arXiv:2305.01942v1 [cs.DS] 3 May 2023

任意 p-范数的实验设计

LapChi Lau*、Robert Wang†、Hong Zhou‡

摘要

我们考虑实验设计问题中的一般 p-范数目标函数，其中涵盖了一些研究充分的目标函数（D/A/E-设计）作为特殊情况。我们证明，随机局部搜索方法提供了一种统一算法，可用于求解所有 p 的这一问题。这为一般 p-范数目标函数提供了首个近似算法，并对特殊情况中已知的最佳界限作出了巧妙的插值。

*计算机科学学院，滑铁卢大学。电子邮件：lapchi@uwaterloo.ca
†计算机科学学院，滑铁卢大学。电子邮件：robert.wang2@uwaterloo.ca
‡数学与统计学院，福州大学。电子邮箱：hong.zhou@fzu.edu.cn

1 引言

在实验设计问题中，我们给定向量 $\mathbf{v}_1、\ldots、\mathbf{v}_n \in \mathbb{R}^d$ 以及预算 $k \geq d$，目标是从 k 个向量中选择一个子集 S，使得 $\sum_{i\in S} \mathbf{v}_i\mathbf{v}_i^\top$能够优化某个衡量输入数据“多样性”的目标函数。最常见且研究最为深入的目标函数包括：

- D-设计：最大化 $(\det(\sum_{i\in S}\mathbf{v}_i\mathbf{v}_i^\top))^{\frac{1}{d}}$。
- A-设计：最小化 $\frac{1}{d}\operatorname{tr}((\sum_{i\in S}\mathbf{v}_i\mathbf{v}_i)^{-1})$。
- E-设计：最大化 $\lambda_{\text{最小值}}(\sum_{i\in S}\mathbf{v}_i\mathbf{v}_i^\top)$。

实验设计问题有着悠久的历史和广泛的应用，涵盖从统计学、机器学习、数值线性代数到图算法等领域。关于这些应用的更多信息，我们请读者参阅 [13, 11, 2, 6, 7] 及其中的参考文献。

尽管 D/A/E-设计的目标看起来大不相同，但我们观察到，利用特征值可以进行一种自然的推广，将这三个目标函数都作为特殊情形囊括其中。

定义 1.1（逆特征值的 p-范数）。给定一个 d 维实对称矩阵 A，其特征值为 $\lambda_1, \ldots, \lambda_d > 0$，以及一个自然数 $0 \leq p \leq \infty$，我们定义

$$\Phi_p(A) := \left(\frac{1}{d}\operatorname{tr}\left(A^{-p}\right)\right)^{\frac{1}{p}} = \left(\frac{1}{d}\sum_{i=1}^{d}\lambda_i^{-p}\right)^{\frac{1}{p}}, \tag{1.1}$$

其中 $\Phi_0(A) := \lim_{p\to 0^+}\Phi_p(A)$，且 $\Phi_\infty(A) := \lim_{p\to+\infty}\Phi_p(A)$。

给定 $p \geq 0$，我们将目标函数为 Φ_p 的实验设计问题称为 Φ_p-设计。为了说明 Φ_p-设计是 D/A/E-设计的推广，令 $A = \sum_{i\in S}\mathbf{v}_i\mathbf{v}_i^T$，并注意：

- 对于 $p = \infty$，$\Phi_\infty(A) = \lambda_{\text{最大值}}(A^{-1}) = 1/\lambda_{\text{最小值}}(A)$，这就是 E-设计目标函数的倒数；
- 对于 $p = 1$，$\Phi_1(A) = \frac{1}{d}\operatorname{tr}(A^{-1})$ 恰好就是 A-设计目标函数；
- 对于 $p = 0$，$\Phi_0(A)$ 是 D-设计目标函数的倒数，因为

$$\Phi_0(A) = \lim_{p\to 0^+}\left(\frac{1}{d}\sum_{i=1}^{d}\lambda_i^{-p}\right)^{1/p} = \left(\prod_{i=1}^{d}\lambda_i^{-1}\right)^{1/d} = \det(A)^{-1/d},$$

其中 the 第二个等式是一个众所周知的事实（例如，参见 [16]第5章中的练习28 ）.

已知，对于任意给定的 $0 \le p \le \infty$， Φ_p(A)关于A是凸的，因此，以下是 Φ_p-设计的自然凸规划松弛：

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & \Phi_p\left(\sum_{i=1}^{n} x(i)\cdot v_i v_i^\top\right) \\ \text{subject to} \quad & \sum_{i=1}^{n} x(i) \le k, \\ & 0 \le x(i) \le 1, \qquad \text{for } 1 \le i \le n. \end{aligned} \tag{1.2}$$

据我们所知，除以下特殊情形 $p = 0,1,\infty$ 外，目前尚无适用于一般 Φ_p-设计问题的已知近似算法。我们总结如下（记号 $x \gtrsim y$ 表示，对于某个足够大的常数 c，有 $x \ge cy$ ）：

1. 当 $k \gtrsim d/\varepsilon$ [13, 9, 11, 7]时，存在一个针对 Φ_0-设计（D-设计）的（1+ε）近似算法。2. 当 $k \gtrsim d/\varepsilon$ [9, 11, 7]时，存在一个针对 Φ_1-设计（A-设计）的（$1+\varepsilon$）近似算法。3. 当 $k \gtrsim d/\varepsilon^2$ [1, 2, 6]时，存在一个针对 Φ_∞-设计（E-设计）的（$1+\varepsilon$）近似算法。

这些结果是紧的，因为它们与凸规划松弛的已知整数间隙下界相符（1.2）（整数间隙示例见 [11] ）。

注意， d/ε 与 d/ε^2 之间存在差距：D/A-设计的松弛（$p = 0,1$）与 E-设计的松弛（$p = \infty$）之间存在差距。本文研究的主要问题是：凸规划松弛（1.2）的整数间隙如何随 p的取值变化？特别是，从 d/ε 到 d/ε^2 的转变发生在何处?

1.1 主要结果

我们的主要结果是，当 $k \gtrsim \min\{dp/\varepsilon, d/\varepsilon^2\}$ 时，存在一个用于Φ_p-设计的 $(1+\varepsilon)$-近似算法。

定理 1.2。给定一个整数 $p \ge 1$，令 x $\in [0,1]^n$ 为 (1.2) 的一个最优分数解。对于任意 $\varepsilon \in$(0, 1), 令 $\gamma = \max\{\varepsilon, 1/p\}$，若 $k \gtrsim d/(\gamma\varepsilon)$，则存在一个随机多项式时间算法，返回一个整数解 Z $= \sum_{i=1}^{n}$ z(*i*) $\cdot$ v$_i$v$_i^\top$，且对于满足以下条件的 $1 \le i \le n$，有 z(*i*) $\in \{0,1\}$

$$\Phi_p\left(\sum_{i=1}^{n} z(i)\cdot v_i v_i^\top\right) \le (1+\varepsilon)\cdot\Phi_p\left(\sum_{i=1}^{n} x(i)\cdot v_i v_i^\top\right) \quad \textit{and} \quad \sum_{i=1}^{n} z(i) \le k,$$

其概率至少为 $1 - O\left(\frac{(d^2+d/\gamma)^2}{\varepsilon^2 p^2}\cdot e^{-\Omega(\gamma\sqrt{d})}\right) - e^{-\Omega(\varepsilon d/\gamma)}$。

备注。 定理中的结果 *1.2* 可以推广到所有实数 $p \ge 1$，详见备注 *3.2*，位于第 *3*节。

这是首个用于 Φ_p-设计的、适用于一般 p的近似算法。定理1.2 表明，Φ_p-设计对于常数 p具有与D/A-设计同样好的近似算法，并且有一种统一算法可以实现这一保证。

注意，当 $p \to +\infty$， Φ_p 就变成了 E-设计目标函数，而 $\gamma = \max\{\varepsilon, 1/p\} = \varepsilon$。因此，定理 1.2 在 D/A-设计的 d/ε 界与 E-设计的 d/ε^2界之间实现了很好的插值。

我们进一步指出，我们的结果可以推广到加权情形，以处理 [7], 中的多个预算/背包约束；但为使表述更简洁，我们省略了细节，因为它们与 [7]中的相同。

定理 1.2 的证明建立在 [7] 和 [6], 中的随机局部搜索方法之上，但为了处理高阶p-范数引入的高阶矩，还需要一些新的技术思路。在第 2节，我们将回顾背景和先前的工作，介绍算法和分析的整体结构，并阐释本工作的新思路。随后，在第 3节，我们将详细介绍 p-范数实验设计。

1.2 相关工作

本文着重于将 D/A/E-设计推广为采用 p-范数目标函数的形式（见 1.1）。不过，推广 D/A/E-设计还有其他不同方式。例如，实验设计的贝叶斯框架 [3] 通过在应用目标函数之前，向协方差矩阵中加入一个固定矩阵 B （该矩阵编码了一些先验信息，例如单位矩阵的倍数），从而扩展了这一问题。当 $B = 0$时，我们便得到经典的实验设计问题。Tantipongpi- pat [14] 针对 B 为单位矩阵倍数的贝叶斯 A-设计问题，提出了一种近似算法。Derezi´nski 等人 [4]研究了采用一般半正定矩阵 B的贝叶斯 A/D-设计（以及另外两个目标函数，即 C/V-设计）的近似算法。另一种推广 A/D-设计的方式是使用初等对称多项式 [10]。Nikolov 等人 [11] 将其用于 A-设计的近似算法推广到这一类广义比率目标函数。

这些都是 D/A/E-设计有趣的推广。值得探究的是， [6, 7] 和本文中的随机局部搜索方法能否应用于这些情形，以取得更好的结果。

2 框架

本节首先回顾第 2.1 节中的随机局部搜索方法。然后，我们在第 2.2 节给出完整算法，并陈述主要技术定理。接着，在第 2.3 节中，我们概述证明计划，并给出分析随机交换算法的精确陈述，随后将这些陈述综合起来证明主要技术定理。最后，我们在第 Φ_p 节讨论分析 3 中目标函数所需面对的主要困难和新思路。

2.1 随机局部搜索方法

定理 1.2 的证明建立在 [7], 中的随机局部搜索方法之上。该方法基于 [2] 为实验设计问题提出的遗憾最小化框架，以及 [6] 中的随机谱舍入技术。本小节将回顾这一方法。

在 [7], 第一步是求解（1.2）以进行 D/A/E-设计，从而得到一个解$x \in \mathbb{R}^n$，然后对向量v_i进行归一化，使得 $\sum_{i=1}^n x(i) \cdot v_i v_i^\top = I$ 以便在 [2]中使用遗憾最小化框架。然后，舍入算法从一个随机初始解集 S_0 开始，该解集按照x独立采样得到。算法利用遗憾最小化框架在每一步维护的密度矩阵A_t，随机选择一对向量v_{i_t} 和v_{j_t}，其概率分布如下：

$$\Pr(i_t = i) \propto \left(1 - x(i)\right) \cdot \left(1 - \alpha \langle v_i v_i^\top, A_t^{1/2} \rangle\right) \quad \text{and} \quad \Pr(j_t = j) \propto x(j) \cdot \left(1 + \alpha \langle v_j v_j^\top, A_t^{1/2} \rangle\right), \tag{2.1}$$

并设置 $S_{t+1} \leftarrow S_t - \mathbf{v}_{i_t} + \mathbf{v}_{j_t}$。利用上述随机局部搜索策略，可以证明解的大小预期会保持在 k 附近，而遗憾最小化框架中与最小特征值相关的势函数 [2] 预期会得到改善，只要最小特征值小于 $1 - \varepsilon$。在 [6] 中，使用 Freedman 鞅不等式和一种新的非鞅集中不等式，证明所有这些量以高概率接近其期望值。该随机局部搜索算法为 D/A/E-设计提供了一个 $(1 \pm \varepsilon)$-近似解，其中 $k \gtrsim d/\varepsilon^2$。

在 [7] 中的主要贡献是证明，随机局部搜索算法可以进行调整，从而在$1 \pm \varepsilon$时为 D/A-设计实现（$k \gtrsim d/\varepsilon$）-近似，进而为实现 D/A/E-设计的最优整数间隙提供一种统一方法。从本质上讲，该算法与 E-设计所用的算法相同，但只要求解的最小特征值为 $3/4$，而不是 $1 - \varepsilon$。在分析中，随机局部搜索算法在概念上被分为两个阶段。在第一阶段，当 $k \gtrsim d/\varepsilon$时，算法将以高概率找到最小特征值至少为 $3/4$的解。在第二阶段，最小特征值将以高概率保持至少为 $1/4$，并且 D/A-设计的目标值将在多项式时间内以高概率提升至最优目标值的（$1 \pm \varepsilon$）倍。第一阶段的分析直接沿用了 [6] 中关于谱舍入的早期研究。第二阶段的分析包括两个主要部分：（1）证明在期望意义下，E-设计在（2.1）中的概率分布同样有利于推动 D/A-设计目标的改进；（2）利用鞅集中性论证，证明目标值的改进集中在其期望附近。在第二阶段的这两个部分中，最小特征值至少为 $1/4$这一条件都非常重要，并且凸规划（1.2）的最优性条件在第二阶段的第二部分中起着关键作用。

在本文中，我们将扩展上述算法和分析，并证明随机局部搜索算法为所有 p 的 Φ_p-设计提供了一种统一的方法。

2.2 算法

本小节将给出 Φ_p-设计的完整算法。

随机交换算法

Input: m 向量 $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{R}^d$, 预算 $k \geq d$, 精度参数 $\varepsilon \in (0,1)$.

