part 1 以至少 0.99 的概率成立。

对于任意 $2 \leqslant \mathrm{t} \leqslant \mathrm{r}+2$，$\mathrm{t}^{\mathrm{th}}$ 特征向量的重数为 k，因此 $\mathrm{a}_{j,t}$ 的分布是一个 k 维高斯分布的范数。由于 $\mathrm{k} = \Theta\left(\mathrm{n}\,\mathrm{r}\right)$ 且 $\mathrm{r} \leqslant \log\mathrm{n}\sqrt{\varepsilon}$，如果 $\mathrm{n} \gg \varepsilon^{-1}\log^2\varepsilon^{-1}$，那么 $\mathrm{k} \geqslant \sqrt{\mathrm{n}}$。因此，根据标准集中不等式，$0.5\sqrt{\mathrm{k}} \leqslant \mathrm{a}_{s,t} \leqslant 2\sqrt{\mathrm{k}}$ 至少为 $1 - \mathrm{e}^{-\Omega(k)} \geqslant 1 - \mathrm{e}^{-\Omega(\sqrt{n})}$ 的概率为.....。对 s 和 t 的至多 n^2 种选择应用并合界，Part 2 以至少 0.99 的概率成立。

由于每个 $v_{j,t}$ 都已归一化为单位向量，我们可以写成 $V^{(t)} = G^{(t)} \cdot \mathrm{diag}(a_t)^{-1}$，其中 $G^{(t)}$ 的第 s 列为 $a_{j,t}v_{j,t}$，而 a_t 是向量 $\{a_{j,t}\}_{1\leqslant j\leqslant s}$。假设 Part 2 成立，则 $\mathrm{diag}(a_t)^{-1}$ 的所有奇异值都在范围 $\left[0.\sqrt{5}k, \sqrt{2}k\right]$ 内。此外，我们可以将 $G^{(t)}$ 的每一列投影到 λ_t- 特征空间上，此后每一列都是 k 维空间中的随机高斯向量。假设 $k \geqslant 100s$，根据事实 4.11，$G^{(t)}$ 的所有奇异值都必定以至少 $1 - \frac{1}{n^2}$ 的概率落在范围 $\left[0.5\sqrt{k}, 2\sqrt{k}\right]$ 内。因此，以至少 0.98 的概率，对于所有 t，$V^{(t)}$ 的所有奇异值都介于 1 4 和 4 之间。

□

从现在起，我们假设引理 5.5 中的三个事件均成立，且不使用关于高斯向量 g_s的其他性质。根据引理 2.7，只需证明以下引理。

引理 5.6。假设引理5.5 中的三个事件均成立。那么，对于任意单位向量 $v \in$ 张成空间$(\mathcal{K})$,$|\langle v, u_1\rangle| < \sqrt{\frac{\varepsilon}{10}}$。

证明。对于$j \leqslant s, \ell \leqslant r+2$，我们计算向量 $A^\ell g_j = \sum_{t=1}^{r+2} \lambda_t^\ell \cdot a_{j,t}v_{j,t}$。现在，任何 $v \in$ 张成空间$(\mathcal{K})$中的向量，都必须是跨越 $0 \leqslant \ell \leqslant r$ 和$j \leqslant s$的$A^\ell g_j$ 的线性组合。这意味着，任何这样的向量都必须能够表示为以下形式

$$\sum_{j=1}^{s}\sum_{\ell=0}^{r} p_{j,\ell}\cdot\sum_{t=1}^{r+2}\lambda_t^\ell\cdot a_{j,t}v_{j,t} = \sum_{t=1}^{r+2}\sum_{j=1}^{s} a_{j,t}v_{j,t}\cdot\sum_{\ell=0}^{r} p_{j,\ell}\lambda_t^\ell = \sum_{t=1}^{r+2}\underbrace{\sum_{j=1}^{s} a_{j,t}\phi_j(\lambda_t)v_{j,t}}_{x_t},$$

对于某些多项式 $\phi_1, \ldots, \phi_s$，每个的次数至多为 r。此外，注意到 $v_{j,1} = \pm u_1$ 对所有 j都成立，因为 λ_1 的特征子空间维数为 1。

首先，考虑某个固定的$t \geqslant 2$。我们将界定x_t 的范数：$= \sum_{j=1}^{s} a_{s,t}\phi_j(\lambda_t)v_{j,t}$。回忆一下，$V^{(t)} \in \mathbb{R}^{n\times s}$ 是列j 等于$v_{j,t}$的矩阵，引理5.5 的第 3 部分告诉我们，$V^{(t)}$ 的所有奇异值都介于 14 和 4 之间。我们可以写成 $x_t = V^{(t)} \cdot w_t$，其中$w_t \in \mathbb{R}^s$ 的j^{th}坐标等于$a_{j,t}\phi_j(\lambda_t)$。因此，

$$\|x_t\|_2^2 \geqslant \frac{1}{16}\cdot\sum_{j=1}^{s}(a_{j,t}\phi_j(\lambda_t))^2 \geqslant \frac{k}{64}\cdot\sum_{j=1}^{s}\phi_j(\lambda_t)^2.$$

第一个不等式由引理 5.5的第3部分推出，第二个不等式由引理 5.5的第2部分推出。

此外，我们知道每个 X_t 都是特征值为 λ_t 的特征向量，就像每个 $v_{j,t}$ 一样。因此，它们彼此正交。这意味着，如果 $v = \sum_{t=1}^{r+2} X_t$ 的范数为 1，那么 $\sum_{t=1}^{r+2} \|X_t\|_2^2 = 1$。不过，请注意，

$$\sum_{t=2}^{r+2}\|X_t\|_2^2 \geqslant \frac{k}{64}\cdot\sum_{t=2}^{r+2}\sum_{j=1}^{s}\phi_j(\lambda_t)^2 \geqslant \frac{k}{64}\cdot\sum_{j=1}^{s}\max_{2\leqslant t\leqslant r+2}\phi_j(\lambda_t)^2. \tag{7}$$

接下来，我们将推论 4.8 应用于 ϕ_j，其次数至多为 r，从而说明（7）至少为

$$\begin{aligned}\frac{j}{64}\cdot\sum_{j=1}^{s}\max_{0\leqslant i\leqslant r}\phi_j\left(\cos\left(\frac{i}{r}\cdot\pi\right)\right)^2 &\geqslant \frac{k}{64}\cdot\sum_{j=1}^{s}\phi_j(1+2\varepsilon)^2\cdot e^{-2C\cdot\min(\sqrt{2\varepsilon}\cdot r, 2\varepsilon\cdot r^2)}\\ &\geqslant \frac{j}{64}\cdot e^{-4C\cdot\sqrt{\varepsilon}\cdot r}\cdot\sum_{j=1}^{s}\phi_j(\lambda_1)^2.\end{aligned}$$

注意，$\sum^{r+2}{}_{t=1} \|x_t\|_2^2 = 1$，这意味着

$$\sum_{j=1}^{s} \phi_j(\lambda_1)^2 \leqslant \frac{64}{k} \cdot e^{4C\cdot\sqrt{\varepsilon}\cdot r}. \tag{8}$$

反之，由于 $v_{s,\,1} = \pm u_1$ 对所有 s，$x_1 : = \sum_{j=1}^{s} a_{j,\,1}\phi_j(\lambda_1)\, v_{s,\,1}$ 的范数至多为 $5\sqrt{\text{对数} n \cdot \sum_{j=1}^{s} |\phi_j(\lambda_1)|}$，根据三角不等式和引理 5.5 的第 1 部分。因此，

$$\begin{aligned} |\langle v, u_1\rangle| = \|X_1\|_2 &\leqslant 5\sqrt{\log n} \cdot \sum_{j=1}^{s} |\phi_j(\lambda_1)| \\ &\leqslant 5\sqrt{\log n} \cdot \sqrt{s \cdot \sum_{j=1}^{s} \phi_j(\lambda_1)^2} \\ &\leqslant 40\sqrt{\frac{\log n \cdot s}{k}} \cdot e^{2C\cdot\sqrt{\varepsilon}\cdot r}. \end{aligned}$$

上面，第二个不等式由柯西-施瓦茨不等式得出，最后一个不等式由（8）。

由于对于充分小的常数 c，$r = c\log n/\sqrt{\varepsilon}$，因此 $e^{2C\cdot\sqrt{\varepsilon}\cdot r} = e^{2cC\log n} \leqslant n^{0.001}$。此外，假设 n 至少是一个充分大的常数，则 $k \geqslant n\, 2r$，所以

$$\begin{aligned} |\langle v, u_1\rangle| &\leqslant 40 \cdot \sqrt{\frac{\log n \cdot s \cdot 2r}{n}} \cdot n^{0.001} \\ &\leqslant 80 \cdot \frac{(\log n)^{1/2}}{n^{0.499}} \cdot \frac{c\log n}{\varepsilon^{1/2}} \\ &\leqslant \frac{1}{10 \cdot n^{0.498} \cdot \varepsilon^{1/2}}. \end{aligned}$$

因此，如果 $n \geqslant \Omega(\varepsilon^{-2.01})$，这至多为 $\sqrt{\frac{\varepsilon}{10}}$，正如所需。

□

定理 5.2 的证明。要理解引理 5.6 为何蕴含定理5.2，假设存在v ∈张成空间($\mathcal{K}$) 使得 $\|A(I - vv^\top)\|_{op} \leqslant (1+\varepsilon)\min_{\|u\|_2=1} \|A(I - uu^\top)\|_{op}$。然后，根据引理2.7，要么 $\langle v, u_1\rangle^2 > \frac{\varepsilon}{2}$ 要么 $\langle Av, u_1\rangle^2 > \frac{\varepsilon}{2}$。然而，$\|A\|_{op} \leqslant 1 + 2\varepsilon$，这意味着要么$\langle v, u_1\rangle^2 > \frac{\varepsilon}{2}$要么 $\langle Av\|Av\|_2, u_1\rangle^2 > \frac{\varepsilon}{10}$。通过将幂方法迭代次数增加 1，我们可以假设 v 和$Av\|Av\|_2$ 都在 $\mathcal{K}$ 中。因此，我们得到了与引理 5.6 的矛盾，这表明引理 5.6 蕴含定理 5.2。

□

6 针对任意自适应算法的下界

在本节中，我们证明，针对块Krylov算法的下界可以推广为针对任意自适应算法的下界。我们将用到的主要技术结果是一条通用归约定理，它表明，对于一大类矩阵-向量问题，只需证明针对块 Krylov 算法的下界即可。由此，我们将得到针对任意自适应算法的秩一谱低秩近似的最优下界。

6.1 将任意自适应查询规约为块 Krylov

令 $A = U^\top D U$，其中 D 是一个对角矩阵，U 是一个 Haar 随机正交矩阵（参见定义 4.12），并且U 和 D 相互独立。我们考虑以下模型，它是对矩阵-向量乘积模型的加强：

定义 6.1（扩展预言机模型）。给定 $K \in \mathbb{N}$，对于所有 $k \in [K]$，，算法都会选择一个新的查询点 v_k，并接收信息 $\{A^i v_j\}_{(i,j)\in H_k}$，其中 $H_k := \{(i,j) : i + j \leqslant k + 1, i \geqslant 0, 1 \leqslant j \leqslant k\}$ 是非负整数有序对的集合。对于任意集合 S，我们用以下记号 $\{A^i v_j\}_S$ 表示 $\{A^i v_j\}_{(i,j)\in S}$。

请注意，这显然是比通常的矩阵-向量预言机更强的预言机模型，因此，针对扩展预言机模型中算法的下界也意味着针对原始矩阵-向量模型中算法的下界。

定义 6.2（自适应确定性算法）。一个进行 K 次扩展预言机查询（见定义 6.1）的自适应确定性算法 $\mathcal{A}$，由一组确定性函数 $v_1, v_2(\cdot), \ldots, v_K(\cdot)$ 给出，其中 v_1 为常量，而每个 $v_k(\cdot)$ 都是 k 个 $(k+1)^2 - 1$ 输入的函数。这对应于一系列查询：第 k 次查询 $v_k(\{A^i v_j\}_{H_{k-1}})$ 会根据算法在第 k 次迭代开始时可获得的信息自适应地选定。（请注意，v_1没有输入。）当输入的选择可从上下文中清楚得知时，为简洁起见，我们可以直接写作 $v_k = v_k(\{A^i v_j\}_{H_{k-1}})$。

我们所需的主要结果与 [CdDL+23]最近的一项结果密切相关，具体如下。

引理 *6.3*（归约到块 *Krylov*）。设 $v_1, v_2(\cdot), \ldots, v_K(\cdot)$ 是一个进行 K 次查询的自适应确定性算法，其中 $K^2 < d$。递归定义 v_1^{alg}、v_2^{alg}、…、$v^{\text{alg}} K$ 如下：$v_1^{\text{alg}} = v_1$，，以及对于 $k \geqslant 2$，$v^{\text{alg}} k = v_k(\{A^i v^{\text{alg}} j \}_{H_{k-1}})$。设 $z_1, \ldots, z_K$ 是独立同分布的标准高斯向量。那么，根据集合 $\{A^i z_j\}_{i+j\leqslant K}$，我们可以构造一组单位向量 $\tilde{v}1$、$\tilde{v}_2$,…、$\tilde{v}_K$，以及一组旋转矩阵 U_1^{sim}、U_2^{sim}、…、U_K^{sim}，并满足以下性质。

