arXiv:2603.05721v1 [math.NA] 5 Mar 2026

FlexTrace：矩阵函数的可交换随机化迹估计

MADHUSUDAN MADHAVAN* , ALENALEXANDERIAN , AND ARVIND K. SAIBABADEPARTMENT OFMATHEMATICS, NORTHCAROLINA STATEUNIVERSITY*通讯作者：mmadhav@ncsu.edu

我们研究如何估计大型对称半正定矩阵**A**的矩阵函数迹 tr(f(**A**))。这一问题出现在多种应用中，包括核方法和反问题。现有迹估计方法面临的一项关键挑战是需要计算 f(**A**) 的矩阵向量乘积（matvec），而这类计算可能代价高昂。本文提出一种新颖的迹估计器 FLEXTRACE：一种可交换的单遍方法，仅使用与**A**的矩阵向量乘法来估计 tr(f(**A**))。我们考虑 f 是满足 $f(0) = 0$ 的算子单调矩阵函数的情形，其中包括 $\log(1+x)$ 和 $x^{1/2}$ 等函数，并推导了概率界，展示 FLEXTRACE 的理论优势。针对合成示例和各类应用领域开展的数值实验表明，与现有方法相比，FLEXTRACE 对 f(**A**)的迹估计准确得多。

关键词：随机化迹估计；矩阵函数；随机 Nyström 近似；方差缩减；可交换性。

1. 引言

矩阵函数是众多科学与工程应用中不可或缺的组成部分，包括网络分析、偏微分方程（PDEs）和统计学习 [9, 23, 33]。这些应用中的一项常见任务是估计矩阵函数的迹 [56, 57]。对于对称正定矩阵$\mathbf{A} \in \mathbb{R}^{n\times n}$和函数 $f\colon [0,\infty) \to [0,\infty)$，矩阵函数的迹可以表示为**A**的特征值 $\{\lambda_i\}_{i=1}^n$的函数，即 $\mathrm{tr}(f(\mathbf{A})) = \sum_{i=1}^n f(\lambda_i)$。然而，一项根本性的挑战是，在大规模场景下，获取**A**的特征值可能代价高昂得难以承受，因此无法显式计算 tr(f(**A**))。此外，在许多应用中——例如偏微分方程约束优化或核方法——矩阵**A** 本身并不能显式获得，只能通过矩阵-向量乘法 $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$访问，这给矩阵函数的计算带来了进一步挑战。

本文介绍 FLEXTRACE，这是一种单遍方法，可直接利用 A 的矩阵向量乘法，高效估计 tr(f(A))。我们重点研究满足 $f(0) = 0$ 的算子单调矩阵函数（见第 2.1 节）。算子单调函数可保持矩阵的谱序，并有助于利用 A 的谱性质推导关于 f(A) 的理论界限。估计这类函数的 tr(f(A)) 可用于多种重要应用。例如，对数行列式的估计对于高斯过程[19], 贝叶斯推断 [2], 和空间统计 [5, 47]等应用至关重要。同样，矩阵 Schatten p-范数（对于矩形矩阵 X，其定义为 $\|X\|_{(p)} := \mathrm{tr}((X^\top X)^{p/2})^{1/p}$）涉及矩阵函数的迹估计，对于矩阵补全及其他低秩恢复任务 [41, 46, 48]都很重要。具体而言，我们的方法适用于估计 $1 \le p \le 2$ 的 Schatten p-范数，其中最重要的情况或许是核范数（$p = 1$）。我们针对重要的算子单调函数类（见表1）给出了 FLEXTRACE 的理论保证，并在不同应用中展示了相较于现有方法显著的精度与效率提升。

© 作者 2026 年。

相关工作

矩阵函数已在文献中得到充分研究，许多工作探讨了其理论和计算方面 [33, 34]。矩阵函数的迹估计可视为隐式迹估计的一个特例（见 第 2.2 节）。过去十年间，研究者日益关注将随机低秩近似技术用于迹估计 [44, 50, 54]。这些方法利用随机向量进行矩阵向量乘法，以构造矩阵的低秩近似 [26, 28, 45]。然而，这类技术通常需要对 $f(\mathbf{A})$进行矩阵向量乘法，这并非易事。对此，通常采用 f 的多项式近似或 Krylov 子空间方法。后者近年来出现了若干重要进展，包括随机 Lanczos 求积（SLQ） [7, 56] 及其变体 [16, 24, 60]。

在这类方法中，通常使用 Lanczos 方法或块 Lanczos 方法 [33] 来近似将 $f(\mathbf{A})$ 应用于随机向量的结果，再用这些结果估计迹。然而，这类方法通常需要对矩阵 $\mathbf{A}$ 进行多次遍历，例如计算 $\mathbf{A}^2\boldsymbol{x} = \mathbf{A}(\mathbf{A}\boldsymbol{x})$ 这类表达式。这对这些方法在依赖于对 $\mathbf{A}$ 进行离线计算的问题中的适用性造成了显著限制。例如，在计算资源密集型应用中，与 $\mathbf{A}$ 进行矩阵-向量乘法可能需要数天或数周的计算时间，并且在迹估计过程中无法随时调用。此外，对 $f(\mathbf{A})$ 的矩阵-向量乘法通常独立于随机迹估计进行，从而增加了估计所需的矩阵-向量乘法总次数。

我们的方法与 funNyström 方法 [51]，密切相关，并以其为基础。该方法利用 $\mathbf{A}$ 的随机 Nyström 近似来估计 $f(\mathbf{A})$。然而，我们主要关注以单遍方式估计 $f(\mathbf{A})$ 的迹，而非直接估计 $f(\mathbf{A})$。 [51] 提出的 FUNNYSTRÖM++ 方法以及近期的研究 [16, 60] 给出了高效进行矩阵-向量乘法以估计 $\mathrm{tr}(f(\mathbf{A}))$ 的技术，但要有效利用这些技术，需要对矩阵进行多次遍历。 [22]，强调了可交换性（即随机向量置换不变性）在迹估计中的重要性，并提出了一种称为 XNYSTRACE 的留一可交换迹估计技术，我们将在第 2.4 节中讨论。我们的方法 FLEXTRACE 以此为基础，提出了一种可交换方法，用于估计矩阵函数的迹。最后，我们提到，我们的工作也与 [18]，中的单样本探测方法有一些相似之处，不过我们着重开发用于算子单调矩阵函数的单遍迹估计器。

我们的方法与贡献

本文提出了一种新颖的估计量 FLEXTRACE（函数无关低秩可交换迹估计算法），它利用低秩随机 Nyström 近似和对余项的蒙特卡洛估计量，来近似计算 $\mathrm{tr}(f(\mathbf{A}))$。该方法无需计算或近似与 $f(\mathbf{A})$ 的矩阵-向量乘法，因此适用于只能**离线** 访问与$\boldsymbol{A}$ 的矩阵-向量乘法的应用场景（例如，在迹估计过程开始之前）。此外，该方法具有函数无关性，无需额外进行与$\mathbf{A}$的矩阵-向量乘法，即可计算多个函数对应的 $\mathrm{tr}(f(\mathbf{A}))$。除了提出估计量 FLEXTRACE 外，我们还介绍了 i-FLEXTRACE——FLEXTRACE 的理想化变体，有助于开展更深入的理论分析。本文的主要贡献包括：

- 一种可交换单遍方法 FLEXTRACE，用于计算矩阵函数的迹；
- 所提出算法的期望和均方误差的理论保证，以及在任意酉不变范数下 funNyström 近似 [51] 的误差界；
- 注重数值稳定性的 FLEXTRACE 高效实现；以及

- 通过合成示例、贝叶斯推断、矩阵补全和核方法，对所提出的算法进行数值演示。

本文的组织结构如下。在第 2 节中，我们定义符号并介绍必要的背景知识，为所提出的方法奠定基础。在第 3 节中，我们介绍所提出的算法 FLEXTRACE及其理想化变体 i-FLEXTRACE，阐述它们的理论性质，并给出 FLEXTRACE的高效稳定实现。我们在第 4 节中给出数值结果，展示所提出的算法在合成矩阵以及贝叶斯反演、核方法等领域应用中的有效性。关键理论结果的证明见第 5 节。最后，我们在第 6 节中给出总结性评论。

2. 背景

本节介绍矩阵函数随机化迹估计所需的背景知识。

2.1. 关键假设与符号

F对于后续的结果和定理，我们作出以下基本假设 s.

矩阵与特征值 在全文中，假设$\mathbf{A}$ 是一个 $n \times n$ 实对称半正定矩阵，其特征值按非递增顺序排列为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$。$\mathbf{A}$的谱分解表示为 $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$，其中 $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$，且$\mathbf{U} \in \mathbb{R}^{n\times n}$是一个正交矩阵。对于指标 $1 \leq i < j \leq n$，我们将子矩阵定义为 $\mathbf{\Lambda}_{i:j} := \mathrm{diag}(\lambda_i, \ldots, \lambda_j)$。给定两个对称矩阵 $\mathbf{B}$,$\mathbf{C}$，若$\mathbf{C} - \mathbf{B}$ 是半正定矩阵，则称$\mathbf{B} \preceq \mathbf{C}$。这在对称矩阵集合上定义了一种称为 Loewner偏序 [34, p. 493]的偏序关系。

高斯随机矩阵 对于 $k \leq n$，$\mathbf{\Omega} \in \mathbb{R}^{n\times k}$ 表示一个元素相互独立且服从标准正态分布的矩阵。本文用 $\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ 表示均值为 $\boldsymbol{\mu}$ 、协方差为 $\mathbf{\Sigma}$的高斯测度。

矩阵函数 给定函数 $f\colon \mathbb{R} \to \mathbb{R}$，矩阵函数 $f(\mathbf{A})$ 利用谱分解 $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ 定义为 $f(\mathbf{A}) = \mathbf{U}f(\mathbf{\Lambda})\mathbf{U}^T$，其中 $f(\mathbf{\Lambda}) = \mathrm{diag}(f(\lambda_1)\ldots, f(\lambda_n))$。若函数$f$满足 [10, 第 5] 章中的定义，即 $\mathbf{C} \preceq \mathbf{D}$ 蕴含 $f(\mathbf{C}) \preceq f(\mathbf{D})$，则称其为算子单调函数。本文考虑集合 $\mathcal{F}$中的函数，其定义如下

$$\mathcal{F} := \left\{ f : [0,\infty) \to [0,\infty) : \quad f(0) = 0 \text{ and } f \text{ is operator monotone} \right\}, \tag{2.1}$$

一些重要的矩阵函数 $\mathcal{F}$ 列于表1中。

范数 本文中，我们用 $\|\cdot\|_2$ 表示谱范数，用 $\|\cdot\|_{(p)}$ 表示 Schatten p-范数，其中 $p \geq 1$。为更加清晰地表述，提及 Frobenius（$p = 2$）范数和核（$p = 1$）范数时，我们分别使用记号 $\|\cdot\|_F$和 $\|\cdot\|_*$。对于任意酉不变范数，我们记作 $|||\mathbf{A}|||$，其满足 $|||\mathbf{A}||| = |||\mathbf{U}\mathbf{A}\mathbf{V}|||$，其中 $\mathbf{U}$，$\mathbf{V}$ 为正交矩阵。

2.2. 隐式迹估计

本节回顾用于隐式迹估计的随机化方法，即利用$\mathrm{tr}(\mathbf{A})$ 与$\mathbf{A}$进行矩阵-向量乘法来估计其迹。简单却应用广泛的一种随机化迹估计技术是 Girard-Hutchinson迹估计器 [25, 36]。该估计量的核心思想源自定理 2.1。

表1 算子单调函数的重要示例

函数（f）	$\mathbf{tr}$（f（$\mathbf{A}$））的等价形式	应用
$\log(1+x)$	$\log\det(\mathbf{I}+\mathbf{A})$	贝叶斯推断 [2]，机器学习 [40]
$x^p\ (0\le p\le 1)$	$\mathrm{tr}(\mathbf{A}^p)$	Schatten 范数估计 [41]，Rényi 熵 [20]
$\frac{x}{x+\zeta}\ (\zeta>0)$	$\mathrm{tr}(\mathbf{A}(\mathbf{A}+\zeta\mathbf{I})^{-1})$	有效维数 [5]，核学习 [19]

定理 2。1 *[6, 引理 9]* 设$\mathbf{A}\in\mathbb{R}^{n\times n}$ 为对称半正定，且 $\boldsymbol{\omega}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$。则，

$$\mathbb{E}[\boldsymbol{\omega}^\top\mathbf{A}\boldsymbol{\omega}]=\mathrm{tr}(\mathbf{A}),\quad \textit{and}\quad \mathrm{Var}[\boldsymbol{\omega}^\top\mathbf{A}\boldsymbol{\omega}]=2\|\mathbf{A}\|_F^2.$$

Girard-Hutchinson 估计量利用这一性质，通过蒙特卡洛方法估计迹。这需要抽取相互独立的随机向量 $\boldsymbol{\omega}_1$、$\boldsymbol{\omega}_2$、...、$\boldsymbol{\omega}_k$，这些向量来自 $\mathcal{N}(\mathbf{0},\ \mathbf{I})$，并计算

$$\widehat{\mathrm{tr}}_{GH}:=\frac{1}{k}\sum_{i=1}^{k}\boldsymbol{\omega}_i^\top\mathbf{A}\boldsymbol{\omega}_i\approx\mathrm{tr}(\mathbf{A}). \tag{2.2}$$

尽管该估计量无偏，但其方差衰减缓慢，速率为 $\mathcal{O}(1/k)$。近年来，已有多种方法 [22, 44, 50] 先利用随机低秩近似捕获（近似的）主特征子空间，再使用该估计量。这些方法的 $\mathcal{O}(1/k^2)$ 方差衰减速度更快；对于谱衰减迅速的矩阵，其表现显著优于 Girard-Hutchinson 估计量。对于这类方法，针对半正定矩阵的随机 Nyström 近似（第 2.3 节）尤其有效。

2.3. 随机 *Nyström* 近似

测试矩阵$\boldsymbol{\Omega}$ 对应的$\mathbf{A}$的 Nyström 近似 [26]定义为：

$$\widehat{\mathbf{A}}_{\mathrm{nys}}=\mathbf{A}\boldsymbol{\Omega}(\boldsymbol{\Omega}^\top\mathbf{A}\boldsymbol{\Omega})^\dagger(\mathbf{A}\boldsymbol{\Omega})^\top. \tag{2.3}$$

随机 Nyström 近似利用对$\mathbf{A}$的草图（随机嵌入）构造$\mathbf{A}$ 的低秩近似。该近似保证 [17]能够给出最佳的秩-k近似，近似对象为$\mathbf{A}$ ，其值域包含于 $\mathcal{R}ange(\mathbf{A}\boldsymbol{\Omega})$ 且被$\mathbf{A}$ 支配（即 $\mathbf{A}_{\mathrm{nys}}\preceq\mathbf{A}$）[4, 定理 5。3]Nyström 近似可用于高效估计$\mathbf{A}$的主特征对；参见算法 2.1。

在实际应用中，我们避免显式计算 $(\boldsymbol{\Omega}^\top\mathbf{A}\boldsymbol{\Omega})^\dagger$ 因为已知这样做在数值上不稳定。我们转而采用 稳定化 *Nyström* 近似 [11], ，它利用截断主元 Cholesky 分解 [32, 第 10 节。3] 对 $\boldsymbol{\Omega}^\top\mathbf{A}\boldsymbol{\Omega}=\mathbf{C}^\top\mathbf{C}$进行分解；记作 `cholp` ，见算法 2.1第 3 行。我们使用 LAPACK [3] 例程 `dpstrf`；当剩余的最大对角元素在数值上为零时，该例程会终止 Cholesky 分解；详情请参阅 [30, 42] 。

算法 2.1 低秩近似 $[\mathbf{\Lambda}, \mathrm{U}, \mathrm{C}, \mathrm{U}, \mathrm{Z}, \hat{r}] = \mathrm{NYSSVD}(\mathrm{A}, \Omega)$ 输入：对称半正定矩阵 $\mathrm{A} \in \mathbb{R}^{n \times n}$，测试矩阵 $\Omega \in \mathbb{R}^{n \times k}$ 输出：对角矩阵 $\widehat{\mathbf{\Lambda}} \in \mathbb{R}^{k \times k}$ 和 $\widehat{\mathrm{U}} \in \mathbb{R}^{n \times k}$，其列为标准正交列，且满足 $\widehat{\mathrm{A}}_{\mathrm{nys}} = \widehat{\mathrm{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathrm{U}}^{\top}$ 1: 构造 $\mathrm{Y} = \mathrm{A}\Omega$ 2: 执行薄 QR 分解，$\mathrm{Y} = \mathrm{QR}$ 3: 计算截断主元 Cholesky 因子，$\mathrm{C} = \mathrm{cholp}(\Omega^{\top}\mathrm{Y})$，其中 $\mathrm{C} \in \mathbb{R}^{k \times \hat{r}}$ 满足 $\hat{r} \leq k$ 4: 执行回代求解，$\mathrm{Z} = \mathrm{R} \,/\, \mathrm{C}$ 5: 计算薄 SVD，$[\mathrm{U}\ \mathbf{\Sigma}, \sim] = \mathrm{svd_econ}(\mathrm{Z})$ 6: 返回 $\widehat{\mathrm{U}} = \mathrm{Q}\widetilde{\mathrm{U}}$ 和 $\widehat{\mathbf{\Lambda}} = \mathbf{\Sigma}^2$ 7: （可选）返回 C、Z、U 和 $\hat{r}$

这还可以量化 $\mathbf{A}^{\mathrm{nys}}$ 的数值秩，这将在 第 3.3 节中发挥作用。参见[11, 45, 55]，以进一步了解 Nyström 近似的稳定性和实现细节。