1. 求解凸规划松弛 (1.2), 并得到一个最优解 $\mathbf{x} \in [0,1]^m$, 其至多包含 d^2+1 个分数项, 即 $|\{i \in [m] \mid 0 < \mathbf{x}(i) < 1\}| \leq d^2+1$。令 $\mathbf{X} = \sum_{i=1}^m \mathbf{x}(i) \cdot \mathbf{u}_i \mathbf{u}_i^\top$。 2. 预处理: 令 $\mathbf{v}_i \leftarrow \mathbf{X}^{-1/2}\mathbf{u}_i$, 对所有 $i \in [m]$, 使得 $\sum_{i=1}^m \mathbf{x}(i) \cdot \mathbf{v}_i \mathbf{v}_i^\top = \mathbf{I}_d$。 3. 初始化: $t \leftarrow 1$, $S_0 \leftarrow \emptyset$, $\gamma = \max\{\frac{\varepsilon}{6}, \frac{1}{6p}\}$, $\kappa = \max\{\frac{\varepsilon}{2}, \frac{1}{2p}\}$, $M \leftarrow \frac{d}{\gamma} + d^2 + 1$, 以及 $\alpha \leftarrow \frac{\sqrt{d}}{\gamma}$。 4. 对每个 $i \in [m]$, 以概率 $\mathbf{x}(i)$ 独立地将 i 加入 S_0。令 $\mathbf{Y}_1 \leftarrow \sum_{i \in S_0} \mathbf{u}_i \mathbf{u}_i^\top$, $\mathbf{Z}_1 \leftarrow \sum_{i \in S_0} \mathbf{v}_i \mathbf{v}_i^\top$。 5. 当终止条件 $(\mathrm{tr}(\mathbf{Y}_t^{-p}))^{\frac{1}{p}} \leq (1+\varepsilon)(\mathrm{tr}(\mathbf{X}^{-p}))^{\frac{1}{p}}$ 尚未满足且 $t \leq \frac{2M}{\gamma} + \frac{2M}{\varepsilon p}$ 时, 执行以下操作:

   (a) $S_t \leftarrow \mathrm{Exchange}(S_{t-1})$.

   (b) Set $Y_{t+1} \leftarrow \sum_{i \in S_t} u_i u_i^\top$, $Z_{t+1} \leftarrow \sum_{i \in S_t} v_i v_i^\top$ and $t \leftarrow t+1$.

6. 返回 S_{t-1} 作为解。

主算法与 [7], 中的算法几乎相同, 但多了两个参数 γ, κ这两个参数取决于 p的值。对于 A-设计（当 $p = 1$时）, [7] 中的算法只是参数为 $\gamma = \frac{1}{8}$ 和 $\kappa = 1$的随机交换算法的一个特例。本文中, 参数 γ 用于调整遗憾最小化框架的学习率 α , 而参数 κ 将用于交换子程序, 具体如下所述。

交换子程序

输入: 当前解集 S_{t-1}。

1. 计算动作矩阵 $\mathbf{A}_t := (\alpha \mathbf{Z}_t - c_t \mathbf{I})^{-2}$, 其中 $\mathbf{Z}_t = \sum_{i \in S_{t-1}} \mathbf{v}_i \mathbf{v}_i^\top$, 且 c_t是唯一满足 $\mathbf{A}_t \succ 0$ 且 $\mathrm{tr}(\mathbf{A}_t) = 1$的标量。

2. 定义 $S_t' := \left\{ i \in S_{t-1} \mid \alpha \langle \mathbf{v}_i \mathbf{v}_i^\top, \mathbf{A}_t^{1/2} \rangle \leq \frac{1}{2} \text{and } \langle \mathbf{v}_i \mathbf{v}_i^\top, \mathbf{Z}_t^{-1} \rangle \leq \kappa \right\}$。

3. 从以下概率分布中采样 $j_t \in [m] \backslash S_{t-1}$

$$\Pr(j_t = j) = \frac{x(j)}{M} \cdot \left(1 + \alpha \langle v_j v_j^\top, A_t^{1/2} \rangle\right), \text{ for } j \in [m] \backslash S_{t-1} \text{ and}$$
$$\Pr(j_t = \emptyset) = 1 - \sum_{j \in [m] \backslash S_{t-1}} \frac{x(j)}{M} \cdot \left(1 + \alpha \langle v_j v_j^\top, A_t^{1/2} \rangle\right).$$

4. 从以下概率分布中抽取样本 $i_t \in S'_{t-1}$

$$\Pr(i_t = i) = \frac{1 - x(i)}{M} \cdot \left(1 - \alpha \langle v_i v_i^\top, A_t^{1/2} \rangle\right), \text{ for } i \in S'_{t-1} \text{ and}$$
$$\Pr(i_t = \emptyset) = 1 - \sum_{i \in S'_{t-1}} \frac{1 - x(i)}{M} \cdot \left(1 - \alpha \langle v_i v_i^\top, A_t^{1/2} \rangle\right).$$

5. 返回 $S_t := S_{t-1} \cup \{j_t\} \backslash \{i_t\}$。

交换子程序也几乎与 [7]中的相同。关键区别在于使用新参数 κ 进一步限制允许从当前解中移除的向量集合。

备注 2.1。 利用本文中的新分析，*[7]* （无需更改参数 γ和 κ）可实现 $(1+\varepsilon)$-近似的 Φ_p-设计，当 $k \gtrsim 2^{O(p)} d/\varepsilon$时，预算需求对 p 呈指数依赖，远差于定理 $k \gtrsim \min\{dp/\varepsilon, d/\varepsilon^2\}$中的 *1.2*。

2.2.1 主要技术定理

为了证明定理1.2，我们将证明随机交换算法是一种双准则近似算法，使其返回的解在$1+\varepsilon$目标函数上是 (Φ_p)-近似的，并且解的规模不会比 k大太多。

定理 2.2。给定 $\varepsilon \in (0,1)$，如果 $k \gtrsim \frac{d}{\gamma\varepsilon}$，那么随机交换算法会在 $\frac{2M}{\gamma} + \frac{2M}{\varepsilon p}$ 次迭代内返回一个满足 S 的解集。

$$\left(\operatorname{tr}\left(\left(\sum_{i \in S} u_i u_i^\top\right)^{-p}\right)\right)^{\frac{1}{p}} \leq (1+\varepsilon) \cdot \left(\operatorname{tr}\left(X^{-p}\right)\right)^{1/p}$$

概率至少为 $1 - O\left(\frac{M^2}{\varepsilon^2 p^2} \cdot e^{-\Omega}(\gamma\sqrt{d})\right)$，其中 X 是 (1.2) 的最优分数解。此外，解集 S满足 $|S| \leq (1+\varepsilon)k + O\left(\frac{d}{\gamma} + \frac{d}{\kappa}\right)$ 的概率至少为 $1 - e^{-\Omega(\varepsilon d/\min\{\gamma,\kappa\})}$。

利用定理 2.2，我们可以证明定理 1.2 ，首先注意到两者 γ 和 κ 都是按照 Θ(最大$\{\varepsilon, 1/p\}$) 的顺序选取的，然后通过与 [7]中相同的缩放论证，将双标准近似结果转化为真正的近似结果。标准证明从略，详情请参阅 [7] 。

2.3 证明思路

本小节给出分析随机交换算法的整体思路和精确表述，并将这些表述结合起来，得到主要技术定理。

2.3.1 定义良好的算法

首先，我们证明随机交换算法定义良好。具体而言，我们需要证明，可以在多项式时间内找到凸松弛问题（1.2）的一个分数最优解，其中至多有 $O(d^2$ 个分数项；还需要证明，对于 $M = d^2 + \frac{d}{\gamma} + 1$, 交换子程序中的概率分布定义良好。这些结论可通过采用引理 4.2 和命题 4.4 中相同的论证（但取不同的 γ 值）得到。注意，修改后的交换子程序不会影响这些论证。

下面这个简单的观察（[7] 中的观察 4.3）将在算法分析中发挥作用。

观察 2.3. 对于任意 $t \geq 0$，对于所有满足 $\mathbf{x}(i) = 1$ 的 i ，都有 $i \in S_t$ ；对于所有满足 $\mathbf{x}(j) = 0$ 的 j ，都有 $j \in [n] \backslash S_t$。进一步可得，对于所有满足 $\mathbf{x}(i) \in \{0, 1\}$ 的 i ，都有 $\Pr(i_t = i) = 0$ ；对于所有满足 $\mathbf{x}(j) \in \{0, 1\}$ 的 j ，都有 $\Pr(j_t = j) = 0$ 。

2.3.2 解的规模界

接下来，我们将证明该算法以高概率返回一个大小不超过 k 太多的解集 S 。

定理 2.4（ [6] 中定理 3.12 的变体）。令 $\alpha = \sqrt{d}/\gamma$ 和 κ 为随机交换算法中使用的参数。假设随机交换算法的解 S_t 对所有 $t \in [\tau]$都满足 $\lambda_{\min}\left(\sum_{i \in S_t} \mathbf{v}_i \mathbf{v}_i^\top\right) < 1$。那么，对于任意给定的 $\delta \in [0, 1]$,

$$\Pr\left[|S_\tau| \leq (1+\delta) \cdot \sum_{i=1}^{n} x(i) + \left(\frac{12d}{\gamma} + \frac{2d}{\kappa}\right)\right] \geq 1 - \exp\left(-\Omega\left(\frac{\delta d}{\min\{\gamma, \kappa\}}\right)\right).$$

当 $\kappa = 1$时，上述定理可直接由 [6] 中的单侧谱舍入结果推出。较小的 κ 会限制可从当前解中换出的向量集合。这会增加无法移除某个向量的可能性，从而增大解的规模。所幸，我们可以证明，解规模的增加可由一个加法 d/κ 项界定。证明思路与 [6], 中的类似，主要区别在于对规模期望变化的界进行了修改。为完整起见，我们在附录B 中给出证明概要。

2.3.3 近似保证

证明中最具技术性的部分是建立近似保证。我们遵循 [7] 中的分析思路，从概念上将算法的执行分为两个阶段，如第 2.1 节所述。

在第一阶段，我们证明最小特征值以高概率将在 $1-2\gamma$ 次 $O(M/\gamma)$ 迭代内达到该值，这可由 [6] 中的谱舍入结果推出。

命题 2.5（[7] 中命题 4.5 的对应命题）。随机交换算法在 $2M/\gamma$ 次迭代内成功终止，或者存在满足 $\lambda_{\min}(Z_{\tau_1}) \geq 1-2\gamma$ 的 $\tau_1 \leq 2M/\gamma$ 的概率至少为 $1-\exp(-\Omega(\sqrt{d}))$。

在第二阶段，接下来的 $\Theta\left(\frac{M}{\varepsilon p}\right)$ 次迭代中，最小特征值以较高概率至少为 $1-5\gamma$。

命题 2.6（[7] 中命题 4.6 的对应命题）。假设对于某个 τ_1，$\lambda_{\min}(Z_{\tau_1}) \geq 1-2\gamma$。在随机交换算法中，对于所有 $\tau_1 \leq t \leq \tau_1 + \frac{2M}{\varepsilon p}$，$\lambda_{\min}(Z_t) \geq 1-5\gamma$ 的概率至少为 $1-\frac{4M^2}{\varepsilon^2 p^2}\cdot e^{-\Omega(\gamma\sqrt{d})}$。

命题 2.5 和命题 2.6 的证明都沿用了 [7] 中的相同论证（其中引入了一个新参数 γ）。我们指出，采用受限 S'_t 的修改版交换子程序不会影响这些论证，因为移除较少的向量只会有助于提高解的最小特征值。证明从略，读者可参阅 [7]。

本文的主要技术贡献在于证明：在第二阶段，当最小特征值至少为 $1-5\gamma$ 时，Φ_p 目标值将改善至不超过最优值的 $(1+\varepsilon)$ 倍。

定理 2.7. 给定 $\varepsilon \in (0,1)$，若 $p \leq 1/\varepsilon$ 且 $k \gtrsim \frac{pd}{\varepsilon}$ 对某个 $\varepsilon \in (0,1)$ 成立，则在随机交换算法执行期间，以下三个事件同时发生的概率至多为 $\exp(-\Omega(\gamma\sqrt{d}))$。

1. 对某个 τ_1，$\lambda_{\min}(Z_{\tau_1}) \geq 1-2\gamma$；2. 对所有 $\tau_1 \leq t \leq \tau_1 + \frac{2M}{\varepsilon p}$，$\lambda_{\min}(Z_t) \geq 1-5\gamma$；3. 随机交换算法在时间 $\tau_1 + \frac{2M}{\varepsilon p}$ 之前尚未终止。

2.3.4 定理证明2.2

我们将本小节中的各项陈述汇总起来，得到定理 2.2。

定理 *2.2* 的证明。 我们首先分析定理中的近似保证。

首先，考虑较为简单的情形 $p \geq 1/\varepsilon$，这意味着 $\gamma = \varepsilon/6$。根据命题 2.5，存在 $\tau_1 \leq 2M/\gamma$ 使得 $\lambda_{\min}(Z_{\tau_1}) \geq (1-\varepsilon/3)$ 以 $1-\exp(-\Omega(\sqrt{d}))$ 的概率成立。我们注意到 $\lambda_{\min}(Z_{\tau_1}) \geq (1-\varepsilon/3)$ 等价于 $Y_{\tau_1} \succcurlyeq (1-\varepsilon/3)X$，这足以证明

$$(\mathrm{tr}(Y_{\tau_1}^{-p}))^{1/p} \leq (1+\varepsilon)(\mathrm{tr}(X^{-p}))^{1/p},$$