1. $\tilde{v}_k$ 和 U_k^{sim}仅取决于 $\{A^i z_j\}_{i+j\leqslant k}$。此外，$\tilde{v}_k \in$ 张成空间$(\{A^i z_j\}_{i+j\leqslant k})$。
2. $\left((U_{1:K}^{\text{sim}})^\top A (U_{1:K}^{\text{sim}}), \{(U_{1:K}^{\text{sim}})^\top A^i \tilde{v}_j\}_{H_K}\right) \overset{d}{=} \left(A, \{A^i v^{\text{alg}} j \}_{H_K}\right)$，其中 $U_{1:K}^{\text{sim}} := U_1^{\text{sim}} \cdots U_K^{\text{sim}}$。这里，分布等价是关于 A 和 $\{z_i\}_{i\leqslant K}$ 的随机性而言。

重要的是，引理性质2 6.3 大致表明，了解 $A^i \tilde{v}j$ 足以重构自适应算法的查询和响应的分布。

注 6.4。尽管归约引理是针对确定性 算法表述的，但事实证明，移除这一假设相当简单，因为矩阵A 已经随机化。

引理的证明推迟至6.3 附录A。

6.2 通用矩阵-向量下界

现在我们可以结合针对块Krylov算法的信息论下界，以及展示块 Krylov 能够模拟自适应确定性算法的提升陈述（见定义 6.2），从而得到矩阵-向量乘积数量的信息论下界。

定理6.5(谱低秩近似的矩阵-向量下界)。 按定义 5.3固定 D，并假设 $0 < \varepsilon < 0.5$。在分布$A = U^\top DU$ 上，任何（可能经过随机化的）自适应算法都不可能以超过0.2 的概率（关于 A 和算法的随机性）输出一个单位向量 v，使得 $\langle v, u_1(A)\rangle \geqslant \sqrt{\frac{\varepsilon}{10}}$，其中，$u_1(\cdot)$ 表示对称矩阵的最大特征向量。因此，根据引理 2.7，以超过 0.2 的概率输出 A 的 $(1+\varepsilon)$ 近似秩一近似需要 $\Omega\left(\log n_{\sqrt{\varepsilon}}\right)$ 次自适应查询。

证明。 首先我们假设该算法是确定性的，因此其行为由函数 $v_1, v_2(\cdot)$、……、$v_K(\cdot)$刻画，如引理 6.3 中所述。通过额外查询一次 $v_{K+1}(\cdot)$，该算法可以确保 v_{K+1} 与 $\{A^i v_j\}_{H_K}$ 的某个线性组合包含这样一个向量v，其概率至少为 0.2。因此，我们的目标是证明，张成空间$(\{A^i v^{\mathsf{alg}}{}_j\}_{i+j\leqslant K+1})$ 以较高概率无法包含这样一个向量 v。

根据引理 6.3，接收 $\{A^i z_j\}_{i+j\leqslant K+1}$ 的块 Krylov 算法可以生成 $\tilde{v}_1, \ldots, \tilde{v}_{K+1}$，其中每个 $\tilde{v}_k \in$ 都属于张成空间$(\{A^i z_j\}_{i+j\leqslant k})$，并且还能生成 $U_1^{\mathsf{sim}}, \ldots, U_{K+1}^{\mathsf{sim}}$，使得

$$\left((U^{\mathsf{sim}}_{1:(K+1)})^\top A (U^{\mathsf{sim}}_{1:(K+1)}), \{(U^{\mathsf{sim}}_{1:(K+1)})^\top A^i \tilde{v}_j\}_{H_{K+1}}\right) \overset{\mathsf{d}}{=} \left(A, \{A^i v_j^{\mathsf{alg}}\}_{H_{K+1}}\right).$$

那么，如果以至少 0.2 的概率存在一个单位向量 $v \in$ 张成空间$(\{A^i v^{\mathsf{alg}}{}_j\}_{i+j\leqslant K+1})$ 满足$\langle v, u_1(A)\rangle \geqslant \sqrt{\frac{\varepsilon}{10}}$，那么以相同的概率存在一个单位向量

$$v' \in \mathrm{Span}(\{U^\top_{1:(K+1)} A^i \tilde{v}_j\}_{i+j\leqslant K+1}) = U^\top_{1:(K+1)} \cdot \mathrm{Span}(\{A^i z_j\}_{i+j\leqslant K+1})$$

满足 $\langle \mathrm{v}', \mathrm{u}_1((\mathrm{U}_1^{\mathsf{sim}}{:}(\mathrm{K}{+}1))^\top \mathrm{A}(\mathrm{U}_1^{\mathsf{sim}}{:}(\mathrm{K}{+}1)))\rangle \geqslant \sqrt{\frac{\varepsilon}{10}}$。然而，$(\mathrm{U}_1^{\mathsf{sim}}{:}(\mathrm{K}{+}1))^\top \mathrm{A}(\mathrm{U}_1^{\mathsf{sim}}{:}(\mathrm{K}{+}1))$ 的最大特征向量是 $(\mathrm{U}_1^{\mathsf{sim}}{:}(\mathrm{K}{+}1))^\top \mathrm{u}_1(\mathrm{A})$，因此这意味着 $\langle \mathrm{v}', (\mathrm{U}_1^{\mathsf{sim}}{:}(\mathrm{K}{+}1))^\top \mathrm{u}_1(\mathrm{A})\rangle \geqslant \sqrt{\frac{\varepsilon}{10}}$。进而，这说明 $\{\mathrm{A}^i \mathrm{z}_j\}_{i+j\leqslant K+1}$ 的张成空间中存在一个单位向量 v''，满足 $\langle \mathrm{v}'', \mathrm{u}_1(\mathrm{A})\rangle \geqslant \sqrt{\frac{\varepsilon}{10}}$，且该情况发生的概率至少为 0.2，这与引理 5.6 矛盾。因此，任何确定性算法成功的概率都不可能超过 0.2。

如果该算法是随机化的，那么它会使用一个与A无关的随机种子 ξ。因此，在给定随机种子的条件下，该算法成功的概率不会超过 0.2，这意味着随机化算法成功找到此类向量的总体概率也至多为 0.2。

最后，根据引理2.7，对A 的$(1+\varepsilon)$近似秩一近似$Avv^\top$ 要求满足 $\langle u_1, v\rangle^2 \geqslant \varepsilon/2$ 或 $\langle u_1, Av\rangle^2 \geqslant \varepsilon/2$。此外，$\langle u_1, Av\rangle^2 = \langle Au_1, v\rangle^2 = (1+2\varepsilon)^2 \langle u_1, v\rangle^2$。假设 $\varepsilon \leqslant 0.5$，这意味着必须有 $\langle u_1, v\rangle^2 \geqslant \varepsilon/10$。证明至此结束。 □

7 Schatten-p 低秩近似的上界

在本节中，我们将展示，对于所有 $p = O(1)$，只需直接应用以单个起始向量开始的 Krylov 方法即可。查询复杂度仅为 $O(\varepsilon^{-1/3}\log(1/\varepsilon))$，且不会随输入规模增长。更正式地说，我们证明以下定理：

定理 7.1 (更精确的低秩近似算法)。 给定 $A \in \mathbb{R}^{n\times d}$、$\varepsilon > 0$ 以及 $p = O(1)$，存在一种算法，使用 $O\left(\log(1/\varepsilon)/\varepsilon^{1/3}\right)$ 次矩阵-向量乘积，并输出一个单位向量 w，使得以至少0.99的概率，

$$\left\|A\left(I - ww^\top\right)\right\|_{\mathcal{S}_p}^p \leqslant (1+\varepsilon) \min_{\|u\|=1} \left\|A\left(I - uu^\top\right)\right\|_{\mathcal{S}_p}^p.$$

众所周知，在任意 Schatten-p 范数下，A的最优秩一近似具有 $Auu^\top$ 的形式，其中某个 u 的范数为 1。（事实上，最优的u 正是A的顶端奇异向量。）因此，$Aww^\top$ 是A的一个$(1+\varepsilon)$-近似秩一近似。

为了证明定理 7.1，我们注意到以下中间引理。

引理7.2 (矩阵勾股定理)。 对于任意单位向量w， $\|Aww^\top\|_F^2 + \|A(I - ww^\top)\|_F^2 = \|A\|_F^2$。更一般地，对于任意投影矩阵 P， $\|AP\|_F^2 + \|A(I-P)\|_F^2 = \|A\|_F^2$。

更一般地，我们需要以下两个引理，它们将勾股定理推广到 Schatten-p 空间，适用于任意$p \geqslant 1$。

引理7.3(广义勾股不等式引理 5.5, [BCW22])。 设$A \in \mathbb{R}^{n\times d}$，且 P，Q 是等秩的投影矩阵，属于 $\mathbb{R}^{n\times n}$。则，对于任意 $p \geqslant 1$,

$$\|A\|_{\mathcal{S}_p}^p \geqslant \|PAQ\|_{\mathcal{S}_p}^p + \|(I-P)A(I-Q)\|_{\mathcal{S}_p}^p.$$

算法 7.4（Krylov迭代（矩形矩阵））。

输入： 一个 $n \times d$ 矩阵 A，目标秩k，目标精度 $\varepsilon > 0$，以及迭代次数$t = O(p^{-1}\cdot\log(1/\varepsilon)/\varepsilon^{1/3})$。

1. 令 $g \sim \mathcal{N}(0, I)$. 令 $\mathcal{K} = [g; (AA^\top)g; (AA^\top)^2 g; \ldots; (AA^\top)^t g]$ 为所得的$d \times (t+1)$ Krylov 矩阵 。

2. 计算 $\mathcal{K}$ 的列空间的标准正交基 Q。令 $M = Q^\top AA^\top Q$
.3.计算M的最大特征向量向量，并将其记为y_1。 4. 令$v = A^\top Q y_1 / \|A^\top Q y_1\|_2$

输出： $Avv^\top$。

我们将重要地使用引理的以下推论 7.3。

引理 7.5 （引理 5.5 在 [BCW22] 中的推论）。 给定任意单位向量 w 和矩阵 $A \in \mathbb{R}^{n\times d}$，令 $v = A^\top w / \left\|A^\top w\right\|_2$。则，

$$\|A\|_{\mathcal{S}_p}^p \geqslant \left\|ww^\top A\right\|_{\mathcal{S}_p}^p + \left\|A\left(I - vv^\top\right)\right\|_{\mathcal{S}_p}^p$$

证明。令 $\mathrm{P} = \mathrm{ww}^\top$，$\mathrm{Q} = \mathrm{vv}^\top$。注意到 $\mathrm{ww}^\top\mathrm{Avv}^\top = \mathrm{ww}^\top\mathrm{AA}^\top\mathrm{ww}^\top\mathrm{A}\|A^\top w\|^2 = \mathrm{ww}^\top\mathrm{A}$。这也意味着 $\mathrm{ww}^\top\mathrm{A}(\mathrm{I} - \mathrm{vv}^\top) = \mathrm{ww}^\top\mathrm{A} - \mathrm{ww}^\top\mathrm{Avv}^\top = 0$，这意味着 $(\mathrm{I} - \mathrm{ww}^\top)\mathrm{A}(\mathrm{I} - \mathrm{vv}^\top) = \mathrm{A}(\mathrm{I} - \mathrm{vv}^\top)$。应用引理 7.3，我们得到

$$\begin{aligned}\|A\|_{\mathcal{S}_p}^p &\geqslant \left\|(ww^\top)A(vv^\top)\right\|_{\mathcal{S}_p}^p + \left\|\left(I - ww^\top\right)A\left(I - vv^\top\right)\right\|_{\mathcal{S}_p}^p \\ &= \left\|ww^\top A\right\|_{\mathcal{S}_p}^p + \left\|A\left(I - vv^\top\right)\right\|_{\mathcal{S}_p}^p,\end{aligned} \tag{9}$$

如所愿。 □

我们还注意到以下基本引理，它将使我们的目标更加简单。

引理7.6(相关向量到低秩近似)。设 A的奇异值为 $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n \geqslant 0$。对于 $p \geqslant 1$，假设 w 是一个单位向量，使得 $\|A^\top w\|_2^p \geqslant (1+\varepsilon)\sigma_1^p - \varepsilon \cdot \|A\|_{S_p}^p$。此外，令 $v = A^\top w / \|A^\top w\|$。则，$\|A(I - vv^\top)\|_{S_p}^p \leqslant (1+\varepsilon) \cdot$最小值$_{\|u\|_2=1} \|A(I - uu^\top)\|_{S_p}^p$。