2.4. 使用 *Nyström* 近似进行隐式迹估计

$\mathbf{A}$的随机 Nyström 近似为低秩或近似低秩矩阵的迹估计提供了一种灵活且有效的工具。也就是说，$\mathrm{tr}(\mathbf{A}^{\mathrm{nys}})$ 可用作 $\mathrm{tr}(\mathbf{A})$ [51, 54]的估计量。此外，将该方法与传统的随机化迹估计技术结合使用，有望进一步提升性能。例如，我们可以考虑这样一种策略：不使用完整草图进行 Nyström 近似，而是保留一些向量，用于对残差进行 Girard-Hutchinson 迹估计。为形式化这一构造，令 $\mathbf{\Omega}_{-i}$ 和$\mathbf{Y}$分别表示删除第 i 列后的矩阵 $\mathbf{\Omega}$ 和$\mathbf{Y}$，并将 $\mathbf{A}_{-i}$ 定义为使用测试矩阵 $\mathbf{\Omega}_{-i}$（其中 $1 \leq i \leq n$）得到的 $\mathbf{A}$ 的随机 Nyström 近似。随后，我们可以定义 NYSTRÖM++ 估计量的留一法版本 [50]:

$$\widehat{\mathrm{tr}}_{\mathrm{Nys}^{++}} := \mathrm{tr}(\widehat{\mathbf{A}}_{-1}) + \boldsymbol{\omega}_1^{\top}(\mathbf{A} - \widehat{\mathbf{A}}_{-1})\boldsymbol{\omega}_1. \tag{2.4}$$

与单独使用低秩近似或 Girard-Hutchinson 估计量相比，这种方法通常更为均衡；还可以根据矩阵的谱性质，有策略地分配用于两者的矩阵-向量乘法次数。

Nyström 近似还可用于估计矩阵函数及其迹。具体而言，funNyström矩阵近似 $f(\mathrm{A}) \approx f(\mathrm{A}^{\mathrm{nys}})$ [51]通过将矩阵函数作用于 A 的随机 Nyström 近似，来估计 $f(\mathrm{A})$。利用算法 2.1 中得到的特征分解 $\widehat{\mathrm{A}}_{\mathrm{nys}} = \widehat{\mathrm{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathrm{U}}^{\top}$，可以按如下方式高效计算：$f(\widehat{\mathrm{A}}^{\mathrm{nys}}) = \widehat{\mathrm{U}} f(\widehat{\mathbf{\Lambda}})\widehat{\mathrm{U}}^{\top}$。请注意，funNyström 近似完全避免了对 $f(\mathrm{A})$ 进行矩阵-向量乘法，只需要对 A 进行矩阵-向量乘法。该方法还可用于估计迹，$\mathrm{tr}(f(\mathrm{A})) \approx \widehat{\mathrm{tr}}_{\mathrm{FN}} := \mathrm{tr}(f(\widehat{\mathrm{A}}^{\mathrm{nys}}))$ [51, 54]。我们将这种估计量称为 FUNNYS，它对于近似低秩的矩阵 A 尤其有效。与 (2.4) 类似，该方法也可与使用一个样本向量的蒙特卡洛估计量结合，从而得到 FUNNYSTRÖM++ 估计量 [51]:

$$\widehat{\mathrm{tr}}_{\mathrm{funNys}++} := \mathrm{tr}(f(\widehat{\mathbf{A}}_{-1})) + \boldsymbol{\omega}_1^{\top}(f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{-1}))\boldsymbol{\omega}_1. \tag{2.5}$$

尽管 FUNNYSTRÖM++ 估计量是无偏的，但与 FUNNYS不同，它需要对 $f(\mathbf{A})$进行矩阵-向量乘法，而这可能无法实现。我们指出，NYSTRÖM++ 和 FUNNYSTRÖM++ 估计量均与 $(\boldsymbol{\omega}_2,\ldots,\boldsymbol{\omega}_k)$的顺序无关，因为 $\mathbf{A}_{-1}$ 在这种顺序变化下保持不变（见引理 A.1(c)）。

2.5. 可交换迹估计

尽管 NYSTRÖM++ 迹估计算法有效，但它缺少一种称为可交换性的性质，即相对于随机向量的置换不变性 $\{\boldsymbol{\omega}_i\}_{i=1}^k$。我们在定义 2.1中对可交换性作出精确定义。

定义 2.1 设 $Q\colon (\mathbb{R}^n)^k \to \mathbb{R}$ 是一个可测函数。若对于每个置换 $\sigma \in S_k$ 以及每个 $(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_k) \in (\mathbb{R}^n)^k$，$Q$ 均满足以下条件，则称其为可交换的：

$$Q(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_k) = Q(\boldsymbol{\omega}_{\sigma(1)},\ldots,\boldsymbol{\omega}_{\sigma(k)}),$$

这里，S_k 表示 k 阶对称群。

可交换性的重要性可追溯至可交换性原理 [29], ，该原理指出，最小方差无偏估计量必定是可交换的。本文 [22]将这一思想应用于随机化线性代数的背景中。在命题 2.2中，我们基于这一洞见进一步证明，对随机向量的各种排列下的估计量取平均，必定能得到均方误差（MSE）更低的估计量[1]。这是贯穿全文、启发我们研究方法的关键结果。

命题 2.2 设 $Q(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_k)$ 是待研究量 Q 的一个估计量，它由具有有限二阶矩的独立同分布随机向量 $\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_k$ 得到。考虑对称化估计量，

$$\overline{Q}(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_k) := \frac{1}{|S_k|}\sum_{\sigma\in S_k} \widehat{Q}(\boldsymbol{\omega}_{\sigma(1)},\ldots,\boldsymbol{\omega}_{\sigma(k)}), \tag{2.6}$$

其中 S_k 表示 k 阶对称群，其阶为 $|S_k| = k!$。那么，Q 是一个可交换估计量，与 Q 具有相同的偏差，但方差更低。具体而言，

$$(i)\ \mathbb{E}[\overline{Q}] = \mathbb{E}[\widehat{Q}];\quad (ii)\ \mathrm{Var}[\overline{Q}] \le \mathrm{Var}[\widehat{Q}];\quad \textit{and}\quad (iii)\ \mathrm{MSE}[\overline{Q}] \le \mathrm{MSE}[\widehat{Q}].$$

证明见第 5 节。□

我们注意到，命题 2.2 可解释为Rao-Blackwell定理[15, 定理 7.3。17], 的一个特例，其中随机向量的多重集是充分统计量；类似的无偏估计量结果也见于 [53, 第 320] 页。我们强调，与可交换性原理不同，命题 2.2适用于无偏和有偏估计量。近期的工作 [22]利用可交换性的思想，通过对.....进行对称化，开发了迹估计技术 XNYSTRACE,

[1] 参数 θ 的估计量 $\hat{\theta}$ 的均方误差定义为 $\mathrm{MSE}[\hat{\theta}] := \mathbb{E}(\theta-\hat{\theta})^2$。

NYSTRÖM++ 估计量 (2.4)。由于 NYSTRÖM++ 估计量本身已对($\boldsymbol{\omega}_2$、……、$\boldsymbol{\omega}_k$)的排列保持不变，正如第 2.4 节所述，因此对称化可简洁地表示为：

$$\widehat{\mathrm{tr}}_{\mathrm{XN}} := \frac{1}{k}\sum_{i=1}^{k}\Big[\mathrm{tr}(\widehat{\mathbf{A}}_{-i}) + \boldsymbol{\omega}_i^\top(\mathbf{A} - \widehat{\mathbf{A}}_{-i})\boldsymbol{\omega}_i\Big]. \tag{2.7}$$

虽然 XNYSTRACE 及类似方法能够高效估计矩阵迹，但若不计算与 $f(\mathbf{A})$ 的矩阵向量乘法，这些方法并不容易扩展到估计矩阵函数的迹。

3. 使用 FLEXTRACE 进行高效迹估计

本节将展示可交换方法在估计矩阵函数的迹方面的潜力。在第 3.1 节中，我们介绍两种针对矩阵算子单调函数的新型随机化迹估计方法：i–FLEXTRACE 和 FLEXTRACE。随后，我们将在第 3.2 节对这些方法进行理论分析。最后，我们将在LEXTRACE 的第 3.3 节中提供一种高效且稳定的计算实现。

3.1. 方法概述

我们首先介绍一个理想化的估计量 i-FLEXTRACE，它是后续内容的基础，并作为理论工具指导我们对所提出算法 FLEXTRACE 的分析。i-FLEXTRACE 通过对 FUNNYSTRÖM++ 估计量进行对称化来定义。这种对称化与 NYSTRÖM++的对称化类似，借助 FUNNYSTRÖM++在 ($\boldsymbol{\omega}_2$、……、$\boldsymbol{\omega}_k$) 的置换下的不变性得以简化。也就是说，对于相互独立的随机向量 $\boldsymbol{\omega}_1$、……、$\boldsymbol{\omega}_k \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{I})$，i-FLEXTRACE 估计器可表示为：

$$\widehat{\mathrm{tr}}_{\mathrm{iFT}} := \frac{1}{k}\sum_{i=1}^{k}\Big(\mathrm{tr}\, f(\widehat{\mathbf{A}}_{-i}) + \boldsymbol{\omega}_i^\top(f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{-i}))\boldsymbol{\omega}_i\Big). \tag{3.1}$$

由于 i-FLEXTRACE 是对 FUNNYSTRÖM++ 的对称化，i-FLEXTRACE 根据 FUNNYSTRÖM++ 和命题 2.2，是一个可交换的无偏估计量，且方差更低。然而请注意，(3.1) 是一个理想化的估计量，因为计算二次型 $\boldsymbol{\omega}_i^\top f(\mathbf{A})\boldsymbol{\omega}_i$ 需要对$f(\mathbf{A})$ 进行 $1 \le i \le k$次矩阵-向量乘法，而我们希望避免这样做。

为消除这一瓶颈，我们提出了 FLEXTRACE（算法 3.1），它将理想化算法中的 $f(\mathbf{A})$ 替换为 $f(\mathbf{A}^{\mathrm{nys}})$。因此，对于独立随机向量 $\boldsymbol{\omega}_1$，…，$\boldsymbol{\omega}_k \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{I})$，估计量 FLEXTRACE 定义如下：

$$\widehat{\mathrm{tr}}_{\mathrm{FT}} := \frac{1}{k}\sum_{i=1}^{k}\Big(\mathrm{tr}\, f(\widehat{\mathbf{A}}_{-i}) + \boldsymbol{\omega}_i^\top(f(\widehat{\mathbf{A}}_{\mathrm{nys}}) - f(\widehat{\mathbf{A}}_{-i}))\boldsymbol{\omega}_i\Big). \tag{3.2}$$

与 i-FLEXTRACE 类似，FLEXTRACE 也是一种可交换估计量；参见定理 3.1。我们在算法 3.1 中概述了 FLEXTRACE 的相应计算过程。在 f 为线性的特殊情况下，$\boldsymbol{\omega}_i^\top f(\mathbf{A})\boldsymbol{\omega}_i = \boldsymbol{\omega}_i^\top f(\widehat{\mathbf{A}}^{\mathrm{nys}})\boldsymbol{\omega}_i$ 对于 $1 \le i \le k$ 成立，并且两个估计量 i-FLEXTRACE（$\widehat{\mathrm{tr}}_{\mathrm{FT}}$）和 FLEXTRACE（$\widehat{\mathrm{tr}}_{\mathrm{iFT}}$）均与 XNYSTRACE 估计量（$\widehat{\mathrm{tr}}_{\mathrm{XN}}$，见（2.7））相同。此外，如果 f 在 A 的谱上可以用二次函数很好地近似，则根据引理 A.1(e)，$\boldsymbol{\omega}_i^\top f(\mathbf{A})\boldsymbol{\omega}_i$ 可以用 $\boldsymbol{\omega}_i^\top f(\widehat{\mathbf{A}}^{\mathrm{nys}})\boldsymbol{\omega}_i$ 很好地近似，其中 $1 \le i \le k$。

算法 3.1 FLEXTRACE($\mathbf{A}$ k)

输入： 半正定矩阵 $\mathbf{A} \in \mathbb{R}^{n\times n}$，矩阵向量乘积预算 k

输出： 迹近似 $\widehat{\mathrm{tr}}_{\mathrm{FT}}$

1：生成标准高斯随机矩阵 $\boldsymbol{\Omega} \in \mathbb{R}^{n\times k}$

2：构造 $\mathbf{Y} = \mathbf{A}\boldsymbol{\Omega}$

3：计算 $\widehat{\mathbf{A}}^{\mathrm{nys}} = \mathbf{Y}(\boldsymbol{\Omega}^\top\mathbf{Y})^\dagger\mathbf{Y}^\top$

4: **对于** $i = 1, \ldots, k$ **do**

5: 定义 $\widehat{\mathbf{A}}_{-i} := \mathbf{Y}_{-i}(\boldsymbol{\Omega}_{-i}^\top\mathbf{Y}_{-i})^\dagger\mathbf{Y}_{-i}^\top$ ▷ 移除$\boldsymbol{\Omega}$ 和$\mathbf{Y}$的第 - 列

6: 计算 $\widehat{\mathrm{tr}}_i := \mathrm{tr}\ f(\widehat{\mathbf{A}}_{-i}) + \boldsymbol{\omega}_i^\top(f(\widehat{\mathbf{A}}^{\mathrm{nys}}) - f(\widehat{\mathbf{A}}_{-i}))\boldsymbol{\omega}_i$

7: **结束循环**

8: 计算 $\widehat{\mathrm{tr}}_{\mathrm{FT}} := \frac{1}{k}\sum_{i=1}^{k}\widehat{\mathrm{tr}}_i$

定理 3.1 该 FLEXTRACE 估计器 *(3.2)* 是一种可交换估计量，其含义是定义 *2.1*。

证明参见第 5.2 节。 □

由于 FLEXTRACE 不需要进行 $f(\mathbf{A})$的矩阵-向量乘法，因此提供了一种高效且实用的方法来估计 $\mathrm{tr}(f(\mathbf{A}))$。然而，该方法的朴素实现可能计算效率低下，且容易出现数值不稳定性。我们在LEXTRACE 的第 3.3 节中提供了一种更精细的实现，更适合实际使用。

我们强调该方法还有另外两个特点，使 FLEXTRACE 相较于其他方法更具吸引力。首先，该方法是单遍的，这意味着可以同时访问所有与$\mathbf{A}$进行的矩阵-向量乘法。在无法重复访问$\mathbf{A}$ 或重复访问存在延迟的情况下，这一点很有优势。其次，该方法是函数无关 的，即可同时高效估计多个矩阵函数的 tr($f(\mathbf{A})$)，且额外开销可以忽略不计。如果函数 f_θ 依赖于参数，且需要针对参数 θ的多个取值计算 tr($f_\theta(\mathbf{A})$)，那么这一特性会很有帮助。这类情形在谱密度估计 [43] 和核方法 [19]中很常见。

3.2. 关键理论结果

本节将给出 i-FLEXTRACE 和 FLEXTRACE 估计量在偏差和均方误差方面的理论保证。我们将在第 3.2.1 节和第 3.2.2 节分别分析偏差和均方误差，然后在第 3.2.3 节中对这些界限作渐近解释。我们的目标是根据矩阵$\mathbf{A}$的谱性质以及矩阵函数 f的谱性质，以解析方式量化迹估计器的期望误差。我们采用双管齐下的方法。首先，根据 funNyström 近似的期望误差，对所提出的两种估计量的偏差和均方误差进行结构性界定。其次，根据矩阵$\mathbf{A}$的谱性质，推导该期望误差的上界。值得注意的是，针对后者得到的理论结果适用范围更广，并不限于迹估计任务。本节所有定理的证明见第 5 节。

3.2.1. 偏差分析

首先，我们证明 i-FLEXTRACE 是 $f(\mathbf{A})$ 的迹的无偏估计量。

定理 3.2 设 $\mathbf{A}$ 是一个对称半正定矩阵，且 $f \in \mathcal{F}$。那么，*i*-FLEXTRACE 估计量 *(3.1)* 是 $\operatorname{tr}(f(\mathbf{A}))$的无偏估计量。

证明见第 5.1 节。□

尽管 i-FLEXTRACE 对所有函数都是无偏的，但 FLEXTRACE 通常存在偏差。定理 3.3指出，FLEXTRACE 的偏差有一个上界，该上界由使用 $(k-1)$ 次矩阵-向量乘法的 FUNNYS近似的偏差所界定；而后者又可用$\mathbf{A}$的尾部特征值来界定。具体而言，定理 3.3表明，对于具有低秩或近低秩结构的矩阵，FLEXTRACE 的偏差可以很小。

定理 3.3 设有一个对称半正定矩阵 $\mathbf{A}$ 并令 $f \in \mathcal{F}$。对于 $k \geq 4$，FLEXTRACE *(算法 3.1)* 满足：

$$0 \leq \mathbb{E}\left[\operatorname{tr}(f(\mathbf{A})) - \widehat{\operatorname{tr}}_{\mathrm{FT}}\right] \leq \mathbb{E}\left[\|f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{\text{-}1})\|_*\right] \leq (k-3)\|f(\mathbf{\Lambda}_{k\text{-}2:n})\|_*.$$

若 f 是线性的，则 $\widehat{\text{迹}}_{\mathrm{FT}}$ 对任意 $k \geq 1$ 都是无偏的。

证明 见 第 5.2 节。 □

定理 3.3 为 任意 函数 $f \in \mathcal{F}$的迹估计器提供了保证。不过，我们指出，对于某些已知的特定函数，还可以改进这一结果；例如，[51, 第 3.3 节3]中针对矩阵平方根函数给出的 funNyström 近似的改进分析，可用于获得更紧的界。

3.2.2. 估计量均方误差的分析
i-FLEXTRACE 和 FLEXTRACE 估计量的均方误差界分别以低秩近似的残差范数为依据，在定理 3.4和定理 3.5中给出，其中低秩近似针对 $f(\mathbf{A})$。请注意，均方误差的上界同时也是估计量方差的界限，因为估计量的方差不超过其均方误差。对于无偏的 i-FLEXTRACE，估计量的方差就是其均方误差。

定理 3.4 设 $\mathbf{A}$ 为一个对称半正定矩阵，且 $f \in \mathcal{F}$。则，*i*-FLEXTRACE 估计量 *(3.1)*满足：

$$\operatorname{MSE}\left[\widehat{\operatorname{tr}}_{\mathrm{iFT}}\right] \leq \frac{2}{k}\mathbb{E}\left[\|f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{\text{-}1})\|_F^2\right] + 2\mathbb{E}\left[\|f(\widehat{\mathbf{A}}_{\text{-}1}) - f(\widehat{\mathbf{A}}_{\text{-}\{1,2\}})\|_F^2\right],$$

其中 $\widehat{\mathbf{A}}_{\text{-}\{1,2\}}$ 表示相对于移除了前两列的测试矩阵 $\mathbf{\Omega}$ 对$\mathbf{A}$ 所作的 *Nyström* 近似。