也就是说，该定理中的近似保证。我们注意到，在此情形的证明中不需要假设$k \gtrsim d/(\gamma\varepsilon)$。

接下来，我们考虑 $p \leq 1/\varepsilon$ 的情形，并为随机交换算法定义坏事件：

- B_1：算法在 $2M/\gamma$ 次迭代内未能成功终止，并且 $\tau_1 > 2M/\gamma$，其中 τ_1 是满足 $\lambda_{\min}(Z_{\tau_1}) \geq 1-2\gamma$ 的最早时刻。
- B_2：存在某个$\tau_1 \leq t \leq \tau_1 + \frac{2M}{\varepsilon p}$，使得 $\lambda_{\min}(Z_t) < 1-5\gamma$。
- B_3：对于所有 $\tau_1 \leq t \leq \tau_1 + \frac{2M}{\varepsilon p}$，终止条件 $(\mathrm{tr}(Y_t^{-p}))^{\frac{1}{p}} \leq (1+\varepsilon)(\mathrm{tr}(X^{-p}))^{\frac{1}{p}}$均未满足。

如果坏事件均未发生，那么算法要么在$2M/\gamma$次迭代内成功终止，要么终止条件会在某个时刻 $t \leq \tau_1 + \frac{2M}{\varepsilon p} \leq \frac{2M}{\gamma} + \frac{2M}{\varepsilon p}$得到满足。因此，随机交换算法在 $\frac{2M}{\gamma} + \frac{2M}{\varepsilon p}$ 次迭代内尚未满足终止条件的概率上界为

$$\begin{aligned}
\Pr[B_1 \cup B_2 \cup B_3] &= \Pr[B_1] + \Pr[B_2 \cap \neg B_1] + \Pr[B_3 \cap \neg B_2 \cap \neg B_1] \\
&\leq O\left(e^{-\Omega(\sqrt{d})}\right) + O\left(\frac{M^2}{\varepsilon^2 p^2} \cdot e^{-\Omega(\gamma\sqrt{d})}\right) + O\left(e^{-\Omega(\gamma\sqrt{d})}\right) \\
&\leq O\left(\frac{M^2}{\varepsilon^2 p^2} \cdot e^{-\Omega(\gamma\sqrt{d})}\right),
\end{aligned}$$

其中 $\Pr[B_1]$ 由命题 2.5给出界，$\Pr[B_2 \cap \neg B_1]$ 由命题 2.6给出界，而 $\Pr[B_3 \cap \neg B_2 \cap \neg B_1]$ 由定理 2.7 给出界（注意此处需要假设 $p \leq 1/\varepsilon$ 和$k \gtrsim pd/\varepsilon$ ）。终止条件直接推出近似保证。

最后，我们考虑返回解的规模。注意，如果 $\lambda_{\min}(Z_t) \geq 1$，那么 $Y_t \succcurlyeq X$，这进一步意味着在时刻 t满足终止条件。因此，我们可以假设算法终止前 $\lambda_{\min}(Z_t) < 1$。因此，我们可以应用定理 2.4 得出，返回解 S以至少 $1-\exp(-\Omega(\frac{\varepsilon d}{\text{最小值}\{\gamma,\kappa\}}))$ 的概率满足 $|S| \leq (1+\varepsilon)k + O\left(\frac{d}{\gamma} + \frac{d}{\kappa}\right)$。 □

2.4 新思路

证明定理的关键2.7 在于界定交换后目标值的变化。对于 A-设计（$p=1$），存在一个简单的不等式来界定目标值的变化：

$$\mathrm{tr}\left(\left(Y - vv^\top + ww^\top\right)^{-1}\right) \leq \mathrm{tr}\left(Y^{-1}\right) + \underbrace{\frac{v^\top Y^{-2} v}{1 - \langle vv^\top, Y^{-1}\rangle} - \frac{w^\top Y^{-2} w}{1 + \langle ww^\top, Y^{-1}\rangle}}_{\text{progress}}.$$

对于一般的 Φ_p-设计，在秩二更新下， Φ_p 函数的变化要复杂得多。利用 Sherman-Morrison公式和利布–塞林不等式，我们可以如下界定 Φ_p 目标（实际上是 Φ_p 目标的 p次幂）的变化：

$$\begin{aligned}&\operatorname{tr}\left((Y+ww^\top-vv^\top)^{-p}\right)\\&\qquad\le \operatorname{tr}(Y^{-p})+\sum_{i=1}^{p}\binom{p}{i}\left((-1)^i\frac{(w^\top Y^{-1}w)^{i-1}\cdot w^\top Y^{-p-1}w}{(1+w^\top Y^{-1}w)^i}+\frac{(v^\top Y^{-1}v)^{i-1}\cdot v^\top Y^{-p-1}v}{(1-v^\top Y^{-1}v)^i}\right).\end{aligned}$$

p-范数会引入许多高阶项，如何处理这些项是本文主要的技术难点。

如备注 2.1所述，如果我们在 [7], 中使用相同的算法，并进行一些谨慎的处理，包括适当地应用 H¨older 不等式，则可以实现 $(1+\varepsilon)$-近似，但要求会严格得多，即 $k \gtrsim 2^{O(p)}d/\varepsilon$。原因在于，从当前解中移除某些“有影响力”的向量（即使其发生概率相对较小），也会因上述不等式中的高阶项而导致目标函数变化的期望值急剧增大。

为了解决这个问题，我们引入参数 κ ，并通过限制允许从当前解中换出的向量（位于S'_t中）来修改随机交换算法。这有助于我们有效地界定上述关于目标函数变化的不等式中的高阶项。但是， κ 越小，我们限制从当前解中换出的向量集合就越小。这会增加无法移除某个向量的概率，进而增大解的规模。幸运的是，我们可以证明，解的规模增量可以由一个加性的 d/κ 项界定，如定理 2.4所述。

文献 [7] 中对 A-设计的分析包含两个部分：（1）界定期望进展；（2）证明总进展的集中性。最小特征值至少为$1/4$ 这一条件在这两个部分中都非常重要，而凸规划问题（1.2）的最优性条件在集中性论证中起着关键作用。有趣的是，凸规划问题（1.2）的最优性条件对于界定具有更高 Φ_p 的目标函数的期望进展也至关重要 p。本文花费了大量精力来准确求得目标值的期望，而对于 D/A-设计（当 $p=0$，1）而言，这相对容易。

3 Φp 目标函数的分析

本节的目标是证明定理 2.7。由于我们关注定理 $p \le 1/\varepsilon$ 中的情形，因此可以不失一般性地假设

$$\gamma=\max\left\{\frac{\varepsilon}{6},\frac{1}{6p}\right\}=\frac{1}{6p}\qquad\text{and}\qquad\kappa=\max\left\{\frac{\varepsilon}{2},\frac{1}{2p}\right\}=\frac{1}{2p}$$

在本节剩余部分中。为便于记号，我们还假设第二阶段的开始时间为 $\tau_1=1$。为了从目标函数的角度分析算法的进展，我们使用以下引理来给出交换后 $\operatorname{tr}(Y_t^{-p})$ 变化的上界，并将在 3.1 节给出证明。

引理 3.1. 设 $\mathbf{Y} \succ 0$ 为一个 d维正定矩阵， $p \geq 1$ 为整数。对于任意满足 $\mathbf{v}^\top \mathbf{Y}^{-1}\mathbf{v} < 1$的 $\mathbf{w} \in \mathbb{R}^d$和 $\mathbf{v} \in \mathbb{R}^d$，

$$\begin{aligned}&\operatorname{tr}\left((Y + ww^\top - vv^\top)^{-p}\right)\\ &\qquad \leq \operatorname{tr}(Y^{-p}) + \sum_{i=1}^{p} \binom{p}{i} \left((-1)^i \frac{(w^\top Y^{-1} w)^{i-1} \cdot w^\top Y^{-p-1} w}{(1 + w^\top Y^{-1} w)^i} + \frac{(v^\top Y^{-1} v)^{i-1} \cdot v^\top Y^{-p-1} v}{(1 - v^\top Y^{-1} v)^i}\right).\end{aligned}$$

在随机交换算法中，我们在每次迭代中交换向量 u_{i_t}和 u_{j_t}，其中 u_{i_t}属于当前解。因此， $u_{i_t}^\top Y_t^{-1} u_{i_t} \leq 1$始终成立，事实上，稍后将会清楚， $u_{i_t}^\top Y_t^{-1} u_{i_t}$严格小于 1。因此， 3.1 可以重复应用，从而得到对于任意 $\tau \geq 1$,

$$\begin{aligned}\operatorname{tr}(Y_{\tau+1}^{-p}) - \operatorname{tr}(Y_1^{-p}) \leq \sum_{t=1}^{\tau} \Bigg(&\underbrace{\sum_{i=1}^{p} \binom{p}{i} \frac{(u_{i_t}^\top Y_t^{-1} u_{i_t})^{i-1} \cdot u_{i_t}^\top Y_t^{-p-1} u_{i_t}}{(1 - u_{i_t}^\top Y_t^{-1} u_{i_t})^i}}_{\text{loss}} \\ &\underbrace{- \sum_{i=1}^{p} \binom{p}{i} (-1)^{i+1} \frac{(u_{j_t}^\top Y_t^{-1} u_{j_t})^{i-1} \cdot u_{j_t}^\top Y_t^{-p-1} u_{j_t}}{(1 + u_{j_t}^\top Y_t^{-1} u_{j_t})^i}}_{\text{gain}} \Bigg).\end{aligned} \tag{3.1}$$

我们将第 t 次迭代中的增益 g_t、损失 l_t 和进展 Γ_t 定义如下。

$$\begin{aligned}g_t &:= \sum_{i=1}^{p} \binom{p}{i} (-1)^{i+1} \frac{(u_{j_t}^\top Y_t^{-1} u_{j_t})^{i-1} \cdot u_{j_t}^\top Y_t^{-p-1} u_{j_t}}{(1 + u_{j_t}^\top Y_t^{-1} u_{j_t})^i},\\ l_t &:= \sum_{i=1}^{p} \binom{p}{i} \frac{(u_{i_t}^\top Y_t^{-1} u_{i_t})^{i-1} \cdot u_{i_t}^\top Y_t^{-p-1} u_{i_t}}{(1 - u_{i_t}^\top Y_t^{-1} u_{i_t})^i},\\ \Gamma_t &:= g_t - l_t.\end{aligned}$$

在第 3.2 节中，我们将证明，如果 x 是（1.2）的一个分数最优解，且Z_t 具有有下界的最小特征值，并且当前解的目标值远非最优，那么第 t次迭代中的期望进展就会很大。随后，在第 3.3 节中，我们将证明总进展集中在其期望值附近。最后，我们将在第 2.7 节中完成定理的证明，具体见第 3.4 节。

3.1 一步中的目标值变化

在 [7], 中，使用秩二更新公式，在一步内计算目标值的变化，其中 $p = 1$。对于一般的 Φ_p-设计，秩二更新公式会变得复杂得多，因此我们改为分两步进行更新：首先使用秩一更新，将u_{j_t} 加入当前解中，然后再使用另一个秩一更新，将u_{i_t} 从当前解中移除。

引理3.1的证明。设$Y_1 = Y + ww^\top$。根据 Sherman-Morrison公式（引理A.1），有

$$\mathrm{tr}(Y_1^{-p}) = \mathrm{tr}\left(\left(Y^{-1} - \frac{Y^{-1}ww^\top Y^{-1}}{1+w^\top Y^{-1}w}\right)^p\right) = \mathrm{tr}\left(\left(Y^{-1/2}\left(I - \frac{Y^{-1/2}ww^\top Y^{-1/2}}{1+w^\top Y^{-1}w}\right)Y^{-1/2}\right)^p\right).$$

然后，我们可以应用利布–塞林不等式（引理 A.2）来证明

$$\mathrm{tr}(Y_1^{-p}) \le \mathrm{tr}\left(Y^{-p/2}\left(I - \frac{Y^{-1/2}ww^\top Y^{-1/2}}{1+w^\top Y^{-1}w}\right)^p Y^{-p/2}\right) = \mathrm{tr}\left(Y^{-p}\left(I - \frac{Y^{-1/2}ww^\top Y^{-1/2}}{1+w^\top Y^{-1}w}\right)^p\right).$$

根据二项式定理展开，

$$\begin{aligned}\mathrm{tr}(Y_1^{-p}) &\le \sum_{i=0}^{p}(-1)^i\binom{p}{i}\mathrm{tr}\left(Y^{-p}\left(\frac{Y^{-1/2}ww^\top Y^{-1/2}}{1+w^\top Y^{-1}w}\right)^i\right)\\ &= \mathrm{tr}(Y^{-p}) + \sum_{i=1}^{p}(-1)^i\binom{p}{i}\frac{(w^\top Y^{-1}w)^{i-1}\cdot w^\top Y^{-p-1}w}{(1+w^\top Y^{-1}w)^i}. \end{aligned} \tag{3.2}$$