证明。根据Eckardt-Young，$\min \|A(I - uu^\top)\|_{S_p}^p = \sum_{i=2}^n \sigma^p$ i。换言之，$\min_{\|u\|=1} \|A(I - uu^\top)\|_{S_p}^p = \|A\|_{S_p}^p - \sigma_1^p$。因此，如果我们找到一个单位向量w，使得 $\|ww^\top A\|_{S_p}^p \geqslant (1+\varepsilon)\sigma_1^p - \varepsilon \cdot \|A\|_{S_p}^p$，那么根据引理 7.5，

$$\begin{aligned}
\left\|A(I - vv^\top)\right\|_{S_p}^p &\leqslant \|A\|_{S_p}^p - \left\|uu^\top A\right\|_{S_p}^p \\
&\leqslant \|A\|_{S_p}^p - (1+\varepsilon)\sigma_1^p + \varepsilon\|A\|_{S_p}^p \\
&= (1+\varepsilon) \cdot (\|A\|_{S_p}^p - \sigma_1^p) \\
&= (1+\varepsilon) \cdot \min_{\|u\|_2=1} \left\|A(I - uu^\top)\right\|_{S_p}^p,
\end{aligned}$$

其中$v = A^\top w / \|A^\top w\|$。

最后，由于$ww^\top A$ 是一个秩为 1 的矩阵，且w 是一个单位向量，其 Schatten-p 范数就是$\|w^\top A\|_2 = \|A^\top w\|_2$。因此，只需满足 $\|A^\top w\|_2^p \geqslant (1+\varepsilon)\sigma_1^p - \varepsilon \cdot \|A\|_{S_p}^p$。 □

引理 7.7 (Krylov 子空间中优良向量的存在性)。设 $\Lambda = AA^\top \in \mathbb{R}^n$，其特征值为 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$。令 $g \sim \mathcal{N}(0, I) \in \mathbb{R}^n$ 为一个随机高斯向量。接下来，假设 $t \geqslant Cp\varepsilon^{-1/3}\log(p/\varepsilon)$，其中 C 是某个充分大的常数，并定义 $\mathcal{K}$:=张成空间$(g, \Lambda g, \ldots, \Lambda^t g)$。那么，存在一个单位向量 $w \in \mathcal{K}$ 使得

$$(w^\top \Lambda w)^{p/2} \geqslant (1+\varepsilon)\lambda_1^{p/2} - \varepsilon \cdot (\lambda_1^{p/2} + \cdots + \lambda_n^{p/2}) = \lambda_1^{p/2} - \varepsilon \cdot \sum_{i=2}^n \lambda_i^{p/2}.$$

证明。我们不妨设 $\|A\|_{\mathrm{op}} = 1$（通过缩放，注意缩放不会影响 $\mathcal{K}$），因此$\lambda_1 = 1$。此外，我们不妨通过旋转 Λ，假设 Λ 是对角的：这会相应地旋转g，但单位协方差高斯分布具有旋转不变性。我们将分析分为四种情况。

情形 1: $\sum_{i=2}^n \lambda^{p/2}{}_i \geqslant \varepsilon^{-1}$. 在这种情况下，$\lambda_1^{p/2} - \varepsilon \cdot \sum_{i=2}^n \lambda^{p/2}{}_i \leqslant 0$，并且由于 Λ 是半正定的，任何单位向量 $w \in \mathcal{K}$ 都满足 $(w^\top \Lambda w)^{p/2} \geqslant 0 \geqslant \lambda_1^{p/2} - \varepsilon \cdot \sum_{i=2}^n \lambda^{p/2}{}_i$。

在其余情形中，给定初始向量 $g = (g_1, \ldots, g_d) \sim \mathcal{N}(0, I_d)$，我们充分利用了以下事实：子空间张成空间 $\{g, \Lambda g, \ldots, \Lambda^t g\}$ 包含向量 $\left(\phi(\lambda_1^2)g_1, \ldots, \phi(\lambda_d^2)g_d\right)$，对于任意次数不超过 t 的多项式 $\phi(x)$，因为我们不失一般性地假设 Λ 为对角的。

情形 2: $\sum_{i=2}^n \lambda^{p/2}\mathrm{i} < \varepsilon^{-1}$，且范围$[1 - \frac{1}{2}\mathrm{p}, 1]$内的特征值数量 λ_i至少为 $\varepsilon^{-1/3} + 1$。在这种情况下，首先注意到 $\sum_{i=2}^n \lambda^{p/2}\mathrm{i} \geqslant (1 - \frac{1}{2}\mathrm{p})^{p/2} \cdot \varepsilon^{-1/3} \geqslant \frac{\varepsilon^{-1/3}}{2}$，因为除 $\lambda_1 = 1$ 外，Λ 至少有 $\varepsilon^{-1/3}$ 个特征值不小于 $1 - \frac{1}{2}\mathrm{p}$。我们首先考虑推论 4.7 中的平移/缩放切比雪夫多项式 $\mathrm{T}_{t,\varepsilon^{2/3}}/(2\mathrm{p})$，其中 $\mathrm{t} = \mathrm{O}(\mathrm{p}^{1/2}/\varepsilon^{1/3} \cdot \log(\mathrm{p}/\varepsilon))$。注意，$\mathrm{T}_{t,\varepsilon^{2/3}}/(2\mathrm{p})(1) = 1$，并且

$$|T_{t,\varepsilon^{2/3}/(2p)}(x)| \leqslant e^{-c \cdot \min(t^2 \cdot \varepsilon^{2/3}/2p, t \cdot \varepsilon^{1/3}/\sqrt{2p})} \leqslant \frac{\varepsilon^2}{p}$$

对于所有 $x \in [0, 1 - \frac{\varepsilon^{2/3}}{2}p]$。令 $\hat{\phi}(x) = x^{\lceil p/2 \rceil} \cdot T_{t,\varepsilon^{2/3}/(2p)}(x)$，则有 $\hat{\phi}(1) = 1$ and $|\hat{\phi}(x)| \leqslant x^{p/2} \cdot \varepsilon^2 p$，对所有 $x \in [0, 1 - \frac{\varepsilon^{2/3}}{2}p]$ 均成立。对于足够大的常数 C，若 $\hat{t} \geqslant C \cdot p \cdot \varepsilon^{-1/3} \log(1/\varepsilon)$，则 $\hat{\phi}(x)$ 的次数至多为 $\hat{t}$。

现在，以至少 0.99 的概率（关于 $g \sim \mathcal{N}(0, I)$ 的随机性），$|g_1| \geqslant 0.01$，且由于 $\hat{\phi}(\lambda_1) = \hat{\phi}(1) = 1$，我们有 $|p(\lambda_1)g_1| \geqslant 0.01$。

接下来，对于任意 i 满足 $\lambda_i \leqslant 1 - \frac{\varepsilon^{2/3}}{2}p$、$|\hat{\phi}(\lambda_i)| \leqslant \varepsilon^2 p \cdot \lambda^{p/2}{}_i$ 的情形。因此，对于每个这样的 i，

$$\mathop{\mathbf{E}}_{g \sim \mathcal{N}(0,I)} \left[(\hat{\phi}(\lambda_i) g_i)^2 \right] \leqslant \frac{\varepsilon^4}{p^2} \cdot \lambda_i^p.$$

现在，由于对于所有 i 均有 $\lambda_i \leqslant 1$，根据我们的假设，我们有 $\sum_{i=2}^n \lambda^p{}_i \leqslant \sum_{i=2}^n \lambda^{p/2}{}_i < \varepsilon^{-1}$。因此，

$$\begin{aligned}
\mathop{\mathbf{E}}_{g \sim \mathcal{N}(0,I)} \left[\sum_{i:\lambda_i \leqslant 1-(\varepsilon^{2/3}/(2p))} (\hat{\phi}(\lambda_i) g_i)^2 \right] &= \sum_{i:\lambda_i \leqslant 1-(\varepsilon^{2/3}/(2p))} \hat{\phi}(\lambda_i)^2 \leqslant \frac{\varepsilon^4}{p^2} \cdot \sum_{i:\lambda_i \leqslant 1-(\varepsilon^{2/3}/(2p))} \lambda_i^p \\
&\leqslant \frac{\varepsilon^4}{p^2} \cdot \sum_{i=2}^n \lambda_i^{p/2} \\
&\leqslant \frac{\varepsilon^3}{p^2}.
\end{aligned} \tag{10}$$

因此，根据马尔可夫不等式，以至少 0.99 的概率，$\sum_{i:\lambda_i \leqslant 1-\varepsilon^{2/3}/2} \phi(\lambda_i)^2 \leqslant \frac{100\varepsilon^3}{2}$. 因此，以至少 0.98 的概率，在 g 上，$(/(p))(\hat{}()g)/p$.

$$\sum_{i:\lambda_i \leqslant 1-\varepsilon^{2/3}/(2p)} (\hat{\phi}(\lambda_i) g_i)^2 \leqslant \frac{10^6 \cdot \varepsilon^3}{p^2} \cdot (\hat{\phi}(\lambda_1) g_1)^2. \tag{11}$$

接下来，我们定义 $\phi(x)$ 为 $\hat{\phi}(x)$ 的缩放版本（即 $\phi(x) = \gamma \cdot \hat{\phi}(x)$，其中参数为 γ），从而向量 $(\phi(\lambda_i) g_i)_{i=1}^n$ 具有单位范数。由方程（11）可知，以至少 0.98 的概率，

$$\sum_{i:\lambda_i \leqslant 1-\varepsilon^{2/3}/(2p)} (\phi(\lambda_i) g_i)^2 \leqslant \frac{10^6 \cdot \varepsilon^3}{p^2} \cdot (\phi(\lambda_1) g_1)^2 \leqslant \frac{10^6 \cdot \varepsilon^3}{p^2},$$

其中，上面的第一个不等式由（11）得到，第二个不等式成立是因为 $(\phi(\lambda_i) g_i)_{i=1}^n$ 具有单位范数，因此 $|\phi(\lambda_1) g_1| \leqslant 1$。这意味着 $\sum_{i:\lambda_i > 1-\varepsilon^{2/3}/2p} (\phi(\lambda_i) g_i)^2 \geqslant 1 - 10^6\varepsilon^3/p^2$。因此，对于归一化向量 $w = (\phi(\lambda_i) g_i)_{i=1}^n$，我们有

$$\begin{aligned}
w^\top \Lambda w &= \sum_{i=1}^n (\phi(\lambda_i) g_i)^2 \cdot \lambda_i \\
&\geqslant \sum_{i:\lambda_i > 1-\varepsilon^{2/3}/2p} (\phi(\lambda_i) g_i)^2 \cdot \left(1 - \frac{\varepsilon^{2/3}}{2p}\right) \\
&\geqslant \left(1 - \frac{10^6\varepsilon^3}{p^2}\right) \cdot \left(1 - \frac{\varepsilon^{2/3}}{2p}\right) \\
&\geqslant 1 - \frac{\varepsilon^{2/3}}{1.5p},
\end{aligned}$$

其中，最后一行假设 ε 小于某个较小但固定的常数。

因此，$(w^{\top}\Lambda^{p/2} \geqslant 1-\varepsilon^{2/3}/2 \cdot 5^{p/2} \geqslant 1-\varepsilon^{2/3}/2) \cdot \sum_{i=2}^{n}\lambda^{p/2} \geqslant w)(/(.p))$。然而，我们知道$\varepsilon^{-1/3}/2$，这意味着

$$\lambda_1^{p/2} - \varepsilon \cdot \sum_{i=2}^{n}\lambda_i^{p/2} \leqslant 1 - \frac{\varepsilon^{2/3}}{2} \leqslant (w^{\top}\Lambda w)^{p/2},$$

如所需。

情况 3：$_{-}\frac{1}{2} \leqslant \sum_{i=2}^{n}\lambda^{p/2}{}_i \leqslant \varepsilon^{-1}$，且范围$[1-\frac{1}{2}p, 1]$内的特征值数量 λ_i至多为 $\varepsilon^{-1/3}\pm 1$。在这种情况下，我们考虑多项式

$$\hat{\phi}(x) = x^{t/2} \cdot \prod_{i:1-\varepsilon/2p \geqslant \lambda_i \geqslant 1-1/2p}(x-\lambda_i).$$

假设 $t \geqslant C \cdot p \cdot \varepsilon^{-1/3}\log(p/\varepsilon)$，$\hat{\phi}(x)$ 的次数至多为 t。注意，

$$\hat{\phi}(1) = \prod_{i:1-\varepsilon/2p \geqslant \lambda_i \geqslant 1-1/2p}(1-\lambda_i) \geqslant \left(\frac{\varepsilon}{2p}\right)^{\varepsilon^{-1/3}+1} \geqslant 2^{-O(\varepsilon^{-1/3}\log(p/\varepsilon))}, \tag{12}$$