证明参见第 5 节.1。 □

定理 3.5 设有一个对称半正定矩阵 $\mathbf{A}$ 且令 $f \in \mathcal{F}$。则 FLEXTRACE *(算法 3.1)* 估计量满足：

$$\mathrm{MSE}\left[\widehat{\mathrm{tr}}_{\mathrm{FT}}\right] \leq 2\,\mathbb{E}\left[\|f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{k-1})\|_F^2\right] + 2\,\mathbb{E}\left[\|f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{k-1})\|_*^2\right].$$

证明参见第 5.2 节。 □

与第 3.2.1 节中的界类似，定理 3.4 和3.5 分别用秩为FLEXTRACE和 FLEXTRACE 的均方误差来界定 i-F$(k-1)$Nyström 近似的均方误差。特别值得注意的是，FLEXTRACE 的均方误差界与少进行一次矩阵向量乘法的 FUNNYS 迹近似的界相当。定理 3.4 和3.5 通过包含 funNyström 近似在 Frobenius 范数和核范数下均方误差的项，给出了估计量的均方误差界。这促使我们为 funNyström 推导一个一般的酉不变界。 定理 3.6 以$\mathbf{A}$ 的尾部特征值和函数 f给出了该误差在任意酉不变范数下的界。

定理 3.6 设 $\mathbf{A}$ 为一个对称半正定矩阵，并令 $f \in \mathcal{F}$。对于 $k \geq 4$，我们有

$$\mathbb{E}\left[|||f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{\mathrm{nys}})|||^2\right] \leq 2|||f(\mathbf{\Lambda}_{k-3:n})|||^2 + \gamma(\mathbf{\Lambda}_{k-3:n}, k)|||f(\mathbf{\Lambda}_{k-3:n}^{1/2})|||^2,$$

其中， $|||\cdot|||$ 表示任意酉不变范数，且

$$\gamma(\mathbf{D}, k) := \frac{16(k-4)(k-2)}{3}\|\mathbf{D}\|_2 + \frac{16e^4k^2}{125}(\|\mathbf{D}^{1/2}\|_*^2 + 2\|\mathbf{D}\|_*) + 2.$$

证明见附录 A。4 □

请注意，定理 3.6 的意义不止于迹估计。尽管 [51] 给出了 funNyström 在不同范数下的近似误差理论界，定理 3.6的结果因以下原因而具有更广泛的适用性。首先，该界适用于任意酉不变范数，包括核范数和 Frobenius 范数。其次，据我们所知，定理 3.6给出了 funNyström 算法平方范数误差的概率界，且无需像 [51,定理 3.4]中那样假设执行了子空间迭代。

我们根据矩阵函数 FLEXTRACE 和 i-FLEXTRACE 及 f 矩阵 $\mathbf{A}$ 的尾部特征值，对它们的均方误差进行界定，见定理 3.7。均方误差量化了迹估计器的期望平方误差，因此综合考虑其偏差和方差，以评估估计量的总体质量。

定理 3.7 设 **A** 为一个对称半正定矩阵，且 $f \in \mathcal{F}$。估计量 $\widehat{\text{迹}}_{\text{iFT}}$、$\widehat{\text{迹}}_{\text{FT}}$ 和 $\widehat{\text{迹}}_{\text{FN}}$ 的均方误差满足以下界：

$$\text{MSE}[\widehat{\text{tr}}_{\text{iFT}}] \le \frac{k+1}{k}\Big(4\|f(\mathbf{\Lambda}_{k\text{-}5:n})\|_F^2 + 2\gamma(\mathbf{\Lambda}_{k\text{-}5:n}, k-2)\|f(\mathbf{\Lambda}_{k\text{-}5:n}^{1/2})\|_F^2\Big),$$

对于 $k \ge 6$;

$$\text{MSE}[\widehat{\text{tr}}_{\text{FT}}] \le 4\big(\|f(\mathbf{\Lambda}_{k\text{-}4:n})\|_F^2 + \|f(\mathbf{\Lambda}_{k\text{-}4:n})\|_*^2\big) + 2\gamma(\mathbf{\Lambda}_{k\text{-}4:n}, k-1)\Big(\|f(\mathbf{\Lambda}_{k\text{-}4:n}^{1/2})\|_F^2 + \|f(\mathbf{\Lambda}_{k\text{-}4:n}^{1/2})\|_*^2\Big),$$

对于 $k \ge 5$; 且

$$\text{MSE}[\widehat{\text{tr}}_{\text{FN}}] \le 2\|f(\mathbf{\Lambda}_{k\text{-}3:n})\|_*^2 + \gamma(\mathbf{\Lambda}_{k\text{-}3:n}, k)\|f(\mathbf{\Lambda}_{k\text{-}3:n}^{1/2})\|_*^2,$$

for $k \ge 4$,

其中 γ 如定理 *3.6*中所定义。

证明参见第 5.2 节。□

注意到，定理 3.7 中两个估计量的均方根误差（RMSE）上界均由因子 $k\|\mathbf{\Lambda}_{k\text{-}5}^{1/2} : n\|_*\|f($ $\mathbf{\Lambda}_{k\text{-}5}^{1/2} : n)\|_*$ 主导。这表明，正如预期，所提出的算法对于特征值衰减较快的矩阵表现更好。

3.2.3. 理论界的渐近行为

为了进一步深入了解，我们重点展示 FLEXTRACE（定理 3.3 和 3.7）中偏差和均方误差的渐近行为，见 表 2 ，其中矩阵的特征值呈指数衰减， $\lambda_i \le \alpha^i$ 对于$1 \le i \le n$，其中 $0 < \alpha < 1$，并考察矩阵-向量乘法次数 k。对于所分析的所有矩阵函数，我们都观察到偏差和方差呈指数衰减。例如，对于函数 $\log(1+x)$，每增加一次矩阵向量乘法，期望偏差和均方根误差就会渐近地缩小 α 倍。

矩阵函数（f）	偏差界	均方误差界
x	0	$\mathcal{O}(k^2\alpha^{2k})$
对数$(1+x)$	$\mathcal{O}(k\alpha^k)$	$\mathcal{O}(k^2\alpha^{2k})$
$\frac{x}{x+\zeta}$	$\mathcal{O}(k\alpha^k)$	$\mathcal{O}(k^2\alpha^{2k})$
$\sqrt{x}$	$\mathcal{O}(k\alpha^{k/2})$	$\mathcal{O}(\alpha^k)$

表格 2 FLEXTRACE估计量的渐近行为，其中 $0 < \alpha < 1$。

我们指出，对于 $f(x) = x$ 的情况，我们的方差界呈现出与 [22] 推导的 XNYSTRACE 迹估计界相同的渐近衰减。对于表 2中的其他函数，FUNNYS 和 i-FLEXTRACE 也渐近满足其中给出的偏差界和均方误差界。

3.3. *FLEXTRACE* 的高效稳定实现FLEXTRACE

尽管隐式迹估计方法的主要开销通常由与$\mathbf{A}$进行矩阵-向量乘法所主导，但在某些应用中，后续矩阵计算的计算可扩展性也不容忽视。本节将探讨提升计算效率的技术，并提供 FLEXTRACE 更快速且数值稳定的实现。

算法 3.1 的一个关键瓶颈是需要计算 $f(\mathbf{A}_{\text{-}i})$ k 次。若直接进行这些运算，其额外开销最高可达 $\mathcal{O}(nk^3+k^4)$ 次浮点运算（flops），这还不包括 Nyström 近似本身。为避免这种情况，我们利用以下事实：

(a) 若 $\text{rank}(\mathbf{A}^{\text{nys}}) < k$，则对于随机的$\mathbf{A} = \mathbf{A}^{\text{nys}}$，几乎必然有此结论；其中随机矩阵为 $\boldsymbol{\Omega}$ [51]。

(b) 若 $\text{rank}(\mathbf{A}^{\text{nys}}) = k$，则

$$\widehat{\mathbf{A}}_{\text{-}i} = \widehat{\mathbf{A}}_{\text{nys}} - \mathbf{Y}\left(\frac{(\boldsymbol{\Omega}^\top\mathbf{Y})^{-1}\boldsymbol{e}_i\boldsymbol{e}_i^\top(\boldsymbol{\Omega}^\top\mathbf{Y})^{-1}}{\boldsymbol{e}_i^\top(\boldsymbol{\Omega}^\top\mathbf{Y})^{-1}\boldsymbol{e}_i}\right)\mathbf{Y}^\top, \tag{3.3}$$

其中 $\mathbf{Y} = \mathbf{A}\boldsymbol{\Omega}$，根据定理 $1 \le i \le k$ [21, 14.2]。

根据性质 (a)，若 $\text{rank}(A^{\text{nys}}) < k$，则 $\text{tr}[f(A^{\text{nys}})]$ 几乎必然是精确的，此时无需进行额外的计算。若 $\text{rank}(A^{\text{nys}}) = k$，则我们利用性质 (b)。具体而言，采用算法 2.1 中的记号，我们将 $A_{\text{nys}} = U\widehat{\Lambda}U^\top$、$(\Omega^\top Y)^{-1} = C^{-1}C^{-\top}$，以及 $Y = QR = Q(\widetilde{U}\widetilde{U}^\top)R = \widetilde{U}\widetilde{U}^\top R$ 表示出来。将这些量代入 (3.3)，得到

$$\widehat{\mathbf{A}}_{\text{-}i} = \widehat{\mathbf{U}}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{b}_i\boldsymbol{b}_i^\top)\widehat{\mathbf{U}}^\top, \quad \text{where } \boldsymbol{b}_i := \frac{\widetilde{\mathbf{U}}^\top\mathbf{R}\mathbf{C}^{-\top}\boldsymbol{e}_i}{\|\mathbf{C}^{-\top}\boldsymbol{e}_i\|_2}. \tag{3.4}$$

因此，我们有 $f(\widehat{\mathbf{A}}_{\text{-}i}) = \widehat{\mathbf{U}}f(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{b}_i\boldsymbol{b}_i^\top)\widehat{\mathbf{U}}^\top$，适用于 $1 \le i \le k$。一个关键观察是，要使用 $f(\widehat{\mathbf{A}}_{\text{-}i})$，我们只需计算 f 对 $\widehat{\boldsymbol{\Lambda}} - \boldsymbol{b}_i\boldsymbol{b}_i^\top$ 的作用；后者是一个 $k \times k$ 对角矩阵加秩一（DPR1）矩阵。我们采用高效技术计算 DPR1 矩阵的特征分解，其浮点运算次数为 $\mathcal{O}(k^2)$，相比之下，稠密矩阵的标准浮点运算次数为 $\mathcal{O}(k^3)$。一旦获得 $1 \le i \le k$ 的特征分解 $\widehat{\boldsymbol{\Lambda}} - \boldsymbol{b}_i\boldsymbol{b}_i^\top = \mathbf{V}_i\widehat{\boldsymbol{\Lambda}}_i\mathbf{V}_i^\top$，FLEXTRACE 的其余步骤便可自然得出：

$$\begin{aligned}\widehat{\text{tr}}_{FT} &= \frac{1}{k}\sum_{i=1}^{k}\text{tr}[f(\widehat{\mathbf{A}}_{\text{-}i})] + \boldsymbol{\omega}_i^\top(f(\widehat{\mathbf{A}}_{\text{nys}}) - f(\widehat{\mathbf{A}}_{\text{-}i}))\boldsymbol{\omega}_i \\ &= \frac{1}{k}\sum_{i=1}^{k}\text{tr}[f(\widehat{\boldsymbol{\Lambda}}_i)] + \boldsymbol{x}_i^\top(f(\widehat{\boldsymbol{\Lambda}}_i) - \mathbf{V}_i f(\widehat{\boldsymbol{\Lambda}}_i)\mathbf{V}_i^\top)\boldsymbol{x}_i,\end{aligned}$$

其中，对于 $1 \le i \le k$，$\boldsymbol{x}_i := \widehat{\mathbf{U}}^\top\boldsymbol{\omega}_i$。请注意，该过程完全避免了显式计算 $1 \le i \le k$ 的$f(\mathbf{A}_{\text{-}i})$。

因此，我们给出加速版 FLEXTRACE 算法，见 算法 3.2。关于 算法 3.2，有几点需要说明：

• 第 3 行：在有限精度算术中，计算 rank(Anys) 时必须谨慎。在我们的实现中，我们采用 NYSSVD（算法 2.1）中通过截断主元 Cholesky 分解得到的数值秩 $\hat{r}$ 。该数值秩也更为保守，因为当 $\Omega^\top Y$ 病态时，它可避免额外计算。• 第 7 行：注意，通过代入 $\widehat{U} = Q\widetilde{U} = YR^{-1}\widetilde{U}$ 和 $Z = RC^{-1}$，$X = \widehat{U}^\top\Omega$ 可简化为 $X = \widetilde{U}^\top Z^{-\top}C$。

• 第 9 行：DPR1 特征值分解可在 $\mathcal{O}(k^2)$ 浮点运算次数内完成；例如，参见 [37, 算法 1] 或 LAPACK 例程 dlaed9 [3]。

算法 3.2 FLEXTRACE 的高效版本（$\mathbf{A}$ k）

输入： 对称半正定矩阵 $\mathbf{A} \in \mathbb{R}^{n\times n}$，矩阵向量乘积预算 k, $f \in \mathcal{F}$

输出：迹近似 $\mathrm{tr}_{\mathrm{FT}}$1：生成矩阵 $\Omega \in \mathbb{R}^{n\times k}$，其元素为独立同分布的标准正态分布2：获取 $[\Lambda$,U, C, U, Z, $\hat{r}]$ = NYSSVD(A,Ω) 3：如果 $\hat{r} < k$ 则 4：返回 $\widehat{\mathrm{tr}}_{\mathrm{FT}} = \mathrm{tr}\left[f(\widehat{\Lambda})\right]$5：结束 if 6：按列定义 $S \in \mathbb{R}^{k\times k}$ 为 $s_i = C^{-\top}e_i/\|C^{-\top}e_i\|_2$，其中 $1 \le i \le k$ 7：计算 $B = \widetilde{U}^\top ZS$ 和 $X = \widetilde{U}^\top Z^{-\top}C$ 8：对于 $i = 1,\ldots,k$，执行 9：计算特征分解，$\widehat{\Lambda} - b_i b_i^\top = V_i\widehat{\Lambda}_i V_i^\top$10：计算 $\widehat{\mathrm{tr}}_i := \mathrm{tr}\left[f(\widehat{\Lambda}_i)\right] + x_i^\top f(\widehat{\Lambda})x_i - x_i^\top V_i f(\widehat{\Lambda}_i)V_i^\top x_i$11：结束 for 12：计算 $\widehat{\mathrm{tr}}_{\mathrm{FT}} := \frac{1}{k}\sum_{i=1}^{k}\widehat{\mathrm{tr}}_i$

–

我们比较了 F 的朴素实现和加速实现的计算成本LEXTRACE，见表 3。还值得注意的是，加速实现（具体而言，第

方法	与 A 的矩阵-向量乘法次数	计算 $f(\widehat{\mathbf{A}}_{\mathrm{nys}})$ 的成本	额外成本
算法 3.1	k	$\mathcal{O}(k^2(n+k))$	$\mathcal{O}(nk^3 + k^4)$
算法 3.2	k	$\mathcal{O}(k^2(n+k))$	$\mathcal{O}(k^3)$

表 3 FLEXTRACE 使用算法*3.1* 和*3.2* 的计算成本（浮点运算次数）

第6-12行（算法 3.2）允许以三种不同方式进行并行化。首先，矩阵草图可针对不同的随机向量并行计算。其次，tr_i 可以并行计算

对于 $1 \leq i \leq k$。第三，DPR1结构（第9行，算法 3.2）允许独立计算各个特征对，从而实现进一步的循环内并行化并节省存储空间。加速实现算法 3.2在实践中同样数值稳定，因为它避免了显式计算（$\mathbf{\Omega}^\top \mathbf{Y}$）$^{-1}$。

4. 数值结果

在本节中，我们通过若干数值实验展示 FLEXTRACE 的性能。在第 4.1 节中，我们将 FLEXTRACE 应用于一组合成矩阵，以分析其有效性。随后，我们在LEXTRACE 第 4.2 节中将 F 应用于核范数估计，在第 4.2 节中讨论贝叶斯反问题，在第 4.3 节中讨论核方法，并在第 4.4 节中作总结。

4.1. 合成矩阵上的应用

在本节中，我们将 FLEXTRACE （算法 3.2）与现有方法在尺寸为 1000×1000 的合成矩阵上进行比较。为此，考虑如下形式的矩阵：$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$，其特征值选取如下（对于 $1 \leq i \leq 1000$）：

- Flat: $\lambda_i = 3 - 2\frac{i-1}{n-1}$
- Poly: $\lambda_i = i^{-2}$
- 指数: $\lambda_i = 0.9^{i-1}$
- 步长: $\lambda_i = \{1: \ i \leq 50, \ \ 10^{-3}: \ i \geq 51\}$

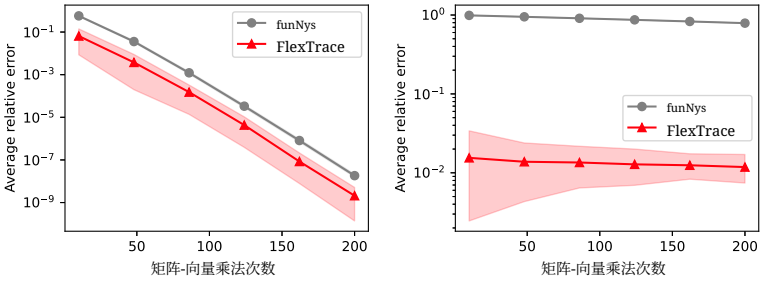
在第 4.1.1 节中，我们比较了 FLEXTRACE 与 FUNNYS 在这些代表性矩阵上的性能。随后，我们在LEXTRACE 中将其与多遍迹估计技术进行比较，详见4.1.2。除非另有说明，本节所示的所有误差均为100次独立试验的平均值，阴影区域表示90% 百分位间距。