对于 $\mathrm{Y}_2 = \mathrm{Y}_1 - \mathrm{vv}^\top$，我们可以应用类似的论证来证明

$$\mathrm{tr}(Y_2^{-p}) \le \mathrm{tr}(Y_1^{-p}) + \sum_{i=1}^{p}\binom{p}{i}\frac{(v^\top Y_1^{-1}v)^{i-1}\cdot v^\top Y_1^{-p-1}v}{(1-v^\top Y_1^{-1}v)^i}.$$

注意到 $\mathrm{Y}_1 = \mathrm{Y} + \mathrm{ww}^\top \succcurlyeq \mathrm{Y}$ 且 $\mathrm{v}^\top\mathrm{Y}^{-1}\mathrm{v} < 1$，因此有

$$\mathrm{tr}(Y_2^{-p}) \le \mathrm{tr}(Y_1^{-p}) + \sum_{i=1}^{p}\binom{p}{i}\frac{(v^\top Y^{-1}v)^{i-1}\cdot v^\top Y^{-p-1}v}{(1-v^\top Y^{-1}v)^i}. \tag{3.3}$$

结合（3.2）和（3.3）即可得到该引理。 □

备注 3.2。 引理 3.1 可以通过在证明中应用牛顿广义二项式定理推广到所有实数 $p \ge 1$ 。为保证广义二项式定理的收敛性，我们需要确保更强的条件 $v^\top Y^{-1}v \le \frac{1}{2}$。对于我们的应用而言，这并不是问题，因为我们的算法总是从受限集合 S_t'中移除向量，从而保证 $v^\top Y^{-1}v \le \frac{1}{2}$成立。给定具有实数 p的新版本引理 3.1，我们可以通过基本相同的分析，将本文的主要结果（即定理 1.2）推广到所有实数 $p \ge 1$ 。

3.2 期望进展

为了分析期望进展，我们需要使用以下两个引理。第一个引理是最小特征值条件具有下界的推论，它类似于 [7] 中的引理 4.13。为完整起见，我们在附录C 中给出了证明。

引理 3.3. 对于 $\gamma \le \frac{1}{6}$，如果 $Z_t \succcurlyeq (1-5\gamma)I$，那么

$$\langle v_i v_i^\top, Z_t^{-1}\rangle \le \alpha \langle v_i v_i^\top, A_t^{\frac{1}{2}}\rangle \le \alpha \lambda_{\min}(Z_t)\langle v_i v_i^\top, Z_t^{-1}\rangle \qquad \forall i \in [n].$$

另一个则是（1.2）的最优性条件的推论。该引理的证明与 [7] 中关于 A-设计的证明类似，为完整起见，我们将其收录于附录C 中。

引理 3.4. 设 $x \in [0,1]^n$ 是针对 p-范数问题的凸规划松弛（(1.2) 的最优分数解。则对于每个 $1 \le i \le n$ 满足 $0 < x(i) < 1$,

$$\langle X^{-p-1}, u_i u_i^\top\rangle \le \frac{1}{k}\cdot \mathrm{tr}(X^{-p}).$$

现在，我们可以对期望进展进行下界估计了。我们先在引理 3.5 和引理 3.6中分别处理期望损失和期望收益。然后，在引理 3.8中结合这两部分，对期望进展进行下界估计。为简化记号，我们将 $\mathbb{E}_t[\cdot]$ 记为给定截至时间 t所发生情况的条件期望，即 $\mathbb{E}[\cdot \mid S_{t-1}]$。

3.2.1 期望损失

最小特征值下界（引理 3.3）、最优性条件（引理 3.4），以及在随机交换算法中引入新参数 κ 对于以下引理都至关重要。

引理 3.5 （期望损失）。 设 S_{t-1} 为时刻 t 的解集，且 $Z_t = \sum_{i\in S_{t-1}} v_i v_i^\top$对于$1 \le t \le \tau$。假设 x是问题 (1.2)的最优解， $\lambda_{\min}(Z_t) \in [1-5\gamma, 1)$， $\gamma = 1/6p$，且$\kappa = 1/2p$。则

$$\mathbb{E}_t[l_t] \le \frac{p}{M}\Big(\mathrm{tr}(Y_t^{-p}) - \langle X_{S_{t-1}}, Y_t^{-p-1}\rangle\Big) + O\Big(\frac{p^2 d}{kM}\Big)\cdot \mathrm{tr}(X^{-p}),$$

其中，对于任意集合 $S \subseteq [n]$，我们记 $\mathbf{X}_S := \sum_{i\in S}\mathbf{x}(i)\mathbf{u}_i\mathbf{u}_i^\top$。

证明。损失项 l_t 中有 p 项。我们分别处理线性项和高阶项。考虑线性项:

$$\begin{aligned}\mathbb{E}_t\left[\frac{p\cdot u_{i_t}^\top Y_t^{-p-1}u_{i_t}}{1-u_{i_t}^\top Y_t^{-1}u_{i_t}}\right] &= \sum_{i\in S'_{t-1}}\frac{1-x(i)}{M}\Big(1-\alpha\langle v_i v_i^\top, A_t^{1/2}\rangle\Big)\cdot\frac{p\cdot u_i^\top Y_t^{-p-1}u_i}{1-u_i^\top Y_t^{-1}u_i}\\ &= \sum_{i\in S'_{t-1}}\frac{1-x(i)}{M}\Big(1-\alpha\langle v_i v_i^\top, A_t^{1/2}\rangle\Big)\cdot\frac{p\cdot u_i^\top Y_t^{-p-1}u_i}{1-\langle v_i v_i^\top, Z_t^{-1}\rangle},\end{aligned}$$

其中第二行依据 $\mathbf{Y}_t$ 和 $\mathbf{Z}_t$的定义得出，这意味着

$$\langle v_i v_i^\top, Z_t^{-1}\rangle = u_i^\top Y_t^{-1}u_i. \tag{3.4}$$

注意到 $\gamma = 1/6p \le 1/6$。因此，我们可以应用引理 3.3 中的第一个不等式，然后将S'_{t-1} 放宽为 S_{t-1}，从而得到

$$\mathbb{E}_t\left[\frac{p\cdot u_{i_t}^\top Y_t^{-p-1}u_{i_t}}{1-u_{i_t}^\top Y_t^{-1}u_{i_t}}\right] \le \frac{p}{M}\sum_{i\in S_{t-1}}(1-x(i))\cdot u_i^\top Y_t^{-p-1}u_i = \frac{p}{M}\Big(\operatorname{tr}(Y_t^{-p}) - \langle X_{S_{t-1}}, Y_t^{-p-1}\rangle\Big). \quad (3.5)$$

接下来，我们考虑剩余的 $p-1$ 高阶损失项。

$$\underbrace{\mathbb{E}_t\left[\sum_{l=2}^{p}\binom{p}{l}\frac{(u_{i_t}^\top Y_t^{-1}u_{i_t})^{l-1}\cdot u_{i_t}^\top Y_t^{-p-1}u_{i_t}}{(1-u_{i_t}^\top Y_t^{-1}u_{i_t})^l}\right]}_{①}$$

$$= \sum_{i\in S'_{t-1}}\frac{1-x(i)}{M}\Big(1-\alpha\langle v_iv_i^\top, A_t^{1/2}\rangle\Big)\sum_{l=2}^{p}\binom{p}{l}\frac{(u_i^\top Y_t^{-1}u_i)^{l-1}\cdot u_i^\top Y_t^{-p-1}u_i}{(1-u_i^\top Y_t^{-1}u_i)^l}.$$

再次应用引理3.3，可得

$$① \le \sum_{i\in S'_{t-1}}\frac{1-x(i)}{M}\cdot u_i^\top Y_t^{-p-1}u_i\cdot\sum_{l=2}^{p}\binom{p}{l}\frac{(u_i^\top Y_t^{-1}u_i)^{l-1}}{(1-u_i^\top Y_t^{-1}u_i)^{l-1}}.$$

注意到

$$\lambda_{\min}(Z_t)\ge 1-5\gamma \qquad \Longleftrightarrow \qquad Y_t \succcurlyeq (1-5\gamma)X. \quad (3.6)$$

因此，根据假设 $\lambda_{\min}(Z_t) \ge 1-5\gamma$，可得

$$① \le \frac{(1-5\gamma)^{-p-1}}{M}\sum_{i\in S'_{t-1}}(1-x(i))\cdot u_i^\top X^{-p-1}u_i\cdot\sum_{l=2}^{p}\binom{p}{l}\frac{(u_i^\top Y_t^{-1}u_i)^{l-1}}{(1-u_i^\top Y_t^{-1}u_i)^{l-1}}.$$

利用x 是 (1.2) 的一个分数最优解这一事实，再应用引理3.4，可得

$$\begin{aligned} ① &\le \frac{(1-5\gamma)^{-p-1}}{kM}\cdot\operatorname{tr}(X^{-p})\sum_{i\in S'_{t-1}}(1-x(i))\cdot\sum_{l=2}^{p}\binom{p}{l}\frac{(u_i^\top Y_t^{-1}u_i)^{l-1}}{(1-u_i^\top Y_t^{-1}u_i)^{l-1}} \\ &\le \frac{(1-5\gamma)^{-p-1}}{kM}\cdot\operatorname{tr}(X^{-p})\sum_{l=2}^{p}\binom{p}{l}\sum_{i\in S'_{t-1}}u_i^\top Y_t^{-1}u_i\cdot\frac{(u_i^\top Y_t^{-1}u_i)^{l-2}}{(1-u_i^\top Y_t^{-1}u_i)^{l-1}}. \end{aligned}$$

根据集合的定义 S'_{t-1} 以及（3.4），可知$u_i^\top Y_t^{-1}u_i \le \kappa$ 对所有 $i\in S'_{t-1}$均成立。因此，

$$① \le \frac{(1-5\gamma)^{-p-1}}{kM}\cdot\operatorname{tr}(X^{-p})\sum_{l=2}^{p}\binom{p}{l}\sum_{i\in S'_{t-1}}u_i^\top Y_t^{-1}u_i\cdot\Big(\frac{\kappa}{1-\kappa}\Big)^{l-2}\cdot\frac{1}{1-\kappa}.$$

由于 $\sum_{i\in S'_{t-1}} u_i^\top Y_t^{-1} u_i \le \langle Y_t, Y_t^{-1}\rangle = d$，我们还可以通过以下方式进一步给出 1 项的上界

$$\begin{aligned}
① &\le \frac{(1-5\gamma)^{-p-1}d}{kM}\cdot \mathrm{tr}(X^{-p})\cdot \sum_{l=2}^{p}\binom{p}{l}\Big(\frac{\kappa}{1-\kappa}\Big)^l\cdot\frac{1-\kappa}{\kappa^2}\\
&= \frac{(1-5\gamma)^{-p-1}d}{kM}\cdot \mathrm{tr}(X^{-p})\cdot \frac{1-\kappa}{\kappa^2}\Big(\frac{1}{(1-\kappa)^p}-1-\frac{p\kappa}{1-\kappa}\Big)\\
&\le \frac{(1-5\gamma)^{-p-1}d}{kM}\cdot \mathrm{tr}(X^{-p})\cdot \frac{1}{\kappa^2}\cdot\frac{1}{(1-\kappa)^p}. \qquad (3.7)
\end{aligned}$$

结合（3.5）和（3.7），期望损失可界定为

$$\begin{aligned}
\mathbb{E}_t[l_t] &\le \frac{p}{M}\Big(\mathrm{tr}(Y_t^{-p})-\langle X_{S_{t-1}}, Y_t^{-p-1}\rangle\Big)+\frac{(1-5\gamma)^{-p-1}d}{\kappa^2(1-\kappa)^p kM}\cdot\mathrm{tr}(X^{-p})\\
&\le \frac{p}{M}\Big(\mathrm{tr}(Y_t^{-p})-\langle X_{S_{t-1}}, Y_t^{-p-1}\rangle\Big)+O\Big(\frac{p^2 d}{kM}\Big)\cdot\mathrm{tr}(X^{-p}),
\end{aligned}$$

其中，最后一个不等式由对 γ 和 κ 的选择得出。 □

3.2.2 期望收益

期望收益的分析稍微复杂一些，因为采样概率不会像损失分析中那样与 g_t 中的分母项很好地相消 l_t，因此我们将分为两种情况，并采用不同的论证方法。同样，以下引理中的最小特征值下界（引理 3.3）和最优性条件（引理 3.4）至关重要。

引理 3.6 （期望收益）。设 S_{t-1} 为时刻 t 的解集，且 $Z_t = \sum_{i\in S_{t-1}} v_i v_i^\top$对于$1 \le t \le \tau$成立。假设 x是（1.2）的一个最优解， $\lambda_{\min}(Z_t) \in [1-5\gamma, 1$，且 $\gamma = 1/6p$。则

$$\mathbb{E}_t[g_t] \ge \frac{p}{M}\Big(\langle X, Y_t^{-p-1}\rangle - \langle X_{S_{t-1}}, Y_t^{-p-1}\rangle\Big) - O\Big(\frac{p^2 d}{kM}\Big)\cdot\mathrm{tr}(X^{-p}).$$

证明。我们考虑两种不同的情况:

$$\begin{aligned}
S_1 &:= \{j\in[m]\backslash S_{t-1} \mid \alpha\langle v_j v_j^\top, A_t^{1/2}\rangle \ge p\cdot u_j^\top Y_t^{-1} u_j\} \quad \text{and}\\
S_2 &:= \{j\in[m]\backslash S_{t-1} \mid \alpha\langle v_j v_j^\top, A_t^{1/2}\rangle < p\cdot u_j^\top Y_t^{-1} u_j\}.
\end{aligned}$$