因为乘积中的每个 $1-\lambda_i$ 项都至少为 $\varepsilon/(2p)$，且这样的项至多有 $\varepsilon^{-1}+1$ 个。此外，对于所有 $\lambda_i \in [1-\frac{1}{2}p, 1-\frac{\varepsilon}{2}p]$，都有 $\hat{\phi}(\lambda_i) = 0$，因为其中一个 $x-\lambda_i$ 项为零。最后，对于所有 $\lambda_i \in [0, 1-\frac{1}{2}p]$，我们知道 $\prod_{i:1-\varepsilon/2p \geqslant \lambda_i \geqslant 1-1/2p}(x-\lambda_i)$ 的绝对值至多为 1；而当 $t \geqslant C \cdot p \cdot \varepsilon^{-1/3}\log(p/\varepsilon)$ 时，对充分大的常数 C，

$$|\lambda_i^{(t-p)/2}| \leqslant (1-1/2p)^{(t-p)/2} \leqslant 2^{-C\varepsilon^{-1/3}\log(p/\varepsilon)/10} \leqslant \hat{\phi}(1) \cdot \frac{\varepsilon^2}{p}.$$

上文中，第一个不等式成立，是因为我们只考虑$\lambda_i \leqslant 1-\frac{1}{2}p$；第二个不等式根据我们的假设成立，即$t \geqslant C \cdot p \cdot \varepsilon^{-1/3}\log(p/\varepsilon)$；第三个不等式由（12）成立。因此，$|\hat{\phi}(\lambda_i)| \leqslant \hat{\phi}(1) \cdot \lambda^{p/2}{}_i \cdot \varepsilon^2{}_p$ 对于所有满足 $\lambda_i \leqslant 1-\varepsilon/2p$ 的i 均成立。

对于$g \sim \mathcal{N}(0, I)$，以至少 0.99 的概率，$|\hat{\phi}(1) \cdot g_1| \geqslant 0.01 \cdot \hat{\phi}(1)$。此外，

$$\begin{aligned}
\mathop{\mathbf{E}}_{g\sim\mathcal{N}(0,I)}\left[\sum_{i:\lambda_i \leqslant 1-\varepsilon/2p}(\hat{\phi}(\lambda_i)g_i)^2\right] &= \sum_{i:\lambda_i \leqslant 1-\varepsilon/2p}\hat{\phi}(\lambda_i)^p \\
&\leqslant \hat{\phi}(1)^2 \cdot \frac{\varepsilon^4}{p^2} \cdot \sum_{i:\lambda_i \leqslant 1-\varepsilon/2p}\lambda_i^p \\
&\leqslant \hat{\phi}(1)^2 \cdot \frac{\varepsilon^4}{p^2} \cdot \sum_{i=2}^{n}\lambda_i^{p/2} \\
&\leqslant \hat{\phi}(1)^2 \cdot \frac{\varepsilon^3}{p^2}.
\end{aligned} \tag{13}$$

因此，根据马尔可夫不等式，对于g，以至少 0.99 的概率有 $\sum_{i:\lambda_i \leqslant 1-\varepsilon}(\hat{\phi}(\lambda_i)g_i)^2 \leqslant 100\,\varepsilon^3\,p^2 \cdot \hat{\phi}(1)^2$。因此，对于$g$，以至少 0.98 的概率有 $\sum_{i:\lambda_i \leqslant 1-\varepsilon}(\hat{P}(\lambda_i)g_i)^2 \leqslant 10^6 \cdot \varepsilon^3\,p^2 \cdot (\hat{P}(\lambda_1)g_1)^2$。—

与第二种情况一样，我们令 $\phi(x)$ 为 $\hat{\phi}(x)$ 的缩放版本，使得 $(\phi(\lambda_i)g_i)_{i=1}^n$ 具有单位范数。于是，$\sum_{i:\lambda_i \leqslant 1-\varepsilon/2p}(\phi(\lambda_i)g_i)^2 \leqslant 10^6 \cdot \varepsilon^3\,p^2$，因此 $\sum_{i:\lambda_i > 1-\varepsilon/2p}(\phi(\lambda_i)g_i)^2 \geqslant 1-10^6\,\varepsilon^3{}_2p$。因此，

$$w^{\top}\Lambda w = \sum_{i=1}^{n}(\phi(\lambda_i)g_i)^2 \cdot \lambda_i$$

$$
\begin{aligned}
&\geqslant \sum_{i:\lambda_i > 1-\varepsilon/2p} (\phi(\lambda_i) g_i)^2 \cdot \left(1 - \frac{\varepsilon}{2p}\right) \\
&\geqslant \left(1 - 10^6 \cdot \frac{\varepsilon^3}{p^2}\right) \cdot \left(1 - \frac{\varepsilon}{2p}\right) \\
&\geqslant 1 - \frac{\varepsilon}{1.5p}
\end{aligned}
$$

因此，$(w^\top \Lambda w)^{p/2} \geqslant (1-\varepsilon/(1.5p))^{p/2} \geqslant 1 - \frac{\varepsilon}{2}$。然而，我们知道 $\sum_{i=2}^n \lambda^{p/2}{}_i \geqslant \frac{1}{2}$，这意味着 _

$$
\lambda_1^{p/2} - \varepsilon \cdot \sum_{i=2}^n \lambda_i^{p/2} \leqslant 1 - \frac{\varepsilon}{2} \leqslant (w^\top \Lambda w)^{p/2},
$$

正如所需。

情况4: $\sum_{i=2}^n \lambda^{p/2}\mathrm{i} < \frac{1}{2}$。在此情况下，对于所有i，都有$\lambda_i \leqslant 1 - \frac{1}{2}\,\mathrm{p} \geqslant 2$。因此，我们可以令 $\hat{}\,\phi(\mathrm{x}) = \mathrm{x}^L$，其中某个 $\mathrm{L} = \mathrm{O}(\log(\mathrm{p}/\varepsilon))$，从而得到 $\hat{}\,\phi(1) = 1$，并且对于所有i，都有 $\hat{}\,\phi(\lambda_i) \leqslant \varepsilon^2\,\mathrm{p} \cdot \lambda^{p/2}\,\mathrm{i} \geqslant 2$。对于 $\mathrm{g} \sim \mathcal{N}(0,\mathrm{I})$，以至少0.99 的概率，有 $|\hat{}\,\phi(1) \cdot \mathrm{g}_1| \geqslant 0.01$，因为 $\hat{}\,\mathrm{P}(1) = 1$，并且

$$
\mathop{\mathbf{E}}_{g\sim\mathcal{N}(0,I)} \left[\sum_{i=2}^n (\hat{\phi}(\lambda_i) g_i)^2\right] = \sum_{i=2}^n \hat{P}(\lambda_i)^2 \leqslant \sum_{i=2}^n \frac{\varepsilon^4}{p^2} \cdot \lambda_i^p \leqslant \frac{\varepsilon^4}{p^2} \cdot \left(\sum_{i=2}^n \lambda_i^{p/2}\right).
$$

定义 $\tau = \sum_{i=2}^n \lambda^{p/2}{}_i$。再次，根据马尔可夫不等式，$\sum_{i=2}^n (\hat{}\,\phi(\lambda_i) g_i)^2 \leqslant 100\,\frac{\varepsilon^4}{p^2} \cdot \tau$ 的概率至少为 0 99，0 98 $\sum_{i=2}^n \phi\ \lambda_i\ {}_i{}^2 \leqslant$。关于 g。$10^6 \cdot \frac{\varepsilon^4}{2} \cdot \tau \cdot \phi\ \lambda_1\ {}_1{}^2$
因此，关于 g，其概率至少为 。$(\hat{}\,()g)_p(\hat{}\,()g)$。
与第二种和第三种情况一样，我们令 $\phi(\mathrm{x})$ 为 $\hat{}\,\phi(\mathrm{x})$ 的缩放版本，使得 $\mathrm{w} := (\phi(\lambda_i)\mathrm{g}_i)_{i=1}^n$具有单位范数。我们有 $\sum_{i=2}^n (\phi(\lambda_i)\mathrm{g}_i)^2 \leqslant 10^6 \cdot \varepsilon^4\,\mathrm{p}^2 \cdot \tau$，这意味着 $(\phi(\lambda_1)\mathrm{g}_1)^2 \geqslant 1 - \mathrm{O}(\varepsilon^4\,\mathrm{p}^2 \cdot \tau)$。因此，$\mathrm{w}^\top \Delta \mathrm{w} \geqslant \lambda_1 \cdot (\phi(\lambda_1)\mathrm{g}_1)^2 \geqslant 1 - \mathrm{O}(\varepsilon^4\,\mathrm{p}^2 \cdot \tau)$，这意味着 $(\mathrm{w}^\top \Lambda \mathrm{w})^{p/2} \geqslant 1 - \mathrm{O}(\varepsilon^4\,\mathrm{p} \cdot \tau)$。然而，$\lambda_1^{p/2} - \varepsilon \cdot \sum_{i=2}^n \lambda^{p/2}\mathrm{i} = 1 - \varepsilon \cdot \tau$。因此，

$$
(w^\top \Lambda w)^{p/2} \geqslant \lambda_1^{p/2} - \varepsilon \cdot \sum_{i=2}^n \lambda_i^{p/2},
$$

由此完成证明。 □

现在我们可以完成定理的证明 7.1。

定理 7.1 的证明。回顾一下，$\mathrm{t} = \mathrm{O}\left(\mathrm{p}^{-1}\log(1/\varepsilon)/\varepsilon^{1/3}\right)$。通过 O(t)次矩阵-向量计算，我们可以计算 $\mathcal{K}$ 并计算$\max_{w\in\mathcal{K}:\|w\|_2=1} \|\mathrm{A}^\top \mathrm{w}\|_2$。这等价于$\mathrm{w} = \mathrm{Q}\mathrm{y}_1$，其中 y_1 是 $\mathrm{A}^\top\mathrm{Q}$的最大奇异向量，其中Q 是 $\mathcal{K}$的列空间的一个标准正交基，或者等价地，是 $\mathrm{Q}^\top\mathrm{A}\mathrm{A}^\top\mathrm{Q}$ 的最大特征向量。如果子空间 $\mathcal{K}$ 中存在一个单位向量 $= \mathrm{Span}(\mathrm{g}, (\mathrm{A}\mathrm{A}^\top)\mathrm{g}, \ldots, (\mathrm{A}\mathrm{A}^\top)^t\mathrm{g})$ 满足引理 7.6的假设，那么 w 也将满足该假设，因为w被定义为在 $\mathrm{w} \in \mathcal{K}$上最大化 $\|\mathrm{A}^\top \mathrm{w}\|_2$。因此，根据引理 7.6，对于 $\mathrm{v} = \mathrm{A}^\top \mathrm{w}/\|\mathrm{A}^\top \mathrm{w}\|$，$\mathrm{A}\mathrm{v}\mathrm{v}^\top$ 是 Schatten-p 范数下的$(1+\varepsilon)$-近似秩一近似。由于 $\|\mathrm{A}^\top \mathrm{w}\|_2 = \sqrt{\mathrm{w}^\top \mathrm{A}\mathrm{A}^\top \mathrm{w}}$，引理 7.7 意味着存在这样的向量，从而证明得证。

□

致谢

作者感谢 Sinho Chewi、Jaume de Dios Pont、Piotr Indyk、JerryLi、ChenLu 和 Erik Waingarten 的有益讨论。AB 获得 AnkurMoitra 的美国海军研究办公室资助。SN 获得 NSF TRIPODS计划、NSF研究生奖学金和 Google奖学金的支持。

参考文献

[AN13] AlexandrAndoni 和 Huy L. Nguyen。在流式模型中的矩阵特征值。载于第二十四届 ACM-SIAM离散算法研讨会论文集，第1729–1737页。工业与应用数学学会，2013 年。12[Bak22] Ainesh Bakshi。潜在模型学习算法：从确立可处理性到逼近最优性。博士论文，卡内基梅隆大学，匹兹堡，宾夕法尼亚州，2022 年。4[BBK+21] AineshBakshi、Chiranjib Bhattacharyya、Ravi Kannan、David P Woodruff 和 Samson Zhou。在输入稀疏时间内学习潜在单纯形。国际学习表征会议，2021 年。3, 12[BBV04] Stephen Boyd、Stephen PBoyd 和LievenVandenberghe。凸优化。剑桥大学出版社，2004 年。1[BCW20] Ainesh Bakshi、NadiiaChepurko和DavidPWoodruff。用于半正定低秩近似的稳健且样本最优算法。载于2020年电气电子工程师学会第 61 届年度计算机科学基础研讨会（计算机科学基础研讨会）论文集，第 506–516 页。电气电子工程师学会，2020 年。12[BCW22] Ainesh Bakshi、Kenneth LClarkson和DavidPWoodruff。利用 $1/\varepsilon^{1/3}$ 矩阵-向量乘积进行低秩近似。载于第 54 届ACM SIGACT计算理论研讨会论文集，第 1130–1143页，2022年。1,2, 3, 6, 10, 11, 20[BDN15] Jean Bourgain、Sjoerd Dirksen 和 JelaniNelson。迈向欧几里得空间中稀疏降维的统一理论。载于第四十七届ACM 计算理论研讨会论文集，STOC 2015，美国俄勒冈州波特兰，2015 年 6月 14–17日，第 499–508 页，2015 年。12[BHSW20] Mark Braverman、Elad Hazan、MaxSimchowitz 和 BlakeWoodworth。线性回归的梯度复杂度。载于学习理论会议，第627–647页。机器学习研究论文集，2020 年。1, 3, 6, 11[BKKS19] Vladimir Braverman、Robert Krauthgamer、Aditya Krishnan和RoiSinoff。矩阵流中的Schatten 范数：稀疏性回归，维度告别。arXiv预印本 arXiv:1907.05457，2019 年。12[BT23] Ainesh Bakshi和 EwinTang。用于量子机器学习的改进型经典奇异值变换。arXiv 预印本 arXiv:2303.01492，2023 年。12