4.1.1. 与 FUNNYS 的比较

本节中，我们比较所提出的 FLEXTRACE 估计器与 FUNNYS 方法在合成矩阵上的性能。FUNNYS 可作为所提出方法的有用基线，因为两种方法都是用于 tr(f(**A**)) 的单遍估计器。我们在这些合成矩阵上进行了两项实验。首先，固定矩阵函数，比较迹估计器在上述矩阵上的准确性。其次，对于选定的矩阵，比较不同矩阵函数 $f \in \mathcal{F}$ 对应的估计器。

在图1中，我们展示了 FLEXTRACE 与 FUNNYS 在估计 tr(f(**A**)) 时的相对误差，其中 f(**A**) $:=$log(**I** $+$**A**)，对象为具有不同谱分布的矩阵。在全部四种谱分布中，我们观察到，就平均相对误差而言，FLEXTRACE 始终优于基线方法 FUNNYS 约一到两个数量级。在 Exp 情形（图1a）中，两种方法在对数-线性尺度上均呈现线性收敛，与第 3.2.3 节中的理论相符。不过，仍值得注意的是，FLEXTRACE 始终保持相对于 FUNNYS 的稳定优势。FLEXTRACE 相对于 FUNNYS 的性能优势在 Step 矩阵和多项式矩阵中最为明显，此时 FLEXTRACE 的表现优于 FUNNYS 数个数量级。这表明，与 FUNNYS 相比，FLEXTRACE 更善于捕捉较小的尾部特征值（长尾）。对于平坦矩阵，其特征值衰减极小，两种方法在增加矩阵-向量乘法后性能几乎都没有提升。尽管如此，FLEXTRACE 的相对误差始终低于 FUNNYS。

图1 还展示了 FUNNYS 和 FLEXTRACE 的不确定性区域（以阴影表示）。由于纵轴采用对数刻度，FUNNYS 的阴影区域看起来比 FLEXTRACE 的小得多。然而，除图1b外，在所展示的所有合成示例中，FLEXTRACE 和 FUNNYS 的百分位间区域大小处于同一量级。例如，在图1d中，在整个矩阵向量乘积预算范围内，FLEXTRACE 的百分位间区域平均比 FUNNYS 小 15%。然而，对于 Flat，当$\mathbf{A}$的谱为.....时，我们确实观察到 FLEXTRACE 的方差高于 FUNNYS。这与我们的理论预期一致；例如，如果$\mathbf{A} = \mathbf{I}$且 $f(x) = x$，FUNNYS 总是返回不准确的估计 $\widehat{\mathrm{tr}}_{FN} = k$，且方差为零，而 FLEXTRACE 则会产生迹的无偏估计，尽管方差较高。这些结果通过实验证实，在各种谱区间内，算法 3.2 能够比 FUNNYS 更有效地捕获迹信息。

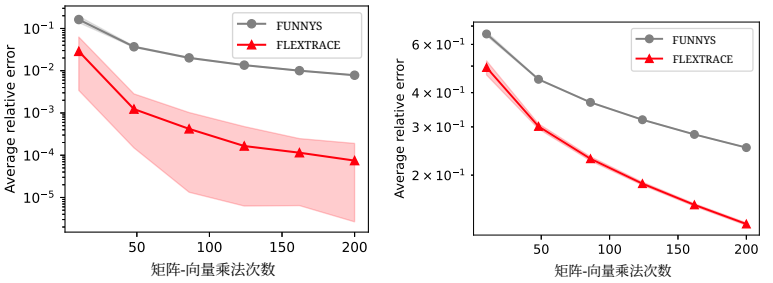


(a) Exp 矩阵 (b) 平坦矩阵

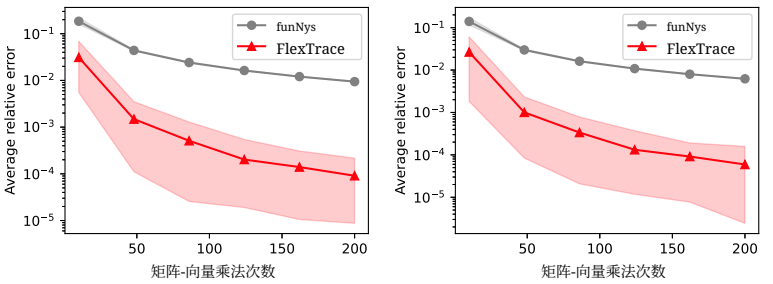
(c) 阶跃矩阵 (d) Poly 矩阵

图1. F 的演示LEXTRACE (算法 3.2) 在合成测试矩阵上的应用。

在图2中，我们在固定矩阵 Poly 上、但针对若干函数 $x,\ x/(1+x)$,$\log(1+x)$ 和 $\sqrt{x}$，比较了 FLEXTRACE 与 FUNNYS 的性能。我们在图2 中观察到，FLEXTRACE 在各种算子单调矩阵函数上始终优于 FUNNYS。对于函数 $x,\ x/(1+x)$, 和 $\log(1+x)$，我们观察到 FLEXTRACE 的相对误差比 FUNNYS 低一个数量级以上，且前者仅用 200 次矩阵-向量乘法便达到 10^{-4} 相对精度。另一方面，我们注意到两种方法对 $f(x) = \sqrt{x}$ 的表现都很差，因为尽管$\mathbf{A}$本身具有快速的谱衰减，$f(\mathbf{A}) = \mathbf{A}^{1/2}$却并非如此。相比之下，对于另外三个函数，$f(\mathbf{A})$ 的谱衰减与$\mathbf{A}$的谱衰减相当。



(a) $f(x) = \text{对数}(1+x)$ (b) $f(x) = \sqrt{x}$



(c) $f(x) = x/(x+1)$ (d) $f(x) = x$

图2。F 的演示LEXTRACE (算法 3.2) 在 Poly矩阵上。

4.1.2. 与多遍技术的比较

尽管 FLEXTRACE 被设计为一种单遍方法，有必要将 FLEXTRACE（算法 3.2）与多遍方法进行比较，以获得更多洞见。具体而言，我们考虑以下用于矩阵函数迹估计的多遍方法：

（i）FUNNYSTRÖM++: [51] 提出的一种用于估计 tr(f(**A**)) 的混合方法。该方法使用$\frown f$(**A**^{nys}) 来近似 f(**A**)，并使用 Hutchinson 方法估计残差矩阵的迹。（ii）SLQ：随机 Lanczos 求积，由 [56], 提出，将高斯求积与 Hutchinson 迹估计相结合。（iii）KA-STE：Krylov 感知随机迹估计，由 [16], 提出，利用 Krylov 子空间的性质，以更高效率计算矩阵函数的迹。

请注意，方法 (i)-(iii) 涉及使用 Lanczos 方法近似 $f(\mathbf{A})$ 的矩阵向量乘法。与 FLEXTRACE不同，这些技术需要以串行或分块串行的方式应用**A** ，并且通常需要进行计算开销较大的操作（例如重新正交化）以确保数值稳定性。由于方法 (i)-(iii) 需要指定若干额外的超参数（块大小、Lanczos 迭代次数、矩阵向量乘积的分配等），因此比较 FLEXTRACE 与上述方法也更加复杂。我们尽可能遵循相关论文 FUNNYSTRÖM++和 SLQ 中建议的超参数选择策略 [16, 51, 56]，力求进行公平比较。我们在表 4中概述了这些选择，其中 k 表示矩阵向量乘积预算。请注意，我们没有在 FUNNYSTRÖM 中进行额外的子空间迭代

FUNNYSTRÖM++。对于 KA-STE，如果矩阵向量乘积预算不足以容纳尝试执行的 Lanczos 迭代次数，则会根据需要减少迭代次数以满足预算。

方法	**超参数设置**（矩阵向量乘积预算为 k）	**Ref.**
FLEXTRACE	—	
FUNNYS	—	
SLQ	Lanczos 迭代：5	[56]
FUNNYSTRÖM++	Nyström 秩：$k/2$；Lanczos 迭代：5	[51]
KA-STE (1)	块大小：2；Krylov 深度：$k/4$；Lanczos 迭代：10	[16]
KA-STE (2)	块大小：4；Krylov 深度：$k/8$；Lanczos 迭代：5	[16]

表 4 图 *3* 中使用的超参数选择图 *3*

在图 3中，我们绘制了使用上述方法以及 FUNNYS 和 FLEXTRACE 估计 $\mathrm{tr}(f(\mathbf{A}))$ 时的估计误差。具体而言，我们关注两个具体情形：Exp 矩阵，其中 $f(x) = \log(1+x)$，以及平坦矩阵，其中 $f(x) = \sqrt{x}$。如第 4.1.1 节所述，这两种情形分别对应 $f(\mathbf{A})$ 谱衰减较快和较慢的情况，因此被选作代表性示例。为使图示清晰，我们未在图 3 中显示百分位间区域。从图 3（左）可以看出，对于 Exp 矩阵，FLEXTRACE 的相对误差在所有对比方法中最低。另一方面，对于平坦矩阵，我们发现多遍方法，尤其是 SLQ，通常比 FLEXTRACE 得到更准确的估计。这些结果表明，在可以应用多遍方法的场景中，FLEXTRACE 仍然具有应用价值，并且对于谱衰减足够快的矩阵，其表现可与多遍迹估计技术相媲美。

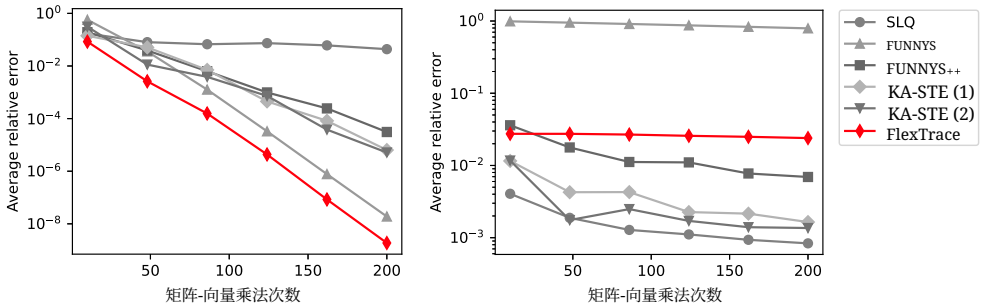


图 3. FLEXTRACE 与现有方法在 $f(x) = \log(1+x)$（使用 Exp，左）和 $f(x) = \sqrt{x}$（使用 Poly，右）上的比较。

4.2. 在核范数估计中的应用

在本节中，我们展示 FLEXTRACE 在估计核范数方面的有效性。核范数在矩阵补全中发挥着关键作用——矩阵补全旨在估计部分观测矩阵中缺失的条目。具体而言，矩阵补全最常见的方法之一是

核范数最小化 [14, 46, 48], , 即通过补全缺失条目，使矩阵的核范数最小化。矩阵补全的一个著名应用是 *Netflix* 问题 [8],, 即利用用户对电影的评分生成有用的推荐。我们采用 `MovieLens100k` 数据集 [31], , 该数据集是推荐系统的基准数据集，包含 943 位用户对 1682 部电影的 100,000 条评分（1 到 5 分）。我们令 $\mathbf{X} \in \mathbb{R}^{943\times 1682}$ 表示评分矩阵，其中行对应用户，列对应电影，缺失条目以零填充。注意，$\mathbf{X}$ 是一个稀疏矩阵，其中约 94% 的条目为零。我们的目标是估计 $\mathbf{X}$ 的核范数，即 $\|\mathbf{X}\|_* = \mathrm{tr}((\mathbf{X}\mathbf{X}^T)^{1/2})$。

核范数的计算通常需要计算奇异值分解（SVD），对于大型矩阵而言，这在计算上可能代价高昂。另一种方法是使用随机化奇异值分解算法 [28], , 该方法可利用与 X 和 $\mathrm{X}^\top$ 的矩阵-向量乘法来估计 X 的前导奇异值。估计奇异值之和可用于近似 X 的核范数。我们将该方法记为 RANDSVD。另一方面，我们可以利用诸如 FLEXTRACE 或 FUNNYS 等迹估计方法来估计 $\|\mathrm{X}\|_* = \mathrm{tr}(\underline{f}(\mathrm{A}))$，其中 $f(x) = \sqrt{x}$ 且 $\mathrm{A} = \mathrm{X}\mathrm{X}^\top$。我们在图 4 中比较这些方法。具体而言，我们绘制了各方法所得相对误差随矩阵-向量乘法总次数（同时涉及 X 与 $\mathrm{X}^\top$）的变化。

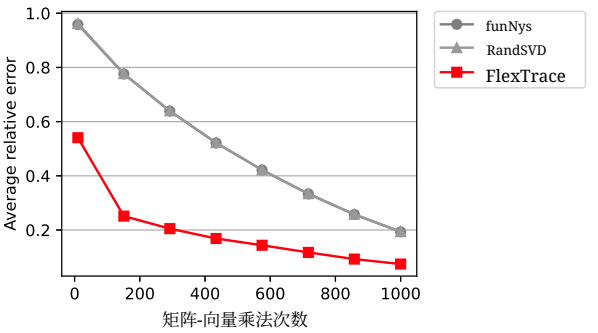


图4. FLEXTRACE、FUNNYS 和 RANDSVD 对$\mathbf{X}$进行核范数估计的比较

如图4所示，FLEXTRACE的效率优于FUNNYS和RANDSVD。需要注意的是，FUNNYS与RANDSVD在图4 中几乎无法区分，因为这两种方法都估计$\mathbf{X}\mathbf{X}^\top$ 的前 k 个特征值，因而表现相似。另一方面，FLEXTRACE仅用300次矩阵-向量乘法即可获得与RANDSVD使用1000次矩阵-向量乘法几乎相同的精度。这一结果表明，FLEXTRACE可以成为实际应用中高效核范数估计的有力工具。

4.3. 贝叶斯反问题的应用

在本节中，我们展示FLEXTRACE在贝叶斯反演背景下的适用性。为此，我们考虑如下对流扩散方程，

$$\begin{aligned} \frac{\partial u}{\partial t} - \kappa \Delta u + \boldsymbol{v} \cdot \nabla u &= f && \text{in } \mathcal{D} \times (0,T), \\ u(0,\boldsymbol{x}) &= m(\boldsymbol{x}) && \text{in } \mathcal{D}, \\ \nabla u \cdot \boldsymbol{n} &= 0 && \text{on } \partial\mathcal{D} \times (0,T), \end{aligned} \tag{4.1}$$

其中， $\mathcal{D} \subseteq \mathbb{R}^3$ 是一个有界区域， $u:\ \mathcal{D} \to \mathbb{R}$ 是状态变量， $f:\ \mathcal{D} \to \mathbb{R}$ 是源项， $\boldsymbol{v}:\ \mathcal{D} \to \mathbb{R}^3$ 是无散速度场， $\kappa > 0$ 是扩散系数。该反问题旨在根据终端时刻 T对 u 的测量数据，恢复未知的初始条件 m 。

我们针对这一反问题采用贝叶斯框架。此类问题的无限维表述、有限元离散化以及数值求解已在 [12],在高斯先验和加性高斯噪声模型的假设下得到充分研究，我们在此也采用这些假设。为使表述简洁，我们聚焦于该问题的离散化版本。具体而言，我们希望根据终端状态 $\boldsymbol{m} \in \mathbb{R}^n$ 的逐点观测 $\boldsymbol{y} \in \mathbb{R}^d$ 来估计 $u(\cdot,\ T)$。这一关系由数据模型指定：

$$\boldsymbol{y} = \mathbf{F}\boldsymbol{m} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}}), \tag{4.2}$$

其中 $\mathbf{F} \in \mathbb{R}^{d\times n}$ 是离散化的参数到观测量映射， $\boldsymbol{\Gamma}_{\text{noise}} = \sigma^2\mathbf{I} \in \mathbb{R}^{d\times d}$ 是噪声协方差矩阵。我们考虑先验分布 $\mu_{\text{pr}} = \mathcal{N}\big(\mathbf{0},\ \boldsymbol{\Gamma}_{\text{pr}}\big)$ 其中 $\boldsymbol{\Gamma}_{\text{pr}}$由椭圆微分算子的平方逆进行离散化得到，[12], 该算子具有尺度参数 $\alpha = 1$ 和相关长度参数 $\beta = 5$。众所周知，在此设置下，状态变量 $\boldsymbol{m}$的后验分布由 $\mu_{\text{post}}^{\boldsymbol{y}} = \mathcal{N}(\boldsymbol{m}_{\text{MAP}},\ \boldsymbol{\Gamma}_{\text{post}})$ 给出，其中

$$\boldsymbol{\Gamma}_{\text{post}} = \left(\mathbf{F}^*\boldsymbol{\Gamma}_{\text{noise}}^{-1}\mathbf{F} + \boldsymbol{\Gamma}_{\text{pr}}^{-1}\right)^{-1}, \quad \boldsymbol{m}_{\text{MAP}} = \boldsymbol{\Gamma}_{\text{post}}\mathbf{F}^*\boldsymbol{\Gamma}_{\text{noise}}^{-1}\boldsymbol{y}.$$

在这个数值示例中，我们令 $\mathcal{D} = (0,1)^3$ ，并考虑离散化的参数维度$n = 100{,}000$。速度场设为 $\boldsymbol{v}([x_1,x_2,x_3])\colon = -(0.5 - x_2, x_1 - 0.5, 0)^\top$ ，终止时间设为 $T = 1.0$。我们通过构造一个真实初始条件 $\boldsymbol{m}_{\text{true}}$来生成合成数据，如图5（左）所示。我们使用 $\boldsymbol{m}_{\text{true}}$ 求解正向模型（4.2），从而在由 $d = 216$ 个传感器位置组成的均匀网格上获得观测数据 $\boldsymbol{y}$ ，如图5（右）所示。为模拟测量误差，我们在观测数据 $\boldsymbol{y}$中加入 1% 的噪声（$\sigma = 0.01\|\boldsymbol{y}\|_2/\sqrt{d}$）。

对于反问题，量化人们预期通过反演过程获得的“信息”量（即期望信息增益（EIG））通常至关重要。在最优实验设计的线性贝叶斯反问题背景下，这一度量也称为 D-最优性准则 [1]。对于当前的线性高斯问题，

$$\Psi_{EIG} = \frac{1}{2}\log\det(\mathbf{I} + \widetilde{\mathbf{H}}), \tag{4.3}$$

