arXiv:2304.04550v1 [cs.DS] 10 Apr 2023

基于梯度预言机的内点法

Adrian Vladu*

摘要

我们提出了一种基于拟牛顿迭代的内点法，只需对强自协调障碍函数进行一阶访问。为此，我们扩展了 Dunagan-Harvey [STOC '07] 的方法，在仅使用一阶信息的同时维护一个预条件器。我们根据该预条件器相对于未知海森矩阵的相对偏心率来衡量其质量，并将这些技术推广到海森矩阵缓慢变化的凸函数。我们将其与内点法结合，证明给定对某个凸集K的适当障碍函数的一阶访问，我们可以在时间$\widetilde{O}\left(\left(\mathcal{T}+n^2\right)\sqrt{n\nu}\log\left(1/\varepsilon\right)\right)$内，以 ε 精度求解定义在 K 上的良态线性优化问题，其中 ν是该障碍函数的自协调参数， $\mathcal{T}$是进行一次梯度查询所需的时间。

由此我们得出：

- 对 n 维凸集进行线性优化，可在 $\widetilde{O}\left(\left(\mathcal{T}n+n^3\right)\log\left(1/\varepsilon\right)\right)$ 时间内求解。这与当前最先进的切割平面法算法所达到的运行时间相当，只需将分离预言机替换为针对适当障碍函数的一阶预言机。
- 我们可以在 $\mathbb{R}^{n\times n}$ in time $\widetilde{O}\left(mn^4+m^{1.25}n^{3.5}\log\left(1/\varepsilon\right)\right)$ 内求解涉及 $m \geq n$ 矩阵的半定规划；当 $m=\Omega\left(n^{\frac{3.5}{\omega-1.25}}\right)$ 时，这优于当前最先进的算法。

在此过程中，我们将借助矩阵分析和舒尔凸性中的技巧，开发一系列工具来控制势函数的演变。

*法国国家科学研究中心与巴黎分校理论与计算机科学研究所，巴黎西岱大学，
vladu@irif.fr

1 引言

梯度方法是连续优化中最基本的要素 [BTN01， Nes18， Bub15]。尽管它们凭借简单性在最小化凸函数理论中的许多重要结果中发挥了核心作用，但除非所使用的函数具有极佳的性质（光滑且强凸），否则它们通常无法提供高精度解。此外，它们的表现会因换基而发生巨大变化，因此找到合适的基是一项艰巨的任务，且可能对其收敛产生剧烈影响 [KOSZ13]。然而，对于许多重要场景而言，单一的良好基可能根本不存在。

特别是，当考虑一般凸集上的线性优化目标函数时，能够提供高精度解的方法包括切割平面法（CPM）或内点法（IPM）。这些方法克服了标准梯度方法的局限性，因为它们都隐式地维护着一个不断变化的基，从而能够非常快速地收敛到目标函数的近似最小值点。然而，这样做的代价非常高。在切割平面法中，需要谨慎地维护当前域的适当中心；添加新约束时，还需要快速更新该中心。而在内点法中，则需要访问适当障碍函数的海森矩阵，而仅仅对其求值就可能代价极高 [NN92， JKL+20]。

在内点法的背景下，一些研究试图借助拟牛顿方法，弥合标准版本（需要访问海森矩阵）与仅能获取梯度的版本之间的差距。这些方法试图模拟内点法中的经典牛顿迭代（可将其视为经海森矩阵预条件的梯度步）：维护一个“伪”海森矩阵，并在利用该矩阵得到的步长被确证指向错误方向时对其进行修正 [GS19， GS22， Tun01]。另一条相关且成果卓著的研究路线，提出了高效的拟牛顿方法 [Bro70， Fle70， Gol70， Sha70， BNS94]，旨在仅凭借梯度预言机，达到二阶方法的性能。尽管这些方法极具潜力，但将其应用于线性规划或半定规划所建模的标准任务时，究竟能否获得良好的收敛界，目前尚不明确。解决这些问题将带来新的改进算法，并开辟高效优化研究的新方向。相关地，Dunagan 和 Harvey [DH07] 提出了一种优美的线性方程组求解方法，该方法部分受共轭梯度法启发，并基于维护动态预条件器。尽管他们明确指出可以在拟牛顿方法中使用其算法，但出于某些令人费解的原因，这一方向至今尚未得到探索。

作为优化算法的重要基准，我们考虑半定规划（半正定规划问题），其在仿射空间与半正定矩阵锥的交集上优化线性目标函数。这些问题广泛应用于多个科学领域，包括理论计算机科学、运筹学和工程学 [VB96]。

We f通常将具有 $n \times n$ 个变量和 m 个约束的半定规划定义为 s.

定义 1（半定规划）。给定对称矩阵 $\boldsymbol{B}$、$\boldsymbol{A}_1$、…、$\boldsymbol{A}_m \in \mathbb{R}^{n\times n}$ 和$c_i \in \mathbb{R}$，对于所有 $i \in \{1,\dots,m\}$，我们旨在求解以下优化问题：

$$\max\{\langle \boldsymbol{B}, \boldsymbol{X}\rangle : \boldsymbol{X} \succeq 0, \langle \boldsymbol{A}_i, \boldsymbol{X}\rangle = c_i, \text{ for all } 1 \le i \le m\}, \tag{1}$$

其中 $\langle \boldsymbol{A}, \boldsymbol{B}\rangle := \sum_{i,\ j} \boldsymbol{A}_{ij}\boldsymbol{B}_{ij}$ 是迹乘积。

高精度求解半正定规划问题的标准方法依赖于切割平面法或内点法。早期研究以及近期对半正定规划求解器运行时间的改进，都采用了切割平面法 [Kha79，LSW15，JLSW20],而近期的一种趋势是转而采用内点法，以期进一步改进[JKL+20，HJS+22]。如前所述，对于某些优化问题，利用内点法实现快速算法的一个主要瓶颈在于计算海森矩阵所需的时间极长，而这正是该障碍出现的情形之一。虽然 [NN92] 采用了一系列巧妙的技巧来加快海森矩阵的计算，[JKL+20] 则基于矩形矩阵乘法开发了一系列复杂的技术。因此，拟牛顿方法无需访问真实的海森矩阵，或许能为半定规划领域实现更快、更实用的算法提供一条可行途径。

因此，本文要探讨的问题是：

我们能否 是否能获得仅依赖梯度信息的高效内点法？

1.1 我们的结果

我们提出了一种快速内点法，用于在只能获取凸集障碍函数梯度的情况下，求解凸集上的线性优化问题。本文采用的障碍函数概念略强于内点法标准文献中使用的概念，即强自协调障碍函数 [LLV20]。我们会在第2.4节中对此作更全面的概述。强自协调性是许多近期研究中隐含使用的一种性质[CLS19], ，标准障碍函数也具有这一性质，例如对数障碍函数 [JKL+20]、通用障碍函数 [Nes18、Gül96], 以及熵障碍函数 [BE15]。正如 [LLV20], 针对后两种情形所证明的，强自协调性这一性质源于关于 KLS猜想 [Che21, JLV22]的最新进展。内点法的迭代次数取决于障碍函数的优劣，而其优劣由自协调参数刻画（见定义 11）。集合在 $\mathbb{R}^n$ 中的通用障碍函数和熵障碍函数的自协调参数为 $\widetilde{O}(n)$。

我们的主要结果如下定理所述。

定理2 （非正式表述）。设 $K \subseteq \mathbb{R}$ 是一个凸集。存在一种内点法，可在 $\widetilde{O}\left(\sqrt{\nu n}\left(n^2+\mathcal{T}_{gradient}\right)\right)$ 时间内求解关于 K 的线性优化问题，其中 $\mathcal{T}_{gradient}$ 是计算 K 的 ν-强自协调障碍函数梯度所需的时间。若可访问通用障碍函数或熵障碍函数的梯度，运行时间为 $\widetilde{O}\left(n^3+n\mathcal{T}_{gradient}\right)$。

内点法的梯度复杂度。有趣的是，将梯度预言机替换为 K 的分离预言机后，我们的运行时间与当前最先进的切割平面法所达到的运行时间相当。尽管这两种预言机无法相互比较，我们的结果为理解不同访问模型下优化问题的查询复杂度这一令人振奋的研究方向做出了贡献 [BBE+22]。注意，朴素地看，人们会预期梯度查询复杂度为 $\widetilde{O}\left(n^{3/2}\right)$，因为运行内点法的 $\widetilde{O}\left(\sqrt{n}\right)$ 次迭代，朴素地需要每次迭代进行 $O(n)$ 次梯度查询，以近似障碍函数的海森矩阵。尽管近期的研究 [CLS19、LSZ19、vdBLSS20、vdB20、JKL+20、HJS+22] 已充分利用海森矩阵在迭代过程中的稳定性，提升了内点法的运行时间，但在我们的设定中如何运用这些方法尚不明确，因为它们的摊销分析关键依赖于底层海森矩阵的结构。

这引发了关于障碍优化的自适应查询复杂度的几个有趣问题。值得注意的是，了解并行查询复杂度是否可以降至低于 $\widetilde{O(n)}$，以及理解并行轮次的数量与查询总数之间的权衡，将会很有意义。特别是，实现 $o(\sqrt{n})$ 的并行轮次复杂度可能会产生广泛影响，因为只要有适用于结构化线性系统的快速求解器，就有可能带来更高效的内点法。

Year	参考文献	方法	迭代次数	每次迭代成本
2015	[LSW15]	CPM	n	$n^2 + \mathcal{T}_{\text{cutting plane}}$
2020	[JLSW20]	CPM	n	$n^2 + \mathcal{T}_{\text{cutting plane}}$
2022	我们的结果	IPM	n	$n^2 + \mathcal{T}_{\text{gradient}}$

表1：凸集上的线性优化算法比较 $K \subseteq \mathbb{R}^n$。CPM代表切割平面法，IPM代表内点法。$\mathcal{T}_{\text{cutting plane}}$ 是计算切割平面所需的时间，而 $\mathcal{T}_{\text{gradient}}$ 表示计算 K的障碍函数梯度所需的时间。

半定规划。 利用定理 2 我们得到以下关于求解一般半定规划的结果，该结果源于在半定规划锥上使用标准对数障碍函数。

定理3 （非正式表述）。 存在一种内点法，可在时间 $\widetilde{O}\left(mn^4 + m^{1.25}n^{3.5}\right)$内求解变量规模为 $n \times n$ 且有 m 个约束的一般半定规划。

我们的运行时间可以解释如下： n 是该方法的迭代次数，mn^2 是输入规模， $mn^2 + n^\omega$ 是计算障碍函数梯度的时间，这需要计算输入矩阵的加权和，并使用快速矩阵乘法计算其逆矩阵。我们注意到，当 $m \geq n$时，这在大多数标准半定规划应用中都成立，运行时间为 $\widetilde{O}\left(mn^4 + m^{1.25}n^{3.5}\right)$，即使在$O\left(n^3\right)$中使用朴素矩阵求逆也是如此。对于那些需要大量调优和专用硬件，才能实现快于 $O\left(n^3\right)$ 理论运行速度的矩阵乘法算法的应用，这可能是一项优势 [Hua18]。

Year	参考文献	方法	迭代次数	每次迭代成本	当 $m = n^4$ 时的成本
1992	[NN92]	IPM	$\sqrt{n}$	$m^2n^2 + mn^\omega + m^\omega$	$n^{10.5}$
2015	[LSW15]	CPM	m	$mn^2 + m^2 + n^\omega$	n^{10}
2020	[JLSW20]	CPM	m	$mn^2 + m^2 + n^\omega$	n^{10}
2020	[JKL+20]	IPM	$\sqrt{n}$	$mn^2 + m^\omega + n^\omega$	$n^{4\omega+1/2}$
2021	[HJS+22]	IPM	$\sqrt{n}$	$m^2+n^4+\left(m^\omega + n^{2\omega}\right) n^{-1/2}$ （摊还）	$n^{4\omega}$
2022	我们的结果	IPM	$n^2 + n^{3/2}m^{1/4}$	$mn^2 + n^\omega$	$n^{8.5}$

表 2：关键半正定规划算法总结。CPM表示切割平面方法，IPM 表示内点法。 n 是变量矩阵的大小， m是约束的数量。运行时间省略了 $n^{o(1)}$、 $m^{o(1)}$ 和多对数（$1/\varepsilon$）因子，其中 ε 是精度参数，以及取决于位复杂度的因子。

1.2 我们的技术

为了实现我们的结果，我们借鉴了 Dunagan 和 Harvey 的思路 [DH07]，为运行内点法时出现的线性方程组维护一个动态预条件器。对于强自协调障碍函数，这些方程组彼此之间有很强的关联，因此我们可以在内点法的多次迭代中重复使用该预条件器。

使用自适应预条件器求解线性系统。在第3 节中，我们回顾 [DH07], 中的主要思想，并以此框架为基础，开发基于梯度的内点法。其基本思路如下：尝试使用 Richardson 迭代求解线性系统$\boldsymbol{H x}=\boldsymbol{b}$ 时，衡量进展的一种标准方法是考察残差范数 $\|\boldsymbol{b}-\boldsymbol{H x}\|_{2}$。执行一步 Richardson 迭代后，残差范数的变化有两种可能。第一种情况是残差范数按一个常数因子减小，这是理想情形，因为这样可以非常快速地收敛到最优解。第二种情况是残差范数仅略有减小。然而，这种情况只能是因为底层矩阵存在非常大或非常小的特征值，而我们尝试采用的下降方向恰好与这些特征值对应的方向对齐。因此，这种情况构成了存在非常大或非常小特征值的证据。我们可以对此进行量化，证明能够利用这些方向构造预条件器，从而避免今后再次出现这种情况。

Dunagan 和 Harvey提出的主要思路是，通过执行涉及该方向的秩一更新来维护这个预条件器，而该方向证明了大/小特征值的存在性。虽然他们无法直接证明这会永久性地降低该方向所证明的极端特征值的影响，但他们取得了进展：某个$\widetilde{}$称为偏心率 的势函数会减小，该势函数衡量我们当前预条件器 H的质量。因此，在每一步中，要么能够将梯度范数降低一个常数倍，要么能够将偏心率降低一个常数倍。这会自动给出迭代次数的上界，而该上界取决于方法初始化时预条件器的质量。

我们在第 3 节中正式还原这一论证，并得到一种完全基于预条件 Richardson 方法的简单算法。为完整起见，我们使用更简单的算法还原 [DH07] 中的收敛保证。

我们强调我们贡献的重要性，因为 [DH07] 采用了一组略有不同的更新规则，对于这些更新规则，在只能近似涉及矩阵的量的情况下（正如我们在第 4节所做的那样），很难证明理论保证。

海森矩阵缓慢变化的凸函数上的牛顿步。 在第 4 节中，我们将第 3 节的分析扩展到对矩阵 $\boldsymbol{H}$ 的访问受到限制的情形。虽然第 4 节的分析只使用矩阵-向量乘积，但在我们所关注的情形中，限制甚至更多，因为该矩阵表示凸函数 g的海森矩阵，而只能通过梯度查询访问。为处理这一难题，我们使用两次梯度查询来近似矩阵-向量乘积，因为直观地说：

$$\boldsymbol{H}_{\boldsymbol{y}} \boldsymbol{v} \approx \frac{1}{\tau}\left(\nabla g\left(\boldsymbol{y}+\tau \boldsymbol{v}\right)-\nabla g\left(\boldsymbol{y}\right)\right),$$

对于某个合适的步长 τ。为此，考虑 H $_{\boldsymbol{y}}$ 的条件数很重要，因为要获得准确的估计，需要将 τ 设得足够大。然而，我们不希望 τ 非常大，因为其大小在其他方面决定了所需的精度位数

来计算该近似值。因此，我们需要提供第 3 节中算法的稳健版本，并证明它们满足类似的保证，尽管我们只能通过梯度查询来访问函数g 。

虽然证明大体上遵循与之前相同的思路，但它要复杂得多，因为即使是为预条件 Richardson 方法选取合适的步长，也需要估计无法直接获得的量。此外，由于我们希望控制数值的位表示大小，因此始终控制预条件器特征值的范围。这些技术难点要求我们非常仔细地分析所生成的偏心率证书。本节的主要结果由稳健步进或更新引理（引理 20）给出，可以将其视为$\widetilde{}$采用自适应预条件器 $\boldsymbol{H}$ 的牛顿步，它要么执行一步以降低“伪”对偶局部范数 $\|\nabla g(\boldsymbol{y})\|_{\widetilde{\boldsymbol{H}}^{-1}}$，要么给出偏心率证书，我们利用该证书来降低偏心势。

使用自适应预条件器的内点法。我们论文的主要贡献在于将自适应预条件应用于内点法中。在第 5 节中，我们描述主要的路径跟踪法。为此，我们近似地跟踪由以下函数族的极小化点构成的中心路径

$$g_\mu(\boldsymbol{y}) = \frac{\langle \boldsymbol{c}, \boldsymbol{y} \rangle}{\mu} + \phi(\boldsymbol{y}) \ ,$$

其中， ϕ 是针对 $K \subseteq \mathbb{R}^n$ 的 ν强自协调障碍函数（见第2.4节）。在经典内点法中，路径跟踪是指生成一系列迭代点 y，使其接近中心路径；中心路径表示对于所有 $\mu \in \in (0,\infty)$，g_μ 的唯一极小化点所构成的集合。与中心路径的接近程度以梯度 $\|\nabla g_\mu(\mathrm{y})\|_{\boldsymbol{H}_\mu^{-1}}$的对偶局部范数来衡量。经典做法是，通过维持不变量 $\|\nabla g_\mu(\mathrm{y})\|_{\boldsymbol{H}_\mu^{-1}} \leq O(1)$，将 μ 按某个因子调低，再通过执行一步（或几步）牛顿步来恢复与中心路径的接近程度，这对应于一系列线性方程组求解。 μ调低的幅度由自协调参数 ν决定。通常，我们从 μ变到 $\mu' = \mu/(1 + O(1/\sqrt{\nu}))$，由此确定路径跟踪$\widetilde{}$方法的总迭代次数为 $O(\sqrt{\nu})$。