[BW18] AineshBakshi 和David Woodruff。距离矩阵的亚线性时间低秩近似。在《神经信息处理系统进展》中，第3782–3792 页，2018年。12[CCHW20] Nadiia Chepurko、Kenneth L Clarkson、LiorHoresh和David P Woodruff。来自随机化数值线性代数的量子启发算法。arXiv预印本 arXiv:2011.04125，2020 年。12[CdDL+23] Sinho Chewi、Jaume de Dios Pont、JerryLi、ChenLu 和Shyam Narayanan。对数凹采样的查询下界。CoRR, abs/2304.02599，2023 年。4, 18, 29[CLR21] Erin Carson、Kathryn Lund 和 MiroslavRozloznik。经典 Gram–Schmidt 方法块变体的稳定性。《SIAM 矩阵分析与应用期刊》，42(3):1365–1380，2021 年。3[CLW18] Nai-Hui Chia、Han-HsuanLin和 Chunhao Wang。用于求解低秩线性系统的量子启发亚线性经典算法。arXiv预印本arXiv:1811.04852, 2018 年。12[Coh16] Michael BCohen。通过迹不等式实现近乎紧的无知子空间嵌入。在《第二十七届 ACM-SIAM离散算法研讨会论文集》中，第278–287 页。SIAM，2016 年。12[CW13] Kenneth LClarkson和David P Woodruff。在输入稀疏时间内进行低秩近似与回归。在《第四十五届 ACM 计算理论研讨会论文集》中，第 81–90 页。ACM，2013 年。11, 12[DM23] PrathameshDharangutte 和Christopher Musco。Hutchinson对角估计器的紧致分析。在《算法简洁性研讨会》（SOSA）中，第 353–364 页。SIAM，2023年。11[Elm96] Howard C Elman。离散 Stokes方程的多重网格与 Krylov 子空间方法。《国际流体数值方法期刊》，22(8):755–770，1996 年。1[ES96] Howard Elman 和DavidSilvester。Navier–Stokes方程的快速非对称迭代与预条件。《SIAM 科学计算期刊》，17(1):33–46，1996 年。1[ESW96] Howard C Elman、DavidSilvester和 AndrewJ Wathen。计算流体力学问题的迭代方法。1996年。1[GBC16] Ian Goodfellow、Yoshua Bengio 和 AaronCourville。深度学习。MIT 出版社，2016 年。1[GLT18] AndrásGilyén、SethLloyd和 EwinTang。对维度呈对数依赖的量子启发低秩随机回归。arXiv预印本arXiv:1811.04909，2018年。12[GSLW19] András Gilyén、YuanSu、GuangHaoLow和Nathan Wiebe。量子奇异值变换及其扩展：量子矩阵运算的指数级改进。在《第 51 届 ACM SIGACT计算理论研讨会论文集》中，第 193–204 页，2019 年。1, 12

[HHL09] Aram W Harrow、Avinatan Hassidim 和 Seth Lloyd。《线性方程组的量子算法》。Physical Review Letters，103(15):150502，2009。1[HMT11] Nathan Halko、Per-Gunnar Martinsson 和 Joel A Tropp。《利用随机性发现结构：构造近似矩阵分解的概率算法》。SIAM Review，53(2):217–288，2011。1, 2[IVWW19] Piotr Indyk、Ali Vakilian、Tal Wagner 和 David Woodruff。《距离矩阵的样本最优低秩近似》。arXiv 预印本 arXiv:1906.00339，2019。12[KK04] Dana A Knoll 和 David E Keyes。《无雅可比矩阵的 Newton–Krylov 方法：方法与应用综述》。Journal of Computational Physics，193(2):357–397，2004。1[KP16] Iordanis Kerenidis 和 Anupam Prakash。《量子推荐系统》。arXiv 预印本 arXiv:1603.08675，2016。12[KW21] Praneeth Kacham 和 David Woodruff。《算子范数误差下的降秩回归》。载于《学习理论会议》，第 2679–2716 页。机器学习研究论文集，2021。3[LNW14a] Yi Li、Huy L Nguyen 和 David P Woodruff。《关于矩阵范数的草图化与顶端奇异向量》。载于第二十五届 ACM-SIAM离散算法研讨会论文集，第 1562–1581 页。SIAM，2014。12[LNW14b] Yi Li、Huy L. Nguyen 和 David P. Woodruff。《转置流模型下的流算法不妨就是线性草图》。载于计算理论研讨会（STOC 2014），美国纽约州纽约市，2014 年 5 月 31 日至 6 月 3 日，第 174–183 页，2014。11[LW16a] Yi Li 和 David P Woodruff。《流中奇异值函数的近似》。载于第四十八届 ACM 计算理论研讨会论文集，第 726–739 页，2016。12[LW16b] Yi Li 和 David P Woodruff。《算子范数、Schatten 范数和子空间嵌入草图化的紧确界》。载于《近似、随机化与组合优化：算法与技术》（APPROX/RANDOM 2016）。Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik，2016。12[LW17] Yi Li 和 David P Woodruff。《Schatten 范数嵌入及其在数据流中的应用》。载于第四十四届国际自动机、语言与程序设计研讨会（ICALP 2017）。Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik，2017。12[LW20] Yi Li 和 David P. Woodruff。《Schatten 范数下的输入稀疏低秩近似》。CoRR，abs/2004.12646，2020。11[MCG04] John Mellor-Crummey 和 John Garvin。《通过展开与融合优化稀疏矩阵-向量乘积计算》。The International Journal of High Performance Computing Applications，18(2):225–236，2004。1[MH02] John C Mason 和 David C Handscomb。《切比雪夫多项式》。CRC 出版社，2002。13[MM13] Xiangrui Meng 和 Michael W. Mahoney。《输入稀疏时间下的低失真子空间嵌入及其在稳健线性回归中的应用》。载于计算理论研讨会

会议，STOC'13，美国加利福尼亚州帕洛阿尔托，2013 年 6 月 1 日至 4 日，第 91–100 页，2013。12[MM15] Cameron Musco 和 Christopher Musco。《用于实现更强、更快速近似奇异值分解的随机化块 Krylov 方法》。载于《神经信息处理系统进展》，第 1396–1404 页，2015。1, 2, 3, 4, 5, 10, 11, 14[MMMW21] Raphael A. Meyer、Cameron Musco、Christopher Musco 和 David P. Woodruff。《Hutch++：最优随机迹估计》。载于第四届算法简洁性研讨会（SOSA 2021），线上会议，2021 年 1 月 11–12 日，第 142–155 页，2021。1, 11[Mut05] S. Muthukrishnan。《数据流：算法与应用》。Found. Trends Theor. Comput. Sci.，1(2)，2005。11[MW17] Cameron Musco 和 David P. Woodruff。《半正定矩阵的次线性时间低秩近似》。载于第五十八届电气电子工程师学会年度计算机科学基础研讨会（FOCS 2017），美国加利福尼亚州伯克利，2017 年 10 月 15–17 日，第 672–683 页，2017。12[Nel11] Jelani Osei Nelson。《高维向量的草图化与流式处理》。博士论文，麻省理工学院，2011。11[NN13] Jelani Nelson 和 Huy L. Nguyen。《OSNAP：通过更稀疏的子空间嵌入实现更快的数值线性代数算法》。载于第五十四届电气电子工程师学会计算机科学基础研讨会（FOCS 2013），美国加利福尼亚州伯克利，2013 年 10 月 26–29 日，第 117–126 页，2013。12[NSW22] Deanna Needell、William Swartworth 和 David P Woodruff。《使用线性测量检验半正定性》。载于 2022 年电气电子工程师学会第六十三届计算机科学基础研讨会（FOCS），第 87–97 页。电气电子工程师学会，2022。1, 11[Rou89] François-Xavier Roux。《通过再正交化加速结构分析问题区域分解方法中的外层共轭梯度法》。载于第三届国际超级计算会议论文集，第 471–476 页，1989。1[RST10] Vladimir Rokhlin、Arthur Szlam 和 Mark Tygert。《一种用于主成分分析的随机化算法》。SIAM Journal on Matrix Analysis and Applications，31(3):1100–1124，2010。1, 2[RWYZ21] Cyrus Rashtchian、David P. Woodruff、Peng Ye 和 Hanlin Zhu。《统计问题的平均情形通信复杂度》，2021。11[RWZ20] Cyrus Rashtchian、David P. Woodruff 和 Hanlin Zhu。《用于求解线性代数、统计学和图问题的向量-矩阵-向量查询》。载于《近似、随机化与组合优化：算法与技术》，APPROX/RANDOM 2020，线上会议，2020 年 8 月 17–19 日，第 26:1–26:20 页，2020。1, 11[Saa81] Yousef Saad。《求解大型非对称线性系统的 Krylov 子空间方法》。Mathematics of Computation，37(155):105–126，1981。1

[SAR18] Max Simchowitz、AhmedElAlaoui和Benjamin Recht。通过有限样本变形Wigner定律给出PCA的紧下界查询复杂度。载于《第50届ACMSIGACT计算理论研讨会论文集》，STOC 2018，美国加利福尼亚州洛杉矶，2018年6月25日至29日，第1249–1259页，2018年。1, 3, 6, 11[Sch09] TamarSchlick。计算化学中的优化方法。《计算化学评论》，3:1–71，2009年。1[SW19] XiaofeiShi和DavidP. Woodruff。结构化矩阵的次线性时间数值线性代数。载于《第三十三届AAAI人工智能会议》，AAAI2019，《第三十一届人工智能创新应用会议》，IAAI 2019，《第九届AAAI人工智能教育进展研讨会》，EAAI 2019，美国夏威夷州檀香山，2019年1月27日至2月1日，第4918–4925页，2019年。12[SWYZ19] Xiaoming Sun、David P. Woodruff、GuangYang和 Jialin Zhang。通过矩阵-向量乘积查询矩阵。载于《第46届自动机、语言与编程国际学术会议》，ICALP 2019，2019年7月9日至12日，希腊帕特雷，第94:1–94:16页，2019年。1,11[Tan19] EwinTang。用于推荐系统的量子启发式经典算法。载于《第51届ACMSIGACT计算理论研讨会论文集》，第217–228页。ACM，2019年。12[TW04] Andrea Toselli 和Olof Widlund。区域分解方法——算法与理论，第34卷。SpringerScience &Business Media，2004年。1[Ver10] RomanVershynin。随机矩阵的非渐近分析导论。arXiv预印本，arXiv:1011.3027，2010年。14[Woo14] David P. Woodruff。作为数值线性代数工具的草图化。《理论计算机科学基础与趋势》，10(1–2):1–157，2014年。1, 3, 11[WWZ14] KarlWimmer、YiWu 和PengZhang。估计矩阵迹的最优查询复杂度。载于《自动机、语言与编程——第41届国际学术会议》，ICALP 2014，丹麦哥本哈根，2014年7月8日至11日，论文集，第一部分，第1051–1062页，2014年。1[WY22] DavidPWoodruff和TaisukeYasuda。利用稀疏性改进低秩近似算法。载于《2022年ACM-SIAM离散算法研讨会论文集》（SODA），第2358–2403页。SIAM，2022年。3[ZVY+15] Dmitry Zuev、Eugene Vecharynski、Chao Yang、NatalieOrms和Anna I Krylov。量子化学中用于迭代矩阵无关特征值求解器的新算法。《计算化学杂志》，36(5):273–284，2015年。1

引理的证明6.3

在本节中，我们证明引理 6.3。证明基本上照搬 [CdDL+23] 中的证明（我们已获得作者明确许可），但有几处细节有所调整，以作小幅推广。

A.1 补充预备知识

我们回顾扩展预言机模型（定义 6.1）和自适应确定性算法（定义6.2）的定义。

首先注意，在扩展预言机模型中，我们可以假设每个 v_k 都是与其输入正交的单位向量。

引理 A.1 （扩展预言机与正交查询）。 对于 $k \in [2,K]$, ，令 v_k 如定义 6.2 所述，并令 $\{A^i v_j\}_{H_{k-1}}$ 如定义 6.1所述。那么，不失一般性，我们可以假设 v_k 与由 $\{A^i v_j\}_{H_{k-1}}$中的向量张成的子空间正交。