其中， $\widetilde{\mathbf{H}} := \boldsymbol{\Gamma}_{\text{pr}}^{1/2}\mathbf{F}^*\boldsymbol{\Gamma}^{-1}$ 噪声$\mathbf{F}\boldsymbol{\Gamma}_{\text{pr}}^{1/2}$是先验预条件数据失配海森矩阵。需要注意的是，该海森矩阵是隐式定义的，每次应用都需要进行两次或更多次偏微分方程求解。因此，对于高维问题，显式计算 EIG 十分困难。另一方面，FLEXTRACE 无需显式计算后验协方差矩阵，即可高效估计 EIG。

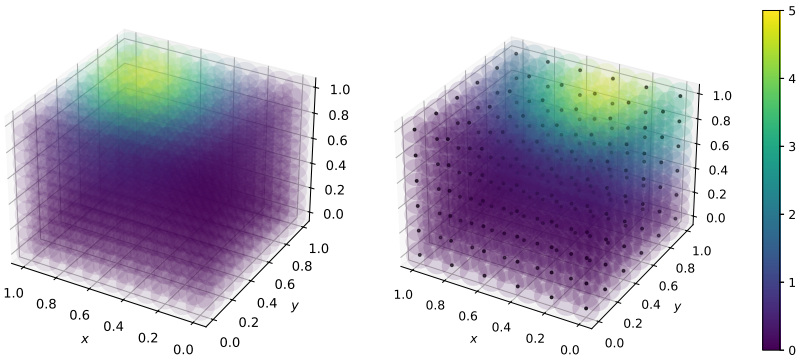


图5. 真实初始条件 $\boldsymbol{m}_{\text{true}}$（左）和终止状态（右）

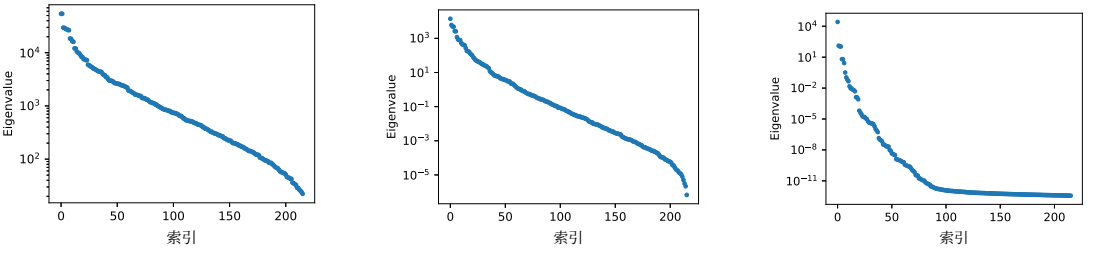
扩散的大小在反问题的控制物理中发挥着重要作用。为对此进行分析，在图6中，我们绘制了 $\widetilde{\mathbf{H}}$ 对于 $\kappa \in \{10^{-3}, 10^{-2}, 10^{-1}\}$ 的特征值。请注意，$\widetilde{\mathbf{H}}$ 的谱衰减速率显著取决于扩散系数 κ。也就是说，κ 的较大值对应于扩散主导区域，此时海森矩阵往往表现出较快的谱衰减；而较小的 κ则对应于对流主导区域，其谱衰减较慢。因此，我们选择分析 FLEXTRACE 在这些场景下的有效性。具体而言，在图7中，我们展示了扩散系数为 $\kappa \in \{10^{-3}, 10^{-2}, 10^{-1}\}$ 时，FLEXTRACE 和 FUNNYS 在 EIG 估计问题中的相对误差。在对流主导区域和扩散主导区域中，我们都观察到 FLEXTRACE 始终优于 FUNNYS。需要注意的是，对于扩散成分较小的问题，由于$\widetilde{\mathbf{H}}$较慢的谱衰减，两种方法在迹估计中都只能达到较低的相对精度。在这种情况下，FLEXTRACE 能更好地捕捉$\widetilde{\mathbf{H}}$的尾部谱行为，因此取得了更大的性能优势。另一方面，在谱衰减较快（κ较大）的情况下，两种方法都能达到较高的精度，其中 FLEXTRACE 略优于 FUNNYS。

4.4. 核方法的应用

本节展示 FLEXTRACE 在核方法中的应用。概括而言，核方法是一类利用核函数将输入数据隐式映射到高维（通常是无限维）希尔伯特空间的技术，从而使线性模型能够捕捉内在的非线性模式。这一框架的核心是核矩阵。对于一组数据点 $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subseteq \mathbb{R}^d$，核矩阵 $\mathbf{K} \in \mathbb{R}^{n\times n}$按元素定义为核函数 κ 的两两求值：$\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $K_{ij} = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$。这类矩阵在多个领域中都十分重要，包括高斯过程（GP）、贝叶斯神经网络、行列式点过程和椭圆图模型。

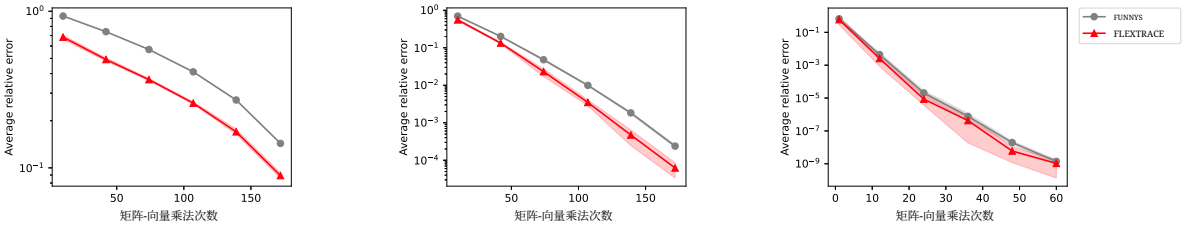
我们聚焦于高斯过程回归，其目标是通过将高斯过程拟合到一组“训练”数据点 $\{(\boldsymbol{x}_i, f(\boldsymbol{x}_i))\}_{i=1}^n$，从而用高斯过程近似函数 $f(\boldsymbol{x})$。核函数 k 有多种可能的选择，例如平方指数核、二次核、Matérn 核等。这些核通常依赖于一组需要优化的超参数。通常通过最大化 GP 模型的对数边际似然来完成优化，其定义如下：

$$\log p(\boldsymbol{y}|X) = -\frac{1}{2}\boldsymbol{y}^\top(\mathbf{K}+\sigma_n^2\mathbf{I})^{-1}\boldsymbol{y} - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log\det(\mathbf{K}+\sigma_n^2\mathbf{I}),$$



(a) $\kappa = 10^{-3}$ (b) $\kappa = 10^{-2}$ (c) $\kappa = 10^{-1}$

图6. 不同 κ取值下先验预条件海森矩阵的谱衰减



(a) $\kappa = 10^{-3}$ (b) $\kappa = 10^{-2}$ (c) $\kappa = 10^{-1}$

图7. 不同 κ取值下FLEXTRACE在特征值估计中的性能

其中，$\boldsymbol{y} \in \mathbb{R}^n$ 表示在数据点 $f(\boldsymbol{x})$ 处的（含噪）观测值 X，噪声方差为 σ_n^2。请注意，显式计算对数边际似然，尤其是对数行列式项，需要$\mathcal{O}(n^3)$ 次浮点运算和 $\mathcal{O}(n^2)$存储空间，这对于大规模问题可能会造成严重限制。尽管如此，过去几十年来，人们已开发出针对核矩阵的高效低秩和结构化稀疏近似方法 [35, 59]，从而能够快速近似执行矩阵向量乘法。我们使用 H2Pack[58] 软件包，它提供了计算核矩阵的 $\mathcal{H}^2$ 和 $\mathcal{HSS}$（分层半可分）近似的程序，具有线性复杂度矩阵向量乘。FLEXTRACE 可用于计算核矩阵的对数行列式，而无需显式计算核矩阵本身。

在数值测试中，我们考虑三维_道路北日德兰（丹麦）高程数据集 [39]，该数据集来自 UCI 机器学习资料库，包含434,874 个训练数据点。这些数据利用覆盖185 km $\times$ 130 km区域、精度为 $\pm$20cm的大规模航空激光扫描数据构建，旨在服务于高级驾驶辅助系统（ADAS）；见图 8。我们向训练数据拟合高斯过程，并加入合成相对噪声5%。

我们的目标是隐式估计 $\log\det(\mathbf{K}+\sigma_n^2\mathbf{I})$。为了能够针对精确的对数行列式计算相对误差，我们首先考虑一个较小的训练数据集子集，其中包含 10,000 个数据点，并将所有核函数缩放 0.1 倍（固定输出方差超参数）。我们在图 9中观察到，对于平方指数、有理二次和 Matérn 核，FLEXTRACE 的表现始终优于 FUNNYS。FLEXTRACE 的优势在于，它能够以计算上可行的方式扩展训练数据点规模。为此，我们在近 380,000 个数据点的完整数据集上测试了我们的方法。请注意，由于对于如此大规模的问题，显式计算真实值以供比较已变得难以处理，因此我们仅在图 10中展示 FLEXTRACE 和 FUNNYS 得到的迹估计。鉴于计算成本较高，该实验的结果也未在多次独立试验间汇总。尽管如此，我们观察到，FLEXTRACE 用不到一半数量的矩阵-向量乘法，就达到了 FUNNYS 使用 10

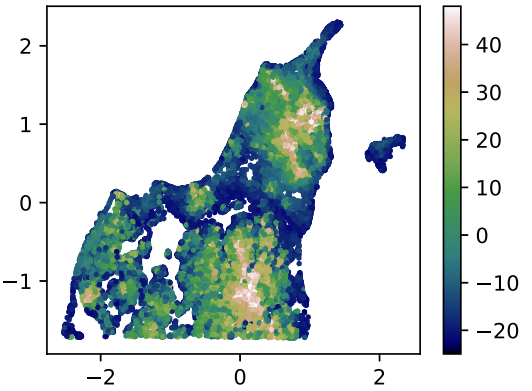
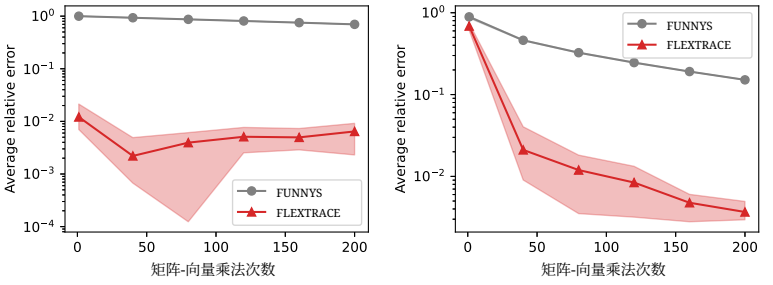


图 8。 北日德兰（丹麦）的 3D 道路高程数据集 [39]。



(a) 平方指数 ($\ell = 0.025$)

(b) 有理二次 ($\ell = 0.05, \alpha = 0.25$)

(c) Matérn 3/2 ($\ell = 0.10$)

(d) Matérn 5/2（$\ell = 0.10$）

图 9. 使用 FLEXTRACE 估计 $\log\det(\mathbf{K}+\sigma_n^2\mathbf{I})$ 的相对误差。

000 次矩阵-向量乘法得到的迹估计。结合平均相对误差的降低，这表明 FLEXTRACE 可以成为在大规模问题中对矩阵函数进行迹估计的强大而有效的工具。

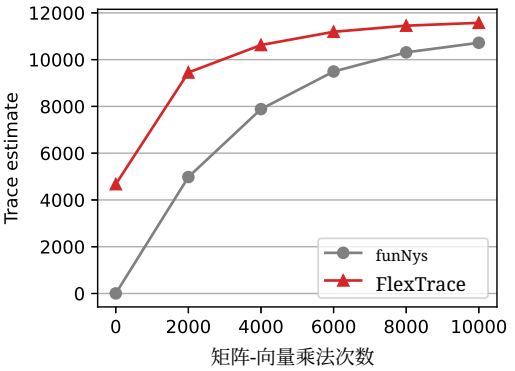


图10. FLEXTRACE 在大规模核问题中的应用

5. 理论结果的证明

本节详细给出第 3.2 节中理论结果所需的证明。首先，我们在下文给出命题 2.2 的证明。随后，分别在第i-FLEXTRACE 和 FLEXTRACE 5.1 节和5.2节中证明相应的理论结果。关于 Nyström 近似的补充背景知识及相关辅助结果的证明见附录 A。

命题*2.2*的证明。由于随机向量 $\{\boldsymbol{\omega}_i\}_{i=1}^k$ 独立同分布，

$$\mathbb{E}[\overline{Q}] = \frac{1}{|S_k|}\sum_{\sigma\in S_k}\mathbb{E}[\widehat{Q}(\boldsymbol{\omega}_{\sigma(1)},\ldots,\boldsymbol{\omega}_{\sigma(k)})] = \mathbb{E}[\widehat{Q}]. \tag{5.1}$$

因此，只需证明 Q 的方差小于 Q 的方差。利用詹森不等式[38, 将方程 4] 应用于 $\varphi(x)=x^2$，可得

$$\overline{Q}^2 = \left(\frac{1}{|S_k|}\sum_{\sigma\in S_k}\widehat{Q}(\boldsymbol{\omega}_{\sigma(1)},\ldots,\boldsymbol{\omega}_{\sigma(k)})\right)^2 \le \frac{1}{|S_k|}\sum_{\sigma\in S_k}\left(\widehat{Q}(\boldsymbol{\omega}_{\sigma(1)},\ldots,\boldsymbol{\omega}_{\sigma(k)})\right)^2.$$

对两边取期望，并利用 $Q(\boldsymbol{\omega}_{\sigma(1)}、\ldots\ldots、\boldsymbol{\omega}_{\sigma(k)})$ 对于$\sigma\in S_k$均同分布，可得

$$\mathbb{E}[\overline{Q}^2] \le \frac{1}{|S_k|}\sum_{\sigma\in S_k}\mathbb{E}\left[\widehat{Q}(\boldsymbol{\omega}_{\sigma(1)},\ldots,\boldsymbol{\omega}_{\sigma(k)})^2\right] = \mathbb{E}[\widehat{Q}^2].$$

注意到 $\mathbb{E}[X^2]=\mathrm{Var}[X]+\mathbb{E}[X]^2$ 对每个随机变量 X 均成立，并代入(5.1) 可得$\mathrm{Var}[Q]\le\mathrm{Var}[\widehat{Q}]$。最后，由于均方误差$[\theta]=\mathrm{Var}[\theta]+\mathbb{E}[(\theta-Q)]^2$ 对每个 θ 关于 Q的估计量均成立，因此可得均方误差$[Q]\le$ 均方误差$[Q]$。注意， Q 的可交换性由构造可直接推出，因此不再单独证明。□

5.1. *i*-FLEXTRACE的理论结果

我们首先证明，i-FLEXTRACE 是 $\mathrm{tr}(f(\mathbf{A}))$ 的无偏估计量。

定理 定理 *3.2*的证明。 为证明 $\widehat{\mathrm{tr}}_{\mathrm{iFT}}$ 是 $\mathrm{tr}(f(\mathbf{A}))$的无偏估计量，只需证明 $\widehat{\mathrm{tr}}_i$ 对所有 tr_i 均为无偏估计量 $1 \le i \le k$。不失一般性，考虑 tr_1。于是，

$$\begin{aligned}\mathbb{E}\left[\widehat{\mathrm{tr}}_1 - \mathrm{tr}(f(\mathbf{A}))\right] &= \mathbb{E}\left[\mathrm{tr}\, f(\widehat{\mathbf{A}}_{-1}) + \boldsymbol{\omega}_1^\top (f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{-1}))\boldsymbol{\omega}_1 - \mathrm{tr}(f(\mathbf{A}))\right] \\ &= \mathbb{E}\left[\boldsymbol{\omega}_1^\top f(\mathbf{A})\boldsymbol{\omega}_1 - \mathrm{tr}(f(\mathbf{A}))\right] + \mathbb{E}\left[\mathrm{tr}\, f(\widehat{\mathbf{A}}_{-1}) - \boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{-1})\boldsymbol{\omega}_1\right] \\ &= 0 + \mathbb{E}\left[\mathrm{tr}\, f(\widehat{\mathbf{A}}_{-1}) - \boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{-1})\boldsymbol{\omega}_1\right] \end{aligned} \tag{5.2}$$

在最后一个等式中，第一项为零，这是因为 Girard-Hutchinson 估计量对于 $\mathrm{tr}(f(\mathbf{A}))$ (定理 2.1是无偏的。对于剩余项，我们使用全期望公式，

$$\mathbb{E}\left[\mathrm{tr}\, f(\widehat{\mathbf{A}}_{-1}) - \boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{-1})\boldsymbol{\omega}_1\right] = \mathbb{E}_{\boldsymbol{\Omega}_{-1}}\left[\mathbb{E}_{\boldsymbol{\omega}_1}\left[\mathrm{tr}\, f(\widehat{\mathbf{A}}_{-1}) - \boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{-1})\boldsymbol{\omega}_1 \,\middle|\, \boldsymbol{\Omega}_{-1}\right]\right]$$

由于 $\mathbf{A}_{-1}$ 不依赖于 $\boldsymbol{\omega}_1$，我们再次利用 定理 2.1，得到$\mathbb{E}_{\boldsymbol{\omega}_1}\left[\mathrm{tr}\, f(\widehat{\mathbf{A}}_{-1}) - \boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{-1})\boldsymbol{\omega}_1 \,\middle|\, \boldsymbol{\Omega}_{-i}\right] = 0$。因此，(5.2)可化简为

$$\mathbb{E}\left[\widehat{\mathrm{tr}}_1 - \mathrm{tr}(f(\mathbf{A}))\right] = \mathbb{E}\left[\mathrm{tr}\, f(\widehat{\mathbf{A}}_{-1}) - \boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{-1})\boldsymbol{\omega}_1\right] = 0.$$

T因此，估计量 $\widehat{\mathrm{tr}}_{\mathrm{iFT}}$ 的平均值也是 $\mathbf{tr}(f(\mathbf{A}))$ 的无偏估计量 . □