[{"id":0,"output":"在这项工作中，我们通过执行明显更短的步长来放慢内点法。具体而言，我们仅将中心性参数降低 $1+O(1/\sqrt{n\nu})$ 倍，并保持不变量 $\|\nabla g_\mu(\mathrm{y})\|_{\boldsymbol{H}_\mu^{-1}} \leq O(1/\sqrt{n})$。这样做的原因是，在路径跟随方法产生的两个相继迭代点 y 和 y′ 之间，海森矩阵会发生显著变化。不幸的是，过大的变化（使用标准步长时很可能发生）会使预条件器显著变差，因此，预条件器 $\widetilde{}$H 与两个海森矩阵 H $_y$ 和 H $_{y'}$ 之间的偏心率可能会大幅增加。使用非常短的步长可以保证情况不会如此，并使我们能够将改变迭代点时偏心率的增加，与预条件器秩一更新所导致的偏心率降低相匹配。"}]

偏心势的一个重要特点是，预条件器的秩一更新所引起的变化很容易控制，而修改预条件矩阵所引起的变化通常难以控制。具体而言，使用第2 节中定义的记号，我们旨在比较$\mathcal{E}\left(\widetilde{}\mathrm{H}_{-1}\mathrm{H}_{y}\right)$与 $\mathcal{E}\left(\widetilde{}\mathrm{H}_{-1}\mathrm{H}_{y'}\right)$。

我们利用强自协调障碍函数的性质，证明了偏心率增量的上界。为此，我们运用主序技术和舒尔凸性，证明偏心率的变化取决于$\boldsymbol{H}^{-1}{}_{\boldsymbol{y}}\boldsymbol{H}_{\boldsymbol{y}'}$的特征值。借助强自协调性，可以对这些特征值进行精确控制。我们在第5.2节证明这一点。

我们得出结论：沿中心路径的每一步都会使偏心率至多增加一个常数因子，而预条件 Richardson 迭代找到偏心率证书后所引发的每次秩一更新，都会使偏心率至少减少一个常数因子。因此，在经过适当初始化后，我们的内点法只需对预条件器 $\widetilde{\mathrm{H}}$ 进行 $\widetilde{O}\left(\sqrt{n\nu}\log\frac{1}{\varepsilon}\right)$ 次秩一更新。由于对 $\widetilde{\mathrm{H}}$ 和 $\widetilde{\mathrm{H}}^{-1}$ 的所有秩一更新都通过 Sherman-Morrison 公式显式维护，且每次更新耗时为 $O\left(n^2\right)$，我们将总运行时间界定为

$$\widetilde{O}\left(\left(n^2+\mathcal{T}_{\text{gradient}}\right)\sqrt{n\nu}\log\frac{1}{\varepsilon}\right),$$

其中 $\mathcal{T}_{\text{gradient}}$ 是对梯度预言机执行查询所需的时间；该预言机以 y 为输入并返回 $\nabla\phi(\mathrm{y})$。

定理的正式表述见定理 25。

作为推论，我们看到，当可以访问定义域 K 上一个 $O(n)$-强自协调障碍函数的梯度预言机时，我们可以在$\widetilde{O}\left(n^3+n\mathcal{T}_{\text{gradient}}\log\frac{1}{\varepsilon}\right)$ 时间内进行线性优化。正式表述见推论 26。

通过内点法求解半正定规划。 对于我们的半正定规划应用，我们不直接求解目标函数（1），而是像 [JKL+20]中一样，考虑对偶问题

$$\min\left\{\langle\boldsymbol{c},\boldsymbol{y}\rangle:\boldsymbol{B}-\sum_{i=1}^{m}\boldsymbol{y}_i\boldsymbol{A}_i\succeq 0\right\}, \tag{2}$$

我们通过求解障碍函数形式来优化该问题

$$g_\mu(\boldsymbol{y})=\frac{\langle\boldsymbol{c},\boldsymbol{y}\rangle}{\mu}+\phi(\boldsymbol{y}), \tag{3}$$

其中，ϕ 是集合 $\{\mathrm{y}:\ \boldsymbol{B}-\sum_{i=1}^{m}\mathrm{y}_i\boldsymbol{A}_i\succeq 0\}$上的强自协调障碍函数。不幸的是，标准的对数行列式障碍函数不满足我们的势函数分析所要求的强自协调性。我们通过将其缩放 $\sqrt{m}$ 倍来解决这一问题，但这也会进一步降低内点法的速度。具体来说，在我们的情形中，我们使用

$$\phi(\boldsymbol{y})=-\sqrt{m}\cdot\log\det\left(\boldsymbol{B}-\sum_{i=1}^{m}\boldsymbol{y}_i\boldsymbol{A}_i\right),$$

对于该函数，我们可以证明它满足所要求的强自协调性，其中 $\nu=n\sqrt{m}$。要计算障碍函数的梯度，只需计算

$$[\nabla\phi(\boldsymbol{y})]_i=\left\langle\boldsymbol{A}_i,\left(\boldsymbol{B}-\sum_{i=1}^{m}\boldsymbol{y}_i\boldsymbol{A}_i\right)^{-1}\right\rangle$$

这可以在 $O\left(mn^2+n^\omega\right)$ 时间内完成。与以往快速求解半正定规划的工作相比，我们的关键优势在于，无须计算 ϕ 的海森矩阵，因为这涉及代价极高的

矩阵乘法，因此运行时间还会额外依赖于 m。事实上， [NN92] 采用了一系列巧妙的技巧，在 $O\left(mn^{\omega}+m^{2}n^{2}+m^{\omega}\right)$时间内计算海森矩阵，而 [JKL+20]则提出了一系列基于矩形矩阵乘法的技术，将运行时间缩短至计算一个 $m\times n^{2}$ 矩阵与一个 $n^{2}\times m$ 矩阵相乘所需的时间。幸运的是，借助我们的动态预条件器，我们可以完全避开这些瓶颈。

结合我们对基于梯度的内点法的分析，我们得到了最终运行时间 $\widetilde{O}\left(mn^{4}+m^{1.25}n^{3.5}\right)$。定理 32 给出了形式化表述。

1.3 相关工作

优化方法的相关文献极为丰富。这里我们总结其中一些相关成果。

割平面法。 一类称为割平面法的优化技术，通过向分离预言机查询，反复细化一个包含目标解的凸集。自 20 世纪 50 年代问世以来，设计高效的割平面算法一直是一项长期工作 [Sho77、YN76、Kha80、KTE88、NN89、Vai89、AV95、BV04、LSW15、JLSW20]。

内点法。 内点法是用于在凸集上最小化线性函数的基础优化技术，前提是能够获取该集合障碍函数的一阶和二阶信息。Nesterov 和 Nemirovski 对早期大量研究 [Fri55、KMY89、Meg89、FM90]作了精彩阐述 [NN94], ，他们发展了自协调障碍函数理论，并将内点法的应用拓展至更一般的情形，包括半定规划以及更广义的锥优化。近期研究利用内点法理论，在降低线性规划和半定规划的理论运行时间方面取得了重要进展 [LS14、CLS19、LSZ19、vdBLSS20、vdB20、JKL+20、HJS+22]。

拟牛顿方法。 一系列重要的发展集中于拟牛顿方法，这些方法试图仅通过访问梯度，获得与牛顿方法相当的收敛保证。一组著名的结果由 BFGS 和 L-BFGS 方法给出[Bro70、 Fle70、 Gol70、Sha70、 BNS94], ，它们被用于实际场景。另请参见 [GS19、 GS22、Tun01]，了解拟牛顿方法在内点求解器中的应用。我们注意到 [GS19、GS22] 使用与标准牛顿步交错进行的拟牛顿步。尽管它们的结果并不完全依赖梯度信息，但由于减少了线性方程组求解的总次数，它们在实践中带来了显著的加速。在 [Tun01], 中，线性规划的零阶和一阶方法均有所提供，但它们仅被证明能够在有限时间内收敛，而没有给出迭代复杂度的明确界。

半正定规划求解器。早期的半正定规划求解器依赖于切割平面法。从 Nesterov-Nemirovski 的工作开始 [NN92] 以及 Anstreicher [Ans00], 已经开发出了几种基于内点法的半正定规划求解器。最近，基于逆矩阵维护的新技术为半正定规划问题提供了快速算法 [JKL+20，HJS+22]。

2 预备知识

2.1 符号约定

我们用粗体表示矩阵和向量。我们用 $\langle \cdot, \cdot \rangle$ 表示内积。给定一个对称矩阵 $\boldsymbol{A}$，我们用 $\|\boldsymbol{A}\|$ 表示 $\boldsymbol{A}$的谱范数，即 $\|\boldsymbol{A}\| = \max\{-\lambda_{\min}(\boldsymbol{A}), \lambda_{\max}(\boldsymbol{A})\}$，其中 $\lambda_{\min}(\boldsymbol{A})$ 和 $\lambda_{\max}(\boldsymbol{A})$ 分别表示 $\boldsymbol{A}$的最小和最大特征值。我们用 $\|\boldsymbol{A}\|_F$ 表示其 Frobenius 范数，即其特征值的 ℓ_2 范数。类似地，我们用 $\|\boldsymbol{A}\|_1$ 表示其特征值的 ℓ_1 范数。对于任意对称矩阵 $\boldsymbol{A}$，我们用 $\boldsymbol{A}_{\geq 1}$ 表示由 $\boldsymbol{A}$ 得到的矩阵，其中所有小于 1 的特征值均增至 1，并以类似方式定义 $\boldsymbol{A}_{<1}$。

给定函数 $g\colon K \to \mathbb{R}$，其中 $K \subseteq \mathbb{R}^n$ 是凸集，我们用 $\nabla g(\mathrm{y})$ 表示 g的梯度，用 $\nabla^2 g(\mathrm{y})$ 表示 y 处的海森矩阵。为简化记号，只要上下文含义清楚，我们就用 $\mathrm{H}_{\mathrm{y}} := \nabla^2 g(\mathrm{y})$ 表示。求解线性方程组时，我们会维护一个预条件器，通常记作 H。

2.2 偏心率势函数

用于分析该方法的主要工具之一是偏心势函数，该函数由 Dunagan 和 Harvey 提出 [DH07]，用于分析共轭梯度方法的某个版本，且不依赖基于最佳插值多项式的论证。广义而言，偏心率衡量矩阵 $\boldsymbol{H}$ 作为另一个矩阵 $\boldsymbol{H}$的预条件器的优劣。在标准分析中，一个关键目标是界定谱范数

$$\|\boldsymbol{X} - \boldsymbol{I}\|$$

其中

$$\boldsymbol{X} = \boldsymbol{H}^{-1/2} \widetilde{\boldsymbol{H}} \boldsymbol{H}^{-1/2}.$$

自然地，这个值越接近 0，H 作为 H 的预条件器就越好。由于这个函数不一定最容易处理，Dunagan 和 Harvey 引入了如下定义的偏心势：

$$\mathcal{E}(\boldsymbol{X}) = \det\left(\frac{\boldsymbol{X}^{1/2} + \boldsymbol{X}^{-1/2}}{2}\right) = \frac{1}{2^n} \frac{\det(\boldsymbol{X} + \boldsymbol{I})}{\sqrt{\det(\boldsymbol{X})}},$$

对于该函数，可以很容易验证其在 $\boldsymbol{X} = \boldsymbol{I}$ 时取得最小值，此时 $\boldsymbol{H}$ 是一个完美的预条件器。该势函数具有若干有利性质，其中包括：对 1 或 $\boldsymbol{H}$ 进行秩 $\boldsymbol{H}$-1 更新后，其值易于跟踪。在下面的引理中，我们将看到进行秩 1 更新后偏心率如何演变（证明见附录 A.1）。

引理 4. 一个秩-1更新

$$\boldsymbol{X}' = \boldsymbol{X} + \boldsymbol{u}\boldsymbol{u}^\top$$

决定偏心率的乘法变化：

$$\frac{\mathcal{E}(\boldsymbol{X}')}{\mathcal{E}(\boldsymbol{X})} = \frac{1 + \boldsymbol{u}^\top (\boldsymbol{I} + \boldsymbol{X})^{-1} \boldsymbol{u}}{\sqrt{1 + \boldsymbol{u}^\top \boldsymbol{X}^{-1} \boldsymbol{u}}}.$$

我们还证明了，如果能够获得一个与 X 的极小/极大特征向量良好对齐的向量 u，那么便可以利用它进行秩一更新，从而显著降低偏心率。

引理 5。 如果单位向量 $\boldsymbol{u}$ 满足

1. $\boldsymbol{u}^{\top}\boldsymbol{X}^{-1}\boldsymbol{u} \geq \gamma$, *then* $\frac{\mathcal{E}\left(\boldsymbol{X}+\boldsymbol{u}\boldsymbol{u}^{\top}\right)}{\mathcal{E}(\boldsymbol{X})} \leq \frac{2}{\sqrt{1+\gamma}}$,
2. $\mathrm{u}^{\top}\mathrm{X}\,\mathrm{u} \geq \gamma$，则 $\frac{\mathcal{E}\left(\mathrm{X}-\frac{\mathrm{X}\boldsymbol{u}\boldsymbol{u}^{\top}\mathrm{X}}{1+\mathrm{u}^{\top}\mathrm{X}\,\mathrm{u}}\right)}{\mathcal{E}(\mathrm{X})} \leq \frac{2}{\sqrt{1+\gamma}}$。

我们在附录 A.2中展示了这一点。关键是，证明和我们的算法都依赖于在执行秩一更新后求矩阵逆的 Sherman-Morrison 公式。

引理 6 (Sherman-Morrison)。 假设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 是一个可逆矩阵，且 $\boldsymbol{u}$、 $\boldsymbol{v} \in \mathbb{R}^{n}$ 是列向量。那么 $\boldsymbol{A}+\boldsymbol{u}\boldsymbol{v}^{\top}$ 可逆，当且仅当 $1+\boldsymbol{v}^{\top}\boldsymbol{A}^{-1}\boldsymbol{u} \neq 0$。在这种情况下，

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

为了简化我们的证明，我们还陈述偏心率的一个有用性质。

引理 7。 偏心率在相似变换下是不变量。给定可逆的 $\boldsymbol{X}$、 $\boldsymbol{Y}$，我们有

$$\mathcal{E}(\boldsymbol{X}) = \mathcal{E}\left(\boldsymbol{Y}\boldsymbol{X}\boldsymbol{Y}^{-1}\right).$$

此外

$$\mathcal{E}\left(\boldsymbol{X}^{-1}\right) = \mathcal{E}(\boldsymbol{X}).$$

最后，我们非常依赖以下引理，该引理依赖于一个特征值界，我们将在第5 节通过主序技术证明。它使我们能够控制相关矩阵发生微小变化时偏心率的变化。

引理 8。 设 $\boldsymbol{A}$， $\boldsymbol{B}$ 为可逆矩阵。则

$$\mathcal{E}(\boldsymbol{A}\boldsymbol{B}) \leq \mathcal{E}(\boldsymbol{A}) \cdot \sqrt{\frac{\det\left(\boldsymbol{B}_{\geq 1}\right)}{\det\left(\boldsymbol{B}_{<1}\right)}}.$$

证明。 利用引理 27，我们写出：

$$\mathcal{E}(\boldsymbol{A}\boldsymbol{B}) = \frac{1}{2^n}\frac{\det(\boldsymbol{A}\boldsymbol{B}+\boldsymbol{I})}{\sqrt{\det(\boldsymbol{A}\boldsymbol{B})}} \leq \frac{1}{2^n}\cdot\frac{\det(\boldsymbol{A}+\boldsymbol{I})\cdot\det\left(\boldsymbol{B}_{\geq 1}\right)}{\sqrt{\det(\boldsymbol{A})}\cdot\sqrt{\det(\boldsymbol{B})}} = \mathcal{E}(\boldsymbol{A})\cdot\sqrt{\frac{\det\left(\boldsymbol{B}_{\geq 1}\right)}{\det\left(\boldsymbol{B}_{<1}\right)}}.$$

□

2.3 理查德森迭代

Richardson 迭代可能是求解线性方程组最重要的迭代方法。一般来说，给定一个线性方程组 $\mathrm{H}\,\mathrm{x} = \mathrm{b}$，该迭代由若干步骤组成，通过执行如下形式的迭代来尝试改进当前迭代点：$\mathrm{x}' = \mathrm{x} + \eta(\mathrm{b} - \mathrm{H}\,\mathrm{x})$，其中步长为某个适当的 η。记残差为 $\mathrm{r} = \mathrm{b} - \mathrm{H}\,\mathrm{x}$，该迭代通过向解中加入一小部分残差来更新解。为了提高这种迭代方法的收敛速度，人们通常会使用预条件。具体而言，使用一个在某种适当意义下近似于 H 的矩阵 H，且其逆矩阵可用，我们便可以改为运行迭代 $\mathrm{x}' = \mathrm{x} + \eta\widetilde{\mathrm{H}}^{-1}(\mathrm{b} - \mathrm{H}\,\mathrm{x})$。这等价于先通过令 $\mathrm{y} = \widetilde{\mathrm{H}}^{1/2}\mathrm{x}$ 进行基变换，并考虑等价方程组，再对原方程组运行标准 Richardson 迭代。

$$\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}\cdot\boldsymbol{y} = \widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{b}.$$

2.4 强自协调障碍函数与内点法

我们求解以下形式的一般优化问题。

$$\min_{\boldsymbol{y}\in K}\langle \boldsymbol{c},\boldsymbol{y}\rangle \tag{4}$$

其中 K 是一个凸域。为此，我们采用一种路径跟踪内点法，求解一系列障碍目标函数，这些目标函数是如下形式的凸最小化问题：

$$g_\mu(\boldsymbol{y})=\frac{\langle \boldsymbol{c},\boldsymbol{y}\rangle}{\mu}+\phi(\boldsymbol{y}),$$

其中 ϕ 是定义域 K的障碍函数。每个函数 g_μ 都有唯一的极小化点。对于所有 $\boldsymbol{y}^\star{}_\mu=\arg\min(\boldsymbol{y})$， $\mu\in(0,\ \infty)$所构成的极小化点集合，表示与该障碍目标函数相对应的中心路径。要解决（4），经典内点法文献（引理 21）表明，只需在中心性参数 g_μ 足够小时，求得 μ的一个近似极小化点。

为此，我们采用一种路径跟踪过程。在该过程的每次迭代中，对于足够小的标量 δ，我们求解形式为 $g_{\mu/(1+\delta)}$的最小化问题，并以 g_μ的一个（近似）最优解作为热启动。 δ 的取值取决于障碍函数的性质，因此决定了该方法的收敛速度。障碍函数的性质对于获得高效算法至关重要。标准内点法理论利用自协调障碍函数的性质 [NN94], ，而本文采用一种稍强的性质，即强自协调性 [LLV。