证明。反设情况并非如此。那么，我们可以将 $\mathrm{v}_k = \sum_{(i,j)\in H_{k-1}}$ 分解为 $\mathrm{c}_{i,j}\mathrm{A}^i\mathrm{v}_j +$ 加上 $\mathrm{c}^{\perp}\mathrm{v}_k^{\perp}$，其中 $\mathrm{v}_k^{\perp}$是与 $\{\mathrm{A}^i\mathrm{v}_j\}_{H_{k-1}}$ 正交的单位向量，且 $\mathrm{c}_{i,j}$和 $\mathrm{c}^{\perp}$ 均为标量。在第 k 次迭代结束时，算法获得的新信息为$\{\mathrm{A}^i\mathrm{v}_j\}_{i+j=k+1,j\leqslant k}$。对于所有 (i, j) $\neq$ (1, k)，新信息都不依赖于 v_k。此外，$\mathrm{Av}_k = \sum_{(i,j)\in H_{k-1}} \mathrm{c}_{i,j}\mathrm{A}^{i+1}\mathrm{v}_j +$ 加上 $\mathrm{c}^{\perp}\mathrm{Av}_k^{\perp}$，其中每个 $\mathrm{A}^{i+1}\mathrm{v}j$都是算法在第 k 次迭代结束时获得的信息（无论如何，依据我们的扩展查询模型）。由于若 (i, j) $\in \mathrm{H}_{k-1}$，则 (i + 1, j) $\in \mathrm{H}_k$，且 (1, k) $\in \mathrm{H}_k$，该表达式表明，若算法查询 $\mathrm{v}_k^{\perp}$而非 v_k，它将获得相同数量的信息（若 $\mathrm{c}^{\perp} = 0$，则会获得更多信息）。对这一推理进行归纳，即可证明所述结论。

□

我们将其与一个采用独立同分布标准高斯分布查询的块 Krylov算法进行比较 $z_1,z_2,\ldots,z_K$然后对所有 $i,j \leqslant K$ 接收 $\{A^i z_j\}$ 。回想一下，块 Krylov算法不进行自适应 查询；针对块 Krylov算法证明下界更加容易。我们的目标现在是展示块 Krylov 算法可以模拟自适应确定性算法。

A.2 条件引理

我们首先证明一个一般的条件引理，该引理将在化归至块Krylov算法的过程中被反复调用。我们隐含地假设所有映射都是可测的，以避免不必要的技术问题。该引理大致表明，如果自适应算法知道 $\{A^i v_j\}_{H_k}$，那么 A 给定 $\{A^i v_j\}_{H_k}$ 的后验分布确实在正交补空间 $\{A^i v_j\}_{H_k}$ 上具有旋转对称性。

我们将使用记号$\mathrm{d} =$ to 表示两个随机变量在概率分布上相等（可能以其他信息为条件）。

引理*A.2*（条件引理，初步版本）。设 U 为 *Haar*随机正交矩阵，且 $A = U^{\top}DU$，其中 D 是固定的正对角矩阵。假设 $\mathcal{A}$ 是一个自适应确定性算法，它生成扩展预言机查询 $v_1,\ldots,v_K$，并且在第 k次查询后，对于所有 $(i, j) \in H_k$，知道 $A^i v_j$ 。对于任意整数$m \geqslant 1$，令k 为满足$k(k+1)/2 \leqslant m < (k+1)(k+2)/2$ 的整数，即m 至少为第 k个三角数但小于第 $(k+1)$个三角数。考虑向量 $v_1, Av_1, v_2, A^2v_1, Av_2, v_3, A^3v_1,\ldots$ 的顺序（这按$i+j$的顺序枚举 $A^i v^j$，若相等则优先取 j 值较小者）。令 W_m 为这些向量中的前 m 个向量组成的集合，令X_k 为集合 $\{v_1,\ldots,v_k\}$。令 V 为一个固定 W_m 且作用于$W_m^{\perp}$的正交补空间上的 *Haar*随机正交矩阵。那么，$(X_k, U) \mathrm{d} = (X_k, UV)$。

在证明这个引理之前，我们注意到，由于算法是确定性的，且D 固定，W_m和X_k 都是A的确定性函数，因此也是U的确定性函数。因此，我们可以将它们记为$v_k(U')$、$W_m(U')$和$X_k(U')$

表示如果我们以$A' = (U')^\top DU'$为起点，就会生成的v_k、W_m和X_k。（如果没有给出参数，则v_k、W_m和X_k分别表示$v_k(U)$、$W_m(U)$和$X_k(U)$。）我们注意到以下命题。

命题 A.3 (固定前 m 次查询和响应)。 假设 V 是任意一个保持 $W_m(U)$不变的正交矩阵。那么，$W_m(U) = W_m(UV)$。

证明。 我们证明 $W_{m'}(U) = W_{m'}(UV)$ 对所有 $m' \leqslant m$。$k = 1$ 的基本情形很简单，因为v_1 是固定的。现在我们证明m'的归纳步骤。

如果 $m' \leqslant m$ 是三角数，$m' = {}_{k(k+1)}\,2$，那么m'个向量在W_m 中是v_k。但请注意，$v_k(U)$ 是$W_{m'-1}(U)$的一个确定性函数，而$v_k(UV)$ 是$W_{m'-1}(UV)$的同一个确定性函数。因此，如果对$m'-1$的归纳法假设成立，那么对m也成立。

如果 $\mathrm{m}' \leqslant \mathrm{m}$ 不是三角数，那么 $\mathrm{W}_m(\mathrm{U})$ 中第 m' 个数是 $\mathrm{A}^i\mathrm{v}_j$，其中 $\mathrm{i} \geqslant 1$。同样，$\mathrm{W}_m(\mathrm{UV})$ 中第 m' 个数是 $\mathrm{V}^\top\mathrm{A}^i\mathrm{V} \cdot \mathrm{v}_j(\mathrm{UV})$。由于 $\mathrm{i} \geqslant 1$，根据关于 $\mathrm{j}(\mathrm{j}+1)_{\,2} < \mathrm{m}'$ 的归纳假设，我们知道 $\mathrm{v}_j(\mathrm{U}) = \mathrm{v}_j(\mathrm{UV})$。但是，我们知道 V 保持 W_m 不变，这意味着它保持 v_j 和 $\mathrm{A}^i\mathrm{v}_j$ 不变。因此，$\mathrm{V}^\top\mathrm{A}^i\mathrm{V}\mathrm{v}_j(\mathrm{UV}) = \mathrm{V}^\top\mathrm{A}^i\mathrm{V}\mathrm{v}_j = \mathrm{A}^i\mathrm{v}_j$。 □

现在我们准备证明引理A.2。

引理 *A.2* 的证明。 我们对 m 使用归纳法证明。对于基础情形 $m = 1$，U 是一个随机矩阵，而 V 是一个保持 v_1 不变的随机矩阵。注意，v_1 的选取独立于 A（因而也独立于U），所以 U 和 V 相互独立。即使对任意固定的 V，分布 UV 也是均匀随机的正交矩阵，因此总体而言，$U \,\mathrm{d}= UV$。此外，v_1 是确定性的，因此 $(v_1, U) \,\mathrm{d}= (v_1, UV)$。

对于归纳步骤，我们将证明分为 2 种情形。这两种情形的证明非常相似，但有一些细微差别。

情形 1： m **是一个三角数**。 这意味着，第m个添加的向量是v_k，其中$m = {}_{k(k+1)}\,2$。令V_1 为一个固定W_{m-1} 的随机正交矩阵，令V_2 为一个固定W_m的随机正交矩阵。于是我们的目标是证明$(X_k, U) \,\mathrm{d}= (X_k, UV_2)$。

为了使论证严谨，我们说明随机变量的生成顺序。首先，随机生成U：W_m 和X_k 由U确定。接着，我们定义V_1 为一个固定W_{m-1}的随机旋转。最后，我们定义V_2 为一个固定W_m的随机旋转，其中V_1, V_2 在给定U的条件下相互独立。

首先，我们证明 $(\mathrm{X}_k, \mathrm{U}) \,\mathrm{d}= (\mathrm{X}_k, \mathrm{UV}_1)$。注意，根据归纳假设，$\mathrm{U} \,\mathrm{d}= \mathrm{UV}_1$。此外，由于 V_1 固定 $\mathrm{W}_{m-1}(\mathrm{U})$，根据命题 A.3，$\mathrm{W}_{m-1}(\mathrm{U}) = \mathrm{W}_{m-1}(\mathrm{UV}_1)$。由于 $\mathrm{m} = \mathrm{k}(\mathrm{k}+1)\ 2$ 是一个三角数，$\mathrm{X}_k(\cdot)$ 是 $\mathrm{W}_{m-1}(\cdot)$ 的确定性函数，这意味着 $\mathrm{X}_k(\mathrm{U}) = \mathrm{X}_k(\mathrm{UV}_1)$。因此，$(\mathrm{X}_k, \mathrm{U}) \,\mathrm{d}= (\mathrm{X}_k(\mathrm{UV}_1), \mathrm{UV}_1) = (\mathrm{X}_k, \mathrm{UV}_1)$。

接下来，我们证明 $(X_k, UV_2) \,\mathrm{d}= (X_k, UV_1V_2)$。只需证明

$$(X_k, U, V_2) \overset{\mathrm{d}}{=} (X_k, UV_1, V_2).$$

为此，我们首先证明 $\mathrm{V}_2 = \mathrm{f}(\mathrm{U}, \mathrm{R})$，其中 f 是一个确定性函数，R 表示一个作用于 $\mathrm{d} - \dim(\mathrm{W}_m)$ 维空间的随机正交矩阵，且与 U 独立。（回想一下，W_m 是 U 的确定性函数。）为定义 $\mathrm{f}(\mathrm{U}, \mathrm{R})$，我们考虑一个确定性映射，它将每个 W_m 映射到 $\mathrm{W}_m^\perp$ 中的一组 $\mathrm{d} - \dim(\mathrm{W}_m)$ 个基向量。接着，我们定义 $\mathrm{V}_2 = \mathrm{f}(\mathrm{U}, \mathrm{R})$ 为

作用于$W_m^{\perp}$，使用R 以及基向量的对应关系。由于W_m 和X_k 根据U确定，这意味着$f(U、R)$ 是良定义的。现在我们将证明，

$$V_2 = f(U,R) = f(UV_1,R) \quad \text{and} \quad X_k = X_k(UV_1).$$

由于$U \overset{d}{=} UV_1$ 根据我们的归纳假设，

$$(X_k,U,V_2) \overset{d}{=} (X_k(UV_1),UV_1,f(UV_1,R)) = (X_k,UV_1,V_2).$$

根据命题A.3，$W_{m-1}(U) = W_{m-1}(UV_1)$，并且由于当$m = k(k+1)/2$ 时，给定$W_{m-1}(\cdot)$ 后$X_k(\cdot)$ 是确定的，~~$X_k(U) = X_k(UV_1)$~~。这意味着$W_m(U) = W_m(UV_1)$，也就是说 $f(UV_1,R) = f(U,R)$，因为 $f(\cdot,R)$ 只取决于 $W_m(\cdot)$ 和 R。这就完成了证明。

接下来，我们证明 $(X_k, UV_1V_2) \overset{d}{=} (X_k, UV_1)$。由于我们选择的顺序是先定义 U，因此可以对 U 进行条件化。由于 X_k 由 U 唯一确定，只需证明 $V_1V_2|U \overset{d}{=} V_1|U$。由于给定 U 时 W_{m-1}、W_m 也都是确定的，注意到 V_1 是一个固定 W_{m-1}的均匀随机正交矩阵，而 V_2 是一个固定 $W_m \supset W_{m-1}$的随机正交矩阵。由于给定 U 时 V_1 和 V_2 条件独立，这意味着 $V_1V_2|U$ 是一个固定 W_{m-1}的均匀随机正交矩阵，因此 $V_1V_2|U \overset{d}{=} V_1|U$。

综上，我们得到

$$\begin{aligned}(X_k,U) &\overset{d}{=} (X_k,UV_1)\\ &\overset{d}{=} (X_k,UV_1V_2)\\ &\overset{d}{=} (X_k,UV_2).\end{aligned}$$

情形 2： m **不是三角数**。再次，令 V_1 是一个保持W_{m-1} 不变的随机正交矩阵，且V_2 是一个保持W_m不变的随机正交矩阵。我们的目标仍然是证明（X_k，U）$\overset{d}{=}$（X_k，UV_2）。