我们现在证明定理 3.4，该定理利用可交换性以及 Girard-Hutchinson 估计量的性质，得到 i-FLEXTRACE 方差的一个界。

证明定理 *3.4*。 以下证明仿照 [22, 命题 5。3]。为控制方差，我们注意到，由可交换性，

$$\mathrm{Var}\left[\widehat{\mathrm{tr}}_{\mathrm{iFT}}\right] = \mathrm{Var}\left[\frac{1}{k}\sum_{i=1}^{k}\widehat{\mathrm{tr}}_i\right] = \left(\frac{1}{k}\right)\underbrace{\mathrm{Var}\left[\widehat{\mathrm{tr}}_1\right]}_{\mathsf{A}} + \left(1 - \frac{1}{k}\right)\underbrace{\mathrm{Cov}\left[\widehat{\mathrm{tr}}_1, \widehat{\mathrm{tr}}_2\right]}_{\mathsf{B}}. \tag{5.3}$$

我们首先关注项 A。根据全方差公式，

$$\mathrm{Var}\left[\widehat{\mathrm{tr}}_1\right] = \mathbb{E}\left[\mathrm{Var}\left[\widehat{\mathrm{tr}}_1 \,\middle|\, \boldsymbol{\Omega}_{-1}\right]\right] + \mathrm{Var}\left[\mathbb{E}\left[\widehat{\mathrm{tr}}_1 \,\middle|\, \boldsymbol{\Omega}_{-1}\right]\right].$$

正如对定理 3.2的证明中所观察到的， $\mathbb{E}\left[\widehat{\mathrm{tr}}_1 \,\middle|\, \boldsymbol{\Omega}_{-1}\right] = \mathrm{tr}\, f(\mathbf{A})$，因此第二项为零。于是剩下

$$\begin{aligned}\mathrm{Var}\left[\widehat{\mathrm{tr}}_1\right] &= \mathbb{E}\left[\mathrm{Var}\left[\widehat{\mathrm{tr}}_1 \,\middle|\, \boldsymbol{\Omega}_{-1}\right]\right] \\ &= \mathbb{E}\left[\mathrm{Var}\left[\mathrm{tr}\, f(\widehat{\mathbf{A}}_{-1}) + \boldsymbol{\omega}_1^\top (f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{-1}))\boldsymbol{\omega}_1 \,\middle|\, \boldsymbol{\Omega}_{-1}\right]\right] \\ &= \mathbb{E}\left[\mathrm{Var}\left[\boldsymbol{\omega}_1^\top (f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{-1}))\boldsymbol{\omega}_1 \,\middle|\, \boldsymbol{\Omega}_{-1}\right]\right]. \end{aligned} \tag{5.4}$$

式 (5.4) 中的最后一个等式源于如下事实：$\mathrm{tr}\, f(\mathbf{A}_{-1})$ 在以$\boldsymbol{\Omega}_{-1}$为条件时为常数。根据定理 2.1，方差项可简化为

$$\mathsf{A} = \mathrm{Var}\left[\widehat{\mathrm{tr}}_1\right] = 2\mathbb{E}\left[\|f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{-1})\|_F^2\right].$$

接下来，我们展开协方差项 B 在 (5.3)中。为此，我们首先通过分离出依赖于 以及 $\boldsymbol{\omega}_1$ 和 $\boldsymbol{\omega}_2$的项，重新表述 B。具体而言，我们考虑矩阵 $\mathbf{X} := f(\mathbf{A}) - f(\mathbf{A}^{-\{1,2\}})$，

它与 $\boldsymbol{\omega}_1$ 和 $\boldsymbol{\omega}_2$相互独立，并注意到

$$\mathbb{E}\Big[\Big(\widehat{\mathrm{tr}}_1 - \mathrm{tr}\, f(\mathbf{A}) + \mathrm{tr}\,\widehat{\mathbf{X}} - \boldsymbol{\omega}_1^\top \widehat{\mathbf{X}}\boldsymbol{\omega}_1\Big)\Big(\widehat{\mathrm{tr}}_2 - \mathrm{tr}\, f(\mathbf{A}) + \mathrm{tr}\,\widehat{\mathbf{X}} - \boldsymbol{\omega}_2^\top \widehat{\mathbf{X}}\boldsymbol{\omega}_2\Big)\Big] \tag{5.5}$$

$$= \mathbb{E}\Big[\big(\widehat{\mathrm{tr}}_1 - \mathrm{tr}\, f(\mathbf{A})\big)\Big(\widehat{\mathrm{tr}}_2 - \mathrm{tr}\, f(\mathbf{A}) + \mathrm{tr}\,\widehat{\mathbf{X}} - \boldsymbol{\omega}_2^\top \widehat{\mathbf{X}}\boldsymbol{\omega}_2\Big)\Big] \tag{5.6}$$

$$= \mathbb{E}\big[\big(\widehat{\mathrm{tr}}_1 - \mathrm{tr}\, f(\mathbf{A})\big)\big(\widehat{\mathrm{tr}}_2 - \mathrm{tr}\, f(\mathbf{A})\big)\big] = \mathrm{Cov}\big[\widehat{\mathrm{tr}}_1, \widehat{\mathrm{tr}}_2\big]. \tag{5.7}$$

为了从 (5.6)得到 (5.5)，我们对 $\boldsymbol{\Omega}_{-1}$进行条件化，并对 $\boldsymbol{\omega}_1$取平均，利用 $\mathbb{E}\big[\widehat{\ }\mathrm{tr}_1 \,\big|\boldsymbol{\Omega}_{-1}\big] = \mathrm{tr}\, f$ $(\mathbf{A})$这一点。同理，为了从 (5.7)得到 (5.6)，我们对 $\boldsymbol{\Omega}_{-2}$进行条件化，并对 $\boldsymbol{\omega}_2$取平均。最后，我们将 B 表示为 (5.5)，并应用柯西–施瓦茨不等式，得到：

$$\begin{aligned}\mathsf{B} = \mathrm{Cov}\big[\widehat{\mathrm{tr}}_1, \widehat{\mathrm{tr}}_2\big] &= \mathbb{E}\big[\big(\widehat{\mathrm{tr}}_1 - \mathrm{tr}\, f(\mathbf{A}) + \mathrm{tr}\,\widehat{\mathbf{X}} - \boldsymbol{\omega}_1^\top \widehat{\mathbf{X}}\boldsymbol{\omega}_1\big)\big(\widehat{\mathrm{tr}}_2 - \mathrm{tr}\, f(\mathbf{A}) + \mathrm{tr}\,\widehat{\mathbf{X}} - \boldsymbol{\omega}_2^\top \widehat{\mathbf{X}}\boldsymbol{\omega}_2\big)\big] \\ &\le \mathbb{E}\Big[\big|\widehat{\mathrm{tr}}_1 - \mathrm{tr}\, f(\mathbf{A}) + \mathrm{tr}\,\widehat{\mathbf{X}} - \boldsymbol{\omega}_1^\top \widehat{\mathbf{X}}\boldsymbol{\omega}_1\big|^2\Big].\end{aligned}$$

注意，$\widehat{\ }\mathrm{tr}_1 - \mathrm{tr}\, f(\mathbf{A})$和 $\mathrm{tr}\,\widehat{\ }\mathbf{X} - \boldsymbol{\omega}_1^\top \widehat{\ }\mathbf{X}\boldsymbol{\omega}_1$的均值为零，这是 定理 3.2 和定理 2.1 分别导致的。因此，

$$\mathbb{E}\Big[\big|\widehat{\mathrm{tr}}_1 - \mathrm{tr}\, f(\mathbf{A}) + \mathrm{tr}\,\widehat{\mathbf{X}} - \boldsymbol{\omega}_1^\top \widehat{\mathbf{X}}\boldsymbol{\omega}_1\big|^2\Big] = \mathrm{Var}\Big[\widehat{\mathrm{tr}}_1 - \mathrm{tr}\, f(\mathbf{A}) + \mathrm{tr}\,\widehat{\mathbf{X}} - \boldsymbol{\omega}_1^\top \widehat{\mathbf{X}}\boldsymbol{\omega}_1\Big]$$

因此，

$$\begin{aligned}\mathsf{B} &\le \mathrm{Var}\Big[\widehat{\mathrm{tr}}_1 - \mathrm{tr}\, f(\mathbf{A}) + \mathrm{tr}\,\widehat{\mathbf{X}} - \boldsymbol{\omega}_1^\top \widehat{\mathbf{X}}\boldsymbol{\omega}_1\Big] \\ &= \mathrm{Var}\Big[\mathrm{tr}[f(\widehat{\mathbf{A}}_{-1}) - f(\widehat{\mathbf{A}}_{-\{1,2\}})] - \boldsymbol{\omega}_1^\top [f(\widehat{\mathbf{A}}_{-1}) - f(\widehat{\mathbf{A}}_{-\{1,2\}})]\boldsymbol{\omega}_1\Big],\end{aligned}$$

$\widehat{\ }$ $\widehat{\ }$其中我们代入 tr_1 和 $\mathbf{X}$的定义。类似地，将全方差公式应用于 (5.4)，可得

$$\mathsf{B} \le \mathbb{E}\Big[\mathrm{Var}\big[\boldsymbol{\omega}_1^\top \big[f(\widehat{\mathbf{A}}_{-\{1,2\}}) - f(\widehat{\mathbf{A}}_{-1})\big]\boldsymbol{\omega}_1 \,\big|\, \boldsymbol{\Omega}_{-1}\big]\Big] = 2\mathbb{E}\Big[\big\|f(\widehat{\mathbf{A}}_{-1}) - f(\widehat{\mathbf{A}}_{-\{1,2\}})\big\|_{\mathrm{F}}^2\Big].$$

结合 A、B 并对 $1 - 1/k < 1$ 进行界定，即完成证明。 □

5.2. *FLEXTRACE* 的理论结果FLEXTRACE

本节证明了 第 3.2 节中提出的 FLEXTRACE 相关结果。首先，我们证明 FLEXTRACE 是可交换的。

定理 3.1 的证明。为证明 FLEXTRACE 是可交换的，我们证明 (3.2) 是 $\widehat{\ }\mathrm{tr}_1 = \mathrm{tr}\ f(\widehat{\ }\mathrm{A}_{-1})$ $+ \boldsymbol{\omega}_1^\top(f(\widehat{\ }\mathrm{A}^{\mathrm{nys}}) - f(\widehat{\ }\mathrm{A}_{-1}))\boldsymbol{\omega}_1$ 的对称化形式。为此，注意到，根据引理 A.1(c)，$\widehat{\ }\mathrm{A}^{\mathrm{nys}}and\widehat{\ }$ A_{-1} 在 $(\boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_k)$ 的置换下保持不变，因此 $\widehat{\ }\mathrm{tr}_1$ 也在 $(\boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_k)$ 的置换下保持不变。因此，要对 $\widehat{\ }\mathrm{tr}_1$, 进行对称化，只需对满足 $\sigma(1) = i$ for $1 \le i \le k$ 的置换 $\sigma \in S_k$ 求平均即可。该简化得到估计量 (3.2)；根据命题 2.2，该估计量因而是可交换的。□

证明 定理 3.3。 考虑 tr_i ，它们来自 算法 3.1 的第 5 行。由于这些估计量同分布， $\mathbb{E}[\widehat{\mathrm{tr}}_{\mathrm{FT}}] = \mathbb{E}[\widehat{\mathrm{tr}}_i]$ 对所有 $1 \le i \le k$ 均成立。不失一般性，考虑 $\widehat{\mathrm{tr}}_1$。注意到

$$\begin{aligned}\mathbb{E}\,[\widehat{\mathrm{tr}}_1] &= \mathbb{E}[\mathrm{tr}\, f(\widehat{\mathbf{A}}_{\text{-}1}) + \boldsymbol{\omega}_1^\top (f(\widehat{\mathbf{A}}_{\mathrm{nys}}) - f(\widehat{\mathbf{A}}_{\text{-}1}))\boldsymbol{\omega}_1] \\ &= \mathbb{E}\Big[\mathbb{E}[\mathrm{tr}\, f(\widehat{\mathbf{A}}_{\text{-}1}) - \boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{\text{-}1})\boldsymbol{\omega}_1 \mid \boldsymbol{\Omega}_{\text{-}1}]\Big] + \mathbb{E}[\boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{\mathrm{nys}})\boldsymbol{\omega}_1] \\ &= \mathbb{E}[\boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{\mathrm{nys}})\boldsymbol{\omega}_1]. \end{aligned} \tag{5.8}$$

由于 f 是算子单调的且 $\mathbf{A}_{\text{-}1} \preceq \mathbf{A}_{\mathrm{nys}} \preceq \mathbf{A}$ （见引理 A.1(d)），$f(\mathbf{A}_{\text{-}1}) \preceq f(\mathbf{A}^{\mathrm{nys}}) \preceq f(\mathbf{A})$。因此，

$$0 \le \boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{\text{-}1})\boldsymbol{\omega}_1 \le \boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{\mathrm{nys}})\boldsymbol{\omega}_1 \le \boldsymbol{\omega}_1^\top f(\mathbf{A})\boldsymbol{\omega}_1.$$

取期望并利用定理 2.1，

$$\mathbb{E}[\mathrm{tr}(f(\widehat{\mathbf{A}}_{\text{-}1}))] \le \mathbb{E}[\boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{\mathrm{nys}})\boldsymbol{\omega}_1] \le \mathrm{tr}(f(\mathbf{A})).$$

整理后，我们得到

$$\mathbb{E}[\mathrm{tr}(f(\mathbf{A})) - \widehat{\mathrm{tr}}_1] = \mathrm{tr}(f(\mathbf{A})) - \mathbb{E}[\boldsymbol{\omega}_1^\top f(\widehat{\mathbf{A}}_{\mathrm{nys}})\boldsymbol{\omega}_1] \le \mathbb{E}\Big[\mathrm{tr}(f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{\text{-}1}))\Big].$$

至此完成了定理 3.3 的第一个不等式。定理 3.3 中的第二个不等式 $\mathbb{E}[\|f(\mathbf{A}) - f(\mathbf{A}_{\text{-}1})\|_*] \le (k-3)\|f(\boldsymbol{\Lambda}_{k\text{-}2:n})\|_*$ 由 [51, 定理 3.9] 给出。我们还注意到，若 f 是线性的，则 FLEXTRACE 退化为 i-FLEXTRACE，因而无偏；见定理 3.2。□

定理 3.5 的证明。 根据 命题 2.2，对称化估计量 $\mathrm{tr}_{\mathrm{FT}}$ 的均方误差 由 tr_1 的均方误差界定。根据 tr_1，

$$\begin{aligned}\mathrm{MSE}[\widehat{\mathrm{tr}}_1] &\le \mathbb{E}\Big[(\mathrm{tr}\, f(\widehat{\mathbf{A}}_{\text{-}1}) + \boldsymbol{\omega}_1^\top (f(\widehat{\mathbf{A}}_{\mathrm{nys}}) - f(\widehat{\mathbf{A}}_{\text{-}1}))\boldsymbol{\omega}_1 - \mathrm{tr}(f(\mathbf{A})))^2\Big] \\ &= \mathbb{E}\Big[(\mathrm{tr}(f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{\text{-}1})) - \boldsymbol{\omega}_1^\top (f(\widehat{\mathbf{A}}_{\mathrm{nys}}) - f(\widehat{\mathbf{A}}_{\text{-}1}))\boldsymbol{\omega}_1)^2\Big] \\ &= \mathbb{E}\Big[(\mathrm{tr}(\mathbf{B}) - \boldsymbol{\omega}_1^\top \mathbf{G}\boldsymbol{\omega}_1)^2\Big], \end{aligned}$$

其中我们定义 $\mathbf{B}$： $= f \quad - f \quad _{\text{-}1}$ $= f \quad _{\mathrm{nys}} - f \quad _{\text{-}1}$ $\alpha - \beta\ ^2 \le$ (A) (A) 和 $\mathbf{G}$: (A) (A)。利用对所有， 均成立的性质，

$$\mathrm{MSE}[\widehat{\mathrm{tr}}_1] \le \mathbb{E}\Big[(\mathrm{tr}(\mathbf{B}) - \boldsymbol{\omega}_1^\top \mathbf{G}\boldsymbol{\omega}_1)^2\Big] \le \mathbb{E}\big[(\mathrm{tr}(\mathbf{B}))^2\big] + \mathbb{E}\Big[(\boldsymbol{\omega}_1^\top \mathbf{G}\boldsymbol{\omega}_1)^2\Big].$$

注意， $\mathbf{0} \preceq \mathbf{G} \preceq \mathbf{B}$ 根据 引理 A.1(d)。因此， $\mathrm{tr}(\mathbf{B}) = \|\mathbf{B}\|_*$ 且 $\boldsymbol{\omega}_1^\top \mathbf{G}\boldsymbol{\omega}_1 \preceq \boldsymbol{\omega}_1^\top \mathbf{B}\boldsymbol{\omega}_1$。所以，

$$\mathrm{MSE}[\widehat{\mathrm{tr}}_1] \le \mathbb{E}\big[\|\mathbf{B}\|_*^2\big] + \mathbb{E}\Big[(\boldsymbol{\omega}_1^\top \mathbf{B}\boldsymbol{\omega}_1)^2\Big].$$

利用 $\mathbf{B}$ 与 $\boldsymbol{\omega}_1$ 相互独立这一事实，我们可以简化右侧的加项。根据定理 2.1，我们注意到

$$\mathbb{E}_{\boldsymbol{\omega}_1}\Big[(\boldsymbol{\omega}_1^\top \mathbf{B}\boldsymbol{\omega}_1)^2\Big] = \mathrm{Var}_{\boldsymbol{\omega}_1}\Big[\boldsymbol{\omega}_1^\top \mathbf{B}\boldsymbol{\omega}_1\Big] + \Big[\mathbb{E}[\boldsymbol{\omega}_1^\top \mathbf{B}\boldsymbol{\omega}_1]\Big]^2 = 2\|\mathbf{B}\|_F^2 + \|\mathbf{B}\|_*^2.$$

根据条件期望的塔式法则，$\mathbb{E}\big[(\boldsymbol{\omega}_1^\top \mathbf{B}\boldsymbol{\omega}_1)^2\big] = \mathbb{E}\big[\mathbb{E}_{\boldsymbol{\omega}_1}\big[(\boldsymbol{\omega}_1^\top \mathbf{B}\boldsymbol{\omega}_1)^2 \mid \mathbf{B}\big]\big] = \mathbb{E}\big[2\|\mathbf{B}\|_F^2 + \|\mathbf{B}\|_*^2\big]$。因此，

$$\mathrm{MSE}[\widehat{\mathrm{tr}}_1] \le 2\,\mathbb{E}\big[\|\mathbf{B}\|_F^2\big] + 2\,\mathbb{E}\big[\|\mathbf{B}\|_*^2\big].$$