这给出了一个比普通自协调性稍强的条件，用于描述在迭代点之间移动时海森矩阵的变化。尽管经典文献中没有明确提及这一点，但许多标准障碍函数都具备这一性质，包括对数障碍函数、通用障碍函数和熵障碍函数。

定义 9（强自协调函数）。给定一个凸域 K 和 $g:\ K\to\mathbb{R}$，如果对于任意y $\in K$ 和 h：凸函数 g 满足以下条件，则称其为强自协调函数：

$$\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\frac{d}{dt}\boldsymbol{H}_{\boldsymbol{y}+t\boldsymbol{h}}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right\|_F\le 2\|\boldsymbol{h}\|_{\boldsymbol{H}_{\boldsymbol{y}}}.$$

我们回顾一下，标准自协调性将涉及 Frobenius 范数的界替换为较弱的谱范数。

引理 10。给定任意强自协调函数 $g:\ K\to\mathbb{R}$，对于任意 $\boldsymbol{x}\ \boldsymbol{\delta}$，满足$\boldsymbol{x}$，$\boldsymbol{x}+\boldsymbol{\delta}\in K$, $\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}<1$,

$$\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\left(\boldsymbol{H}_{\boldsymbol{y}+\boldsymbol{\delta}}-\boldsymbol{H}_{\boldsymbol{y}}\right)\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right\|_F\le\frac{\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}}{\left(1-\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\right)^2}.$$

此外，如果 g 仅是自协调的，那么

$$\boldsymbol{H}_{\boldsymbol{y}}\cdot\left(1-\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\right)^2\preceq\boldsymbol{H}_{\boldsymbol{y}+\boldsymbol{\delta}}\preceq\boldsymbol{H}_{\boldsymbol{y}}\cdot\frac{1}{\left(1-\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\right)^2}.$$

除了这些性质之外，我们还使用 ν-强自协调障碍函数这一概念，它类似于自协调障碍函数，但还满足强自协调性。

定义 11 （ν-强自协调障碍函数）。 给定一个凸域 K 和一个强自协调函数 $g\colon K \to \mathbb{R}$，如果当 $\mathrm{y} \to \partial K$ 时 $g(\mathrm{y}) \to \infty$，则称 g 是一个 ν-强自协调障碍函数，且

$$\|\nabla g(\boldsymbol{y})\|^2_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}} \leq \nu\,,$$

对于所有 $\mathrm{y} \in K$。

更强的性质有助于在仅对海森矩阵进行微小改变时，获得偏心率增幅的更紧界限。我们依赖以下重要引理，并将在第 5 节中证明：

引理 12。 设 $K \subseteq \mathbb{R}^n$ 为一个凸集。给定任意强自协调函数 $g\colon K \to \mathbb{R}$，对于任意 $\boldsymbol{y}$ $\boldsymbol{\delta}$，满足 $\boldsymbol{y},\ \boldsymbol{y}+\boldsymbol{\delta} \in K$, $\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}} < \left(1-\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\right)^2 < 1$，以及任意预条件器 $\widetilde{\ }\boldsymbol{H}$，

$$\mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{\boldsymbol{y}+\boldsymbol{\delta}}\right) \leq \mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{\boldsymbol{y}}\right)\cdot\exp\left(\frac{1}{2}\sqrt{n}\cdot\frac{\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}}{\left(1-\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\right)^2-\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}}\right)\,.$$

此外，我们还要求中心路径邻域中各点具有良好条件性的一些保证。以下定义对此进行了刻画。

定义 13 ($\kappa(\mu)$-条件目标函数)**.** 若障碍目标函数 (5) 满足 $\kappa(\mu)$-条件，则令$\boldsymbol{y}^{\mu'}$为 $g_{\mu'}$ 的最小化点，有

$$\max\left\{\left\|\boldsymbol{H}_{\boldsymbol{y}_{\mu'}}\right\|,\left\|\boldsymbol{H}^{-1}_{\boldsymbol{y}_{\mu'}}\right\|\right\} \leq \kappa(\mu)$$

对所有 $\mu' \geq \mu$ 均成立。

这刻画了中心路径上点对应的海森矩阵特征值可能有多大或多小，适用于所有高于给定阈值 μ 的中心性参数。尽管标准文献中仅隐式地使用了这一量，但它与内点法的大多数具体实现都有关，因为即使使用快速矩阵乘法来求解相关线性方程组，障碍目标函数 $\kappa(\mu)$ 的条件数也决定了存储海森矩阵及其逆矩阵所需的精度位数。这个上界的一个标准性质是，它对中心路径邻域内的点同样成立；这一点我们将在附录A.4中证明。

引理 14. 设 $g_\mu : K \to \mathbb{R}$ 是一个障碍目标函数，其自协调障碍函数如 *(5)* 所示，并设 $\boldsymbol{y} \in$ 属于 $\boldsymbol{i}$ $\boldsymbol{nt}\,(K)$，且满足 $\|\nabla g_\mu(\boldsymbol{y})\|_{\boldsymbol{H}^{-1}\boldsymbol{y}} \leq 1/3$。则

$$\max\left\{\|\boldsymbol{H}_{\boldsymbol{y}}\|,\left\|\boldsymbol{H}^{-1}_{\boldsymbol{y}}\right\|\right\} \leq 4\kappa(\mu)\,.$$

这些事实决定了 ε障碍函数形式的条件数概念，我们将在定义22中对其进行形式化定义和讨论，并将利用它来给出我们的收敛保证。

3 使用自适应预条件器求解线性系统

我们首先考虑求解以下形式的线性方程组：$\boldsymbol{H}\boldsymbol{x} = \boldsymbol{b}$，并提供一种基于 [DH07]所使用技术的算法。我们强调我们工作的贡献的重要性，因为 [DH07]使用了一组略有不同的更新规则；在矩阵相关的量只能像我们在下面第 4 节中所做的那样进行近似的情况下，很难为这些更新规则证明理论保证。

虽然本节的分析假设可以访问 H，而后续各节并非如此，但我们注意到，对 H 的访问仅通过矩阵向量乘积进行。这一点很重要，因为只需进行常数次梯度查询，就可以估算海森矩阵与向量的乘积。

本节将说明，如果在 Richardson 迭代的单步中未能取得显著进展，就会产生一个证书，使我们能够改进当前的预条件器。这一分析大体沿用了 Dunagan-Harvey 中的思路，但从技术角度而言略作简化，因为它并不试图与共轭梯度算法保持一致。相反，它只执行 Richardson 迭代，并选择使残差范数最小的步长。我们先给出关键引理，再说明如何利用这些引理推导出 [DH07]中的主要结果。下一节将把这些结果推广到海森矩阵非常数的情形，为此我们会运用矩阵分析中的其他技术。

3.1 使用 Richardson 迭代最小化残差范数

我们给出以下引理：当（未预条件化的）Richardson 更新的单步迭代未能显著降低梯度范数时，该引理提供偏心率的证书。在附录 B.1 中，我们首先分析未预条件化的情形（见引理 34）。作为推论，我们得到了与执行预条件化步骤相对应的引理的一般版本。

引理 15. 勒 $\widetilde{t}\ \boldsymbol{H}, \boldsymbol{H} \in \mathbb{R}^{n\times n}$，以及向量 $\boldsymbol{b}, \boldsymbol{x} \in \mathbb{R}^n$。令 $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{H}\boldsymbol{x}$，并考虑这一步。

$$\boldsymbol{x}' = \boldsymbol{x} + \frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}{\left\|\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\,.$$

设 $\beta \in (0,1)$ 为一个标量。若新的残差 $\boldsymbol{r}' = \boldsymbol{b} - \boldsymbol{H}\boldsymbol{x}'$ 满足

$$\left\|\boldsymbol{r}'\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2 \geq (1-\beta)\left\|\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2\,,$$

我们至少可以获得以下偏心度证书之一：

1. $\dfrac{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}}^2}{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2^2} \geq \frac{1}{\sqrt{\beta}}$,
2. $\dfrac{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}^{1/2}}^2}{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2^2} \geq \frac{1}{\sqrt{\beta}}$.

证明可见于附录 B.2。最后，根据引理15 提供的证书，我们可以设计一个例程来更新我们的伪海森矩阵 $\widetilde{\boldsymbol{H}}_{-1}$。其对偏心率的影响如下方引理所述，我们将在附录 B.3中证明。

引理 16。 设 $\boldsymbol{H} \in \mathbb{R}^{n\times n}$，并假设逆矩阵 $\widetilde{\boldsymbol{H}}^{-1}$ 可用。给定由引理 *15*所提供的类型 *1* 或类型 *2* 偏心率证书，存在一个算法（算法 *1*）$\widetilde{\boldsymbol{H}}$，该算法对 $\boldsymbol{H}$ 及其逆矩阵执行秩一更新，以获得一个新的预条件器 $\boldsymbol{H}'$，使得

$$\mathcal{E}\left(\widetilde{\boldsymbol{H}'}^{-1}\boldsymbol{H}\right) \le \mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}\right)\cdot\frac{2}{\sqrt{1+\frac{1}{\sqrt{\beta}}}}.$$

此更新可在 $O\left(n^2\right)$ 时间内实现。此外，或者

1. $1 \le \left\|\widetilde{\boldsymbol{H}}^{-1/2}\widetilde{\boldsymbol{H}'}\widetilde{\boldsymbol{H}}^{-1/2}\right\| \le \left(2\cdot\frac{\mathcal{E}(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1})}{\mathcal{E}(\widetilde{\boldsymbol{H}'}\boldsymbol{H}^{-1})}\right)^2$，*or*
2. $1 \le \left\|\widetilde{\boldsymbol{H}}^{1/2}\widetilde{\boldsymbol{H}'}^{-1}\widetilde{\boldsymbol{H}}^{1/2}\right\| \le \left(2\cdot\frac{\mathcal{E}(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1})}{\mathcal{E}(\widetilde{\boldsymbol{H}'}\boldsymbol{H}^{-1})}\right)^2$.

算法 1 给定偏心率证书时，预条件器秩一更新的伪代码。

确保： 更新预条件器 $\widetilde{\boldsymbol{H}}$ 及其逆矩阵可在 $O\left(n^2\right)$ 时间内完成更新，并且只需进行常数次梯度查询。

过程 UpdatePreconditioner($\widetilde{\mathrm{H}}$, $\widetilde{\mathrm{H}}^{-1}$, r, type)
  if type $=1$ **then**
    $\widetilde{\boldsymbol{H}'} = \widetilde{\boldsymbol{H}} - \frac{\boldsymbol{r}\boldsymbol{r}^\top}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2+\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}$, $\widetilde{\boldsymbol{H}'}^{-1} = \widetilde{\boldsymbol{H}}^{-1} + \frac{\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\boldsymbol{r}^\top\widetilde{\boldsymbol{H}}^{-1}}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}$
  else ▷ type $=2$
    $\widetilde{\boldsymbol{H}'} = \widetilde{\boldsymbol{H}} + \frac{\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\boldsymbol{r}^\top\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}$, $\widetilde{\boldsymbol{H}'}^{-1} = \widetilde{\boldsymbol{H}}^{-1} - \frac{\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\boldsymbol{r}^\top\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2+\left\|\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}$
  end if
  return $\left(\widetilde{\boldsymbol{H}'}, \widetilde{\boldsymbol{H}'}^{-1}\right)$
end procedure

3.2 总结回顾

结合引理 15 和引理 16 （算法 1），我们得到一个过程（算法 2），它要么执行一步，使残差 $\|\mathrm{b}-\mathrm{H}\,\mathrm{x}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2$ 按一个常数乘法因子降低；要么在 $O\left(n^2\right)$ 时间内更新 $\widetilde{\mathrm{H}}$ 和 $\widetilde{\mathrm{H}}^{-1}$，使偏心率按一个常数因子降低。这将是下一节所述路径跟踪方法的主要驱动过程。

虽然这一过程足以用于我们的内点法，但作为预热，我们首先分析这样一种情形：我们打算求解一个单独的线性系统，从而恢复 [DH07] 的主要结果。其完整证明可见附录 B.4。

引理17。 考虑线性系统 $\boldsymbol{H}\boldsymbol{x}=\boldsymbol{b}$，其中 $\boldsymbol{H}$ 是一个对称正定矩阵，令 $\boldsymbol{x}_0$ 为初始解，并令 $\boldsymbol{H}_0=\boldsymbol{I}$ 为初始预条件器。运行迭代

$$\left(\widetilde{\boldsymbol{H}}_{t+1}, \widetilde{\boldsymbol{H}}_{t+1}^{-1}, \boldsymbol{x}_{t+1}\right) = \textsc{StepOrUpdate}\left(\boldsymbol{H}, \widetilde{\boldsymbol{H}}_t, \widetilde{\boldsymbol{H}}_t^{-1}, \boldsymbol{b}, \boldsymbol{x}_t\right)$$

对于 $T=100\left(\text{对数}\,\mathcal{E}(\boldsymbol{H})+\text{对数}\,\frac{1}{\varepsilon}\right)$ 步骤，我们得到一个向量 $\boldsymbol{x}_T$，使得

$$\|\boldsymbol{b}-\boldsymbol{H}\boldsymbol{x}_T\| \le \varepsilon\cdot\|\boldsymbol{b}-\boldsymbol{H}\boldsymbol{x}_0\|.$$

Algorithm 2 Pseudocode for the step problem.

Ensure: Returns a new iterate $\boldsymbol{x'}$ such that $\|\boldsymbol{b}-\boldsymbol{H}\boldsymbol{x'}\|_{\widetilde{\boldsymbol{H}}^{-1}} \leq (1-\beta)\|\boldsymbol{b}-\boldsymbol{H}\boldsymbol{x}\|_{\widetilde{\boldsymbol{H}}^{-1}}$, or returns a new preconditioner $\widetilde{\boldsymbol{H}}'$ together with its inverse such that $\mathcal{E}\left(\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1}\right) \leq \mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right)\cdot\frac{2}{\sqrt{1+\frac{1}{\sqrt{\beta}}}}$, and either $1 \leq \left\|\widetilde{\boldsymbol{H}}^{-1/2}\widetilde{\boldsymbol{H}}'\widetilde{\boldsymbol{H}}^{-1/2}\right\| \leq \left(2\cdot\frac{\mathcal{E}(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1})}{\mathcal{E}(\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1})}\right)^2$, or $1 \leq \left\|\widetilde{\boldsymbol{H}}^{1/2}\widetilde{\boldsymbol{H}}'^{-1}\widetilde{\boldsymbol{H}}^{1/2}\right\| \leq \left(2\cdot\frac{\mathcal{E}(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1})}{\mathcal{E}(\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1})}\right)^2$.

procedure StepOrUpdate($\boldsymbol{H}, \widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{b}, \boldsymbol{x}$)
  $\boldsymbol{r} = \boldsymbol{b}-\boldsymbol{H}\boldsymbol{x}$
  $\boldsymbol{x'} = \boldsymbol{x} + \frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}{\left\|\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}$, $\boldsymbol{r'} = \boldsymbol{b}-\boldsymbol{H}\boldsymbol{x'}$
  if $\|\boldsymbol{r'}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2 \leq (1-\beta)\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2$ **then**
    return $\left(\widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{x'}\right)$ ▷ return new iterate
  else
    if $\frac{\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2} \geq \frac{1}{\sqrt{\beta}}$ **then** ▷ test for type 1 excentricity certificate
      $\left(\widetilde{\boldsymbol{H}}', \widetilde{\boldsymbol{H}}'^{-1}\right)$ =UpdatePreconditioner($\widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{r}$, type $= 1$)
    else ▷ if test fails, then we must have a type 2 excentricity certificate
      $\left(\widetilde{\boldsymbol{H}}', \widetilde{\boldsymbol{H}}'^{-1}\right)$ =UpdatePreconditioner($\widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{r}$, type $= 2$)
    end if
    return $\left(\widetilde{\boldsymbol{H}}', \widetilde{\boldsymbol{H}}'^{-1}, \boldsymbol{x}\right)$ ▷ return new preconditioner
  end if
end procedure

4 估计海森矩阵-向量乘积

尽管第3 节中的分析为预条件器质量的变化提供了紧确界，但其中描述的算法几乎完全依赖于能够访问矩阵 H。在我们的设定中，矩阵是某个凸函数的海森矩阵，我们无法直接访问它。相反，我们注意到，我们与 H 的所有交互都以矩阵向量乘积的形式进行。因此，我们应该可以不用计算形如 H $_y$v 的乘积，而是对其进行近似。

$$\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v} \approx \frac{1}{\tau}\left(\nabla g\left(\boldsymbol{y}+\tau \boldsymbol{v}\right)-\nabla g\left(\boldsymbol{y}\right)\right),$$

对于某个适当的步长 τ。为此，考虑 H $_y$ 的条件数十分重要，因为要获得准确的估计，需要将 τ 设置得足够大。然而，我们不希望 τ 极大，因为其大小在一定程度上决定了计算该近似所需的精度位数。

本节将给出第 3 节中算法的鲁棒版本，并证明它们满足类似的保证，尽管我们只能通过梯度查询访问函数 g。附录 C.2 中给出了关于我们所用近似 p_y (v) ≈H $_y$(v) 和 $n_y(v) \approx \|v\|_{H_y}$质量的正式陈述。

此时，我们可以给出对第 3 节中算法所作的适当修改。

引理 *18.* 设 $g : K \to \mathbb{R}$ 为一个自协调函数，设 $y \in int\,(K)$，$Hy = \nabla^2 g(y)$，$\widetilde{H} \in \mathbb{R}^{n\times n}$，以及向量 $b, x \in \mathbb{R}^n$。此外，假设 $max\left\{\|Hy\|, \|H^{-1}y\|, \left\|\widetilde{H}\right\|, \left\|\widetilde{H}^{-1}\right\|, \|b\|_2\right\} \le B$，其中某个标量为 $B \ge 1000$。设 $r = b - p_y(x)$，并考虑这一步。

$$\boldsymbol{x}' = \boldsymbol{x} + \frac{n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2}{\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}}\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right).$$

令 $\beta \in \left(\frac{1}{B}, 1\right)$ 为一个标量。令新的残差 $r' = b - p_y(x')$。已知$\min\left\{\|r\|_{\widetilde{H}^{-1}}, \|r'\|_{\widetilde{H}^{-1}}\right\} \ge \frac{1}{B}$，并且

$$\left\|\boldsymbol{r}'\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}} \ge (1-\beta)\left\|\boldsymbol{r}\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}},$$

我们至少可以得到以下偏心度证书之一：

1. $\dfrac{\left\|\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|^2_{\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\tilde{H}^{-1}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}}}{\left\|\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|^2_2} \ge \dfrac{1}{\sqrt{\frac{20}{9}\beta}}$,
2. $\dfrac{\left\|\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|^2_{\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\tilde{H}^{-1}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}}}{\left\|\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|^2_2} \ge \dfrac{1}{\sqrt{\frac{20}{9}\beta}}$.