然后，期望收益可以写成

$$\begin{aligned}
\mathbb{E}_t[g_t] = &\sum_{j\in S_1}\frac{x(j)}{M}\cdot\Big(1+\alpha\langle v_j v_j^\top, A_t^{1/2}\rangle\Big)\cdot\sum_{l=1}^{p}\binom{p}{l}(-1)^{l+1}\frac{(u_j^\top Y_t^{-1}u_j)^{l-1}\cdot u_j^\top Y_t^{-p-1}u_j}{(1+u_j^\top Y_t^{-1}u_j)^l} \qquad (3.8)\\
&+\sum_{j\in S_2}\frac{x(j)}{M}\cdot\Big(1+\alpha\langle v_j v_j^\top, A_t^{1/2}\rangle\Big)\cdot\sum_{l=1}^{p}\binom{p}{l}(-1)^{l+1}\frac{(u_j^\top Y_t^{-1}u_j)^{l-1}\cdot u_j^\top Y_t^{-p-1}u_j}{(1+u_j^\top Y_t^{-1}u_j)^l}. \qquad (3.9)
\end{aligned}$$

我们分别分析（3.8）和（3.9）。首先，我们考虑（3.8），并稍微重新整理一下 g_t 项，这样更便于分析（3.8）。

$$\begin{aligned} g_t &= \frac{u_{j_t}^\top Y_t^{-p-1} u_{j_t}}{u_{j_t}^\top Y_t^{-1} u_{j_t}} \cdot \sum_{i=1}^{p} \binom{p}{i} (-1)^{i+1} \left(\frac{u_{j_t}^\top Y_t^{-1} u_{j_t}}{1 + u_{j_t}^\top Y_t^{-1} u_{j_t}} \right)^i \\ &= \frac{u_{j_t}^\top Y_t^{-p-1} u_{j_t}}{u_{j_t}^\top Y_t^{-1} u_{j_t}} \left(1 - \left(1 - \frac{u_{j_t}^\top Y_t^{-1} u_{j_t}}{1 + u_{j_t}^\top Y_t^{-1} u_{j_t}} \right)^p \right) \qquad (3.10) \\ &= \frac{u_{j_t}^\top Y_t^{-p-1} u_{j_t}}{u_{j_t}^\top Y_t^{-1} u_{j_t}} \left(1 - \frac{1}{(1 + u_{j_t}^\top Y_t^{-1} u_{j_t})^p} \right). \qquad (3.11) \end{aligned}$$

因此，我们可以将（3.8）改写为

$$(3.8) = \sum_{j \in S_1} \frac{x(j)}{M} \cdot \left(1 + \alpha \langle v_j v_j^\top, A_t^{1/2} \rangle \right) \cdot \frac{u_j^\top Y_t^{-p-1} u_j}{u_j^\top Y_t^{-1} u_j} \left(1 - \frac{1}{(1 + u_j^\top Y_t^{-1} u_j)^p} \right).$$

根据 S_1 的定义，有

$$(3.8) \geq \sum_{j \in S_1} \frac{x(j)}{M} \cdot u_j^\top Y_t^{-p-1} u_j \cdot \frac{1 + p \cdot u_j^\top Y_t^{-1} u_j}{u_j^\top Y_t^{-1} u_j} \left(1 - \frac{1}{(1 + u_j^\top Y_t^{-1} u_j)^p} \right).$$

令 $x = \mathrm{u}_j^\top \mathrm{Y}_t^{-1} \mathrm{u}_j > 0$ （as $\mathrm{Y}_t \succ 0$）。则有

$$\frac{1 + px}{x} \cdot \left(1 - \frac{1}{(1+x)^p} \right) = \frac{1 + px}{x} \cdot \frac{(1+x)^p - 1}{(1+x)^p} = \frac{px(1+x)^p + (1+x)^p - 1 - px}{x(1+x)^p} \geq p.$$

因此，

$$(3.8) \geq \frac{p}{M} \sum_{j \in S_1} x(j) \cdot u_j^\top Y_t^{-p-1} u_j. \qquad (3.12)$$

接下来，我们考虑（3.9）。与引理 3.5的证明一样，我们将（3.9）分成两部分，2 涉及线性项，3 涉及剩余的 $p-1$ 高阶项。

$$\begin{aligned} (3.9) = &\underbrace{\sum_{j \in S_2} \frac{x(j)}{M} \cdot \left(1 + \alpha \langle v_j v_j^\top, A_t^{1/2} \rangle \right) \cdot \frac{p \cdot u_j^\top Y_t^{-p-1} u_j}{1 + u_j^\top Y_t^{-1} u_j}}_{②} \\ &\underbrace{- \sum_{j \in S_2} \frac{x(j)}{M} \cdot \left(1 + \alpha \langle v_j v_j^\top, A_t^{1/2} \rangle \right) \cdot \sum_{l=2}^{p} \binom{p}{l} (-1)^l \frac{(u_j^\top Y_t^{-1} u_j)^{l-1} \cdot u_j^\top Y_t^{-p-1} u_j}{(1 + u_j^\top Y_t^{-1} u_j)^l}}_{③}. \end{aligned}$$

线性项 2 ◯ 很容易进行界定，我们可以通过引理 3.3 (结合（3.4）) 来控制它，使得 t

$$② \geq \frac{p}{M} \sum_{j \in S_2} x(j) \cdot u_j^\top Y_t^{-p-1} u_j. \tag{3.13}$$

接下来，我们对高阶项 3 取上界（注意，3 不包含负号）。为了对 3 取上界，对于每个 $j \in S_2$，我们不失一般性地假设 $\sum_{l=2}^{p} \binom{p}{l}(-1)^l \frac{(u_j^\top Y_t^{-1} u_j)^{l-1} \cdot u_j^\top Y_t^{-p-1} u_j}{(1+u_j^\top Y_t^{-1} u_j)^l} \geq 0$，否则我们可以直接忽略第 j 项。根据这一假设以及 S_2 的定义，可得

$$\begin{aligned} ③ &\leq \frac{1}{M} \sum_{j \in S_2} x(j)(1 + p u_j^\top Y_t^{-1} u_j) \sum_{l=2}^{p} \binom{p}{l} (-1)^l \frac{(u_j^\top Y_t^{-1} u_j)^{l-1} \cdot u_j^\top Y_t^{-p-1} u_j}{(1 + u_j^\top Y_t^{-1} u_j)^l} \\ &= \frac{1}{M} \sum_{j \in S_2} x(j) u_j^\top Y_t^{-p-1} u_j \cdot (1 + p u_j^\top Y_t^{-1} u_j) \sum_{l=2}^{p} \binom{p}{l} (-1)^l \frac{(u_j^\top Y_t^{-1} u_j)^{l-1}}{(1 + u_j^\top Y_t^{-1} u_j)^l}. \end{aligned}$$

根据假设 $\lambda_{\min}(Z_t) \geq 1 - 5\gamma$ 和 (3.6)，可得

$$③ \leq \frac{(1-5\gamma)^{-p-1}}{M} \sum_{j \in S_2} x(j) u_j^\top X^{-p-1} u_j \cdot (1 + p u_j^\top Y_t^{-1} u_j) \sum_{l=2}^{p} \binom{p}{l} (-1)^l \frac{(u_j^\top Y_t^{-1} u_j)^{l-1}}{(1 + u_j^\top Y_t^{-1} . u_j)^l}.$$

由于 x 是（1.2）的最优解，我们可以应用引理 3.4 并推出

$$\begin{aligned} ③ &\leq \frac{(1-5\gamma)^{-p-1}}{kM} \cdot \operatorname{tr}(X^{-p}) \sum_{j \in S_2} x(j) \cdot (1 + p u_j^\top Y_t^{-1} u_j) \sum_{l=2}^{p} \binom{p}{l} (-1)^l \frac{(u_j^\top Y_t^{-1} u_j)^{l-1}}{(1 + u_j^\top Y_t^{-1} u_j)^l} \\ &= \frac{(1-5\gamma)^{-p-1}}{kM} \cdot \operatorname{tr}(X^{-p}) \sum_{j \in S_2} x(j) \cdot \frac{1 + p u_j^\top Y_t^{-1} u_j}{u_j^\top Y_t^{-1} u_j} \left(\left(1 - \frac{u_j^\top Y_t^{-1} u_j}{1 + u_j^\top Y_t^{-1} u_j} \right)^p - 1 + \frac{p u_j^\top Y_t^{-1} u_j}{1 + u_j^\top Y_t^{-1} u_j} \right). \end{aligned}$$

令 $x = \mathrm{u}_j^\top \mathrm{Y}_t^{-1} \mathrm{u}_j$，我们要对其取上界

$$\frac{1+px}{x} \left(\left(1 - \frac{x}{1+x} \right)^p - 1 + \frac{px}{1+x} \right) = \frac{1}{x} \left(\left(1 - \frac{x}{1+x} \right)^p - 1 + \frac{px}{1+x} \right) + p \left(\left(1 - \frac{x}{1+x} \right)^p - 1 + \frac{px}{1+x} \right).$$

对于任意的 $y \in [0,1]$，都有 $(1-y)^p \leq 1 - py + \binom{p}{2} y^2$。因此，可以得出

$$\begin{aligned} &\frac{1}{x} \left(\left(1 - \frac{x}{1+x} \right)^p - 1 + \frac{px}{1+x} \right) + p \left(\left(1 - \frac{x}{1+x} \right)^p - 1 + \frac{px}{1+x} \right) \\ \leq\ & \frac{1}{x} \cdot \binom{p}{2} \cdot \frac{x^2}{(1+x)^2} + p \cdot \frac{px}{1+x} = \binom{p}{2} \cdot \frac{x}{1+x} + \frac{p^2 x}{1+x} \leq 2p^2 x. \end{aligned}$$

因此，第 3 项还可以进一步界定为

$$③ \leq \frac{2(1-5\gamma)^{-p-1} \cdot p^2}{kM} \cdot \operatorname{tr}(X^{-p}) \sum_{j \in S_2} x(j) \cdot u_j^\top Y_t^{-1} u_j \leq \frac{2(1-5\gamma)^{-p-1} \cdot p^2}{kM} \cdot \operatorname{tr}(X^{-p}) \langle X, Y_t^{-1} \rangle.$$

使用（3.6），可得

$$③ \le \frac{2(1-5\gamma)^{-p-2}\cdot p^2 d}{kM}\cdot \mathrm{tr}(X^{-p}) \le O\Big(\frac{p^2 d}{kM}\Big)\cdot \mathrm{tr}(X^{-p}), \tag{3.14}$$

其中，最后一个不等式是根据 γ 的选取而得出的。

结合（3.12）、（3.13）和（3.14），我们可以得到期望收益的一个下界：

$$\begin{aligned}\mathbb{E}_t[g_t] &\ge \frac{p}{M}\Big(\sum_{j\in[m]\setminus S_{t-1}} x(j)u_j^\top Y_t^{-p-1}u_j\Big) - O\Big(\frac{p^2 d}{kM}\Big)\cdot \mathrm{tr}(X^{-p})\\ &= \frac{p}{M}\Big(\langle X, Y_t^{-p-1}\rangle - \langle X_{S_{t-1}}, Y_t^{-p-1}\rangle\Big) - O\Big(\frac{p^2 d}{kM}\Big)\cdot \mathrm{tr}(X^{-p}).\end{aligned}$$

□

3.2.3 期望进展

最后，我们应用了适当应用 H¨older 不等式来比较增益项和损失项 .