代入$\widehat{\mathbf{B}}$的定义即可得到所需结果。 □

定理 *3.7*的证明。$\widehat{\mathrm{tr}}_{\mathrm{iFT}}$ 和 $\widehat{\mathrm{tr}}_{\mathrm{FT}}$ 的均方误差界可直接将 定理 3.6 中的不等式分别代入定理 3.4 和 3.5得到。最后，界关于 $\widehat{\mathrm{tr}}_{\mathrm{FN}}$ 的均方误差，可直接通过应用 定理 3.6 并采用核范数得到。 □

6. 结论

本文提出了 FLEXTRACE，这是一种新颖的随机迹估计算法，可高效估计 tr(f(**A**))。该算法利用低秩随机 Nyström 近似以及经典迹估计方法，通过对**A** 进行随机草图来估计矩阵函数的迹。与传统技术相比，所提出的算法在性能和实用性方面均具优势，因为 FLEXTRACE 无超参数、单遍、函数无关且易于并行。我们为 FLEXTRACE 及其理想化变体 i-FLEXTRACE 在算子单调函数上的表现提供了理论保证，并观察到对于谱快速衰减的矩阵，其误差呈渐近衰减。此外，我们还提供了计算可扩展例程，以加速并行化 FLEXTRACE。数值结果表明，所提出的算法在合成示例和实际应用中均效果显著。具体而言，对于具有不同谱性质的矩阵，FLEXTRACE 通常优于现有的同类方法。此外，我们还证明，FLEXTRACE 能够加速贝叶斯反问题、矩阵补全和核方法中关注量的计算。

未来研究有若干值得探索的方向。由于 FLEXTRACE 是一种单遍算法，对于在谱范围内表现出高度非线性的矩阵函数，可能需要对所提出的算法进行修改。在这类情形下，可以利用函数的性质扩展 FLEXTRACE，或将可交换算法与多遍 Krylov 方法融合。尽管本文主要关注 FLEXTRACE 在算子单调函数 f 上的应用，但也值得研究 FLEXTRACE 在非算子单调递增函数上的应用。例如，对于 $f(x) = x^2$，由于引理 A.1(e)，FLEXTRACE 是一个无偏估计量；而且我们在数值测试中观察到，它通常优于 FUNNYS。对于更广泛函数类的 funNyström 近似，也需要建立理论保证，以便对 FLEXTRACE 进行概率分析。其他值得探索的方向包括采用混合精度方法、对这些方法进行稳定性分析，以及将该方法应用于不同领域。

致谢

本研究部分由美国国家科学基金会通过资助项目 DMS-2026830，以及美国能源部科学办公室先进科学计算研究项目通过资助项目 DE-SC0023188 和 DE-SC0026310 提供支持。

参考文献

1. A. Alexanderian。受偏微分方程支配的无限维贝叶斯反问题的最优实验设计：综述。《反问题》，37(4):043001，2021年。

2. A. Alexanderian 和 A. K. Saibaba。无限维贝叶斯线性反问题的高效 D-最优实验设计。*SIAM*科学计算杂志，40(5):A2956–A2985，2018年1月。3. E. Anderson、Z. Bai、C. Bischof、S. Blackford、J. Demmel、J. Dongarra、J. Du Croz、A. Greenbaum、S. Hammarling、A. McKenney 和 D. Sorensen。*LAPACK*用户指南。工业与应用数学学会，宾夕法尼亚州费城，第三版，1999年。4. T. Ando。舒尔补与矩阵不等式：算子理论方法。载于舒尔补及其应用，第137–162页。施普林格美国，马萨诸塞州波士顿，2005年。5. H. Avron、K. L. Clarkson 和 D. P. Woodruff。利用草图化和预条件化加速核岭回归。*SIAM* 矩阵分析与应用期刊，38(4):1116–1138，2017年。6. H. Avron 和 S. Toledo。用于估计隐式对称半正定矩阵的迹的随机算法。*ACM*杂志，58(2)，2011年4月。7. Z. Bai、G. Fahey 和 G. Golub。一些大规模矩阵计算问题。计算与应用数学杂志，74(1):71–89，1996年11月。8. J. Bennett 和 S. Lanning。Netflix奖。载于*2007*年*KDD*杯研讨会论文集，第3–6页，纽约，2007年8月。ACM。9. M. Benzi 和 P. Boito。网络分析中的矩阵函数。*GAMM*通讯，43(3):e202000012，2020年9月。10. R. Bhatia。矩阵分析，数学研究生教材第169卷。施普林格出版社，纽约，1997年。11. A. Bucci、Y. Nakatsukasa 和 T. Park。Nyström方法的数值稳定性，2025年11月。arXiv:2511.15583 [math]。12. T. Bui-Thanh、O. Ghattas、J. Martin 和 G. Stadler。无限维贝叶斯反问题的计算框架（第一部分）：线性化情形及其在全球地震反演中的应用。*SIAM*科学计算杂志，35(6):A2494–A2523，2013年。13. J. R. Bunch、C. P. Nielsen 和 D. C. Sorensen。对称特征值问题的秩一修正。数值数学，31(1):31–48，1978年。14. E. J. Candès 和 B. Recht。通过凸优化实现精确矩阵补全。计算数学基础，9(6):717–772，2009年12月。15. G. Casella 和 R. L. Berger。统计推断。Duxbury，加利福尼亚州太平洋丛林市，第2版，2002年。16. T. Chen 和 E. Hallman。Krylov 感知随机迹估计。*SIAM* 矩阵分析与应用期刊，44(3):1218–1244，2023年。17. Y. Chen、E. N. Epperly、J. A. Tropp 和 R. J. Webber。随机主元Cholesky分解：只需少量矩阵元素求值即可实现核矩阵的实用近似。纯粹与应用数学通讯，78(5):995–1041，2025年。18. A. Cortinovis 和 D. Toni。预条件对数行列式近似：一个高斯探测向量几乎总是足够的，2026年1月。19. K. Dong、D. Eriksson、H. Nickisch、D. Bindel 和 A. G. Wilson。高斯过程核学习的可扩展对数行列式。载于神经信息处理系统进展，2017年。20. Y. Dong、T. Gong、H. Chen 和 C. Li。序列数据上基于矩阵的 Rényi 熵的高效近似。*IEEE*神经网络与学习系统汇刊，36(2):2040–2050，2025年2月。21. E. N. Epperly。物尽其用：用于矩阵计算的资源高效随机算法。博士论文，加州理工学院，2025年6月。22. E. N. Epperly、J. A. Tropp 和 R. J. Webber。XTrace：充分利用随机迹估计中的每个样本。*SIAM* 矩阵分析与应用期刊，45(1):1–23，2024年。23. E. Estrada 和 D. J. Higham。通过矩阵函数揭示网络性质。*SIAM* 评论，52(4):696–714，2010年1月。24. R. D. Fuentes、M. Donatelli、C. Fenu 和 G. Mantica。矩阵函数迹的估计及其在复杂网络中的应用。数值算法，92(1):503–522，2023年1月。25. A. Girard。一种适用于含噪数据下大型最小二乘问题的快速“蒙特卡洛交叉验证”过程。数值数学

，56(1)：1–23 ，1989年。

26. A. Gittens 和 M. W. Mahoney. 重新审视 Nyström 方法，以改进大规模机器学习. 机器学习研究期刊, 17:117:1–117:65, 2013 年 . 27. M. Gu 和 S. C. Eisenstat. 对称特征值问题秩一修正的稳定高效算法 . *SIAM* 矩阵分析与应用期刊, 15(4):1266–1276, 1994. 28. N. Halko, P. G. Martinsson 和 J. A. Tropp. 借助随机性发现结构：用于构造近似矩阵分解的概率算法. *SIAM* 评论, 53(2):217–288, 2011. 29. P. R. Halmos. 无偏估计理论. 数学统计年刊, 17(1):34–43, 1946年3月. 30. S. Hammarling, N. J. Higham 和 C. Lucas. 用于带主元 Cholesky 分解和 QR 更新的 LAPACK 风格代码. 数值分析报告 442，曼彻斯特数学科学研究所，曼彻斯特大学，2007. ISSN 1749-9097. 31. F. M. Harper 和 J. A. Konstan. MovieLens 数据集：历史与背景. *ACM* 交互式智能系统汇刊 *(TiiS)*, 5(4):19:1–19:19, 2015年12月. 32. N. J. Higham. 数值算法的精度与稳定性. 工业与应用数学学会，宾夕法尼亚州费城，美国，第二版，校订重印版，2002年. 33. N. J. Higham. 矩阵函数：理论与计算. 工业与应用数学学会，2008年1月. 34. R. A. Horn 和 C. R. Johnson. 矩阵分析. 剑桥大学出版社，纽约州纽约市，第二版，校订重印版，2017年. 35. H. Huang, X. Xing 和 E. Chow. H2Pack：使用代理点方法处理核矩阵的高性能 H^2 矩阵软件包. *ACM* 数学软件汇刊, 47(1):1–29, 2020年12月. 36. M. F. Hutchinson. 用于估计拉普拉斯平滑样条影响矩阵之迹的随机估计量. 统计通讯：模拟与计算, 18:1059–1076, 1989. 37. N. Jakovčević Stor, I. Slapničar 和 J. Barlow. 对角矩阵秩一修正的前向稳定特征值分解. 线性代数及其应用, 487:301–315, 2015. 38. J. L. W. V. Jensen. 关于凸函数及均值之间的不等式. 数学学报, 30(0):175–193, 1906. 39. M. Kaul. 三维道路网络（北日德兰，丹麦）. UCI 机器学习资料库，2013. DOI: https://doi.org/10.24432/C5GP51. 40. N. Krämer, P. Moreno-Muñoz, H. Roy 和 S. Hauberg. 大矩阵函数的梯度. 载于第*38*届神经信息处理系统国际会议论文集，第37卷*NIPS '24*，第49484–49518页，2024年12月. 41. Y. Li, H. L. Nguyen 和 D. P. Woodruff. 矩阵范数草图估计与最大奇异向量. 载于*2014*年 *ACM-SIAM* 离散算法年度研讨会论文集 *(SODA)*，第 1562–1581 页 . 工业与应用数学学会，2013 年 12 月 . 42. C. Lucas. 用于第 2 级和第 3 级带主元 Cholesky 分解的 LAPACK 风格代码. 数值分析报告 442，曼彻斯特计算数学中心，曼彻斯特大学，2004年2月. ISSN 1360-1725. 43. F. Matti, H. He, D. Kressner 和 H. Y. Lam. 应用于谱密度近似的参数相关矩阵随机迹估计，2025年2月. 44. R. A. Meyer, C. Musco, C. Musco 和 D. P. Woodruff. Hutch++: 最优随机迹估计. *SIAM* 算法简明性研讨会论文集, 2021:142–155, 2020. 45. Y. Nakatsukasa. 快速稳定的随机化低秩矩阵近似，2025年9月. arXiv：. math . 46. L. T. Nguyen, J. Kim 和 B. Shim. 低秩矩阵补全：当代综述. *IEEE Access*, 7:94215–94237, 2019. 47. R. K. Pace 和 J. P. LeSage. 一种用于估计空间似然问题中对数行列式的抽样方法. 地理系统期刊, 11(3):209–225, 2009年9月. 48. C. Panish 和 L. Villani. 矩阵补全综述：理论、算法与实证评估，2025年12月。

49. D. Persson、N. Boullé 和 D. Kressner。非负自伴算子的随机 Nyström 近似。*SIAM* 数据科学数学杂志，7(2)：670–698，2025年6月。50. D. Persson、A. Cortinovis 和 D. Kressner。改进的 Hutch++ 算法及其在迹估计中的应用。*SIAM* 矩阵分析与应用期刊，43：1162–1185，2021年。51. D. Persson 和 D. Kressner。单调矩阵函数的随机化低秩近似。*SIAM* 矩阵分析与应用期刊，44(2)：894–918，2023年。52. D. Persson、R. A. Meyer 和 C. Musco。算子单调矩阵函数的与算法无关的低秩近似。*SIAM* 矩阵分析与应用期刊，46(1)：1–21，2025年。53. C. R. Rao。线性统计推断及其应用：第二版。Wiley 概率与统计丛书。Wiley，第1版，1973年4月。54. A. K. Saibaba、A. Alexanderian 和 I. C. F. Ipsen。随机化无矩阵迹与对数行列式估计量。数值数学，137：353–395，2017年。55. J. A. Tropp 和 R. J. Webber。低秩矩阵近似的随机化算法：设计、分析与应用，2023年。56. S. Ubaru、J. Chen 和 Y. Saad。通过随机 Lanczos 求积快速估计 $tr(f(A))$。*SIAM* 矩阵分析与应用期刊，38(4)：1075–1099，2017年。57. S. Ubaru 和 Y. Saad。迹估计技术的应用。载于 科学与工程中的高性能计算，第19–33页。Springer，2018年。58. X. Xing 和 E. Chow。利用代理点对核矩阵进行插值分解。*SIAM* 矩阵分析与应用期刊，41：221–243，2020年。59. X. Xing、H. Huang 和 E. Chow。对称正定 $\mathcal{H}^2$ 矩阵的 HSS 预条件子的高效构造。*SIAM* 矩阵分析与应用期刊，42(2)：683–707，2021年。60. K. Yeon、P. Ghosal 和 M. Anitescu。BOLT：用于矩阵函数迹估计的块正交 Lanczos，2025年12月。61. J. Yi、J. wu Gao、T. Wang、X. Wu 和 W. Xu。利用具有理论性能保证的生成模型进行异常值检测。*IEEE* 信息论汇刊，71：4012–4031，2018年。

A. funNyström 的理论保证

本节将推导 funNyström 近似的理论保证 [51]。具体而言，$\widehat{}$本节的主要目标是推导 $f(\mathbf{A}) - f(\mathbf{A}^{\text{nys}})$ 的期望平方范数界，$\widehat{}$其中 $\mathbf{A}^{\text{nys}}$ 是 $\mathbf{A}$ 的 Nyström 近似（参见第 2.3 节），且 $f \in \mathcal{F}$。在 附录 A.1中，我们列出 Nyström 近似的一些重要性质。随后在附录 A.2 概述理论方法，并在附录 A.3介绍随机矩阵理论的一些结果。最后，我们在附录 A.4证明 f unNyström 近似的关键结果。

A.1. *Nyström* 近似的性质

我们首先在引理 A.1中列出 Nyström 近似的一些有用性质，这些性质将在证明我们的主要结果时发挥作用。

引理 A.1 *Nyström* 近似*(2.3)*满足以下性质：

(a) 投影： 令 $\mathbf{\Pi}_{\mathbf{A}^{1/2}\mathbf{\Omega}} = (\mathbf{A}^{1/2}\mathbf{\Omega})(\mathbf{A}^{1/2}\mathbf{\Omega})^{\dagger}$ 表示到$\mathbf{A}^{1/2}\mathbf{\Omega}$值域上的投影算子。则， $\widehat{}\mathbf{A}^{\text{nys}} = \mathbf{A}^{1/2}\mathbf{\Pi}_{\mathbf{A}^{1/2}\mathbf{\Omega}}\mathbf{A}^{1/2}$.$\widehat{}$ *(b)* 插值性： 对于每个向量 $\boldsymbol{\nu} \in \mathcal{R}ange(\mathbf{\Omega})$，$\mathbf{A}_{\text{nys}}\boldsymbol{\nu} = \mathbf{A}\boldsymbol{\nu}$。

(c) 不变性：$\mathbf{A}^{\text{nys}}$ 在$\boldsymbol{\Omega}$的列置换下保持不变。 *(d)* 单调性：$\mathbf{A}_{-i} \preceq \mathbf{A}_{\text{nys}} \preceq \mathbf{A}$ 对所有 $1 \le i \le k$ 均成立。*(e)* 二次型的精确性：令q $\boldsymbol{\omega}_i^\top q \quad \boldsymbol{\omega}_i \qquad 1 \le i \le k$ $_{\text{nys}}\ \boldsymbol{\omega}_i =$ ： $\mathbb{R} \to \mathbb{R}$ 为一个二次函数。则，$\boldsymbol{\omega}_i^\top q(\mathbf{A})$对所有 均成立。

请注意，这些性质并不要求$\boldsymbol{\Omega}$ 是随机矩阵。

证明 (a) 的证明见 [26, 引理 1], ，由此可立即推出 (b)，因为：

$$\widehat{\mathbf{A}}_{\text{nys}}\boldsymbol{\Omega} = \mathbf{A}^{1/2}\boldsymbol{\Pi}_{\mathbf{A}^{1/2}\boldsymbol{\Omega}}\mathbf{A}^{1/2}\boldsymbol{\Omega} = \mathbf{A}\boldsymbol{\Omega}.$$

性质 (c) 也可由 (a) 推出，因为对于任意置换矩阵 P，$\mathcal{R}\text{ange}(\text{A}^{1/2}\Omega) = \mathcal{R}\text{ange}(\text{A}^{1/2}\Omega\text{P})$。最后，我们证明性质 (d)。由于 $\mathcal{R}\text{ange}(\text{A}^{1/2}\Omega_{-i}) \subseteq \mathcal{R}\text{ange}(\text{A}^{1/2}\Omega) \subseteq \mathbb{R}^n$，根据 [28, 命题 8.5]，有 $\boldsymbol{\Pi}_{\mathbf{A}^{1/2}\boldsymbol{\Omega}_{-i}} \preceq \boldsymbol{\Pi}_{\mathbf{A}^{1/2}\boldsymbol{\Omega}} \preceq \mathbf{I}$ 对于 $1 \le i \le k$成立。因此，由性质 (a)，对于 $1 \le i \le k$，$\text{A}_{-i} \preceq \text{A}_{\text{nys}} \preceq \text{A}$。最后，为证明性质 (e)，令 $q(x) = \alpha x^2 + \beta x + \gamma$，其中 $\alpha, \beta, \gamma \in \mathbb{R}$。注意，

$$\boldsymbol{\omega}_i^\top \widehat{\mathbf{A}}_{\text{nys}}^2 \boldsymbol{\omega}_i = (\widehat{\mathbf{A}}_{\text{nys}}\boldsymbol{\omega}_i)^\top (\widehat{\mathbf{A}}_{\text{nys}}\boldsymbol{\omega}_i) = (\mathbf{A}\boldsymbol{\omega}_i)^\top (\mathbf{A}\boldsymbol{\omega}_i) = \boldsymbol{\omega}_i^\top \mathbf{A}^2 \boldsymbol{\omega}_i,$$