该证明过程较为繁琐，但并无特别深刻的洞见，因此我们将其推迟至附录 C.3。该方法基于为涉及的所有量提供乘法误差界

Algorithm 3 Pseudocode for preconditioner rank-1 updates, given excentricity certificates.

Ensure: Updates the preconditioner $\widetilde{\boldsymbol{H}}$ and its inverse in $O\left(n^2\right)$ time, plus a constant number of gradient queries.

procedure RobustUpdatePreconditioner($\boldsymbol{y}, \widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{r}$, type)
  if type $= 1$ **then**
    $\widetilde{\boldsymbol{H}'} = \widetilde{\boldsymbol{H}} - \frac{\boldsymbol{r}\boldsymbol{r}^\top}{n_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2 + \|\boldsymbol{r}\|^2_{\widetilde{\boldsymbol{H}}^{-1}}}$, $\widetilde{\boldsymbol{H}'}^{-1} = \widetilde{\boldsymbol{H}}^{-1} + \frac{\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\boldsymbol{r}^\top\widetilde{\boldsymbol{H}}^{-1}}{n_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2}$
  else ▷ type $= 2$
    $\widetilde{\boldsymbol{H}'} = \widetilde{\boldsymbol{H}} + \frac{p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^\top}{\left(1+\frac{1}{400\cdot B^{20}}\right)^2 n_y\left(\widetilde{\boldsymbol{H}}^{-1}\nabla g(\boldsymbol{y})\right)^2}$
    $\widetilde{\boldsymbol{H}'}^{-1} = \widetilde{\boldsymbol{H}}^{-1} - \frac{\widetilde{\boldsymbol{H}}^{-1}p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^\top\widetilde{\boldsymbol{H}}^{-1}}{\left(1+\frac{1}{400\cdot B^{20}}\right)^2 n_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2 + \left\|p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}}$
  end if
  return $\left(\widetilde{\boldsymbol{H}'}, \widetilde{\boldsymbol{H}'}^{-1}\right)$
end procedure

procedure $p_{\boldsymbol{y}}(\boldsymbol{v})$
  $\tau = \frac{1}{1000\|\boldsymbol{v}\|B^{21}}$
  return $\frac{1}{\tau}\left(\nabla g\left(\boldsymbol{y}+\tau\boldsymbol{v}\right) - \nabla g\left(\boldsymbol{y}\right)\right)$
end procedure

procedure $n_{\boldsymbol{y}}(\boldsymbol{v})$
  $\tau = \frac{1}{1000\|\boldsymbol{v}\|B}$
  return $\left(\frac{1}{1-\frac{1}{1000}}\right)\cdot\sqrt{\frac{1}{\tau}\cdot\langle\boldsymbol{v}, \nabla g\left(\boldsymbol{y}+\tau\boldsymbol{v}\right) - \nabla g\left(\boldsymbol{y}\right)\rangle}$
end procedure

Algorithm 4 Pseudocode for the step problem.

Ensure: Returns a new iterate $\boldsymbol{x'}$ such that $\|\boldsymbol{b}-\boldsymbol{H_y x'}\|_{\widetilde{\boldsymbol{H}}^{-1}} \leq (1-\beta)\|\boldsymbol{b}-\boldsymbol{H_y x}\|_{\widetilde{\boldsymbol{H}}^{-1}}$, or returns a new preconditioner $\widetilde{\boldsymbol{H'}}$ together with its inverse such that $\mathcal{E}\left(\widetilde{\boldsymbol{H'}}\boldsymbol{H_y}^{-1}\right) \leq \mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H_y}^{-1}\right)\cdot\frac{2}{\sqrt{1+\frac{1}{\sqrt{\beta}}}}$, and either $1 \leq \left\|\widetilde{\boldsymbol{H}}^{-1/2}\widetilde{\boldsymbol{H'}}\widetilde{\boldsymbol{H}}^{-1/2}\right\| \leq \left(2\cdot\frac{\mathcal{E}(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1})}{\mathcal{E}(\widetilde{\boldsymbol{H'}}\boldsymbol{H}^{-1})}\right)^2$, or $1 \leq \left\|\widetilde{\boldsymbol{H}}^{1/2}\widetilde{\boldsymbol{H'}}^{-1}\widetilde{\boldsymbol{H}}^{1/2}\right\| \leq \left(2\cdot\frac{\mathcal{E}(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1})}{\mathcal{E}(\widetilde{\boldsymbol{H'}}\boldsymbol{H}^{-1})}\right)^2$.

procedure RobustStepOrUpdate($\boldsymbol{y}, \widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{b}, \boldsymbol{x}$)
  $\boldsymbol{r} = \boldsymbol{b} - p_{\boldsymbol{y}}(\boldsymbol{x})$
  if $\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}} \leq 1/B$ **then**
    return $\left(\widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{x}\right)$ ▷ return the original iterate, since the residual is small
  end if
  $\boldsymbol{x'} = \boldsymbol{x} + \frac{n_y(r)^2}{\left\|p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r},\ \boldsymbol{r'} = \boldsymbol{b} - p_{\boldsymbol{y}}(\boldsymbol{x'})$
  if $\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}} \leq 1/B$ **then**
    return $\left(\widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{x'}\right)$ ▷ return the new iterate, since the residual is small
  end if
  if $\|\boldsymbol{r'}\|^2_{\widetilde{\boldsymbol{H}}^{-1}} \leq (1-\beta)\|\boldsymbol{r}\|^2_{\widetilde{\boldsymbol{H}}^{-1}}$ **then**
    return $\left(\widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{x'}\right)$ ▷ return new iterate
  else
    if $\frac{\|\boldsymbol{r}\|^2_{\widetilde{\boldsymbol{H}}^{-1}}}{n_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2} \geq \frac{1}{\sqrt{\frac{20}{9}\beta}}$ **then** ▷ test for type 1 excentricity certificate
      $\left(\widetilde{\boldsymbol{H'}}, \widetilde{\boldsymbol{H'}}^{-1}\right)$ =RobustUpdatePreconditioner($\widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{r}$, type $= 1$)
    else ▷ if test fails, then we must have a type 2 excentricity certificate
      $\left(\widetilde{\boldsymbol{H'}}, \widetilde{\boldsymbol{H'}}^{-1}\right)$ =RobustUpdatePreconditioner($\widetilde{\boldsymbol{H}}, \widetilde{\boldsymbol{H}}^{-1}, \boldsymbol{r}$, type $= 2$)
    end if
    return $\left(\widetilde{\boldsymbol{H'}}, \widetilde{\boldsymbol{H'}}^{-1}, \boldsymbol{x}\right)$ ▷ return new preconditioner
  end if
end procedure

H $_y$及其利用 p_y或 ^{n}y得到的近似。我们通过证明近似会引入较小的加性误差来得到这些结果，并利用矩阵范数的上界证明这些误差远小于涉及的其他量。为此，我们还需要假设残差$\|\mathbf{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}$，$\|\mathbf{r}'\|_{\widetilde{\boldsymbol{H}}^{-1}}$足够大。不过，根据假设，这一点自然成立。采用这一假设不会带来问题，因为一旦范数足够小，我们就会停止迭代方法。

最后，我们需要证明，算法 3中执行的秩一更新虽然涉及近似 p_y 和 ^{n}y ，仍能使偏心率充分降低。为此，我们需要引理 16的一个鲁棒版本。

引理 *19.* 设 $g\colon K \to \mathbb{R}$ 是一个自协调函数，令 $y \in int\,(K)$，令 $Hy = \nabla^2 g(y)$，令 $\widetilde{H} \in \mathbb{R}^{n\times n}$，并假设逆矩阵 $\widetilde{H}^{-1}$ 可用。此外，假设 $max\left\{\|Hy\|, \|H^{-1}y\|, \left\|\widetilde{H}\right\|, \left\|\widetilde{H}^{-1}\right\|\right\} \le B$，其中某些标量为 $B \ge 1000$。给定引理 *18* 提供的类型 *1* 或类型 *2* 偏心率证书，存在一个算法（算法 *3*），它对 $\widetilde{H}$ 及其逆矩阵执行秩一更新，从而得到新的预条件器 H'，使得

$$\mathcal{E}\left(\widetilde{H'}^{-1}H\right) \le \mathcal{E}\left(\widetilde{H}^{-1}H\right)\cdot\frac{2}{\sqrt{1+\frac{99}{100}\cdot\frac{1}{\sqrt{\frac{20}{9}\beta}}}}.$$

此更新可在 $O\left(n^2\right)$ 时间内实现，此外还需为 g 进行常数次梯度查询所需的时间。另外，$\max\left\{\left\|\widetilde{H'}\right\| - \left\|\widetilde{H}\right\|, \left\|\widetilde{H'}^{-1}\right\| - \left\|\widetilde{H}^{-1}\right\|\right\} \le 2\max\left\{\|Hy\|, \|H^{-1}y\|\right\}^2$。

证明基于我们为引理 16所给出的证明，并在附录 C.4中呈现。最后，结合引理 18和 19 我们得到本节的主要结果。

引理*20*（稳健步进或更新）。设 $g: K \to \mathbb{R}$ 为自协调函数，令 $y \in int\,(K)$，令 $Hy = \nabla^2 g(y)$，令 $\widetilde{H} \in \mathbb{R}^{n\times n}$，且令向量 $b, x \in \mathbb{R}^n$。此外，假设 $max\left\{\|Hy\|, \|H^{-1}y\|, \left\|\widetilde{H}\right\|, \left\|\widetilde{H}^{-1}\right\|, \|b\|_2\right\} \le B$，其中某个标量为 $B \ge 1000$。令 $\beta \in \left(\frac{1}{B}, 1\right)$ 为标量。存在一个算法（算法*4*），其结果恰好是以下两种之一：

1. 返回 n *ew iterate* x' *such that* $\|b - H_y x'\|^2_{\widetilde{H}^{-1}} \le \max\left\{\frac{2}{B}, \left(1-\frac{\beta}{2}\right)\|b - H_y x\|^2_{\widetilde{H}^{-1}}\right\}$，

2. 更新 $\widetilde{H}$ 并以获得新的预条件器 H'，使得

$$\mathcal{E}\left(\widetilde{H'}^{-1}H\right) \le \mathcal{E}\left(\widetilde{H}^{-1}H\right)\cdot\frac{2}{\sqrt{1+\frac{99}{100}\cdot\frac{1}{\sqrt{\frac{20}{9}\beta}}}}.$$

该步骤可在 $O\left(n^2\right)$ 时间内实现，外加针对 g进行常数次梯度查询所需的时间。此外，最大$\left\{\left\|\widetilde{H'}\right\| - \left\|\widetilde{H}\right\|, \left\|\widetilde{H'}^{-1}\right\| - \left\|\widetilde{H}^{-1}\right\|\right\} \le 2$最大$\left\{\|H_y\|, \|H^{-1}y\|\right\}^2$。

证明。 除了调用引理 18 和 19 之外，我们还进行了两项检查。首先，我们检查$\widetilde{\boldsymbol{H}^{-1}}$范数下原始残差 $\boldsymbol{r} = \boldsymbol{b} - p_{\boldsymbol{y}}(\boldsymbol{x})$ 是否过小；如果是，则原样返回迭代点。随后，在计算出新的迭代点及对应残差后，我们再次检查 $\boldsymbol{r}'$ 是否过小；如果是，则再次返回该值。完成这些检查后，我们检查范数 $\|\boldsymbol{r}'\|_{\widetilde{\boldsymbol{H}}^{-1}} \leq (1-\beta)\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}$。如果未通过此项检查，则调用算法 3 来更新预条件器。运行此例程所需的总时间仍主要由估算过程中使用的梯度查询以及执行秩一更新所需的时间决定，其时间为 $O\left(n^2\right)$。

最后，我们需要将涉及残差 $\mathrm{r}' = \mathrm{b} - p_{\boldsymbol{y}}(\mathrm{x}')$ 的保证转换为涉及线性函数 $\mathrm{b} - \mathrm{H}_{\boldsymbol{y}}\mathrm{x}'$ 的保证。

注意，在前者中，使用我们在引理 18（方程（15））中证明的相同近似，我们有界限 $\|\mathrm{b} - p_{\boldsymbol{y}}(\mathrm{x}')\|^2_{\widetilde{\boldsymbol{H}^{-1}}} \leq \frac{1}{B}$意味着

$$\left\|\boldsymbol{b} - \boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}'\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \leq \left(\left\|\boldsymbol{b} - p_{\boldsymbol{y}}(\boldsymbol{x})\right\|_{\widetilde{\boldsymbol{H}}^{-1}} + \frac{1}{400 \cdot B^{16}}\right) \frac{1}{1 - \frac{1}{400 \cdot B^{18}}} \leq \frac{2}{B}.$$

在另一种情况下，使用(16)我们有

$$\begin{aligned}
&\left\|\boldsymbol{b} - \boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}'\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \left(1 - \frac{1}{200 \cdot B^{15}}\right) \leq \left\|\boldsymbol{b} - p_{\boldsymbol{y}}(\boldsymbol{x}')\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}} \\
&\leq (1-\beta)\left\|\boldsymbol{b} - p_{\boldsymbol{y}}(\boldsymbol{x})\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}} \leq (1-\beta)\left\|\boldsymbol{b} - \boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \left(1 + \frac{1}{200 \cdot B^{15}}\right),
\end{aligned}$$

因此

$$\begin{aligned}
\left\|\boldsymbol{b} - \boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}'\right\|_{\widetilde{\boldsymbol{H}}^{-1}} &\leq (1-\beta)\left\|\boldsymbol{b} - \boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \left(1 + \frac{1}{200 \cdot B^{15}}\right) \frac{1}{1 - \frac{1}{200 \cdot B^{15}}} \\
&\leq (1 - \beta/2)\left\|\boldsymbol{b} - \boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right\|_{\widetilde{\boldsymbol{H}}^{-1}},
\end{aligned}$$

从而完成证明。 □

处理具有零空间的矩阵。虽然我们之前假设底层矩阵 $\mathrm{H}_{\boldsymbol{y}}$ 是满秩的，但如果它具有非平凡的零空间 $\mathcal{S}$，我们的分析也可延续到右侧向量 $\mathrm{b} \in \mathcal{S}^{\perp}$满足以下条件的情况，即它与 $\mathcal{S}$正交。为说明这一点，我们注意到，算法 4的分析可以限制在 $\mathcal{S}^{\perp}$上。注意，根据自协调性，底层函数 g 的海森矩阵始终以 $\mathcal{S}$ 为零空间。如果事实并非如此，通过对该自变量进行轻微扰动，我们就可以移动到一个新点，在那里新的海森矩阵无法近似原海森矩阵，因为张成空间中的一个完整方向被置零了，因此两个海森矩阵无法彼此近似，这与自协调性相矛盾。这表明，$p_{\boldsymbol{y}}(\mathrm{v})$ 始终输出一个与 $\mathcal{S}$正交的向量，因此估计的残差 r 始终与 $\mathcal{S}$正交。类似地，$n_{\boldsymbol{y}}(\mathrm{v})$ 会将 v 中平行于 $\mathcal{S}$的分量置零，因此它所进行的所有计算都发生在 $\mathcal{S}^{\perp}$中。虽然$\widetilde{\ }$我们没有明确对预条件器H施加限制，但在分析中可以将其替换$\widetilde{\ }$为投影 $\boldsymbol{\Pi}_{\mathcal{S}^{\perp}}\mathrm{H}\,\boldsymbol{\Pi}_{\mathcal{S}^{\perp}}$，而不会影响所给出的保证，因为这不会影响偏心率证书。

数值精度方面。 到目前为止，分析的阐述都基于精确算术这一假设。实际上，我们的算法确实能够容忍规模为

问题的自然参数（n， ε， B）。假设我们使用的数字位数关于这些参数是多对数级的，就可以确保所有舍入误差都不超过 $1/\mathrm{poly}\,(n,\ \varepsilon,\ B)$。这些误差只会给我们对残差范数的计算带来多项式级的小加性误差，而这些误差已被分析所涵盖。至关重要的是，我们需要确保这些误差不会影响偏心势。首先，多项式级的小加性误差不会影响偏心率证书——它们仍能证明存在一个条件不佳的方向，只是引理 18 所给界的质量可能出现多项式级的小乘性下降。其次，更重要的是，我们需要论证，尽管引入了舍入误差，对虚拟预条件器及其逆矩阵进行秩一更新仍能按需降低偏心率。这是因为我们可以将舍入误差视为对预条件矩阵 $\mathrm{X} = \mathrm{H}_{y}^{-1/2}\widetilde{\mathrm{H}}\,\mathrm{H}^{-1/2}{}_{y}$ 施加的对抗性扰动 $\boldsymbol{\Delta}$，使得 $\boldsymbol{\Delta}$的所有元素都为多项式级的小量。我们证明，这类扰动不会显著增大偏心率。事实上，由于行列式在小扰动下具有乘性稳定性，

$$\begin{aligned}\det\left(\boldsymbol{A}+\boldsymbol{\Delta}\right)=\det\left(\boldsymbol{A}\right)\det\left(\boldsymbol{I}+\boldsymbol{A}^{-1/2}\boldsymbol{\Delta}\boldsymbol{A}^{-1/2}\right)&\leq\det\left(\boldsymbol{A}\right)\cdot\left(1+\frac{\|\boldsymbol{\Delta}\|}{\left\|\boldsymbol{A}^{-1}\right\|}\right)^{n}\\&\leq\det\left(\boldsymbol{A}\right)\cdot\left(1+\frac{n\max_{i,j}\left|\boldsymbol{\Delta}_{ij}\right|}{\left\|\boldsymbol{A}^{-1}\right\|}\right)^{n},\end{aligned}$$