首先，根据我们的归纳假设，我们再次有 $(X_k,UV_1) \overset{d}{=} (X_k,U)$。

接下来，我们证明$(X_k,UV_2) \overset{d}{=} (X_k,UV_2V_1)$。只需证明，

$$(X_k,U,V_2) \overset{d}{=} (X_k,UV_1,V_1^{\top}V_2V_1),$$

由于 $(UV_1)(V_1^{\top}V_2V_1) = UV_2V_1$。我们回顾随机变量 R，并使用相同的函数 $V_2 = f(U,R)$。由于我们已经证明 $Ud = UV_1$，，这意味着 $(X_k, U, V_2)d = (X_k(UV_1),UV_1, f(UV_1,R))$。由于 m 不是三角形的，$X_k(\cdot)$包含于 $W_{m-1}(\cdot)$ 中，因此根据命题 A.3，$X_k(U) = X_k(UV_1)$。所以，我们有

$$(X_k,U,V_2) \overset{d}{=} (X_k(UV_1),UV_1,f(UV_1,R)) = (X_k,UV_1,f(UV_1,R)).$$

现在，如果我们固定 U 和 V_1,$W_{m-1}(UV_1) = W_{m-1}(U)$，根据命题 A.3。然而，由于当 m 不是三角数时，第 m 个 (i, j) 对具有 $i \geqslant 1$，$W_m(UV_1)$ 中的最终向量将为 $V_1^{\top}A^iV_1v_j = V_1^{\top}(A^iv_j)$。对于固定的 U，$V_1, f(U,R)$ 是一个固定 W_{m-1} 和 A^iv_j 的随机旋转，但 $f(UV_1,R)$ 是一个固定 W_{m-1} 和 $V_1^{\top}(A^iv_j)$ 的随机旋转。由于根据我们定义 V_1，的方式，$V_1^{\top}$固定 W_{m-1}，这意味着对于固定的 U，$V_1, f(U,R)$ 是一个固定 W_m 的随机旋转，但 $f(UV_1,R)$ 是一个随机旋转

旋转固定 $V_1^\top W_m$。因此，在给定 U 的条件下，$V_1, f(UV_1, R)$ 与 $V_1^\top f(U、R)V_1$具有相同的分布。由于X_k 由U确定，因此这意味着

$$(X_k, UV_1, f(UV_1, R))|U, V_1 \overset{\mathrm{d}}{=} (X_k, UV_1, V_1^\top f(U, R)V_1)|U, V_1.$$

我们可以消去条件限制，从而证明 $(X_k, UV_1, f(UV_1, R)) \mathrm{d} = (X_k, UV_1, V_1^\top f(U, R)V_1) = (X_k, UV_1, V_1^\top V_2 V_1)$，由此完成证明。

接下来，我们证明 $(X_k, UV_2V_1) \mathrm{d} = (X_k, UV_1)$。证明与 m 为三角形矩阵时的情形基本相同。我们再次对 U 条件化，并且有 $V_2V_1|U \mathrm{d} = V_1|U$，它们具有相同的分布，都是固定 $W_{m-1}(U)$ 的均匀正交矩阵。由于 X_k 是 U 的确定性函数，这意味着 $(X_k, UV_2V_1)|U \mathrm{d} = (X_k, UV_1)|U$，消去条件限制即可完成证明。

总之，

$$\begin{aligned}(X_k, U) &\overset{\mathrm{d}}{=} (X_k, UV_1)\\ &\overset{\mathrm{d}}{=} (X_k, UV_2V_1)\\ &\overset{\mathrm{d}}{=} (X_k, UV_2).\end{aligned}$$

□

现在我们证明主要的条件引理，它将是对引理 A.2 的修改。

引理 A.4（条件引理）。设所有记号均与引理A.2中的相同，并设 V_0 是一个固定的正交矩阵，且保持 W_m不变。重要的是，V_0 是仅依赖于 W_m（而不直接依赖于 U）的确定性函数。于是，$(X_k, U) \mathrm{d} = (X_k, UV_0)$。

证明。首先请注意，由于 V_0 是W_m 的确定性函数，它也是 U 的确定性函数。我们可以将 $V_0(\cdot)$ 写作该函数，并 $V_0 = V_0(U)$。

现在，引理 A.2 证明了 $(X_k, U) \mathrm{d} = (X_k, UV)$。注意，给定 U，V 是一个固定 W_m 的随机矩阵，而 V_0 是一个固定 W_m 的矩阵，这意味着 $VV_0|U \mathrm{d} = V|U$。因此，$(X_k, UV) \mathrm{d} = (X_k, UVV_0)$。但根据命题 A.3，$X_k(UV) = X_k(U)$，且 $W_m(UV) = W_m(U)$，这意味着 $V_0(\cdot)$（它只依赖于 $W_m(\cdot)$）满足 $V_0(UV) = V_0(U)$。因此，由于 $U \mathrm{d} = UV$，我们有 $(X_k, UVV_0) = (X_k(UV), UV \cdot V_0(UV)) \mathrm{d} = (X_k(U), U \cdot V_0(U)) = (X_k, UV_0)$。

总之，我们有 $(X_k, U) \mathrm{d} = (X_k, UV) \mathrm{d} = (X_k, UVV_0) \mathrm{d} = (X_k, UV_0)$，由此完成证明。

□

A.3 从查询算法到块 Krylov 算法

现在我们旨在证明引理 6.3，它意味着扩展预言机模型中的任何自适应确定性算法都可以通过旋转块 Krylov 算法的输出来模拟。

首先，我们描述如何构造 $\tilde{v}_k$。令 $\tilde{v}_1 = z_1 / \|z_1\|$，对于$k \geqslant 2$，令 $\tilde{v}_k$ 为与z_k 的下述分量平行的单位向量：该分量正交于 $\{A^i z_j\}_{H_{k-1}}$的张成空间。（以概率 1，这一定义良好。）等价地，我们可以令 Q_k 为$\{A^i z_j\}_{H_{k-1}}$的正交基，并定义 $\tilde{v}_k := (I - Q_k Q_k^\top) z_k / \| (I - Q_k Q_k^\top) z_k \|_2$。

我们注意到 $\tilde{v}_k$ 可以表示为 $A^i z_j$ 的形式，同样地， $\tilde{z}_k$ 可以表示为$A^i \tilde{v}_j$的形式。形式化地说，有如下结论。

命题 A.5. 对于每个 $k \geqslant 1$, $\tilde{v}_k$ 都是 $\{A^i z_j\}_{i+j\leqslant k}$的线性组合，而 z_k 是 $\{A^i \tilde{v}_j\}_{i+j\leqslant k}$的线性组合。

证明。 根据定义， $\tilde{v}_k$ 是 $\{A^i z_j\}_{H_{k-1}}$ 和 z_k的线性组合，因此它也是 $\{A^i z_j\}_{i+j\leqslant k}$的线性组合。因此，对于所有 $k \leqslant K$，我们可以将集合 $\{A^i \tilde{v}_j\}_{H_k}$ 中的元素构造为集合$\{A^i z_j\}_{H_k}$中元素的线性组合。

现在我们通过归纳法证明，z_k 是 $\{\mathrm{A}^i \tilde{\mathrm{v}}_j\}_{i+j\leqslant k}$的线性组合。$\mathrm{k} = 1$时的基础情形显然成立。现在，假设 $\mathrm{k}-1$时归纳假设成立。以概率 1， $\tilde{\mathrm{v}}_k$ 非零，因此 z_k是 $\tilde{\mathrm{v}}_k$和 $\{\mathrm{A}^i \mathrm{z}_j\}_{H_{k-1}}$的线性组合。根据归纳假设，每个 z^j都是 $\{\mathrm{A}^{i'} \mathrm{z}_{j'}\}_{i'+\mathrm{j}'\leqslant \mathrm{j}}$（其中 $\mathrm{j} < \mathrm{k}$）的线性组合，这意味着对于 $(\mathrm{i}, \mathrm{j}) \in \mathrm{H}_{k-1}$，$\mathrm{A}^i \mathrm{z}^j$可以写成 $\mathrm{A}^{i+i'} \mathrm{z}'_j$（其中 $\mathrm{i} + \mathrm{i}' + \mathrm{j}' \leqslant \mathrm{i} + \mathrm{j} \leqslant \mathrm{k}$）的线性组合。因此，$(\mathrm{i} + \mathrm{i}', \mathrm{j}') \in \mathrm{H}_{k-1}$，这意味着 z_k是 $\tilde{\mathrm{v}}_k$ 和 $\{\mathrm{A}^i \tilde{\mathrm{v}}_j\}_{H_{k-1}}$的线性组合，或者等价地，是 $\{\mathrm{A}^i \tilde{\mathrm{v}}_j\}_{i+j\leqslant k}$的线性组合。

□

现在我们构造旋转矩阵 U_k^{sim}。首先，我们定义矩阵值函数$U_k(\cdot)$，对于$k = 1, \ldots, K$，定义如下。

定义 A.6。 对于 $1 \leqslant k \leqslant K$，函数$U_k(\cdot)$ 接受参数 $\{x_{i,j}\}_{H_{k-1}}, y_k, z_k$，其中向量$y_k$ 和z_k 的单位范数均为 1，且都与集合 $\{x_{i,j}\}_{H_{k-1}}$ 正交。

定义 $\mathrm{U}_1(\cdot)$：由于 H_0 为空，第一个函数 U_1 只接受参数 $\mathrm{y}_1, \mathrm{z}_1,$ ，并且 $\mathrm{U}_1(\mathrm{y}_1, \mathrm{z}_1)$ 是一个确定性的正交矩阵，满足 $\mathrm{U}_1(\mathrm{y}_1, \mathrm{z}_1)^\top \mathrm{y}_1 = \mathrm{z}_1$。注意，$\mathrm{U}_1(\cdot)$ 存在，因为 y_1 和 z_1 的单位范数均为 1；例如，我们可以将 y_1 和 z_1 补充为标准正交基 $(\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_d)$、$(\mathrm{z}_1, \mathrm{z}_2, \ldots, \mathrm{z}_d)$，并令 $\mathrm{U}_1(\mathrm{y}_1, \mathrm{z}_1) = \sum_{i=1}^d \mathrm{y}_i \mathrm{z}_i^\top$。

定义 $U_k(\cdot)$：$U_k(\{x_{i,j}\}_{H_{k-1}}, y_k, z_k)$ 是一个满足以下条件的确定性正交矩阵

$$\begin{aligned} U_k^\top x_{i,j} &= x_{i,j}\,, \qquad \text{for all } (i,j) \in H_{k-1}\,, \\ U_k^\top y_k &= z_k\,. \end{aligned} \tag{14}$$

这样的U_k 总是可以取到，因为 $k^2 < d$，并且y_k 和z_k 都与$x_{i,j}$正交；例如，我们可以先取子空间 $\{x_{i,j}\}_{H_{k-1}}$上的恒等矩阵，再加上由y_k 和z_k 分别补全为正交补的两组标准正交基后形成的外积之和。

接下来，我们描述如何构造 U_k^{sim}。我们将定义U_k^{sim}，以及辅助序列$\{v_k^{\mathsf{sim}}\}_{k=1,2,\ldots,K-1}$。

定义 A.7。 我们令$v_1^{\mathsf{sim}} = v_1$,和$U_1^{\mathsf{sim}} = U_1(\tilde{v}_1, v_1^{\mathsf{sim}})$。对于$k \geqslant 2$，$v_k^{\mathsf{sim}}$和$U_k^{\mathsf{sim}}$按如下方式递归定义：

$$\begin{aligned} v_k^{\mathsf{sim}} &= v_k\big(\{(U_{1:(k-1)}^{\mathsf{sim}})^\top A^i \tilde{v}_j\}_{H_{k-1}}\big) \\ U_k^{\mathsf{sim}} &= U_k\big(\{(U_{1:(k-1)}^{\mathsf{sim}})^\top A^i \tilde{v}_j\}_{H_{k-1}},\ (U_{1:(k-1)}^{\mathsf{sim}})^\top \tilde{v}_k,\ v_k^{\mathsf{sim}}\big)\,. \end{aligned} \tag{15}$$

直观地说，可以将v_k^{sim}视为模拟器认为算法正在查询的第*k</style}*个向量，而U_k^{sim}则是与v_k^{sim}相对应的一种旋转，这个v_k^{sim}是块 Krylov 已知的随机单位向量。

命题 A.8. 每个 U_k^{sim}都有明确定义。

证明。为了说明这种对 U_k^{sim}的选择是可行的，我们需要验证 $(U_1^{\mathsf{sim}}:(k-1))^\top \tilde{v}_k$ 和 v_k^{sim} 都是单位范数，且都与由 $(U_1^{\mathsf{sim}}:(k-1))^\top A^i \tilde{v}_j$（其中 $(i,j) \in H_{k-1}$）张成的子空间 S_k 正交。它们都是单位范数，因为 $\tilde{v}_k$ 和 v_k^{sim} 的构造均使其具有单位范数；并且根据归纳假设，可以认为 $U_1^{\mathsf{sim}}:(k-1)$ 是正交的。注意，根据我们对函数 $v_k(\cdot)$ 的假设，v_k^{sim} 与 S_k 正交；而 $(U_1^{\mathsf{sim}}:(k-1))^\top \tilde{v}_k$ 也与 S_k 正交，因为

$$\langle (U_{1:(k-1)}^{\mathsf{sim}})^\top A^i \tilde{v}_j, (U_{1:(k-1)}^{\mathsf{sim}})^\top \tilde{v}_k \rangle = \langle A^i \tilde{v}_j, \tilde{v}_k \rangle = 0\,,$$