引理 3.7。 给定正定矩阵 A、B $\in \mathbb{S}_{++}^d$ 和整数 $p \ge 1$，有 t

$$\langle A, B^{-p-1}\rangle \ge \left(\frac{\mathrm{tr}(B^{-p})}{\mathrm{tr}(A^{-p})}\right)^{1/p}\cdot \mathrm{tr}(B^{-p}).$$

证明。设 A $= \sum_{i=1}^d a_i \mathrm{v}_i \mathrm{v}_i^\top$ 为 A 的特征分解，B $= \sum_{i=1}^d b_i \mathrm{w}_i \mathrm{w}_i^\top$ 为 B 的特征分解。那么，

$$\begin{aligned}\mathrm{tr}(B^{-p}) = \sum_{i=1}^d \frac{1}{b_i^p} = \sum_{1\le i,j\le d}\frac{1}{b_i^p}\langle v_i, w_j\rangle^2 &= \sum_{1\le i,j\le d}\frac{a_j^{p/(p+1)}}{b_i^p}\langle v_i, w_j\rangle^{2p/(p+1)}\cdot\frac{1}{a_j^{p/(p+1)}}\langle v_i, w_j\rangle^{2/(p+1)}\\ &\le \Big(\sum_{1\le i,j\le d}\frac{a_j}{b_i^{p+1}}\langle v_i, w_j\rangle^2\Big)^{p/(p+1)}\cdot\Big(\sum_{1\le i,j\le d}\frac{1}{a_j^p}\langle v_i, w_j\rangle^2\Big)^{1/(p+1)}\\ &= \langle A, B^{-p-1}\rangle^{p/(p+1)}\cdot \mathrm{tr}(A^{-p})^{1/(p+1)},\end{aligned}$$

其中，不等式由 H¨older 不等式得出。然后，对两边取 $(p+1)/p$ 次方并重新整理各项，即可得到该引理。 □

现在，我们 通过结合引理 3.5 和引理 3.6，为期望进展给出下界。

引理 3.8. 设 S_{t-1}为时间 t的解集，且 $\mathrm{Z}_t = \sum_{i\in S_{t-1}} \mathrm{v}_i \mathrm{v}_i^\top$对于 $1 \le t \le \tau$。假设 x 是 (1.2) 的分数最优解，并且 $\lambda_{\min}(\mathrm{Z}_t) \in [1-5\gamma, 1)$，$(\mathrm{tr}(\mathrm{Y}_t^{-2}))^{1/p} \ge \lambda\cdot(\mathrm{tr}\ (\mathrm{X}^{-2}))^{1/p}$，其中 $\lambda \ge 1$ 且对于 $1 \le t \le \tau$。那么

$$\mathbb{E}_t[\Gamma_t] \ge \left(\frac{p(\lambda-1)\lambda^p}{M} - O\Big(\frac{p^2 d}{kM}\Big)\right)\cdot \mathrm{tr}(X^{-p}).$$

特别是，当 $\lambda > 1+\varepsilon$ 和 $k \gtrsim \frac{pd}{\varepsilon}$ 时，期望进展为正。

证明。结合　　期望损失引理 3.5 和期望收益引理 3.6，可得

$$\mathbb{E}_t[\Gamma_t] \geq \frac{p}{M}\Big(\langle X, Y_t^{-p-1}\rangle - \mathrm{tr}(Y_t^{-p})\Big) - O\Big(\frac{p^2 d}{kM}\Big)\cdot \mathrm{tr}(X^{-p}).$$

应用引理3.7，我们得到

$$\mathbb{E}_t[\Gamma_t] \geq \frac{p}{M}\left(\Big(\frac{\mathrm{tr}(Y_t^{-p})}{\mathrm{tr}(X^{-p})}\Big)^{1/p}\cdot \mathrm{tr}(Y_t^{-p}) - \mathrm{tr}(Y_t^{-p})\right) - O\Big(\frac{p^2 d}{kM}\Big)\cdot \mathrm{tr}(X^{-p}).$$

根据假设 $(\mathrm{tr}(Y_t^{-p}))^{1/p} \geq \lambda(\mathrm{tr}(X^{-p}))^{1/p}$，或者等价地 $\mathrm{tr}(Y_t^{-p}) \geq \lambda^p \mathrm{tr}(X^{-p})$，我们得出最终界

$$\mathbb{E}_t[\Gamma_t] \geq \frac{p(\lambda-1)}{M}\cdot \mathrm{tr}(Y_t^{-p}) - O\Big(\frac{p^2 d}{kM}\Big)\cdot \mathrm{tr}(X^{-p}) \geq \left(\frac{p(\lambda-1)\lambda^p}{M} - O\Big(\frac{p^2 d}{kM}\Big)\right)\cdot \mathrm{tr}(X^{-p}). \qquad \square$$

3.3 鞅集中性论证

在本小节中，我们证明总进展集中在其期望附近。证明利用最小特征值假设以及引理 3.4中的最优性刻画来界定随机过程的方差。证明思路与 [7], 中的类似，但我们还需要额外处理由更高p-范数引入的高阶项。

引理 3.9。对于任意 $\eta > 0$，有

$$\Pr\left[\sum_{t=1}^{\tau}\Gamma_t \leq \sum_{t=1}^{\tau}\mathbb{E}_t[\Gamma_t] - \eta \bigcap \min_{1\leq t\leq\tau}\lambda_{\min}(Z_t) \geq 1-5\gamma\right] \leq \exp\left(-\Omega\left(\frac{\eta^2 kM}{\tau p^3\sqrt{d}(\mathrm{tr}(X^{-p}))^2 + \eta p M\,\mathrm{tr}(X^{-p})}\right)\right).$$

证明。 我们定义两列随机变量 $\{X_t\}_t$ 和 $\{Y_t\}_t$，其中 $X_t := \mathbb{E}_t[\Gamma_t] - \Gamma_t$ 以及$Y_t := \sum_{l=1}^{t} X_l$。容易验证， $\{Y_t\}_t$ 是关于 $\{S_t\}_t$的鞅。我们将使用 Freedman不等式（定理 A.3）来界定概率 $\Pr[Y_\tau \geq \eta \cap \min_{1\leq t\leq\tau}\lambda_{\min}(Z_t) \geq 1-5\gamma]$。

下面，我们首先证明：如果事件 $\min_{1\leq t\leq\tau}\lambda_{\min}(Z_t) \geq 1-5\gamma$ 发生，那么我们可以对 X_t 和 $\mathbb{E}_t[X_t^2]$ 给出上界，从而应用 Freedman不等式。为了给出 X_t的上界，我们首先证明 g_t 和 l_t的一个上界。

注意，如果事件 $\lambda_{\min}(Z_t) \geq 1-5\gamma$ 发生，那么 $Y_t \succcurlyeq (1-5\gamma)X$，这意味着

$$u_{i_t}^\top Y_t^{-p-1} u_{i_t} \leq (1-5\gamma)^{-p-1} u_{i_t}^\top X^{-p-1} u_{i_t} \quad \text{and} \quad u_{j_t}^\top Y_t^{-p-1} u_{j_t} \leq (1-5\gamma)^{-p-1} u_{j_t}^\top X^{-p-1} u_{j_t}.$$

根据观察 2.3，对于所有交换对 i_t, j_t，都有 $x(i_t), x(j_t) \in (0,1)$。因此，我们可以应用引理 3.4 得出 $\langle X^{-p-1}, u_{i_t}u_{i_t}^\top\rangle \leq \frac{1}{k}\cdot\mathrm{tr}(X^{-p})$ 和 $\langle X^{-p-1}, u_{j_t}u_{j_t}^\top\rangle \leq \frac{1}{k}\cdot\mathrm{tr}(X^{-p})$。因此，由于 $\gamma = 1/6p$，

$$u_{i_t}^\top Y_t^{-p-1} u_{i_t} \leq O\Big(\frac{1}{k}\Big)\cdot\mathrm{tr}(X^{-p}) \quad \text{and} \quad u_{j_t}^\top Y_t^{-p-1} u_{j_t} \leq O\Big(\frac{1}{k}\Big)\cdot\mathrm{tr}(X^{-p}). \tag{3.15}$$

我们首先给出关于 g_t 的一个确定性界。根据（ $x = u_{j_t}^\top Y_t^{-p-1} u_{j_t}$，根据（3.10），增益项$g_t$ 可以写为

$$g_t = u_{j_t}^\top Y_t^{-p-1} u_{j_t} \cdot \frac{1}{x} \cdot \left(1 - \left(1 - \frac{x}{1+x}\right)^p\right).$$

由于对于 $y \in [0,1]$ 和 $p \geq 1$，$(1-y)^p \geq 1 - py$，我们可以将 g_t 上界为

$$0 \leq g_t \leq u_{j_t}^\top Y_t^{-p-1} u_{j_t} \cdot \frac{1}{x} \cdot \frac{px}{1+x} \leq p \cdot u_{j_t}^\top Y_t^{-p-1} u_{j_t} \leq O\left(\frac{p}{k}\right) \cdot \mathrm{tr}(X^{-p}),$$

其中，最后一个不等式由（3.15）得出。

接下来，我们给出 l_t 的一个确定性界。根据 S'_{t-1} 的定义以及（3.4），可得$0 < u_{i_t}^\top Y_t^{-1} u_{i_t} = \langle v_{i_t} v_{i_t}^\top, Z_t^{-1} \rangle \leq \kappa$。因此，我们可以将 l_t 上界为

$$\begin{aligned}
0 \leq l_t &= \sum_{i=1}^{p} \binom{p}{i} \frac{(u_{i_t}^\top Y_t^{-1} u_{i_t})^{i-1} \cdot u_{i_t}^\top Y_t^{-p-1} u_{i_t}}{(1 - u_{i_t}^\top Y_t^{-1} u_{i_t})^i} \\
&\leq u_{i_t}^\top Y_t^{-p-1} u_{i_t} \cdot \sum_{i=1}^{p} \binom{p}{i} \frac{\kappa^{i-1}}{(1-\kappa)^i} \\
&= u_{i_t}^\top Y_t^{-p-1} u_{i_t} \cdot \frac{1}{\kappa} \cdot \left(\frac{1}{(1-\kappa)^p} - 1\right).
\end{aligned}$$

对于 $\kappa = \frac{1}{2p}$，我们可以控制 l_t，使其满足

$$l_t \leq O(p) \cdot u_{i_t}^\top Y_t^{-p-1} u_{i_t} \leq O\left(\frac{p}{k}\right) \cdot \mathrm{tr}(X^{-p}),$$

其中，最后一个不等式由（3.15）得出。

利用上述关于 g_t 和 l_t 的界，我们可以将鞅增量的大小控制为

$$|X_t| = |\mathbb{E}_t[\Gamma_t] - \Gamma_t| \leq g_t + l_t \leq O\left(\frac{p}{k}\right) \cdot \mathrm{tr}(X^{-p}).$$

接下来，我们对 $\mathbb{E}_t[X_t^2]$ 给出上界，得到

$$\mathbb{E}_t[X_t^2] \leq |X_t| \cdot \mathbb{E}_t[|X_t|] \leq O\left(\frac{p}{k}\right) \cdot \mathrm{tr}(X^{-p}) \cdot \left(\mathbb{E}_t[g_t] + \mathbb{E}_t[l_t]\right).$$

利用引理 3.5，我们将期望损失项界定为

$$\begin{aligned}
\mathbb{E}_t[l_t] &\leq \frac{p}{M}\left(\mathrm{tr}(Y_t^{-p}) - \langle X_{S_{t-1}}, Y_t^{-p-1} \rangle\right) + O\left(\frac{p^2 d}{kM}\right) \cdot \mathrm{tr}(X^{-p}) \\
&\leq O\left(\frac{p}{M}\right) \cdot \mathrm{tr}(X^{-p}) + O\left(\frac{p^2 d}{kM}\right) \cdot \mathrm{tr}(X^{-p}) \\
&\leq O\left(\frac{p}{M}\right) \cdot \mathrm{tr}(X^{-p}),
\end{aligned}$$

其中，第二行的不等式由以下假设得出： $\lambda_{\min}(Z_t) \geq 1 - 5\gamma$成立（并且满足（3.6）），最后一行的不等式则由关于 k的假设得出。

然后，结合（ g_t （3.10）中的表达式，我们将期望收益写为

$$\mathbb{E}_t[g_t] = \sum_{j\in[m]\setminus S_{t-1}} \frac{x(j)}{M}\cdot\Big(1+\alpha\langle v_j v_j^\top, A_t^{\frac{1}{2}}\rangle\Big)\cdot\frac{u_j^\top Y_t^{-p-1}u_j}{u_j^\top Y_t^{-1}u_j}\left(1-\left(1-\frac{u_j^\top Y_t^{-1}u_j}{1+u_j^\top Y_t^{-1}u_j}\right)^p\right).$$

利用 $(1-y)^p \geq 1-py$ 对于 $y\in[0,1]$ 和 $p\geq 1$ 成立这一事实，可将期望收益界定为

$$\mathbb{E}_t[g_t] \leq \sum_{j\in[m]\setminus S_{t-1}} \frac{p}{M}\cdot x(j)\cdot u_j^\top Y_t^{-p-1}u_j\cdot\frac{1+\alpha\langle v_j v_j^\top, A_t^{\frac{1}{2}}\rangle}{1+u_j^\top Y_t^{-1}u_j}.$$

根据（3.4），$u_j^\top Y^{-}{}_t{}^1 u_j = \langle vjv_j^\top,\ Z^{-1}\rangle$。因此，根据引理 3.3中的第二个不等式，有

$$\mathbb{E}_t[g_t] \leq \frac{p}{M}\cdot\alpha\lambda_{\min}(Z_t)\cdot\sum_{j=1}^{m} x(j)\cdot u_j^\top Y_t^{-p-1}u_j \leq \frac{p}{M}\cdot\alpha\cdot\langle X, Y_t^{-p-1}\rangle,$$

其中，最后一个不等式成立是因为在算法终止之前 $\lambda_{\min}（Z_t）< 1$ 。根据假设， $\lambda_{\min}（Z_t）\geq 1-5\gamma$ 发生（且（3.6）成立），并结合 $\alpha=\sqrt{d}/\gamma$ 和 $\gamma = 1/6p$的选取，可得界

$$\mathbb{E}_t[g_t] \leq \frac{p\sqrt{d}}{\gamma(1-5\gamma)^{p+1}M}\cdot\operatorname{tr}(X^{-p}) \leq O\Big(\frac{p^2\sqrt{d}}{M}\Big)\cdot\operatorname{tr}(X^{-p}).$$

因此，

$$\mathbb{E}_t[X_t^2] \leq O\Big(\frac{p}{k}\Big)\cdot O\Big(\frac{p^2\sqrt{d}}{M}\Big)\cdot\big(\operatorname{tr}(X^{-p})\big)^2 = O\Big(\frac{p^3\sqrt{d}}{kM}\Big)\cdot\big(\operatorname{tr}(X^{-p})\big)^2,$$

这意味着

$$W_t := \sum_{l=1}^{t}\mathbb{E}_l[X_l^2] \leq O\Big(\frac{\tau p^3\sqrt{d}}{kM}\Big)\cdot\big(\operatorname{tr}(X^{-p})\big)^2,\quad \forall t\in[\tau].$$