因此，我们可以使噪声导致的偏心率增加仅使 $\boldsymbol{\Delta}$ only 受到一个 $1+$ ^{1} 多项式(n,ε,B) ~~乘法因子~~的影响。因此，多项式级小的误差就足够了，只需要对数级数量的比特。

5 使用伪海森矩阵的路径跟踪

引理 15 和16 （引理18 和19）的鲁棒版本为我们提供了一种方法，用于维护自协调函数 g 在固定点y 处的海森矩阵的预条件器。要在路径跟踪方法中使用它们$\widetilde{\ }$，我们需要能够控制预条件器 H 的质量，因为我们会移动到中心路径上的新点。幸运的是，强自协调性恰好是确保偏心率 $\mathcal{E}\left(\widetilde{\mathrm{H}}_{-1}\mathrm{H}_{y}\right)$ 不会随着我们移动到新点 y′而大幅增加所需的性质。

这里我们关键地依赖引理 12，其证明将在本节末尾给出。为了证明引理 12 ，我们使用一个行列式不等式（引理 27），该不等式通过对两个矩阵之和的特征值使用主序界，并结合与初等对称多项式的舒尔凸性相关的一些性质来证明。

5.1 设置

我们考虑障碍函数形式

$$g_{\mu}\left(\boldsymbol{y}\right)=\frac{\left\langle\boldsymbol{c},\boldsymbol{y}\right\rangle}{\mu}+\phi\left(\boldsymbol{y}\right),\tag{5}$$

其中， $\phi\colon K\to\mathbb{R}$ 是凸集 K 的 ν-强自协调障碍函数。我们的目标是当 μ 足够小时，求得 g_{μ} 的近似极小值点。以下引理对这一点进行了定量描述，是内点法中的标准结果。

引理 21 (近似最优性 [Ren01])。 假设 $\varepsilon_N \leq 1/10$，一个可行解 $\boldsymbol{y} \in \mathbb{R}^m$以及参数 $\mu \leq 1$ 满足以下关于牛顿步长的界

$$\|\nabla g_\mu(\boldsymbol{y})\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}} \leq \varepsilon_N .$$

设 $\boldsymbol{y}^\star$是原目标函数 $\min\{\langle \boldsymbol{c}, \boldsymbol{y}\rangle : \boldsymbol{y} \in K\}$ 的一个最优解。于是 $\langle \boldsymbol{c}, \boldsymbol{y}\rangle \leq \langle \boldsymbol{c}, \boldsymbol{y}^\star\rangle + \mu\nu(1+2\varepsilon_N)$。

根据引理 21 ，要将问题求解到精度 ε ，我们需要找到 g_μ 在 $\mu \lesssim \frac{\varepsilon}{2\nu}$ 下的极小化点。由此可知，在标准内点法的整个执行过程中，中心路径邻域内的矩阵，其特征值（及其逆）都受障碍函数形式的 ε-条件数

定义 22 (ε-条件数)。 给定一个障碍函数形式（5），其中 ϕ 是一个 ν-强自协调障碍函数，我们将该形式的 ε-条件数定义为

$$\kappa_\star := 4\kappa\left(\frac{\varepsilon}{2\nu}\right) . \tag{6}$$

该参数决定了我们估算海森矩阵-向量乘积所需达到的精度，以及整个算法的迭代次数。后者之所以成立，是因为我们无法测量梯度的真实对偶局部范数，因此通常无法判断自己是否已充分接近中心路径，从而可以通过调低 μ 继续前进。对海森矩阵特征值的保证，以及我们对预条件器及其逆的特征值所维持的上界，使我们能够近似计算 $\|\nabla g(\mathrm{y})\|_{\boldsymbol{H}^{-1}\mathrm{y}}$。

在算法的整个执行过程中，我们将始终维持以下关于中心路径邻域内预条件器和海森矩阵的不变量。

不变量 1. 对于 $B_{\boldsymbol{H}} = 4\kappa_\star$和 $B_{\widetilde{\boldsymbol{H}}} = 10^9 \cdot \kappa_\star^2 \cdot \left(n\log\kappa_\star + \sqrt{\nu n}\log\frac{n\nu}{\varepsilon}\right)$，我们始终有 $\max\{\|\mathrm{H}_{\mathrm{y}}\|, \|\mathrm{H}^{-1}{}_{\mathrm{y}}\|\} \leq B_{\boldsymbol{H}}$and $\max\left\{\left\|\widetilde{\mathrm{H}}\right\|, \left\|\widetilde{\mathrm{H}}^{-1}\right\|\right\} \leq B_{\widetilde{\boldsymbol{H}}}$。

虽然 $B_{\boldsymbol{H}}$ 的上界在我们始终接近中心路径时会自动成立（根据引理 14 ，前提是我们在 $\mu = \frac{\varepsilon}{2\nu}$处停止内点法），但我们还需要证明$\widetilde{\ }$预条件器 $\boldsymbol{H}$ 的条件不会变得太差。我们将通过追踪内点法执行期间它的演化过程来证明这一点。

引理 23. 令 $\boldsymbol{y}$满足 $\|\nabla g_\mu(\boldsymbol{y})\|_{\boldsymbol{H}^{-1}\mathrm{y}} \leq \frac{1}{20\sqrt{n}}$。那么，算法 5会返回一个新的迭代点 $\boldsymbol{y'}$，满足 $\left\|\nabla g_{\mu/\left(1+\frac{1}{20\sqrt{\nu n}}\right)}(\boldsymbol{y'})\right\|_{\boldsymbol{H}^{-1}\mathrm{y}'} \leq \frac{1}{20\sqrt{n}}$。

证明。该算法本质上是执行一次牛顿步以恢复中心性，随后调整中心性参数 μ。在执行过程中，它可能会多次调用 RobustStepOrUpdate 例程，因为可能需要调整预条件器。调用该例程时，我们使用参数 $B = B_{\boldsymbol{H}}^{10}$ 和 $\beta = 1/100$。首先证明，执行结束时，返回的迭代点确实仍具有较小的对偶局部范数。

在一次成功调用后，当预条件器不发生变化时，根据引理20的保证，我们有

$$\left\|\nabla g_\mu(\boldsymbol{y}) - \boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x'}\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \leq \max\left\{\frac{2}{B_{\widetilde{\boldsymbol{H}}}^{10}}, \left(1-\frac{\beta}{2}\right)\|\nabla g_\mu(\boldsymbol{y}) - \boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2\right\} .$$

且计数器 t 会增加。由于计数器在 $T=\frac{100}{\beta}\log\left(B_{\boldsymbol{H}}B_{\widetilde{\boldsymbol{H}}}\right)$ 次成功迭代后停止，执行结束时，返回的变量 x 满足

$$\begin{aligned}\left\|\nabla g_{\mu}(\boldsymbol{y})-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right\|_{\widetilde{\boldsymbol{H}}^{-1}} &\leq \max\left\{\frac{2}{B_{\widetilde{\boldsymbol{H}}}^{10}},\left(1-\frac{\beta}{2}\right)^{T}\left\|\nabla g_{\mu}(\boldsymbol{y})\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^{2}\right\}\\ &\leq \max\left\{\frac{2}{B_{\widetilde{\boldsymbol{H}}}^{10}},\left(1-\frac{\beta}{2}\right)^{T}\left\|\nabla g_{\mu}(\boldsymbol{y})\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}^{2}\cdot\|\boldsymbol{H}_{\boldsymbol{y}}\|\left\|\widetilde{\boldsymbol{H}}^{-1}\right\|\right\}\leq\frac{2}{B_{\widetilde{\boldsymbol{H}}}^{10}}.\end{aligned}$$

因此

$$\left\|\nabla g_{\mu}(\boldsymbol{y})-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}\leq\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1}\right\|\left\|\widetilde{\boldsymbol{H}}\right\|\frac{2}{B_{\widetilde{\boldsymbol{H}}}^{10}}\leq\frac{2}{B_{\widetilde{\boldsymbol{H}}}^{8}}.$$

根据引理 38 中的标准论证，我们得到

$$\left\|\nabla g_{\mu}(\boldsymbol{y}')\right\|_{\boldsymbol{H}_{\boldsymbol{y}'}^{-1}}\leq\frac{4}{B_{\widetilde{\boldsymbol{H}}}^{8}}+7\cdot\left\|\nabla g_{\mu}(\boldsymbol{y})\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}^{2}.$$

因此，从中心性 μ 到 $\mu/\left(1+\frac{1}{200\sqrt{\nu n}}\right)$，通过引理 39 可得

$$\begin{aligned}\left\|\nabla g_{\mu/\left(1+\frac{1}{200\sqrt{\nu n}}\right)}(\boldsymbol{y}')\right\|_{\boldsymbol{H}_{\boldsymbol{y}'}^{-1}} &=\left\|\left(1+\frac{1}{200\sqrt{\nu n}}\right)\nabla g_{\mu}(\boldsymbol{y}')-\frac{1}{200\sqrt{\nu n}}\nabla\phi(\boldsymbol{y}')\right\|_{\boldsymbol{H}_{\boldsymbol{y}'}^{-1}}\\ &\leq\left(1+\frac{1}{200\sqrt{\nu n}}\right)\left\|\nabla g_{\mu}(\boldsymbol{y}')\right\|_{\boldsymbol{H}_{\boldsymbol{y}'}^{-1}}+\frac{1}{200\sqrt{\nu n}}\left\|\nabla\phi(\boldsymbol{y}')\right\|_{\boldsymbol{H}_{\boldsymbol{y}'}^{-1}}\\ &\leq\left(1+\frac{1}{200\sqrt{\nu n}}\right)\left\|\nabla g_{\mu}(\boldsymbol{y}')\right\|_{\boldsymbol{H}_{\boldsymbol{y}'}^{-1}}+\frac{1}{200\sqrt{n}}\\ &\leq\left(1+\frac{1}{200\sqrt{\nu n}}\right)\left(\frac{4}{B_{\widetilde{\boldsymbol{H}}}^{8}}+7\cdot\left\|\nabla g_{\mu}(\boldsymbol{y})\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}^{2}\right)+\frac{1}{200\sqrt{n}}\\ &\leq\left(1+\frac{1}{200\sqrt{\nu n}}\right)\left(\frac{4}{B_{\widetilde{\boldsymbol{H}}}^{8}}+7\cdot\left(\frac{1}{20\sqrt{n}}\right)^{2}\right)+\frac{1}{200\sqrt{n}}\\ &\leq\frac{1}{20\sqrt{n}}.\end{aligned}$$

接下来，我们证明预条件器更新例程调用总次数的上界。

□

引理 24。设 $\boldsymbol{y}_0$ 满足 $\left\|\nabla g_{\mu}(\boldsymbol{y}_0)\right\|_{\boldsymbol{H}^{-1}\boldsymbol{y}_0}\leq\frac{1}{20\sqrt{n}}$，其中 $\mu=n^{O(1)}$，并设 $\widetilde{\boldsymbol{H}}_0=\boldsymbol{I}$ 为初始化时的预条件器。在调用算法 5 $\widetilde{O}\left(\sqrt{\nu n}\log\left(\frac{n\nu}{\varepsilon}\right)\right)$ 次后，我们得到新点 $\boldsymbol{y}'$，使得 $\langle\boldsymbol{c},\boldsymbol{y}'\rangle-\langle\boldsymbol{c},\boldsymbol{y}^{\star}\rangle\leq\varepsilon$。在整个执行过程中，算法最多进行 $O\left(\log\mathcal{E}\left(\boldsymbol{H}_{\boldsymbol{y}_0}\right)+\sqrt{\nu n}\log\frac{n\nu}{\varepsilon}\right)$ 次预条件器更新。

证明。首先，根据引理 23，在一次调用算法 5后，中心性参数 μ 会缩小至原来的 $1+\frac{1}{20\sqrt{\nu n}}$，同时对偶局部范数保持不超过 $1/(20\sqrt{n})$。根据引理 21，当 $\mu\leq\frac{\varepsilon}{\nu(1+2\cdot\frac{1}{20})}$ 时停止，可确保目标函数误差至多为 ε，因此结论成立。所以，经过 $O\left(\sqrt{\nu n}\log\frac{n\nu}{\varepsilon}\right)$ 次调用后，算法可以返回一个具有 ε 加性

误差。我们通过分析偏心率的变化来界定预条件器更新的次数。在每次调用算法 5时，偏心率恰好会增加一次，即从 y 移动到 y′时，因此我们将 $\mathcal{E}\left(\widetilde{\mathrm{H}}^{-1}\mathrm{H}_{y}\right)$ 替换为 $\mathcal{E}\left(\widetilde{\mathrm{H}}^{-1}\mathrm{H}_{y'}\right)$。

由此，根据引理 38，我们得到迭代点的变化 $\|\boldsymbol{\delta}\|_{\boldsymbol{H}_y}=\|\boldsymbol{y}-\boldsymbol{y}'\|_{\boldsymbol{H}_y}\leq\frac{2}{B^{10}\bar{H}}+\|\nabla g_{\mu}(\boldsymbol{y})\|_{\boldsymbol{H}^{-1}y}$ $\leq\frac{1}{10\sqrt{n}}$。因此，根据引理 12，

$$\mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{y'}\right)\leq\mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{y}\right)\cdot\exp\left(\frac{1}{2}\sqrt{n}\cdot\frac{\|\boldsymbol{\delta}\|_{\boldsymbol{H}_y}}{\left(1-\|\boldsymbol{\delta}\|_{\boldsymbol{H}_y}\right)^{2}-\|\boldsymbol{\delta}\|_{\boldsymbol{H}_y}}\right)\leq\mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{y}\right)\cdot\exp\left(\frac{1}{10}\right).$$

此外，偏心率的其他所有变化都仅由预条件器更新造成。根据引理 20 ，我们知道这些更新会降低偏心率，即

$$\mathcal{E}\left(\widetilde{\boldsymbol{H}'}^{-1}\boldsymbol{H}_{y}\right)\leq\mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{y}\right)\cdot\frac{2}{\sqrt{1+\frac{99}{100}\cdot\frac{1}{\sqrt{\frac{20}{9}\beta}}}}\leq\mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{y}\right)\cdot\frac{3}{4}.$$

由于整个算法过程中偏心率最多增加 $\exp\left(O\left(\sqrt{\nu n}\log\frac{n\nu}{\varepsilon}\right)\right)$，且每次其他预条件器更新至少会使其降低 3/4倍（偏心率始终至少为 1），因此预条件器更新的总次数不会太多。于是，我们将预条件器更新的总次数界定为

$$O\left(\log\mathcal{E}\left(\widetilde{\boldsymbol{H}}_{0}^{-1}\boldsymbol{H}_{y_0}\right)+\sqrt{\nu n}\log\frac{n\nu}{\varepsilon}\right)=O\left(\log\mathcal{E}\left(\boldsymbol{H}_{y_0}\right)+\sqrt{\nu n}\log\frac{n\nu}{\varepsilon}\right),$$

由此完成证明。 □

最后，我们需要说明在整个执行过程中，不变量 1 都得以保持。为此，我们需要对 $\left\|\widetilde{\mathrm{H}}\right\|,\left\|\widetilde{\mathrm{H}}^{-1}\right\|$ 进行界定。显然，根据引理 20，每一步都会使这两个范数各自最多增加 2 $\max\left\{\|\mathrm{H}_{y}\|,\|\mathrm{H}^{-1}{}_{y}\|\right\}^{2}\leq 2\kappa_{\star}^{2}$，因此到最后我们有

$$\begin{aligned}\max\left\{\left\|\widetilde{\boldsymbol{H}}\right\|,\left\|\widetilde{\boldsymbol{H}}^{-1}\right\|\right\}&\leq 2\kappa_{\star}^{2}\cdot\left(\log\mathcal{E}\left(\boldsymbol{H}_{y_0}\right)+O\left(\sqrt{\nu n}\log\frac{n\nu}{\varepsilon}\right)\right)\\&\leq 2\kappa_{\star}^{2}\cdot\left(n\log\kappa_{\star}+O\left(\sqrt{\nu n}\log\frac{n\nu}{\varepsilon}\right)\right),\end{aligned}$$

这表明不变量在整个执行过程中始终成立。

结合所有结果，并运用初始化内点法的经典技巧， [Nes98]即可得到本节的主定理。

定理 25。 给定一个凸集 $K\subseteq\mathbb{R}^{n}$，其具有一个 ν-强自协调障碍函数 $\phi:K\to\mathbb{R}$，我们可以通过梯度预言机访问该障碍函数；再给定一个初始点 $\boldsymbol{y}_0\in int\,(K),\|\nabla\phi(\boldsymbol{y}_0)\|_{\boldsymbol{H}^{-1}y_0}\leq\frac{1}{20\sqrt{n}}$，在时间

$$O\left(\left(\log\mathcal{E}\left(\boldsymbol{H}_{y_0}\right)+\sqrt{\nu n}\log\frac{n\nu}{\varepsilon}\right)\cdot\left(n^{2}+\mathcal{T}_{gradient}\right)\right)$$

我们可以得到一个迭代点 $\boldsymbol{y}$，使得 $\langle\boldsymbol{c},\boldsymbol{y}\rangle\leq\langle\boldsymbol{c},\boldsymbol{y}^{\star}\rangle+\varepsilon$，其中 $\boldsymbol{y}^{\star}$ 是原始目标函数的最小化解 $\min\{\langle\boldsymbol{c},\boldsymbol{y}\rangle:\boldsymbol{y}\in K\}$。此外，我们遇到的所有矩阵及其逆矩阵的特征值都被 $\left(\kappa_{\star}n\nu\log\frac{1}{\varepsilon}\right)^{O(1)}$所界定，其中 $\kappa_{\star}$ 是障碍函数形式的 ε-条件数（定义22）。

请注意，对障碍梯度的对偶局部范数所作的条件对应于接近中心路径。通常通过适当修改优化问题来处理这一点，从而确保该保证成立。

需要注意的是，只要 $\kappa_\star$ 是准多项式有界的，所有涉及的量都是准多项式的，因此我们可以使用每个数仅具有多对数位精度来实现内点法。

此外，这意味着对于在线性函数上进行优化$\widetilde{}$凸域，还可得到以下重要推论。这源于 $O(n)$-强自协调性的通用障碍函数和熵障碍函数 [LLV20], ，而这又得益于近期关于KLS猜想的研究进展 [Che21、JLV22]。尽管 [LLV20] 证明了通用障碍函数和熵障碍函数满足定义9的一个近似版本，即在右侧前乘以一个多对数常数，但通过将它们略微缩小，可以使其严格满足该定义。作为交换，它们的自协调参数会小幅增加一个多对数因子，但这不会影响我们的界。