其中，第二行由 $\tilde{v}_k$的定义可得。 □

我们在下面的引理中总结v_k^{sim}和U_k^{sim}的一些重要附加性质。

引理 A.9 （模拟序列的性质）。 上述定义的变量 U_k^{sim}和 v_k^{sim}（$k = 1,\dots,K$）满足以下性质：

(P1) v_k^{sim}仅取决于 $\{A^i \tilde{v}_j\}_{H_{k-1}}$，而 U_k^{sim}仅取决于 $\{A^i \tilde{v}_j\}_{i+j\leqslant k}$。

(P2) 对任意 $k \geqslant j$，我们有

$$\tilde{v}_j = U_{1:k}^{\mathsf{sim}} v_j^{\mathsf{sim}}\,.$$

(P3) 对于 $k \geqslant 2$，v_k^{sim} 满足

$$v_k^{\mathsf{sim}} = v_k\big(\{(U_{1:(k-1)}^{\mathsf{sim}})^\top A^i U_{1:(k-1)}^{\mathsf{sim}} v_j^{\mathsf{sim}}\}_{H_{k-1}}\big)\,.$$

(P4) 对于 $k \geqslant 2$，U_k^{sim} 满足

$$U_k^{\mathsf{sim}} = U_k\big(\{(U_{1:(k-1)}^{\mathsf{sim}})^\top A^i U_{1:(k-1)}^{\mathsf{sim}} v_j^{\mathsf{sim}}\}_{H_{k-1}},\ (U_{1:(k-1)}^{\mathsf{sim}})^\top \tilde{v}_k,\ v_k^{\mathsf{sim}}\big)\,.$$

在给出证明之前，我们先强调 (P2) $k = K(U_{1:K}^{\mathsf{sim}})^\top$ 的重要性，它大致说明该映射实际上会将每个由块 Krylov 生成的向量 $\tilde{v}_j$ 映射为模拟向量 v_j^{sim}。

*Pro*的。 (P1) 由定义可立即得到，因为 $\{(i,j) : i + j \leqslant k\} = H_{k-1} \cup \{(0,k)\}$.

T展示 (P2)，注意到函数U_k 由（14）可知

$$v_j^{\mathsf{sim}} = (U_j^{\mathsf{sim}})^\top (U_{1:(j-1)}^{\mathsf{sim}})^\top \tilde{v}_j = (U_{1:j}^{\mathsf{sim}})^\top \tilde{v}_j\,. \tag{16}$$

这证明了 (P2) 对于 $k = j$。要证明 (P2) 对于 $k > j$，我们对 k进行归纳法。如果 (P2)对$k - 1 \geqslant j$成立，那么

$$(U_{1:k}^{\mathsf{sim}})^\top \tilde{v}_j = (U_k^{\mathsf{sim}})^\top (U_{1:(k-1)}^{\mathsf{sim}})^\top \tilde{v}_j = (U_{1:(k-1)}^{\mathsf{sim}})^\top \tilde{v}_j = v_j^{\mathsf{sim}}. \tag{17}$$

上面的中间等式由（14）的第一个性质成立，因为 U_k^{sim}固定 $(U_1^{\mathsf{sim}}{}_{:(k-1)})^\top \tilde{v}_j$这是因为 $j \leqslant k - 1$。最后一个等式由归纳假设成立。因此，(P2) 对 k成立。

最后，(P3) 和 (P4) 由 (P2)推出，因为 $k - 1 \geqslant j$ 当 $j \in H_{k-1}$。 □

在证明引理6.3之前，我们还必须给出一个基本定义。

定义 A.10. 对于$k \geqslant 2$，给定矩阵A和集合$\{v_j\}_{1\leqslant j\leqslant k-1}$，定义$\mathfrak{C}_k$为满足$\mathfrak{C}_k(A, \{v_j\}_{1\leqslant j\leqslant k-1}) = \{A^i v_j\}_{H_{k-1}}$的函数。此外，定义$\mathfrak{D}_k = v_k \circ \mathfrak{C}_k$。

我们现在可以证明引理6.3了。尽管证明的记号较为繁杂，但核心思想是，我们可以通过反复应用条件引理（引理 A.4），归纳地证明分布相等；该引理正是为当前情形设计的。

引理 6.3 的证明。显然，$\tilde{v}_1, \ldots, \tilde{v}_K$ 和 $U_1^{\mathsf{sim}}, \ldots, U_K^{\mathsf{sim}}$ 满足引理 6.3 中的性质 1。我们着重证明第二个性质。对于 $1 \leqslant k \leqslant K$，令 $A_k := (U_{1:k}^{\mathsf{sim}})^\top A U_{1:k}^{\mathsf{sim}}$。由于根据引理 A.9 的 (P2)，对于任意 $k \geqslant j$，我们可以写成 $(U_{1:k}^{\mathsf{sim}})^\top A^i \tilde{v}_j = (U_{1:k}^{\mathsf{sim}})^\top A^i (U_{1:k}^{\mathsf{sim}}) v_j^{\mathsf{sim}} = A_k^i v_j^{\mathsf{sim}}$，因此只需归纳地证明对所有 $1 \leqslant k \leqslant K$，

$$(A_k, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k}) \overset{\mathsf{d}}{=} (A, \{v_j^{\mathsf{alg}}\}_{1\leqslant j\leqslant k}). \tag{18}$$

对于 $k = 1$ 的基础情形，只需证明 $(A_1, v_1^{\mathsf{sim}}) \mathsf{d} = (A, v_1^{\mathsf{alg}})$。然而，注意到 $v_1^{\mathsf{sim}} = v_1^{\mathsf{alg}} = v_1$, and $A_1 = (U_1^{\mathsf{sim}})^\top A (U_1^{\mathsf{sim}}) = U_1(\tilde{v}_1, v_1)^\top A U_1(\tilde{v}_1, v_1)$。由于 v_1 是一个确定性向量，$\tilde{v}_1$ 与 A 独立，且 A 的分布具有旋转不变性，因此结论成立。

对于归纳步骤，假设我们已知 $(A_k, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k}) \mathsf{d} = (A, \{v^{\mathsf{alg}} j \}_{1\leqslant j\leqslant k})$。接着注意到 $v^{\mathsf{alg}} k_{+1} = v_{k+1}(\{A^i v^{\mathsf{alg}} j \}_{H_k})$ and $v_{k+1}^{\mathsf{sim}} = v_k(\{A_k^i v_j^{\mathsf{sim}}\}_{H_k})$。因此，我们有 $v^{\mathsf{alg}} k_{+1} = \mathfrak{D}_{k+1}(A, \{v^{\mathsf{alg}} j \}_{1\leqslant j\leqslant k})$ and $v_{k+1}^{\mathsf{sim}} = \mathfrak{D}_{k+1}(A_k, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k})$。此外，由于根据 (P4)，对所有 $(i, j) \in H_k$，U_{k+1}^{sim} fixes $A_k^i v_j^{\mathsf{sim}}$，我们也有对所有 $(i, j) \in H_k$，$A_{k+1}^i v_j^{\mathsf{sim}} = A_k^i v_j^{\mathsf{sim}}$，这意味着 $v_{k+1}^{\mathsf{sim}} = \mathfrak{D}_{k+1}(A_{k+1}, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k})$。因此，只需证明

$$(A_{k+1}, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k}) \overset{\mathsf{d}}{=} (A, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k}), \tag{19}$$

由于这一点 成立$(A_{k+1}, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k+1}) \mathsf{d} = (A, \{v^{\mathsf{alg}} j \}_{1\leqslant j\leqslant k+1})$，从而完成归纳步骤 ep.

接下来，我们证明U_{k+1}^{sim} 将 $\tilde{v}_{k+1}$ 映射为一个与迄今为止的模拟查询正交的随机单位向量。注意，$A_{k+1} = (U_{k+1}^{\mathsf{sim}})^\top A_k (U_{k+1}^{\mathsf{sim}})$，其中根据 (P4)，

$$U_{k+1}^{\mathsf{sim}} = U_{k+1}(\{A_k^i v_j^{\mathsf{sim}}\}_{H_k}, (U_{1:k}^{\mathsf{sim}})^\top \tilde{v}_{k+1}, v_{k+1}^{\mathsf{sim}}). \tag{20}$$

注意，$\tilde{v}_{k+1}$ 是一个与 $\{A^i z_j\}_{H_k}$正交的随机单位向量，或者等价地说，它是一个与 $\{A^i \tilde{v}_j\}_{H_k}$正交的随机单位向量。然而，由于对于所有$(i, j) \in H_k$，$(U_{1:k}^{\mathsf{sim}})^\top A^i \tilde{v}_j = (U_{1:k}^{\mathsf{sim}})^\top A^i (U_{1:k}^{\mathsf{sim}}) v_j^{\mathsf{sim}} = A_k^i v_j^{\mathsf{sim}}$成立（根据(P2)），这意味着$(U_{1:k}^{\mathsf{sim}})^\top \tilde{v}_{k+1}$ 与 $\{A_k^i v_j^{\mathsf{sim}}\}_{H_k}$正交。此外，根据(P1)以及$v_k^{\mathsf{sim}}$的定义，$U_k^{\mathsf{sim}}$(公式(15))，我们知道$U_{k+1}^{\mathsf{sim}}$ 的第一个和第三个参数仅取决于 $\{A^i \tilde{v}_j\}_{H_k}$。因此，$\tilde{v}_{k+1}$ 的随机方向与$\{A^i \tilde{v}_j\}_{H_k}$ 无关，除了与它们正交这一点之外；由此根据(P1)，$(U_{1:k}^{\mathsf{sim}})^\top \tilde{v}_{k+1}$ 是一个与 $\{A_k^i v_j^{\mathsf{sim}}\}_{H_k}$正交的均匀随机单位向量。

回顾 $v_{k+1}^{\mathsf{sim}} = \mathfrak{D}_{k+1}(A_k, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k})$，这意味着我们可以将 (20) 重写为

$$U_{k+1}^{\mathsf{sim}} = U_{k+1}\left(\{A_k^i v_j^{\mathsf{sim}}\}_{H_k}, \hat{v}^{\mathsf{sim}}, \mathfrak{D}_{k+1}(A_k, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k})\right), \tag{21}$$

其中 $\hat{v}^{\mathsf{sim}}$ 是一个与 $\{A_k^i v_j^{\mathsf{sim}}\}_{H_k}$正交的随机单位向量。因此，如果我们定义 e

$$U_{k+1}^{\mathsf{alg}} := U_{k+1}\left(\{A^i v_j^{\mathsf{alg}}\}_{H_k}, \hat{v}^{\mathsf{alg}}, \mathfrak{D}_{k+1}(A, \{v_j^{\mathsf{alg}}\}_{1\leqslant j\leqslant k})\right), \tag{22}$$

其中 $\hat{v}^{\mathsf{alg}}$ 是一个与 $\{A^i v^{\mathsf{alg}}{}_j \quad\}_{H_k}$ 正交的随机单位向量，然后

$$\begin{aligned}(A_{k+1}, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k}) &= ((U_{k+1}^{\mathsf{sim}})^\top A_k (U_{k+1}^{\mathsf{sim}}), \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k}) \\ &\overset{\mathsf{d}}{=} \left((U_{k+1}^{\mathsf{alg}})^\top A (U_{k+1}^{\mathsf{alg}}), \{v_j^{\mathsf{alg}}\}_{1\leqslant j\leqslant k}\right).\end{aligned}$$

上面，第一个等式由定义可得，第三个等式则由我们的归纳假设成立，即 $_k$、$\{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k}) \,\mathsf{d}\, =$ $(A$、$\{v^{\mathsf{alg}}{}_j \quad\}_{1\leqslant j\leqslant k})$，再结合式（21）和式（22）。

现在我们可以应用条件化引理（引理 A.4）。注意，$U^{\mathsf{alg}}{}_{k+1}$ 仅依赖于 $\{A^i v^{\mathsf{alg}}{}_j \quad\}_{H_k}$（以及 $\hat{}$ v^{alg} 中的一些随机性；但给定 $\{A^i v^{\mathsf{alg}}{}_j \quad\}_{H_k}$ 后，该随机性与其他所有因素独立，因此我们可以放心地对其进行条件化）。因此，我们可以对 $U^{\mathsf{alg}}{}_{k+1}$ 应用条件化引理，从而得到

$$(A_{k+1}, \{v_j^{\mathsf{sim}}\}_{1\leqslant j\leqslant k}) \overset{\mathsf{d}}{=} \left((U_{k+1}^{\mathsf{alg}})^\top A (U_{k+1}^{\mathsf{alg}}), \{v_j^{\mathsf{alg}}\}_{1\leqslant j\leqslant k}\right) \overset{\mathsf{d}}{=} \left(A, \{v_j^{\mathsf{alg}}\}_{1\leqslant j\leqslant k}\right),$$

由此得出式（19），并完成证明。 □