最后，我们可以应用弗里德曼鞅不等式定理 A.3 结合

$$R = O\Big(\frac{p}{k}\Big)\cdot\operatorname{tr}(X^{-p}) \qquad \text{and} \qquad \sigma^2 = O\Big(\frac{p^3\sqrt{d}\tau}{kM}\Big)\cdot\big(\operatorname{tr}(X^{-p})\big)^2$$

从而得出

$$\begin{aligned}\Pr\left[Y_\tau\geq\eta\bigcap\min_{1\leq t\leq\tau}\lambda_{\min}(Z_t)\geq 1-5\gamma\right] &\leq \Pr[\exists t\in[\tau]: Y_t\geq\eta\cap W_t\leq\sigma^2]\\ &\leq \exp\left(-\frac{\eta^2/2}{\sigma^2+R\eta/3}\right)\\ &= \exp\left(-\Omega\left(\frac{\eta^2 kM}{\tau p^3\sqrt{d}(\operatorname{tr}(X^{-p}))^2+\eta pM\operatorname{tr}(X^{-p})}\right)\right).\end{aligned}$$

注意到 $\sum_{t=1}^{\tau}\Gamma_t \leq \sum_{t=1}^{\tau}\mathbb{E}_t[\Gamma_t]-\eta$ 等价于 $Y_\tau\geq\eta$，引理得证。 □

定理 **2.7** 的证明

下面开始证明定理 2.7。设 $\tau = \frac{2M}{\varepsilon p}$。我们要对以下三个事件同时发生的概率给出上界：

- E_1：随机交换算法进入了第二阶段，即 $\lambda_{\min}(Z_1) \geq 1 - 2\gamma$ 使用本小节中的记号。
- E_2: $\min_{1 \leq t \leq \tau} \lambda_{\text{最小值}}(Z_t) \geq 1 - 5\gamma$.
- E_3：到时间 τ，算法的第二阶段尚未终止。

假设事件 E_3 发生了。那么 $\lambda = \min_{1 \leq t \leq \tau+1} (\mathrm{tr}(Y_t^{-p}))^{1/p} / (\mathrm{tr}(X^{-p}))^{1/p} > (1+\varepsilon)$。如果事件 E_2 也发生了，那么引理 3.8 推出

$$\sum_{t=1}^{\tau} \mathbb{E}_t [\Gamma_t] \geq \tau \cdot \left(\frac{p(\lambda-1)\lambda^p}{M} - O\left(\frac{p^2 d}{kM}\right) \right) \cdot \mathrm{tr}(X^{-p}) \geq \frac{2M}{\varepsilon p} \cdot \left(\frac{\varepsilon p}{M} - \frac{\varepsilon p}{2M} \right) \cdot \mathrm{tr}(X^{-p}) \geq \mathrm{tr}\left(X^{-p}\right), \tag{3.16}$$

其中，第二个不等式在 $k \gtrsim pd/\varepsilon$ 取足够大的常数时成立。

另一方面，第二阶段的初始解满足 $Z_1 \succcurlyeq (1-2\gamma)I$（由 E_1可得），这意味着 $Y_1 \succcurlyeq (1-2\gamma)X$。因此，$\mathrm{tr}(Y_1^{-p}) \leq (1-2\gamma)^{-p} \mathrm{tr}(X^{-p}) \leq (1-1/3p)^{-p} \mathrm{tr}(X^{-p}) \leq 3\,\mathrm{tr}(X^{-p})/2$，其中 $p \geq 1$。当事件 E_2 发生时，根据引理 3.3，我们知道 $\langle v_{i_t} v_{i_t}^\top, Z_t^{-1} \rangle \leq 1\alpha \cdot \langle v_{i_t} v_{i_t}^\top, A_t^{\frac{2}{}} \rangle < \frac{1}{2}$，因此可以应用 (3.1) 推得

$$\mathrm{tr}(Y_{\tau+1}^{-p}) \leq \mathrm{tr}(Y_1^{-p}) - \sum_{t=1}^{\tau} \Gamma_t \leq \frac{3}{2} \cdot \mathrm{tr}\left(X^{-p}\right) - \sum_{t=1}^{\tau} \Gamma_t.$$

由于算法截至时间 τ尚未终止，

$$\mathrm{tr}(X^{-p}) \leq \mathrm{tr}(Y_{\tau+1}^{-p}) \quad \Longrightarrow \quad \sum_{t=1}^{\tau} \Gamma_t \leq \frac{1}{2} \cdot \mathrm{tr}\left(X^{-p}\right). \tag{3.17}$$

结合（3.16）和（3.17），$E_1 \cap E_2 \cap E_3$ 意味着进展相对于期望值出现了较大的偏差，因此

$$\sum_{t=1}^{\tau} \Gamma_t - \sum_{t=1}^{\tau} \mathbb{E}_t[\Gamma_t] < -\frac{1}{2} \cdot \mathrm{tr}(X^{-p}).$$

Th我们可以应用引理 3.9，并结合 $\eta = \frac{1}{2} \cdot \mathrm{tr}(X^{-p})$ 和 $\tau = \frac{2M}{\varepsilon p}$，从而得出

$$\begin{aligned}
\Pr\left[E_1 \cap E_2 \cap E_3\right] &\leq \Pr\left[\sum_{t=1}^{\tau} \Gamma_t < \sum_{t=1}^{\tau} \mathbb{E}_t\left[\Gamma_t\right] - \frac{1}{2} \cdot \mathrm{tr}(X^{-p}) \bigcap E_2 \right] \\
&\leq \exp\left(-\Omega\left(\frac{(\mathrm{tr}(X^{-p}))^2 \cdot kM}{\left(\frac{2M}{\varepsilon p}\right) p^3 \sqrt{d} \cdot (\mathrm{tr}(X^{-p}))^2 + pM \cdot (\mathrm{tr}(X^{-p}))^2} \right) \right) \\
&\leq \exp\left(-\Omega\left(\frac{\varepsilon k}{p^2 \sqrt{d}} \right) \right) \leq \exp\left(-\Omega\left(\frac{\sqrt{d}}{p} \right) \right) = \exp\left(-\Omega\left(\gamma \sqrt{d} \right) \right),
\end{aligned}$$

其中，最后一个不等式由假设 $k \gtrsim pd/\varepsilon$ 得出，最后一个等式则由 $\gamma = 1/6p$ 得出。

致谢

我们感谢 Mohit Singh 提请我们注意 Φ_p 目标函数。我们也感谢早期版本稿件的匿名审稿人提出的宝贵建议。

参考文献

[1] Zeyuan Allen-Zhu、Yuanzhi Li、Aarti Singh 和 Yining Wang。通过遗憾最小化实现近最优实验设计。载于《国际机器学习大会论文集》，ICML'17 第70卷，第126–135页。JMLR.org，2017年。3[2] Zeyuan Allen-Zhu、Yuanzhi Li、Aarti Singh 和 Yining Wang。实验设计的近最优离散优化：一种遗憾最小化方法。《数学规划》，186(1-2，A辑)：439–478，2021年。2、3、5、27[3] Kathryn Chaloner 和 Isabella Verdinelli。贝叶斯实验设计：综述。《统计科学》，10(3)：273–304，1995年。4[4] Michal Derezinski、Feynman T. Liang 和 Michael W. Mahoney。使用正则化行列式点过程进行贝叶斯实验设计。载于 Silvia Chiappa 和 Roberto Calandra（编），第23届国际人工智能与统计学大会，AISTATS 2020，2020年8月26日至28日，在线 /意大利西西里岛巴勒莫/, ，《机器学习研究论文集》第108卷，第3197–3207页。PMLR，2020年。4[5] David A. Freedman。关于鞅的尾概率。《概率年刊》，3：100–118，1975年。26[6] Lap Chi Lau 和 Hong Zhou。网络设计的谱方法。载于第52届年度 ACM SIGACT 计算理论研讨会论文集，STOC '20，第826–839页，纽约州纽约，美国，2020年。计算机协会。2、3、4、5、8、9、26[7] Lap Chi Lau 和 Hong Zhou。实验设计的局部搜索框架。载于第三十二届年度 ACM-SIAM 离散算法研讨会论文集，SODA '21，第1039–1058页，美国，2021年。工业与应用数学学会。2、3、4、5、6、7、8、9、11、12、13、14、20、27[8] Elliott H. Lieb 和 Walter E. Thirring。关于 Schr ¨odinger 哈密顿量特征值矩的若干不等式及其与索博列夫不等式的关系，第269–304页。普林斯顿大学出版社，1976年。25

[9] Vivek Madan、Mohit Singh、Uthaipon Tantipongpipat 和 Weijun Xie。最优设计的组合算法。载于 Alina Beygelzimer 和 Daniel Hsu（编），第三十二届学习理论会议论文集，第99卷，《机器学习研究论文集》，第2210–2258页，美国凤凰城，2019年6月25日至28日。PMLR。3[10] Zelda Mariet 和 Suvrit Sra。用于最优试验设计的初等对称多项式。载于第31届神经信息处理系统国际会议论文集，NIPS'17，第2136–2145页，美国纽约州雷德胡克，2017年。Curran Associates Inc。4[11] Aleksandar Nikolov、Mohit Singh 和 Uthaipon Tao Tantipongpipat。比例体积抽样与 A-最优设计的近似算法。载于第三十届年度 ACM-SIAM 离散算法研讨会论文集，SODA '19，第1369–1386页，美国宾夕法尼亚州费城，2019年。工业与应用数学学会。2，3，4

[12] Jack Sherman 和 Winifred J. Morrison。关于给定矩阵中一个元素发生变化所对应的逆矩阵调整。《数学统计年刊》，21:124–127，1950年。25

[13] Mohit Singh 和 Weijun Xie。近似正相关分布及 D-最优设计的近似算法。载于第二十九届ACM-SIAM离散算法年会论文集，SODA '18，美国，2018年，第2240–2255页。工业与应用数学学会。2，3

[14]Uthaipon Tantipongpipat。λ-正则化 A-最优设计及其通过 λ-正则化比例体积抽样的近似。*CoRR*，abs/2006.11182，2020年。4

[15] Joel A. Tropp。矩阵鞅的 Freedman不等式。《电子通信概率》，16:262–270，2011年。26

[16] Richard L. Wheeden 和 Antoni Zygmund。测度与积分。纯粹与应用数学（博卡拉顿）。CRC出版社，佛罗里达州博卡拉顿，第二版，2015年。实分析导论。2

背景

我们将使用以下两个结果来分析 Φ_p 目标函数。

引理 A.1 （Sherman-Morrison公式[12]）。假设$A \in \mathbb{R}^{d\times d}$ 是一个可逆矩阵，并且u、$v \in \mathbb{R}^d$。那么，$A + uv^\top$可逆，当且仅当 $1 + v^\top A^{-1}u \neq 0$，在这种情况下

$$\left(A + uv^\top\right)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}.$$

引理 A.2（利布–塞林不等式 [8]）。对于任意 A、B $\succcurlyeq 0$ 和 $p \geq 1$，均有以下不等式：

$$\mathrm{tr}((BAB)^p) \leq \mathrm{tr}(B^pA^pB^p).$$

我们将使用鞅的集中不等式来分析我们的随机化算法。回顾一下，如果对于所有 $t > 0$ 都有，随机变量序列 $Y_1, \ldots, Y_\tau$ 关于随机变量序列 $Z_1, \ldots, Z_\tau$ 是一个鞅，则有

1. Y_t 是 Z_1、……、Z_{t-1}的函数；
2. $\mathbb{E}[|Y_t|] < \infty$；3.
   $\mathbb{E}[Y_{t+1}|Z_1, \ldots, Z_t] = Y_t$。

我们将使用利用 Freedman 的下述定理来界定 Y_τ 较大的概率。 .

定理 A.3 ([5, 15])。 设 $\{Y_t\}_t$ 是关于 $\{Z_t\}_t$的实值鞅，且 $\{X_t = Y_t - Y_{t-1}\}_t$ 为其差分序列。假设 $X_t \le R$ 对 $1 \le t \le \tau$确定成立。令$W_t := \sum_{j=1}^{t} \mathbb{E}[X_j^2|Z_1, ..., Z_{j-1}]$ 对 $1 \le t \le \tau$成立。则对所有 $\delta \ge 0$和 $\sigma^2 > 0$,

$$\Pr\left(\exists t \in [\tau] : Y_t \ge \delta \text{ and } W_t \le \sigma^2\right) \le \exp\left(\frac{-\delta^2/2}{\sigma^2 + R\delta/3}\right).$$

B 第2节中省略的证明2

定理 2.4 的证明概要。 与 [6], 中定理 3.12 的证明类似，证明定理 2.4 分为两个主要部分：(1) 界定解规模的期望变化；(2) 界定随机过程的条件方差。证明定理 2.4 的主要区别在于第一部分，即由于新的 κ 参数，解规模的期望变化会变得更差。

令 1_{i_t} 和 1_{j_t} 分别表示 i_t 和 j_t 是否为空的指示变量。更新后的证明的关键是证明以下不等式。

$$\frac{1}{M}\left(\sum_{i=1}^{n} x(i) - |S_{t-1}|\right) \leqslant \mathbb{E}_t[1_{j_t} - 1_{i_t}] \leqslant \frac{1}{M}\left(\sum_{i=1}^{n} x(i) - |S_{t-1}| + \left(\frac{11d}{\gamma} + \frac{d}{\kappa}\right)\right). \tag{B.1}$$

我们将概述如何得到这个界，并指出 $\kappa <$所造成的损失 1.