推论26。给定一个凸集 $K \subseteq \mathbb{R}^n$，可查询关于通用障碍函数或熵障碍函数 ϕ： $K \to \mathbb{R}$ 的梯度预言机，并给定一个初始点 $\boldsymbol{y}_0 \in int\,(K)$， $\|\nabla\phi(\boldsymbol{y}_0)\|_{\boldsymbol{H}^{-1}\boldsymbol{y}_0} \leq \frac{1}{20\sqrt{n}}$，在时间

$$O\left(\log\left(\frac{n\kappa_\star}{\varepsilon}\right)\cdot\left(n^3 + n\mathcal{T}_{gradient}\right)\right)$$

我们可以得到一个迭代点 $\boldsymbol{y}$，使得 $\langle \boldsymbol{c},\ \boldsymbol{y}\rangle \leq \langle \boldsymbol{c},\ \boldsymbol{y}^\star\rangle + \varepsilon$，其中 $\boldsymbol{y}^\star$ 是原始目标函数 $\min\{\langle \boldsymbol{c}, \boldsymbol{y}\rangle :\ \boldsymbol{y} \in K\}$。此外，我们遇到的所有矩阵及其逆矩阵的特征值都被 $\left(\kappa_\star n\nu\right.$ 对数 $\left.\frac{1}{\varepsilon}\right)^{O(1)}$所界定，其中 $\kappa_\star$ 是 ε-障碍函数形式的条件数（定义 22）。

算法 5 路径跟踪。要求： $\|\nabla g_\mu(\mathbf{y})\|_{\boldsymbol{H}^{-1}\mathbf{y}} \leq \frac{1}{20\sqrt{n}}$, $B_{\boldsymbol{H}} = 4\kappa_\star$, $B_{\widetilde{\boldsymbol{H}}} = 10^9\cdot\kappa_\star^2\cdot\left(n\log\kappa_\star + \sqrt{\nu n}\log\frac{n\nu}{\varepsilon}\right)$, $B = B_{\widetilde{\boldsymbol{H}}}^{10}$, $\beta = 1/100$**确保：返回一个新的迭代点 $\mathbf{y}'$，使得** $\left\|\nabla g_{\mu/}(1+\frac{1}{10\nu})(\mathbf{y}')\right\|_{\boldsymbol{H}^{-1}}\mathbf{y}'$ $\leq \frac{1}{20\sqrt{n}}$。 **1: 过程 PathFollowing($\mathbf{y}$ $\widetilde{}$ $\mathbf{H}$ $\widetilde{}$ $\mathbf{H}_{-1}$, μ) 2:** $t = 0, T = \frac{100}{\beta}\log\left(B_{\boldsymbol{H}}B_{\widetilde{\boldsymbol{H}}}\right)$, $\mathbf{x} = \mathbf{0}$ **3: 当** $t < T$ **时 4:** $\left(\widetilde{}\,\mathbf{H}',\ \widetilde{}\,\mathbf{H}'^{-1}, \mathbf{x}'\right)$ = **RobustStepOrUpdate($\mathbf{y}$** $\widetilde{}$ $\mathbf{H}$ $\widetilde{}$ $\mathbf{H}_{-1}$, $\nabla g_\mu(\mathbf{y})$, $\mathbf{x}$**)** $\sim$ $\sim$ **5: 如果 $\mathbf{H}' = \mathbf{H}$ 则** ▷ **预条件器未发生变化，因此取得了进展 6:** $\mathbf{x} = \mathbf{x}'$, $t = t+1$ **7: 结束 if 8: 结束 while 9: 返回** $\left(\mathbf{y} - \mathbf{x}\right.$ $\widetilde{}$ $\mathbf{H}$ $\widetilde{}$ $\left.\mathbf{H}_{-1}, \mu/\left(1+\frac{1}{200\sqrt{n\nu}}\right)\right)$ **10: 结束过程**

5.2 Excentricity Proofs

首先，我们给出引理12的证明。

引理的证明 *12*。 利用引理 8 可写为

$$\mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{\boldsymbol{y}+\boldsymbol{\delta}}\right)=\mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{\boldsymbol{y}}\cdot\boldsymbol{H}_{\boldsymbol{y}}^{-1}\boldsymbol{H}_{\boldsymbol{y}+\boldsymbol{\delta}}\right)\leq\mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{\boldsymbol{y}}\right)\cdot\sqrt{\frac{\det\left(\left(\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}+\boldsymbol{\delta}}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right)_{\geq 1}\right)}{\det\left(\left(\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}+\boldsymbol{\delta}}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right)_{<1}\right)}}\,.$$

设 $\lambda_1,\ldots,\lambda_n$ 为 $\mathrm{H}^{-1/2}{}_{\boldsymbol{y}}\mathrm{H}_{\boldsymbol{y}+\boldsymbol{\delta}}\mathrm{H}^{-1/2}{}_{\boldsymbol{y}}$的特征值。由于 g是强自协调的，根据引理 10，对于 $\alpha=\frac{\|\boldsymbol{\delta}\|_{H_{\boldsymbol{y}}}}{\left(1-\|\boldsymbol{\delta}\|_{H_{\boldsymbol{y}}}\right)^2}$，我们也有

$$\sqrt{\sum_{i=1}^{n}\left(\lambda_i-1\right)^2}\leq\alpha\,.$$

因此，每当 $0\leq x\leq 1-\frac{1}{\beta}$ 时，使用$\frac{1}{1-x}\leq 1+\beta x$，我们有

$$|\lambda_i-1|\leq\alpha=1-\frac{1}{(1-\alpha)^{-1}}\,,$$

以便

$$\frac{1}{1-|\lambda_i-1|}\leq 1+\frac{1}{1-\alpha}\,|\lambda_i-1|\,.$$

因此,

$$\begin{aligned}
\prod_{i=1}^{n}\max\left\{\lambda_i,\frac{1}{\lambda_i}\right\}&=\prod_{i=1}^{n}\max\left\{1+\left(\lambda_i-1\right),\frac{1}{1+\left(\lambda_i-1\right)}\right\}\leq\prod_{i=1}^{n}\max\left\{1+|\lambda_i-1|\,,\frac{1}{1-|\lambda_i-1|}\right\}\\
&\leq\prod_{i=1}^{n}\max\left\{1+|\lambda_i-1|\,,1+\frac{1}{1-\alpha}\,|\lambda_i-1|\right\}\\
&\leq\prod_{i=1}^{n}\left(1+\frac{1}{1-\alpha}\,|\lambda_i-1|\right)\\
&\leq\left(1+\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-\alpha}\,|\lambda_i-1|\right)^n && \text{(AM-GM)}\\
&\leq\left(1+\frac{1}{1-\alpha}\sqrt{\frac{1}{n}\sum_{i=1}^{n}|\lambda_i-1|^2}\right)^n && \text{(QM-AM)}\\
&\leq\left(1+\frac{\alpha}{1-\alpha}\sqrt{\frac{1}{n}}\right)^n\\
&\leq\exp\left(\sqrt{n}\cdot\frac{\alpha}{1-\alpha}\right)\,.
\end{aligned}$$

将其代回新偏心率的上界，可得

$$\begin{aligned}\mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{\boldsymbol{y}+\boldsymbol{\delta}}\right) &\leq \mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{\boldsymbol{y}}\right)\cdot\exp\left(\frac{1}{2}\sqrt{n}\cdot\frac{\alpha}{1-\alpha}\right)\\ &= \mathcal{E}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}_{\boldsymbol{y}}\right)\cdot\exp\left(\frac{1}{2}\sqrt{n}\cdot\frac{\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}}{\left(1-\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\right)^{2}-\|\boldsymbol{\delta}\|_{\boldsymbol{H}_{\boldsymbol{y}}}}\right),\end{aligned}$$

由此得证。 □

为了证明引理 12 我们依赖以下重要不等式。

引理 27。设 $\boldsymbol{A}$ 和 $\boldsymbol{D}$ 为对称正定矩阵。令 $\boldsymbol{D}_{\geq 1}$ 表示由 $\boldsymbol{D}$ 得到的矩阵，即将所有小于 *1* 的特征值增大至 1，并以类似方式定义 $\boldsymbol{D}_{<1}$ 。那么

$$\det\left(\boldsymbol{A}\boldsymbol{D}+\boldsymbol{I}\right)\leq\det\left(\boldsymbol{A}+\boldsymbol{I}\right)\det\left(\boldsymbol{D}_{\geq 1}\right)$$

该证明遵循 两个半正定矩阵之和的特征值主序界所导出的结果 .

定义 28 （主序）。给定 x、y $\in\mathbb{R}^d$，我们称 y 主序于 x（记作 x $\prec$ y）当且仅当

$$\sum_{i=1}^{k}\boldsymbol{x}_i\leq\sum_{i=1}^{k}\boldsymbol{y}_i,\ \text{for all } 1\leq k\leq d,$$

and

$$\sum_{i=1}^{d}\boldsymbol{x}_i=\sum_{i=1}^{d}\boldsymbol{y}_i.$$

给定两个对称正定矩阵，可以证明它们的特征值之间存在一种主序关系。以下引理由[和94]推出。

引理 29 ([和94]). 给定一个对称半正定矩阵 $\boldsymbol{M}$，令 $\lambda^{\uparrow}(\boldsymbol{M})$ 和 $\lambda^{\downarrow}(\boldsymbol{M})$ 分别表示 $\boldsymbol{M}$ 按升序和降序排列的特征值。则

$$\lambda^{\downarrow}(\boldsymbol{A})+\lambda^{\uparrow}\left(\boldsymbol{D}^{-1}\right)\prec\lambda^{\downarrow}\left(\boldsymbol{A}+\boldsymbol{D}^{-1}\right).$$

这类界很有用，因为舒尔凸性理论可用于证明涉及对满足主序关系的向量施加函数的不等式。

定义 30 (舒尔凸性). 若对任意满足 x $\prec$ y 的 x、y $\in\mathbb{R}^d$，都有 f(x) $\leq f$(y)，则称函数 $f\colon\mathbb{R}^d\to\mathbb{R}$ 为舒尔凸函数。若反向不等式成立，则称函数 f为舒尔凹函数；也就是说，x $\prec$ y 蕴含 f(x) $\geq f$(y)。

关于舒尔凸/舒尔凹函数已有大量理论 [VR73]。此处我们只关注一个特定函数，即初等对称多项式。

Lemma 31. 该函数 $f\colon\ \mathbb{R}^d_{>0}\to\mathbb{R}$ 由 $f(\boldsymbol{x})=\prod_{i=1}^{d}\boldsymbol{x}_i$ 定义，是 *Schur*-凹的 *e.*

利用引理 31，结合 29 我们现在可以证明引理 27。

引理的证明27。令 $\lambda^{\uparrow}(\boldsymbol{M})$ 和 $\lambda^{\downarrow}(\boldsymbol{M})$ 分别表示 $\boldsymbol{M}$ 的特征值，按升序和降序排列。根据引理29，这些特征值满足主序关系：

$$\lambda^{\downarrow}(\boldsymbol{A})+\lambda^{\uparrow}\left(\boldsymbol{D}^{-1}\right) \prec \lambda^{\downarrow}\left(\boldsymbol{A}+\boldsymbol{D}^{-1}\right) .$$

利用31可知，映射 $(x_1,..,x_n) \to \prod_{i=1}^{n} x_i$ 是一个初等对称函数；只要所有 $x_i > 0$，它就是Schur凹函数。将其分别应用于 $\lambda^{\downarrow}(\boldsymbol{A}) + \lambda^{\uparrow}\left(\boldsymbol{D}^{-1}\right)$ 和 $\lambda^{\downarrow}\left(\boldsymbol{A}+\boldsymbol{D}^{-1}\right)$，得到：

$$\begin{aligned}
\det\left(\boldsymbol{A}+\boldsymbol{D}^{-1}\right) &= \prod_j \lambda_j^{\downarrow}\left(\boldsymbol{A}+\boldsymbol{D}^{-1}\right) \leq \prod_j \left(\lambda_j^{\downarrow}(\boldsymbol{A})+\lambda_j^{\uparrow}\left(\boldsymbol{D}^{-1}\right)\right) \\
&\leq \prod_j \begin{cases} \left(\lambda_j^{\downarrow}(\boldsymbol{A})+1\right)\lambda_j^{\uparrow}\left(\boldsymbol{D}^{-1}\right) & \lambda_j^{\uparrow}\left(\boldsymbol{D}^{-1}\right) \geq 1 \\ \lambda_j^{\downarrow}(\boldsymbol{A})+1 & \lambda_j^{\uparrow}\left(\boldsymbol{D}^{-1}\right) \leq 1 \end{cases} \\
&= \prod_j \left(\lambda_j^{\downarrow}(\boldsymbol{A})+1\right) \cdot \prod_j \max\left\{1, \lambda_j^{\uparrow}\left(\boldsymbol{D}^{-1}\right)\right\} \\
&= \det(\boldsymbol{A}+\boldsymbol{I}) \cdot \prod_j \frac{1}{\min\left\{1, \lambda_j^{\uparrow}(\boldsymbol{D})\right\}} \\
&= \det(\boldsymbol{A}+\boldsymbol{I}) \frac{1}{\det\left(\boldsymbol{D}_{<1}\right)} .
\end{aligned}$$

因此

$$\begin{aligned}
\det(\boldsymbol{A}\boldsymbol{D}+\boldsymbol{I}) = \det(\boldsymbol{D}) \det\left(\boldsymbol{A}+\boldsymbol{D}^{-1}\right) &\leq \det(\boldsymbol{D}) \det(\boldsymbol{A}+\boldsymbol{I}) \frac{1}{\det\left(\boldsymbol{D}_{<1}\right)} \\
&= \det(\boldsymbol{A}+\boldsymbol{I}) \det\left(\boldsymbol{D}_{\geq 1}\right) ,
\end{aligned}$$

由此完成证明。 □

6 求解半正定规划问题

这里给出求解半定规划的主要定理。我们给出的表述与 [JKL+20] 中的表述形式一致。

定理32。 考虑一个变量规模为 $n \times n$ 且有 $m \geq n$ 个约束的半定规划，其形式为

$$\max\left\{\langle \boldsymbol{B}, \boldsymbol{X}\rangle : \boldsymbol{X} \succeq 0, \langle \boldsymbol{A}_i, \boldsymbol{X}\rangle = \boldsymbol{c}_i, \ \textit{for all}\ 1 \leq i \leq m\right\} .$$

假设任意可行解 $\boldsymbol{X} \succeq 0$ 都满足 $\|\boldsymbol{X}\| \leq R$。那么对于任意误差参数 $0 < \varepsilon < 0.01$，存在一种算法，可在以下时间内输出

$$\widetilde{O}\left(mn^4 + m^{1.25}n^{3.5} \log \frac{1}{\varepsilon}\right)$$

一个矩阵 $\boldsymbol{X} \succeq 0$ 使得

$$\langle \boldsymbol{B}, \boldsymbol{X}\rangle \geq \langle \boldsymbol{B}, \boldsymbol{X}^{\star}\rangle - \varepsilon \|\boldsymbol{B}\| R$$

and

$$\sum_{i=1}^{m} |\langle \boldsymbol{A}_i, \boldsymbol{X} \rangle - \boldsymbol{c}_i| \leq 4n\varepsilon \left(R \sum_{i=1}^{m} \|\boldsymbol{A}_i\|_1 + \|\boldsymbol{c}\|_1 \right).$$

此外，若障碍目标函数 *(3))* 具有准多项式 ε-条件数（定义 *22*），那么我们遇到的所有矩阵及其逆矩阵的特征值均有准多项式界。

为证明这一陈述，我们需要将从障碍目标函数的解中获得的解映射回原始目标函数的解。关于此问题的详细处理，请参阅[JKL+20] 中的定理 5.1。同样，关于如何初始化内点法，请参阅 [JKL+20]中的第 9 节。解决这些技术问题后，只需了解求解相应障碍目标函数的速度即可。

为了应用定理 25 我们需要为可行集找到一个合适的强自协调障碍函数。遗憾的是，对于半正定规划问题，我们无法证明标准对数障碍函数是强自协调的。事实上，这似乎并不成立。相反，我们证明，将其放大 $\sqrt{m}$ 倍后，它就会变为强自协调。这个缩放后障碍函数所对应的自协调参数变为 $\nu = n\sqrt{m}$。我们在引理 33 中证明这一点。

请注意，缩放障碍函数的梯度

$$\phi(\boldsymbol{y}) = -\sqrt{m} \cdot \log\det\left(\boldsymbol{B} - \sum_{i=1}^{m} \boldsymbol{y}_i \boldsymbol{A}_i\right)$$

定义为

$$[\nabla\phi(\boldsymbol{y})]_i = \sqrt{m} \cdot \left\langle \boldsymbol{A}_i, \left(\boldsymbol{B} - \sum_{i=1}^{m} \boldsymbol{y}_i \boldsymbol{A}_i\right)^{-1} \right\rangle$$

可在$\mathcal{T}_{\text{gradient}} = O(mn^2 + n^\omega)$时间内计算。为此，我们首先在 $O(mn^2)$ 时间内计算松弛矩阵 $\boldsymbol{S}_{\boldsymbol{y}} = \boldsymbol{B} - \sum_{i=1}^{m} \mathrm{y}_i \boldsymbol{A}_i$，在 $O(n^\omega)$ 时间内将其求逆，然后在 $O(mn^2)$ 时间内计算所有内积 $\langle \boldsymbol{A}_i, \boldsymbol{S}_{\boldsymbol{y}}^{-1} \rangle$。

接下来我们使用定理 25。这里，尽管我们维护的虚拟海森矩阵的维度为 $m \times m$，但我们注意到，其秩始终至多为 n^2。这是因为

$$\nabla^2\phi(\boldsymbol{y}) = \sqrt{m} \cdot \mathcal{A}^\top \left(\boldsymbol{S}_{\boldsymbol{y}}^{-1} \otimes \boldsymbol{S}_{\boldsymbol{y}}^{-1}\right) \mathcal{A},$$