其中，我们利用$\mathbf{A}^{\text{nys}}$ 的对称性和性质 (b)，将$\mathbf{A}_{\text{nys}}\boldsymbol{\omega}_i = \mathbf{A}\boldsymbol{\omega}_i$ 替换为适用于所有 $1 \le i \le k$ 的表达式。因此，

$$\boldsymbol{\omega}_i^\top q(\widehat{\mathbf{A}}_{\text{nys}})\boldsymbol{\omega}_i = \alpha(\boldsymbol{\omega}_i^\top \mathbf{A}^2 \boldsymbol{\omega}_i) + \beta(\boldsymbol{\omega}_i^\top \mathbf{A}\boldsymbol{\omega}_i) + \gamma(\boldsymbol{\omega}_i^\top \boldsymbol{\omega}_i) = \boldsymbol{\omega}_i^\top q(\mathbf{A})\boldsymbol{\omega}_i,$$

对所有 $1 \le i \le k$ 均成立。□

A.2. 理论方法概述

本节将介绍后续推导 funNyström 近似误差所需的理论框架。为说明我们的方法，首先简要讨论随机数值线性代数中的一个相关问题，即估计矩阵的最优秩 r 近似，其中$1 \le r \le k$。对于此类问题，利用 k 随机向量 (2.3) 得到的随机 Nyström 近似可以截断，忽略 $(k-r)$ A^{nys}的尾部特征值。由此得到的近似在此记为 $[\![\mathbf{A}_{\text{nys}}]\!]_r$。为随机草图分配多于所需秩的向量并进行截断，这一过程也称为过采样 [28]。基于随机低秩近似的迹估计技术通常不会忽略低秩近似中的任何特征值。尽管如此，我们仍可借鉴针对截断随机低秩逼近所建立的理论基础。这基于以下观察：对于任意酉不变范数 $|||\cdot|||$,

$$|||f(\mathbf{A}) - f(\widehat{\mathbf{A}}_{\text{nys}})||| \le |||f(\mathbf{A}) - f([\![\widehat{\mathbf{A}}_{\text{nys}}]\!]_r)|||, \tag{A.1}$$

对所有 $1 \le r \le k$ 均成立。这是因为 $[\![\mathbf{A}_{\text{nys}}]\!]_r \preceq \mathbf{A}_{\text{nys}} \preceq \mathbf{A}$，因此，根据 f 的算子单调性，有 $f([\![\mathbf{A}_{\text{nys}}]\!]_r) \preceq f(\mathbf{A}^{\text{nys}}) \preceq f(\mathbf{A})$，而范数的大小关系由 [51, 引理 3.1] 得出。因此，要界定 Nyström 近似 $\mathbf{A}^{\text{nys}}$ 的误差，只需界定截断Nyström近似 $[\![\mathbf{A}_{\text{nys}}]\!]_r$ 的误差。以下推导均假设采用下述设定。

设定 A.1。 对于 $1 \le r \le k$，考虑其谱分解

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \begin{bmatrix}\mathbf{U}_1 & \mathbf{U}_2\end{bmatrix}\begin{bmatrix}\mathbf{\Lambda}_1 & \\ & \mathbf{\Lambda}_2\end{bmatrix}\begin{bmatrix}\mathbf{U}_1^T \\ \mathbf{U}_2^T\end{bmatrix}, \tag{A.2}$$

其中 $\mathbf{\Lambda}_1 = diag(\lambda_1, \ldots, \lambda_r) \in \mathbb{R}^{r\times r}$, $\mathbf{\Lambda}_2 = diag(\lambda_{r+1}, \ldots, \lambda_n) \in \mathbb{R}^{(n-r)\times(n-r)}$, $U_1 \in \mathbb{R}^{n\times r}$ 和 $U_2 \in \mathbb{R}^{n\times(n-r)}$。我们将 *(A.2)* 中对特征值/向量的划分称为 A 的 r-划分。令 $\Omega \in \mathbb{R}^{n\times k}$ 为一个标准高斯随机矩阵，并令 $\Omega_1 = U_1^T\Omega \in \mathbb{R}^{r\times k}$ 且 $\Omega_2 = U_2^T\Omega \in \mathbb{R}^{(n-r)\times k}$。我们假设 $rank(\Omega_1) = r$，因为该性质几乎必然成立（例如，见引理3.1 [61,]）。最后，我们使用简写 $\rho := k - r \ge 0$ 表示过采样参数。

A.3. 随机矩阵理论的一些结果

本节概述高斯随机矩阵的一些预备结果。引理 A.2 给出了标准高斯随机矩阵的范数四阶矩期望界，涵盖不同的范数。接下来，引理 A.3 和 引理 A.4给出了随机正态矩阵乘积的范数性质。

引理 A.2 令 $\mathbf{H} \in \mathbb{R}^{r\times k}$ 为标准高斯随机矩阵，其中 $\rho = k - r \ge 4$。则，

$$\mathbb{E}[\|\mathbf{H}^\dagger\|_{(4)}^4] = \frac{r(k-1)}{\rho(\rho-1)(\rho-3)},$$

$$\mathbb{E}[\|\mathbf{H}^\dagger\|_F^4] = \frac{r^2\rho - 2r(r-1)}{\rho(\rho-1)(\rho-3)},$$

$$\mathbb{E}[\|\mathbf{H}^\dagger\|_2^4] \le \frac{e^4k^2}{(\rho+1)^3(\rho-3)}.$$

证明前两个等式由 [49, 引理 A.2]得出。我们着重证明最后一个不等式。由 [28, 命题 A.3],

$$\mathbb{P}\{\|\mathbf{H}^\dagger\|_2 > t\} \le C\, t^{-(\rho+1)}, \quad \text{where} \quad C = \frac{1}{\sqrt{2\pi(\rho+1)}}\left[\frac{e\sqrt{k}}{\rho+1}\right]^{\rho+1}.$$

因此，我们可以表示为

$$\mathbb{P}\{\|\mathbf{H}^\dagger\|_2^4 > t\} = \mathbb{P}\{\|\mathbf{H}^\dagger\| > t^{1/4}\} \le Ct^{-(\rho+1)/4}.$$

非负随机变量均值的积分公式意味着，对于所有 $E > 0$,

$$\begin{aligned}\mathbb{E}[\|\mathbf{H}^\dagger\|_2^4] = \int_0^\infty \mathbb{P}\{\|\mathbf{H}^\dagger\|_2^4 > t\}\, dt &\le E + \int_E^\infty \mathbb{P}\{\|\mathbf{H}^\dagger\|_2^4 > t\}\, dt \\ &\le E + \int_E^\infty Ct^{-(\rho+1)/4}\, dt \\ &\le E + \frac{4C}{\rho-3}E^{-(\rho-3)/4}.\end{aligned}$$

注意到，上界在 $E = C^{4/(\rho+1)}$ 处取得最小值。代入并化简后，我们得到

$$\mathbb{E}[\|\mathbf{H}^\dagger\|_2^4] \le \frac{e^4 k^2}{(\rho+1)^3(\rho-3)} \left[2\pi(\rho+1)\right]^{-2/(\rho+1)}.$$

代入 $\left[2\pi(\rho+1)\right]^{-2/(\rho+1)} \le 1$ 对所有 $\rho \ge 0$ 成立这一事实，即可得到所需的界。□

引理 A.3 设 $\mathbf{S}$ 为固定矩阵，$\mathbf{G}$ 和 $\mathbf{H}$ 为维度适当且相互独立的标准高斯矩阵。那么，

$$\mathbb{E}[\|\mathbf{S}\mathbf{G}\mathbf{H}^\dagger\|_2^4] \le 8\|\mathbf{S}\|_2^4 \mathbb{E}\|\mathbf{H}^\dagger\|_{\mathrm{F}}^4 + 16\|\mathbf{S}\|_2^4 \mathbb{E}\|\mathbf{H}^\dagger\|_{(4)}^4 + 8(\|\mathbf{S}\|_{\mathrm{F}}^4 + 2\|\mathbf{S}\|_{(4)}^4)\mathbb{E}\|\mathbf{H}^\dagger\|_2^4.$$

证明 以下证明主要基于 [55, 引理 B.1], 其内容如下：

$$(\mathbb{E}_{\mathbf{G}}[\|\mathbf{S}\mathbf{G}\mathbf{H}^\dagger\|_2^4])^{1/4} \le \|\mathbf{S}\|_2(\|\mathbf{H}^\dagger\|_{\mathrm{F}}^4 + 2\|\mathbf{H}^\dagger\|_{(4)}^4)^{1/4} + \|\mathbf{H}^\dagger\|_2(\|\mathbf{S}\|_{\mathrm{F}}^4 + 2\|\mathbf{S}\|_{(4)}^4)^{1/4}.$$

两边取四次方，并对 a，$b \ge 0$ 使用不等式 $(a+b)^4 \le 8(a^4+b^4)$，得到

$$\mathbb{E}_{\mathbf{G}}[\|\mathbf{S}\mathbf{G}\mathbf{H}^\dagger\|_2^4] \le 8\|\mathbf{S}\|_2^4(\|\mathbf{H}^\dagger\|_{\mathrm{F}}^4 + 2\|\mathbf{H}^\dagger\|_{(4)}^4) + 8\|\mathbf{H}^\dagger\|_2^4(\|\mathbf{S}\|_{\mathrm{F}}^4 + 2\|\mathbf{S}\|_{(4)}^4).$$

对 $\mathbf{H}$ 取期望，得到

$$\begin{aligned}
\mathbb{E}[\|\mathbf{S}\mathbf{G}\mathbf{H}^\dagger\|_2^4] &= \mathbb{E}_{\mathbf{H}}[\mathbb{E}_{\mathbf{G}}[\|\mathbf{S}\mathbf{G}\mathbf{H}^\dagger\|_2^4]] \\
&\le 8\|\mathbf{S}\|_2^4 \mathbb{E}_{\mathbf{H}}[\|\mathbf{H}^\dagger\|_{\mathrm{F}}^4 + 2\|\mathbf{H}^\dagger\|_{(4)}^4] + 8(\|\mathbf{S}\|_{\mathrm{F}}^4 + 2\|\mathbf{S}\|_{(4)}^4)\mathbb{E}_{\mathbf{H}}[\|\mathbf{H}^\dagger\|^4] \\
&= 8\|\mathbf{S}\|_2^4 \mathbb{E}[\|\mathbf{H}^\dagger\|_{\mathrm{F}}^4] + 16\|\mathbf{S}\|_2^4 \mathbb{E}[\|\mathbf{H}^\dagger\|_{(4)}^4] + 8(\|\mathbf{S}\|_{\mathrm{F}}^4 + 2\|\mathbf{S}\|_{(4)}^4)\mathbb{E}[\|\mathbf{H}^\dagger\|^4].
\end{aligned}$$

□

引理 A.4 考虑一个固定矩阵$\mathbf{S}$以及维数适当的独立标准高斯矩阵$\mathbf{G}$和$\mathbf{H}$，其中$\mathbf{H} \in \mathbb{R}^{r\times k}$ 且 $\rho = k - r \ge 4$。则，

$$\mathbb{E}[\|\mathbf{S}\mathbf{G}\mathbf{H}^\dagger\|_2^4] \le \frac{8r(r+2)}{(\rho-1)(\rho-3)}\|\mathbf{S}\|_2^4 + \frac{8e^4k^2}{(\rho+1)^3(\rho-3)}(\|\mathbf{S}\|_{\mathrm{F}}^4 + 2\|\mathbf{S}\|_{(4)}^4).$$

证明 将 $\mathbb{E}[\|\mathbf{H}^\dagger\|_F^4]$、$\mathbb{E}[\|\mathbf{H}^\dagger\|_{(4)}^4]$和 $\mathbb{E}\|\mathbf{H}^\dagger\|_2^4]$ 在引理 A.2中的界代入引理 A.3即可得到该结果。□

A.4. 低秩近似的概率界

本节给出 funNyström 方法的平方近似误差的概率界。首先借助结构界，引理 A.5，将该量分解为一个确定性量与一个概率量之和。

引理 A.5 考虑 r对$\mathbf{A}$在设定 *A.1*下的划分。则，

$$|||f(\mathbf{A}) - f([\![\widehat{\mathbf{A}}_{\text{nys}}]\!]_r)|||^2 \leq 2|||f(\mathbf{\Lambda}_2)|||^2 + 2|||f((\mathbf{\Omega}_2\mathbf{\Omega}_1^\dagger)^\top \mathbf{\Lambda}_2(\mathbf{\Omega}_2\mathbf{\Omega}_1^\dagger))|||^2,$$

其中 $|||\cdot|||$ 表示任意酉不变范数。

证明 直接由 [51, 引理 3.13] 以及如下事实得到：$(\alpha+\beta)^2 \leq 2\alpha^2 + 2\beta^2$ 对所有非负标量 α, $\beta \geq 0$成立。□

我们希望对引理 A.5中该界的第二项得到期望意义下的界。我们在引理 A.7 中证明这样的界，随后给出本节的主要结果定理 A.8。为此，我们首先在引理 A.6中陈述关于算子单调函数的一个初等性质。

引理 A.6 对于算子单调函数 $f\colon [0,\infty) \mapsto [0,\infty)$,

$$f(tx) \leq \max\{t,1\} f(x),$$

对所有 $t, x \geq 0$。

证明该结果是 [52,引理 2.1-2.2]的直接推论。 □

引理 A.7 考虑 r的分块 $\mathbf{A}$ 在 设定 *A.1* 下，过采样量为 $\rho = k - r \geq 4$。则，

$$\begin{aligned}\mathbb{E}\Big[|||f((\mathbf{\Omega}_2\mathbf{\Omega}_1^\dagger)^\top \mathbf{\Lambda}_2(\mathbf{\Omega}_2\mathbf{\Omega}_1^\dagger))|||^2\Big] \leq \Bigg[1 + \frac{8r(r+2)}{(\rho-1)(\rho-3)}\|\mathbf{\Lambda}_2\|_2 \\ + \frac{8e^4k^2}{(\rho+1)^3(\rho-3)}(\|\mathbf{\Lambda}_2^{1/2}\|_*^2 + 2\|\mathbf{\Lambda}_2\|_*)\Bigg] |||f(\mathbf{\Lambda}_2^{1/2})|||^2.\end{aligned}$$

证明 令 $\mathbf{Z}:=\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{\dagger}\in\mathbb{R}^{(n-r)\times k}$。注意，对于任意对称半正定矩阵 $\mathbf{A}$ 和 $\mathbf{D}$，$\mathbf{D}^{1/2}\mathbf{A}\mathbf{D}^{1/2}\preceq\|\mathbf{A}\|_2\,\mathbf{D}$。由于 f 是算子单调的，

$$f(\mathbf{Z}^\top\boldsymbol{\Lambda}_2\mathbf{Z})=f(\mathbf{Z}^\top\boldsymbol{\Lambda}_2^{1/4}\boldsymbol{\Lambda}_2^{1/2}\boldsymbol{\Lambda}_2^{1/4}\mathbf{Z})\preceq f(\|\boldsymbol{\Lambda}_2^{1/4}\mathbf{Z}\|_2^2\boldsymbol{\Lambda}_2^{1/2}).$$

由于 $\boldsymbol{\Lambda}_2$ 是具有非负元素的对角矩阵，我们可以应用引理 A.6得到，

$$f(\mathbf{Z}^\top\boldsymbol{\Lambda}_2\mathbf{Z})\preceq f(\|\boldsymbol{\Lambda}_2^{1/4}\mathbf{Z}\|_2^2\boldsymbol{\Lambda}_2^{1/2})\preceq\max\{\|\boldsymbol{\Lambda}_2^{1/4}\mathbf{Z}\|_2^2,1\}f(\boldsymbol{\Lambda}_2^{1/2}).$$

因此，根据 [51, 引理 3.1]，

$$|||f(\mathbf{Z}^\top\boldsymbol{\Lambda}_2\mathbf{Z})|||\le\max\{\|\boldsymbol{\Lambda}_2^{1/4}\mathbf{Z}\|_2^2,1\}|||f(\boldsymbol{\Lambda}_2^{1/2})|||.$$

将两边平方，用总和界定最大值，并取期望，可得

$$\mathbb{E}\Big[|||f(\mathbf{Z}^\top\boldsymbol{\Lambda}_2\mathbf{Z})|||^2\Big]\le\Big(\mathbb{E}\Big[\|\boldsymbol{\Lambda}_2^{1/4}\mathbf{Z}\|_2^4\Big]+1\Big)|||f(\boldsymbol{\Lambda}_2^{1/2})|||^2.$$

应用引理 A.4，

$$\mathbb{E}\Big[\|\boldsymbol{\Lambda}_2^{1/4}\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{\dagger}\|_2^4\Big]\le\frac{8r(r+2)}{(\rho-1)(\rho-3)}\|\boldsymbol{\Lambda}_2\|_2+\frac{8e^4k^2}{(\rho+1)^3(\rho-3)}(\|\boldsymbol{\Lambda}_2^{1/2}\|_*^2+2\|\boldsymbol{\Lambda}_2\|_*),$$

并由此得到所需结果。□

定理 A.8 考虑 r 对 $\mathbf{A}$ 在 设定 *A.1* 下进行的过采样 $\rho=k-r\ge4$。那么，

$$\mathbb{E}\Big[|||f(\mathbf{A})-f([\![\widehat{\mathbf{A}}_{\mathrm{nys}}]\!]_r)|||^2\Big]\le2|||f(\boldsymbol{\Lambda}_2)|||^2+\Bigg[2+\frac{16r(r+2)}{(\rho-1)(\rho-3)}\|\boldsymbol{\Lambda}_2\|_2+\frac{16e^4k^2}{(\rho+1)^3(\rho-3)}(\|\boldsymbol{\Lambda}_2^{1/2}\|_*^2+2\|\boldsymbol{\Lambda}_2\|_*)\Bigg]|||f(\boldsymbol{\Lambda}_2^{1/2})|||^2.$$

证明 此结果可由引理 A.5 和 A.7 得出。□

定理 定理 *3.6*的证明。 此结果可通过将 $r=k-4$ 代入 定理 A.8 并结合(A.1) 得出。□