我们首先对1_{j_t}的条件期望进行界定。使用与[6], 相同的论证，我们可以证明

$$\frac{1}{M}\left(\sum_{i=1}^{n} x(i) - \sum_{i \in S_{t-1}} x(i)\right) \leqslant \mathbb{E}_t[1_{j_t}] \leqslant \frac{1}{M}\left(\sum_{i=1}^{n} x(i) - \sum_{i \in S_{t-1}} x(i) + \frac{d\,\|c\|_\infty}{\gamma}\right). \tag{B.2}$$

接下来，我们对 1_{i_t} 的期望进行界定。这是由 κ 导致的主要差异。根据概率

i_t 的分布，

$$\begin{aligned}\mathbb{E}_t[1_{i_t}] &= \frac{1}{M}\sum_{i\in S'_{t-1}}(1-x(i))\big(1-\alpha\langle v_i v_i^\top, A_t^{\frac{1}{2}}\rangle\big)\\ &= \frac{1}{M}\Big(\sum_{i\in S'_{t-1}}(1-x(i)) - \alpha\sum_{i\in S'_{t-1}}(1-x(i))\langle v_i v_i^\top, A_t^{\frac{1}{2}}\rangle\Big)\\ &= \frac{1}{M}\Bigg(|S_{t-1}| - \Big(\sum_{i\in S_{t-1}}x(i)\Big) - \sum_{i\in S_{t-1}\setminus S'_{t-1}}(1-x(i)) - \alpha\sum_{i\in S'_{t-1}}(1-x(i))\langle v_i v_i^\top, A_t^{\frac{1}{2}}\rangle\Bigg).\end{aligned}$$

我们希望对右端的最后两项进行界定。回顾 $S'_{t-1} := \{i \in S_{t-1} \mid \alpha\langle \mathrm{v}_i\mathrm{v}_i^\top, \mathrm{A}_t^{\frac{1}{2}}\rangle \le \frac{1}{2}$且 $\langle \mathrm{v}_i\mathrm{v}_i^\top, \mathrm{Z}_t^{-1}\rangle \le \kappa\}$。因此，对于任意 $i \in S_{t-1}\setminus S'_{t-1}$，有$1 < 2\alpha\langle \mathrm{v}_i\mathrm{v}_i^\top, \mathrm{A}_t^{1/2}\rangle + \kappa^{-1}\langle \mathrm{v}_i\mathrm{v}_i^\top, \mathrm{Z}_t^{-1}\rangle$。这意味着

$$\begin{aligned}|S_{t-1}\setminus S'_{t-1}| &\le \sum_{i\in S_{t-1}\setminus S'_{t-1}} 2\alpha\langle v_i v_i^\top, A_t^{\frac{1}{2}}\rangle + \sum_{i\in S_{t-1}\setminus S'_{t-1}}\kappa^{-1}\langle v_i v_i^\top, Z_t^{-1}\rangle\\ &\le 2\alpha\sum_{i\in S_{t-1}}\langle v_i v_i^\top, A_t^{\frac{1}{2}}\rangle + \kappa^{-1}\sum_{i\in S_{t-1}}\langle v_i v_i^\top, Z_t^{-1}\rangle\\ &\le 2\big(d + \alpha\sqrt{d}\cdot\lambda_{\min}(Z_t)\big) + \frac{d}{\kappa} \le \frac{8d}{\gamma} + \frac{d}{\kappa},\end{aligned}$$

其中，倒数第二个不等式用到了 [2] 中的命题 2.14，而最后一个不等式则由 $\alpha = \frac{\sqrt{d}}{\gamma}$ 以及以下假设得出：$\lambda_{\min}(Z_t) < 1$。由于 $x \in [0,1]^m$，因此倒数第二项为

$$0 \le \sum_{i\in S_{t-1}\setminus S'_{t-1}}(1-x(i)) \le |S_{t-1}\setminus S'_{t-1}| \le \Big(\frac{8d}{\gamma} + \frac{d}{\kappa}\Big)$$

类似地，对于最后一项，

$$0 \le \alpha\sum_{i\in S'_{t-1}}(1-x(i))\langle v_i v_i^\top, A_t^{\frac{1}{2}}\rangle \le \alpha\sum_{i\in S_{t-1}}\langle v_i v_i^\top, A_t^{\frac{1}{2}}\rangle \le (d + \alpha\sqrt{d}\cdot\lambda_{\min}(Z_t)) \le \frac{2d}{\gamma}.$$

将这两个上界和下界代回最后两项，对于 $\kappa < 1$，我们得到

$$\frac{1}{M}\Bigg(|S_{t-1}| - \Big(\sum_{i\in S_{t-1}}x(i)\Big) - \Big(\frac{10d}{\gamma} + \frac{d}{\kappa}\Big)\Bigg) \le \mathbb{E}_t[1_{i_t}] \le \frac{1}{M}\Big(|S_{t-1}| - \sum_{i\in S_{t-1}}x(i)\Big). \tag{B.3}$$

结合（B.2）和（B.3），我们可以得到（B.1）。剩余的集中性论证与 [7] 中定理 3.12 的证明相同。□

C 第 3 节中省略的证明3

引理 3.3 的证明。 回忆$A_t = (\alpha Z_t - c_t I)^{-2}$，其中 c_t 是满足以下条件的唯一值： $A_t \succ 0$ 且 tr$(A_t) = 1$。由于 $Z_t \succcurlyeq \lambda_{\min}(Z_t) \cdot I$，因此可得

$$1 = \operatorname{tr}(A_t) \le (\alpha\lambda_{\min}(Z_t) - c_t)^{-2} \cdot \operatorname{tr}(I) \quad \Longrightarrow \quad \alpha\lambda_{\min}(Z_t) - c_t \le \sqrt{d} \quad \Longrightarrow \quad c_t \ge 0,$$

其中，最后一个蕴含成立，因为 $\alpha = \sqrt{d}/\gamma$, $\lambda_{\min}(Z_t) \ge 1 - 5\gamma$ 且假设 $\gamma \le \frac{1}{6}$。这意味着 $A_t^{\frac{1}{2}} = (\alpha Z_t - c_t I)^{-1} \succcurlyeq \alpha^{-1} Z_t^{-1}$，从而证明了第一个不等式。

对于第二个不等式，考虑 $Z_t = \sum_{j=1}^d \lambda_j w_j w_j^\top$ 的特征分解，其中 $0 < \lambda_1 \le \ldots \le \lambda_d$ 是特征值，$\{w_j\}$ 是对应的标准正交特征向量。于是，

$$\frac{\langle v_i v_i^\top, A_t^{\frac{1}{2}} \rangle}{\langle v_i v_i^\top, Z_t^{-1} \rangle} = \frac{\sum_{j=1}^d \frac{\langle v_i, w_j \rangle^2}{\alpha\lambda_j - c_t}}{\sum_{j=1}^d \frac{\langle v_i, w_j \rangle^2}{\lambda_j}} \le \max_{j \in [d]} \frac{\lambda_j}{\alpha\lambda_j - c_t} \le \frac{\lambda_1}{\alpha\lambda_1 - c_t} \le \lambda_1,$$

其中，第一个不等式成立是因为 $\alpha\lambda_j - c_t > 0$ as $A_t \succ 0$；第二个不等式成立是因为 $c_t \ge 0$，并且函数 $f(x) = \frac{x}{\alpha x - c_t}$ 在 $c_t \ge 0$ 时关于 $x \ge \frac{l_t}{\alpha}$ 递减；最后一个不等式由 $1 = \operatorname{tr}(A_t) \ge (\alpha\lambda_1 - c_t)^{-2}$ 得出，这意味着 $\alpha\lambda_1 - c_t \ge 1$。 □

引理 3.4 的证明。 我们将使用对偶刻画来证明该引理。我们为第一个等式约束引入对偶变量 Y，为预算约束引入对偶变量 $\mu \ge 0$，对应于预算约束 $\sum_{i=1}^m x(i) - k \le 0$，为每个非负约束引入对偶变量 $\beta_i^- \ge 0$，对应于非负约束 $-x(i) \le 0$，并为每个容量约束引入对偶变量 $\beta_i^+ \ge 0$，对应于容量约束 $x(i) - 1 \le 0$。

拉格朗日函数 $L(x, X, Y \quad \mu, \beta^+, \beta^-)$ 定义为

$$\begin{aligned}
&\left(\operatorname{tr}\left(X^{-p}\right)\right)^{1/p} + \left\langle Y, X - \sum_{i=1}^m x(i) \cdot u_i u_i^\top \right\rangle + \mu\left(\sum_{i=1}^m x(i) - k\right) - \sum_{i=1}^m \beta_i^- x(i) + \sum_{i=1}^m \beta_i^+ (x(i) - 1) \\
=&\left(\operatorname{tr}\left(X^{-p}\right)\right)^{1/p} + \langle Y, X \rangle - \mu k - \sum_{i=1}^m \beta_i^+ - \sum_{i=1}^m x(i) \cdot \left(\langle Y, u_i u_i^\top \rangle - \mu + \beta_i^- - \beta_i^+\right).
\end{aligned}$$

拉格朗日对偶规划为

$$\max_{\substack{Y \succ 0, \mu \ge 0, \\ \beta^+ \ge 0, \beta^- \ge 0}} \min_{X \succ 0, x} L(x, X, Y, \mu, \beta^+, \beta^-).$$

注意，我们可以在内层最小化问题中加入约束 $X \succ 0$，不失一般性；否则拉格朗日函数将无界。同样，我们可以假设在外层最大化问题中 $Y \succ 0$；否则内层最大化问题将无下界。

为了简化对偶规划，我们将对内层最小化问题应用一阶最优性条件。首先，我们计算 \((\operatorname{tr}(X\{v1\}))\{v2\}\) 的梯度。

$$\begin{aligned}
\nabla_X (\operatorname{tr}(X^{-p}))^{1/p} = \frac{1}{p} (\operatorname{tr}(X^{-p}))^{\frac{1-p}{p}} \cdot \partial_X (X^p)^{-1} &= -\frac{1}{p} (\operatorname{tr}(X^{-p}))^{\frac{1-p}{p}} \cdot X^{-p} (\partial_X X^p) X^{-p} \\
&= -(\operatorname{tr}(X^{-p}))^{\frac{1-p}{p}} \cdot X^{-p-1}.
\end{aligned}$$

因此，给定任意 $\mathbf{Y} \succ 0, \mu, \beta^+, \beta^- \geq 0$，内层最小化问题的极小化解 $\mathbf{X} \succ 0$,$\mathbf{x}$ 满足如下最优性条件，即

$$\nabla_X L = -(\mathrm{tr}(X^{-p}))^{\frac{1-p}{p}} \cdot X^{-p-1} + Y = 0 \quad \text{and} \quad \nabla_{x(i)} L = -\langle Y, u_i u_i^\top \rangle + \mu - \beta_i^- + \beta_i^+ = 0. \quad \text{(C.1)}$$

另一方面，很容易验证，对于足够小的 δ，$\mathbf{x} = \delta\vec{1}$ 是原始规划的严格可行解。因此，根据斯莱特条件，强对偶性成立。设 $\mathbf{x}$、$\mathbf{X}$ 是原始规划的一个最优解，$\mathbf{Y}$ μ, β^+, β^- 是对偶规划的一个最优解。强对偶性意味着，

$$\left(\mathrm{tr}\left(X^{-p}\right)\right)^{1/p} = \left(\mathrm{tr}\left(X^{-p}\right)\right)^{1/p} + \langle Y, X \rangle - \mu k - \sum_{i=1}^m \beta_i^+ - \sum_{i=1}^m x(i) \cdot \left(\langle Y, u_i u_i^\top \rangle - \mu + \beta_i^- - \beta_i^+\right),$$

其中，左侧为原始规划的最优值，右侧则等于给定最优原始解和对偶解时对偶规划的最优值。

然后，由拉格朗日最优性条件（C.1）可得

$$0 = \mathrm{tr}(X^{-p})^{1/p} - \mu k - \sum_{i=1}^m \beta_i^+ \quad \Longrightarrow \quad \mu k + \sum_{i=1}^m \beta_i^+ = \mathrm{tr}(X^{-p})^{1/p} \quad \Longrightarrow \quad \mu \leq \frac{1}{k} \cdot \mathrm{tr}(X^{-p})^{1/p},$$

其中，最后一个蕴含关系由 $\beta_i^+ \geq 0$ 对所有 $i \in [m]$ 成立可得。
最后，根据互补松弛条件，我们有 $\beta_i^- \cdot x(i) = 0$ 以及 $\beta_i^+ \cdot (1 - x(i)) = 0$对于所有 $i \in [n]$。因此，对于每个 $i \in [n]$ 满足 $0 < x(i) < 1$的 $\beta_i^+ = \beta_i^- = 0$，我们必有。由（C.1），可得

$$\frac{1}{k} \cdot \mathrm{tr}(X^{-p})^{1/p} \geq \mu = \langle Y, u_i u_i^\top \rangle = (\mathrm{tr}(X^{-p}))^{\frac{1-p}{p}} \cdot \langle X^{-p-1}, u_i u_i^\top \rangle \quad \Longrightarrow \quad \langle X^{-p-1}, u_i u_i^\top \rangle \leq \frac{1}{k} \cdot \mathrm{tr}(X^{-p}).$$

□