其中，$\mathcal{A} \in \mathbb{R}^{n^2 \times m}$ 是这样一个矩阵：其 i^{th} 列是将 $\boldsymbol{A}_i$ 展平为长度为 n^2 的向量 vec（$\boldsymbol{A}_i$）后得到的。因此，根据我们在第 4 节中的讨论，我们可以在环境维度为 $\min\{m, n^2\}$ 的子空间中进行完整分析。根据定理 25，在界定 $\log \mathcal{E}(\mathrm{H}_{\boldsymbol{y}_0}) = \widetilde{O}(\min\{m, n^2\})$ 后，我们需要的时间为

$$\begin{aligned} &\widetilde{O}\left(\left(\min\{m, n^2\} + \sqrt{n\sqrt{m} \cdot \min\{m, n^2\}} \log\frac{nm}{\varepsilon}\right) \cdot \left(n^2 + \mathcal{T}_{\text{gradient}}\right)\right) \\ =&\widetilde{O}\left(\left(n^2 + n^{3/2}m^{1/4} \log\frac{1}{\varepsilon}\right)\left(mn^2 + n^\omega\right)\right) \end{aligned}$$

以求解对偶问题，达到 ε 精度。假设 $m \geq n$，这次所需时间为

$$\widetilde{O}\left(mn^4 + m^{1.25}n^{3.5} \log\frac{1}{\varepsilon}\right),$$

由此得到我们所声称的界。

引理 33。障碍函数 $\phi(\boldsymbol{y}) = -\sqrt{m}\cdot\log\det(\boldsymbol{B}-\sum_{i=1}^{m}\boldsymbol{y}_i\boldsymbol{A}_i)$ 是一个强自协调障碍函数，其中 $\nu = n\sqrt{m}$。

证明。令 $\psi(\mathrm{y}) = -\log\det(\boldsymbol{B}-\sum_{i=1}^{m}\mathrm{y}_i\boldsymbol{A}_i)$ 为标准的对数障碍函数。令 $\mathrm{H}_{\boldsymbol{y}} = \nabla^2\psi(\mathrm{y})$。我们将证明，对任意 h，它都满足

$$\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\frac{d}{dt}\boldsymbol{H}_{\boldsymbol{y}+t\boldsymbol{h}}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right\|_F \le \frac{2}{\sqrt{m}}\|\boldsymbol{h}\|_{\boldsymbol{H}_{\boldsymbol{y}}}.$$

因此，将其放大 $\sqrt{m}$ 倍即可得到一个 $n\sqrt{m}$-自协调障碍函数，满足定义9中的要求。为验证这一点，我们首先计算海森矩阵的各个元素

$$(\boldsymbol{H}_{\boldsymbol{y}})_{ij} = \mathrm{tr}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}\boldsymbol{A}_i\boldsymbol{S}_{\boldsymbol{y}}^{-1}\boldsymbol{A}_j\right),$$

这使我们能够写出

$$\boldsymbol{H}_{\boldsymbol{y}} = \mathcal{A}^{\top}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right)\mathcal{A},$$

其中，$\boldsymbol{S}_{\boldsymbol{y}} = \boldsymbol{B}-\sum_{i=1}^{m}\mathrm{y}_i\boldsymbol{A}_i$ 是松弛矩阵，$\mathcal{A}\in\mathbb{R}^{n^2\times m}$ 是这样一个矩阵：其 i^{th} 列由将 $\boldsymbol{A}_i$ 展平为长度为 n^2的向量得到，而 $\otimes$ 表示克罗内克积。对于任意$\mathrm{h}\in\mathbb{R}^m$，令 $\Delta\boldsymbol{S}(\mathrm{h}) = \boldsymbol{S}_{\boldsymbol{y}+\boldsymbol{h}}-\boldsymbol{S}_{\boldsymbol{y}} = -\sum_i \mathrm{h}_i\boldsymbol{A}_i$。因此，利用展开式 $(\mathrm{X}+t\mathrm{Y})^{-1} = \mathrm{X}^{-1} - t\mathrm{X}^{-1}\mathrm{Y}\,\mathrm{X}^{-1} + O_{\boldsymbol{X},\boldsymbol{Y}}(t^2)\,\mathrm{I}$，我们计算：

$$\begin{aligned}
\frac{d}{dt}\boldsymbol{H}_{\boldsymbol{y}+t\boldsymbol{h}} &= \lim_{t\to0}\frac{1}{t}\cdot\mathcal{A}^{\top}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}-\boldsymbol{S}_{\boldsymbol{y}+t\boldsymbol{\delta}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}+t\boldsymbol{\delta}}^{-1}\right)\mathcal{A}\\
&= \lim_{t\to0}\frac{1}{t}\cdot\mathcal{A}^{\top}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}-\left(\boldsymbol{S}_{\boldsymbol{y}}+\Delta\boldsymbol{S}(t\boldsymbol{\delta})\right)^{-1}\otimes\left(\boldsymbol{S}_{\boldsymbol{y}}+\Delta\boldsymbol{S}(t\boldsymbol{\delta})\right)^{-1}\right)\mathcal{A}\\
&= \lim_{t\to0}\frac{1}{t}\cdot\mathcal{A}^{\top}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}-\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}-t\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right)\otimes\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}-t\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right)\right)\mathcal{A}\\
&= \lim_{t\to0}\frac{1}{t}\cdot\mathcal{A}^{\top}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}-\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}-t\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right)\otimes\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}-t\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right)\right)\mathcal{A}\\
&= \mathcal{A}^{\top}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}+\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right)\mathcal{A}.
\end{aligned}$$

Using the inequality

$$\left\|\left(\boldsymbol{M}^{\top}\boldsymbol{X}\boldsymbol{M}\right)^{-1/2}\left(\boldsymbol{M}^{\top}\boldsymbol{Y}\boldsymbol{M}\right)\left(\boldsymbol{M}^{\top}\boldsymbol{X}\boldsymbol{M}\right)^{-1/2}\right\|_F \le \left\|\boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2}\right\|_F = \left\|\boldsymbol{X}^{-1}\boldsymbol{Y}\right\|_F,$$

我们可以对 Frobenius 范数给出上界

$$\begin{aligned}
&\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\frac{d}{dt}\boldsymbol{H}_{\boldsymbol{y}+t\boldsymbol{h}}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right\|_F\\
&\le \left\|\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right)^{-1}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}+\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right)\right\|_F\\
&= \left\|\left(\boldsymbol{S}_{\boldsymbol{y}}\otimes\boldsymbol{S}_{\boldsymbol{y}}\right)\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}+\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right)\right\|_F\\
&= \left\|\boldsymbol{I}_m\otimes\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}+\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}\otimes\boldsymbol{I}_m\right\|_F\\
&\le 2\left\|\boldsymbol{I}_m\otimes\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right\|_F\\
&= 2\left\|\boldsymbol{I}_m\right\|_F\left\|\Delta\boldsymbol{S}(\boldsymbol{\delta})\,\boldsymbol{S}_{\boldsymbol{y}}^{-1}\right\|_F\\
&= 2\sqrt{m}\left\|\boldsymbol{S}_{\boldsymbol{y}}^{-1}\Delta\boldsymbol{S}(\boldsymbol{\delta})\right\|_F. \qquad (7)
\end{aligned}$$

给定矩阵 X，令 vec (X) 表示将其展平为向量。利用恒等式 $\left(\boldsymbol{B}^{\top} \otimes \boldsymbol{A}\right)$ vec (X) =vec ($\boldsymbol{A}$X $\boldsymbol{B}$)，我们得到：

$$
\begin{aligned}
\|\boldsymbol{h}\|_{\boldsymbol{H}_{\boldsymbol{y}}}^{2} &= \boldsymbol{h}^{\top} \mathcal{A}^{\top}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1} \otimes \boldsymbol{S}_{\boldsymbol{y}}^{-1}\right) \mathcal{A} \boldsymbol{h} \\
&= \operatorname{vec}\left(\Delta \boldsymbol{S}\left(\boldsymbol{\delta}\right)\right)^{\top}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1} \otimes \boldsymbol{S}_{\boldsymbol{y}}^{-1}\right) \operatorname{vec}(\Delta \boldsymbol{S}\left(\boldsymbol{\delta}\right)) \\
&= \left\langle \operatorname{vec}\left(\Delta \boldsymbol{S}\left(\boldsymbol{\delta}\right)\right), \operatorname{vec}\left(\boldsymbol{S}_{\boldsymbol{y}}^{-1} \Delta \boldsymbol{S}\left(\boldsymbol{\delta}\right) \boldsymbol{S}_{\boldsymbol{y}}^{-1}\right)\right\rangle \\
&= \left\langle \Delta \boldsymbol{S}\left(\boldsymbol{\delta}\right), \boldsymbol{S}_{\boldsymbol{y}}^{-1} \Delta \boldsymbol{S}\left(\boldsymbol{\delta}\right) \boldsymbol{S}_{\boldsymbol{y}}^{-1}\right\rangle \\
&= \left\|\boldsymbol{S}_{\boldsymbol{y}}^{-1} \Delta \boldsymbol{S}\left(\boldsymbol{\delta}\right)\right\|_{F}^{2} . \qquad (8)
\end{aligned}
$$

结合（7）和（8），我们得到了关于海森矩阵的相对变化的所需界。最后，由于 ψ（$\boldsymbol{y}$）是 n-自协调的，按照标准论证，将其缩放为 $\sqrt{m}$ 可得$\nabla \phi$（$\boldsymbol{y}$）$^{\top}\left(\nabla^{2} \phi\right.$（$\boldsymbol{y}$）$\left.\right)^{-1} \nabla \phi$（$\boldsymbol{y}$） $=\sqrt{m} \cdot \nabla \psi$（$\boldsymbol{y}$）$^{\top}\left(\nabla^{2} \psi\right.$（$\boldsymbol{y}$）$\left.\right)^{-1} \nabla \psi$（$\boldsymbol{y}$） $\leq n \sqrt{m}$。证明至此结束。

□

致谢。

我们感谢法国国家研究署（ANR）通过 ANR-21-CE48-0016 项目资助（COMCOPT 项目）提供支持，也感谢法国国家科学研究中心通过 CoopIntEER IEA 项目资助（ALFRED 项目）提供支持。

A 第 2 节的证明2

A.1 引理的证明 4

证明。 我们使用恒等式

$$\begin{aligned}\mathcal{E}(\boldsymbol{X}) &= \det\left(\frac{\boldsymbol{X}^{1/2}+\boldsymbol{X}^{-1/2}}{2}\right) = \frac{1}{2^n}\det\left(\left(\boldsymbol{X}+\boldsymbol{I}\right)\boldsymbol{X}^{-1/2}\right) \\ &= \frac{1}{2^n}\det(\boldsymbol{X}+\boldsymbol{I})\det\left(\boldsymbol{X}^{1/2}\right)^{-1} = \frac{1}{2^n}\frac{\det(\boldsymbol{X}+\boldsymbol{I})}{\sqrt{\det(\boldsymbol{X})}}.\end{aligned} \tag{9}$$

因此，我们有

$$\begin{aligned}\mathcal{E}\left(\boldsymbol{X}'\right) &= \frac{1}{2^n}\frac{\det\left(\boldsymbol{X}'+\boldsymbol{I}\right)}{\sqrt{\det\left(\boldsymbol{X}'\right)}} = \frac{1}{2^n}\frac{\det\left(\boldsymbol{X}+\boldsymbol{I}+\boldsymbol{u}\boldsymbol{u}^\top\right)}{\sqrt{\det\left(\boldsymbol{X}+\boldsymbol{u}\boldsymbol{u}^\top\right)}} = \frac{1}{2^n}\frac{\det(\boldsymbol{X}+\boldsymbol{I})\cdot\left(1+\boldsymbol{u}^\top(\boldsymbol{I}+\boldsymbol{X})^{-1}\boldsymbol{u}\right)}{\sqrt{\det(\boldsymbol{X})\left(1+\boldsymbol{u}^\top\boldsymbol{X}^{-1}\boldsymbol{u}\right)}} \\ &= \mathcal{E}(\boldsymbol{X})\cdot\frac{1+\boldsymbol{u}^\top(\boldsymbol{I}+\boldsymbol{X})^{-1}\boldsymbol{u}}{\sqrt{1+\boldsymbol{u}^\top\boldsymbol{X}^{-1}\boldsymbol{u}}},\end{aligned}$$

其中用到了恒等式 $\det\left(\mathrm{X}+\mathrm{v}\,\mathrm{v}^\top\right)=\det(\mathrm{X})\left(1+\mathrm{v}^\top\mathrm{X}^{-1}\mathrm{v}\right)$。 □

A.2 引理的证明 5

证明。 根据引理 4我们需要给出上界

$$\frac{1+\boldsymbol{u}^\top(\boldsymbol{I}+\boldsymbol{X})^{-1}\boldsymbol{u}}{\sqrt{1+\boldsymbol{u}^\top\boldsymbol{X}^{-1}\boldsymbol{u}}} \le \frac{1+\boldsymbol{u}^\top\boldsymbol{u}}{\sqrt{1+\gamma}} = \frac{2}{\sqrt{1+\gamma}}.$$

我们利用以下事实分析第二种情况： $\mathcal{E}\left(\boldsymbol{X}\right)=\mathcal{E}\left(\boldsymbol{X}^{-1}\right)$ 因此我们可以对 1 进行秩一更新： $\boldsymbol{X}^{-1}$。 更准确地说，根据 Sherman-Morrison 公式（引理 6），我们有：

$$\frac{\mathcal{E}\left(\boldsymbol{X}-\frac{\boldsymbol{X}\boldsymbol{u}\boldsymbol{u}^\top\boldsymbol{X}}{1+\boldsymbol{u}^\top\boldsymbol{X}\boldsymbol{u}}\right)}{\mathcal{E}(\boldsymbol{X})} = \frac{\mathcal{E}\left(\left(\boldsymbol{X}^{-1}+\boldsymbol{u}\boldsymbol{u}^\top\right)^{-1}\right)}{\mathcal{E}(\boldsymbol{X})} = \frac{\mathcal{E}\left(\boldsymbol{X}^{-1}+\boldsymbol{u}\boldsymbol{u}^\top\right)}{\mathcal{E}\left(\boldsymbol{X}^{-1}\right)} \le \frac{2}{\sqrt{1+\gamma}}.$$

□

A.3 引理证明 7

证明。 我们应用公式（9），并使用 行列式（$\boldsymbol{A}\boldsymbol{B}$） =行列式（$\boldsymbol{A}$） 行列式（$\boldsymbol{B}$）：

$$\begin{aligned}\mathcal{E}(\boldsymbol{X}) &= \frac{1}{2^n}\frac{\det(\boldsymbol{X}+\boldsymbol{I})}{\sqrt{\det(\boldsymbol{X})}} = \frac{1}{2^n}\frac{\det(\boldsymbol{Y})\det(\boldsymbol{X}+\boldsymbol{I})\det\left(\boldsymbol{Y}^{-1}\right)}{\sqrt{\det(\boldsymbol{Y})\det(\boldsymbol{X})\det\left(\boldsymbol{Y}^{-1}\right)}} \\ &= \frac{1}{2^n}\frac{\det\left(\boldsymbol{Y}\boldsymbol{X}\boldsymbol{Y}^{-1}+\boldsymbol{I}\right)}{\sqrt{\det\left(\boldsymbol{Y}\boldsymbol{X}\boldsymbol{Y}^{-1}\right)}} = \mathcal{E}\left(\boldsymbol{Y}\boldsymbol{X}\boldsymbol{Y}^{-1}\right).\end{aligned}$$

第二个性质根据定义成立。 □

A.4 引理的证明14

证明。 根据自协调性性质，如果 $\boldsymbol{y}^\star$ 是 g_μ 的极小值点，那么由标准论证可得 [Ren01] 成立

$$\|\boldsymbol{y}-\boldsymbol{y}^\star\|_{\boldsymbol{H}_{\boldsymbol{y}^\star}} \leq \frac{\|\nabla g_\mu(\boldsymbol{y})\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}}{1-\|\nabla g_\mu(\boldsymbol{y})\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}} \leq \frac{1}{2}.$$

因此

$$\boldsymbol{H}_{\boldsymbol{y}} \cdot \left(1-\|\boldsymbol{y}-\boldsymbol{y}^\star\|_{\boldsymbol{H}_{\boldsymbol{y}}}\right)^2 \preceq \boldsymbol{H}_{\boldsymbol{y}^\star} \preceq \boldsymbol{H}_{\boldsymbol{y}} \cdot \left(\frac{1}{1-\|\boldsymbol{y}-\boldsymbol{y}^\star\|_{\boldsymbol{H}_{\boldsymbol{y}}}}\right)^2$$

这给出

$$\boldsymbol{H}_{\boldsymbol{y}^\star} \cdot \frac{1}{4} \preceq \boldsymbol{H}_{\boldsymbol{y}} \preceq \boldsymbol{H}_{\boldsymbol{y}^\star} \cdot 4.$$

利用假设中的界，我们得到 $\max\left\{\|\mathbf{H}_{\boldsymbol{y}^\star}\|, \left\|\mathbf{H}_{\boldsymbol{y}^\star}^{-1}\right\|\right\} \leq \kappa(\mu)$，从而得出该结论。

□

B 章节证明3

B.1 理查森迭代得到的偏心率证书

引理34。 设 $\boldsymbol{H} \in \mathbb{R}^{n\times n}$，且 $\boldsymbol{b}$、$\boldsymbol{x} \in \mathbb{R}^n$ 为向量。设 $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{H}\boldsymbol{x}$，并考虑以下步长

$$\boldsymbol{x}' = \boldsymbol{x} + \frac{\langle \boldsymbol{r}, \boldsymbol{H}\boldsymbol{r}\rangle}{\|\boldsymbol{H}\boldsymbol{r}\|_2^2}\boldsymbol{r}.$$

令 $\beta \in (0,\ 1)$ 为一个标量。若新的残差 $\boldsymbol{r}' = \boldsymbol{b} - \boldsymbol{H}\boldsymbol{x}'$ 满足

$$\frac{\|\boldsymbol{r}\|_2^2 - \|\boldsymbol{r}'\|_2^2}{\|\boldsymbol{r}\|_2^2} \leq \beta,$$

我们便可得到以下偏心率证书中的至少一个：

1. $\frac{\|\boldsymbol{r}\|_{\boldsymbol{H}}^2}{\|\boldsymbol{r}\|_2^2} \leq \sqrt{\beta} \iff \frac{\left\|\boldsymbol{H}^{1/2}\boldsymbol{r}\right\|_{\boldsymbol{H}^{-1}}^2}{\left\|\boldsymbol{H}^{1/2}\boldsymbol{r}\right\|_2^2} \geq \frac{1}{\sqrt{\beta}}$,
2. $\frac{\|\boldsymbol{H}\boldsymbol{r}\|_2^2}{\|\boldsymbol{r}\|_{\boldsymbol{H}}^2} \geq \frac{1}{\sqrt{\beta}} \iff \frac{\left\|\boldsymbol{H}^{1/2}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}{\left\|\boldsymbol{H}^{1/2}\boldsymbol{r}\right\|_2^2} \geq \frac{1}{\sqrt{\beta}}$.

证明。 我们注意到，我们考虑的步长是使新残差 $\mathbf{r}'$ 的 ℓ_2 范数最小的步长。使用此更新，我们测量新的范数：

$$\begin{aligned}
\left\|\boldsymbol{r}'\right\|_2^2 &= \left\|\boldsymbol{b}-\boldsymbol{H}\boldsymbol{x}'\right\|_2^2 = \left\|\boldsymbol{b}-\boldsymbol{H}\boldsymbol{x}-\boldsymbol{H}\left(\boldsymbol{x}'-\boldsymbol{x}\right)\right\|_2^2 = \left\|\boldsymbol{r}-\boldsymbol{H}\left(\boldsymbol{x}'-\boldsymbol{x}\right)\right\|_2^2 \\
&= \left\|\boldsymbol{r}-\boldsymbol{H}\cdot\frac{\langle \boldsymbol{r}, \boldsymbol{H}\boldsymbol{r}\rangle}{\|\boldsymbol{H}\boldsymbol{r}\|_2^2}\boldsymbol{r}\right\|_2^2 = \|\boldsymbol{r}\|_2^2 + \frac{\langle \boldsymbol{r}, \boldsymbol{H}\boldsymbol{r}\rangle^2}{\|\boldsymbol{H}\boldsymbol{r}\|_2^4}\|\boldsymbol{H}\boldsymbol{r}\|_2^2 - 2\cdot\frac{\langle \boldsymbol{r}, \boldsymbol{H}\boldsymbol{r}\rangle^2}{\|\boldsymbol{H}\boldsymbol{r}\|_2^2} \\
&= \|\boldsymbol{r}\|_2^2 - \frac{\langle \boldsymbol{r}, \boldsymbol{H}\boldsymbol{r}\rangle^2}{\|\boldsymbol{H}\boldsymbol{r}\|_2^2}.
\end{aligned}$$

因此，我们可以将乘法进展写为：

$$\frac{\|\boldsymbol{r}\|_2^2-\|\boldsymbol{r'}\|_2^2}{\|\boldsymbol{r}\|_2^2}=\frac{\langle\boldsymbol{r},\boldsymbol{H}\boldsymbol{r}\rangle^2}{\|\boldsymbol{H}\boldsymbol{r}\|_2^2\|\boldsymbol{r}\|_2^2}\leq\beta\,.$$

因此，进展不大意味着以下两个条件之一必然成立。要么 $\|\mathrm{r}\|_2\geq\|\mathrm{H\,r}\,\|_2$，在这种情况下，乘法进展的上界意味着

$$\frac{\|\boldsymbol{r}\|_{\boldsymbol{H}}^4}{\|\boldsymbol{r}\|_2^4}\leq\beta\iff\frac{\|\boldsymbol{r}\|_{\boldsymbol{H}}^2}{\|\boldsymbol{r}\|_2^2}\leq\sqrt{\beta}\iff\frac{\left\|\boldsymbol{H}^{1/2}\boldsymbol{r}\right\|_{\boldsymbol{H}^{-1}}^2}{\left\|\boldsymbol{H}^{1/2}\boldsymbol{r}\right\|_2^2}\geq\frac{1}{\sqrt{\beta}}\,.$$

否则，我们必有 $\|\mathrm{H\,r}\,\|_2\geq\|\mathrm{r}\|_2$，此时我们有

$$\frac{\|\boldsymbol{H}\boldsymbol{r}\|_2^4}{\|\boldsymbol{r}\|_H^4}\geq\frac{1}{\beta}\iff\frac{\|\boldsymbol{H}\boldsymbol{r}\|_2^2}{\|\boldsymbol{r}\|_H^2}\geq\frac{1}{\sqrt{\beta}}\iff\frac{\left\|\boldsymbol{H}^{1/2}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}{\left\|\boldsymbol{H}^{1/2}\boldsymbol{r}\right\|_2^2}\geq\frac{1}{\sqrt{\beta}}\,.$$

证明到此结束。 □

B.2 引理证明15

证明。我们进行变量代换 $\mathrm{y}=\widetilde{\mathrm{H}}^{1/2}\mathrm{x}$，并考虑等价系统 $\widetilde{\mathrm{H}}^{-1/2}\mathrm{H}\widetilde{\mathrm{H}}^{-1/2}\cdot\mathrm{y}=\widetilde{\mathrm{H}}^{-1/2}\mathrm{b}$。令未预条件化残差为 $\mathrm{r}=\mathrm{b}-\mathrm{Hx}$，并令预条件化残差为 $\widehat{\mathrm{r}}=\widetilde{\mathrm{H}}^{-1/2}\left(\mathrm{b}-\mathrm{H}\widetilde{\mathrm{H}}^{-1/2}\mathrm{y}\right)=\widetilde{\mathrm{H}}^{-1/2}\mathrm{r}$。根据引理34，迭代

$$\boldsymbol{y'}=\boldsymbol{y}+\frac{\left\langle\widehat{\boldsymbol{r}},\left(\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}\right)\widehat{\boldsymbol{r}}\right\rangle}{\left\|\left(\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}\right)\widehat{\boldsymbol{r}}\right\|_2^2}\cdot\widehat{\boldsymbol{r}}$$

映射回

$$\begin{aligned}\boldsymbol{x'}=\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{y'}&=\widetilde{\boldsymbol{H}}^{-1/2}\left(\boldsymbol{y}+\frac{\left\langle\widehat{\boldsymbol{r}},\left(\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}\right)\widehat{\boldsymbol{r}}\right\rangle}{\left\|\left(\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}\right)\widehat{\boldsymbol{r}}\right\|_2^2}\cdot\widehat{\boldsymbol{r}}\right)\\&=\boldsymbol{x}+\frac{\left\langle\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{r},\left(\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}\right)\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{r}\right\rangle}{\left\|\left(\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}\right)\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{r}\right\|_2^2}\cdot\widetilde{\boldsymbol{H}}^{-1/2}\cdot\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{r}\\&=\boldsymbol{x}+\frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}{\left\|\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\cdot\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\,,\end{aligned}$$

在原始参数化中。关于 $\|\hat{\mathrm{r}}\|$ 减小的条件映射回了

$$\frac{\|\widehat{\boldsymbol{r}}\|_2^2-\|\widehat{\boldsymbol{r}}'\|_2^2}{\|\widehat{\boldsymbol{r}}\|_2^2}=\frac{\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2-\|\boldsymbol{r}'\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}{\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\leq\beta\,.$$

最后，所提供的偏心率证书要么是

$$\frac{\|\widehat{\boldsymbol{r}}\|_{\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}}^2}{\|\widehat{\boldsymbol{r}}\|_2^2}=\frac{\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}}^2}{\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{r}\right\|_2^2}=\frac{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2^2}{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}}^2}\leq\sqrt{\beta}$$

or

$$\frac{\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}\cdot\widehat{\boldsymbol{r}}\right\|_2^2}{\|\widehat{\boldsymbol{r}}\|_{\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}}^2}=\frac{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}^{1/2}}^2}{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2^2}\geq\frac{1}{\sqrt{\beta}}\,,$$

由此得证。 □

B.3 引理 16的证明

证明。 我们分别分析这两种类型的证书。

类型 1。 令 $\boldsymbol{X}=\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}$且 $\boldsymbol{u}=\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}$，则有 $\boldsymbol{u}^\top\boldsymbol{X}\boldsymbol{u}\geq\frac{1}{\sqrt{\beta}}\boldsymbol{u}^\top\boldsymbol{u}$。根据引理 5，我们得到，将预条件器改为 $\boldsymbol{H}'$，使得

$$\begin{aligned}\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1/2}&=\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}-\frac{\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}\left(\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\left(\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^\top\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}}{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2^2+\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}}^2}\\&=\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}-\frac{\boldsymbol{H}^{-1/2}\boldsymbol{r}\boldsymbol{r}^\top\boldsymbol{H}^{-1/2}}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2+\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\,,\end{aligned}$$

并等价地令

$$\widetilde{\boldsymbol{H}}'=\widetilde{\boldsymbol{H}}-\frac{\boldsymbol{r}\boldsymbol{r}^\top}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2+\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\,,$$

我们得到

$$\mathcal{E}\left(\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1}\right)\leq\mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right)\cdot\frac{2}{\sqrt{1+\frac{1}{\sqrt{\beta}}}}\,.$$

类型 2。 令 $\boldsymbol{X} = \boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}$和 $\boldsymbol{u} = \boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}$，可得 $\boldsymbol{u}^\top\boldsymbol{X}^{-1}\boldsymbol{u} \geq \frac{1}{\sqrt{\beta}}\boldsymbol{u}^\top\boldsymbol{u}$。使用引理 5 可知，将预条件器改为 $\boldsymbol{H}'$ 使得

$$\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1/2} = \boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2} + \frac{\left(\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\left(\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^\top}{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2^2},$$

并且等价地令

$$\widetilde{\boldsymbol{H}}' = \widetilde{\boldsymbol{H}} + \frac{\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\boldsymbol{r}^\top\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2},$$

我们得到

$$\mathcal{E}\left(\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1}\right) \leq \mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right) \cdot \frac{2}{\sqrt{1+\frac{1}{\sqrt{\beta}}}}.$$

我们描述算法 1中的预条件器更新。逆预条件器的更新$\widetilde{\boldsymbol{H}}^{-1}$ 由应用 Sherman-Morrison 公式（引理 6）得出。

预条件器范数的变化。现在，我们验证与 $\widetilde{\mathrm{H}}$ 和 $\widetilde{\mathrm{H}}^{-1}$ 的范数相比，更新不会使 $\widetilde{\mathrm{H}}'$ 和 $\widetilde{\mathrm{H}}'^{-1}$ 的范数增加过多。每当预条件器范数增加时，都是由于类型 2 证书，此时我们有：

$$\widetilde{\boldsymbol{H}}' = \widetilde{\boldsymbol{H}} + \frac{\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\boldsymbol{r}^\top\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2} = \widetilde{\boldsymbol{H}}^{1/2}\left(\boldsymbol{I} + \frac{\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\boldsymbol{r}^\top\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}\right)\widetilde{\boldsymbol{H}}^{1/2},$$

因此

$$\left\|\widetilde{\boldsymbol{H}}^{-1/2}\widetilde{\boldsymbol{H}}'\widetilde{\boldsymbol{H}}^{-1/2}\right\| \leq 1 + \frac{\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\boldsymbol{r}^\top\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}\widetilde{\boldsymbol{H}}^{-1/2}\right\|}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2} = 1 + \frac{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}^{1/2}}^2}{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}.$$

此时我们利用这样一个事实：偏心率的下降由证书所证明的特征值的大小决定。事实上，将 β 设置为恰好匹配比值 $\beta = \mathrm{u}^\top\mathrm{X}^{-1}\mathrm{u}/\mathrm{u}^\top\mathrm{u}$，我们有：

$$\mathcal{E}\left(\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1}\right) \leq \mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right) \cdot \frac{2}{\sqrt{1+\frac{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{H}^{1/2}}^2}{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}}},$$

结合前面的不等式，我们现在可以得到如下上界：

$$\left\|\widetilde{\boldsymbol{H}}^{-1/2}\widetilde{\boldsymbol{H}}'\widetilde{\boldsymbol{H}}^{-1/2}\right\| \leq \left(2 \cdot \frac{\mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right)}{\mathcal{E}\left(\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1}\right)}\right)^2.$$

另一种情况类似。每当逆预条件器范数增大时，都是因为出现了一个 1 型证书，此时我们有：

$$\widetilde{\boldsymbol{H}'}^{-1} = \widetilde{\boldsymbol{H}}^{-1} + \frac{\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\boldsymbol{r}^\top\widetilde{\boldsymbol{H}}^{-1}}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2} = \widetilde{\boldsymbol{H}}^{-1/2}\left(\boldsymbol{I} + \frac{\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{r}\boldsymbol{r}^\top\widetilde{\boldsymbol{H}}^{-1/2}}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}\right)\widetilde{\boldsymbol{H}}^{-1/2},$$

so

$$\left\|\widetilde{\boldsymbol{H}}^{1/2}\widetilde{\boldsymbol{H}'}^{-1}\widetilde{\boldsymbol{H}}^{1/2}\right\| \le 1 + \frac{\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{r}\right\|_2^2}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}.$$

再次利用偏心率的下降由证书所证明的特征值大小决定这一事实。令 β 恰好匹配比值 $\beta = \mathrm{u}^\top \mathrm{X}\,\mathrm{u}\,/\mathrm{u}^\top\mathrm{u}$，我们得到：

$$\mathcal{E}\left(\widetilde{\boldsymbol{H}'}\boldsymbol{H}^{-1}\right) \le \mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right)\cdot\frac{2}{\sqrt{1+\frac{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}}^2}{\left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2^2}}} = \mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right)\cdot\frac{2}{\sqrt{1+\frac{\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{r}\right\|_2^2}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}}^2}}},$$

结合前一个不等式，我们可以给出如下上界：

$$\left\|\widetilde{\boldsymbol{H}}^{1/2}\widetilde{\boldsymbol{H}'}^{-1}\widetilde{\boldsymbol{H}}^{1/2}\right\| \le \left(2\cdot\frac{\mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right)}{\mathcal{E}\left(\widetilde{\boldsymbol{H}'}\boldsymbol{H}^{-1}\right)}\right)^2.$$

□

B.4 引理的证明 17

~ ~证明。令 $\mathrm{r}_t = \mathrm{b} - \mathrm{H}\,\mathrm{x}_t$。注意，对于任意向量 v 以及任意 H、H′，

$$\|\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2 = \left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{v}\right\|_2^2 = \left\|\widetilde{\boldsymbol{H}}^{-1/2}\widetilde{\boldsymbol{H}'}^{1/2}\cdot\widetilde{\boldsymbol{H}'}^{-1/2}\boldsymbol{v}\right\|_2^2 \le \left\|\widetilde{\boldsymbol{H}'}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\widetilde{\boldsymbol{H}'}^{1/2}\right\|\cdot\|\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}'}^{-1}}^2.$$

利用这个不等式，我们可以写成：

$$\begin{aligned}
\|\boldsymbol{r}_T\|^2 &\le \left(\prod_{t=0}^{T-1}\left\|\widetilde{\boldsymbol{H}}_{t+1}^{1/2}\widetilde{\boldsymbol{H}}_t^{-1}\widetilde{\boldsymbol{H}}_{t+1}^{1/2}\right\|\right)\left\|\widetilde{\boldsymbol{H}}_T^{-1/2}\boldsymbol{r}_T\right\|_2^2 \\
&= \left(\prod_{t=0}^{T-1}\left\|\widetilde{\boldsymbol{H}}_{t+1}^{1/2}\widetilde{\boldsymbol{H}}_t^{-1}\widetilde{\boldsymbol{H}}_{t+1}^{1/2}\right\|\right)\left(\prod_{t=0}^{t-1}\frac{\left\|\widetilde{\boldsymbol{H}}_{t+1}^{-1/2}\boldsymbol{r}_{t+1}\right\|^2}{\left\|\widetilde{\boldsymbol{H}}_t^{-1/2}\boldsymbol{r}_t\right\|^2}\right)\left\|\widetilde{\boldsymbol{H}}_0^{-1/2}\boldsymbol{r}_0\right\|_2^2 \\
&\le \prod_{t=0}^{T-1}\left(\left\|\widetilde{\boldsymbol{H}}_{t+1}^{1/2}\widetilde{\boldsymbol{H}}_t^{-1}\widetilde{\boldsymbol{H}}_{t+1}^{1/2}\right\|\cdot\frac{\|\boldsymbol{r}_{t+1}\|_{\widetilde{\boldsymbol{H}}_{t+1}^{-1}}^2}{\|\boldsymbol{r}_t\|_{\widetilde{\boldsymbol{H}}_t^{-1}}^2}\right)\cdot\|\boldsymbol{r}_0\|_{\widetilde{\boldsymbol{H}}_0^{-1}}^2 \\
&= \Phi_T\cdot\|\boldsymbol{r}_0\|_{\widetilde{\boldsymbol{H}}_0^{-1}}^2,
\end{aligned}$$

其中，我们使用定义为 Φ 的势函数

$$\Phi_t = \begin{cases} 1\,, & \text{if } t = 0\,, \\ \prod_{t=0}^{t-1} \left(\left\| \widetilde{\boldsymbol{H}}_{t+1}^{1/2} \widetilde{\boldsymbol{H}}_t^{-1} \widetilde{\boldsymbol{H}}_{t+1}^{1/2} \right\| \cdot \frac{\|\boldsymbol{r}_{t+1}\|^2_{\widetilde{\boldsymbol{H}}_{t+1}^{-1}}}{\|\boldsymbol{r}_t\|^2_{\widetilde{\boldsymbol{H}}_t^{-1}}} \right), & \text{if } t \geq 1\,. \end{cases}$$

现在我们来看 Φ_t如何演化。有两种情况。在前一种情况下，$\|\boldsymbol{r}_{t+1}\|^2_{\widetilde{\boldsymbol{H}}_{t+1}^{-1}} \leq (1-\beta)\,\|\boldsymbol{r}_t\|^2_{\widetilde{\boldsymbol{H}}_t^{-1}}$，因此预条件器保持不变，所以 $\Phi_t \leq (1-\beta)\,\Phi_{t-1}$。在后一种情况下，我们执行预条件器更新，但保持迭代点不变 $\boldsymbol{r}_{t+1} = \boldsymbol{r}_t$。根据引理 16，有两种可能性。

若 $1 \leq \left\| \widetilde{\mathrm{H}}_t^{-1/2} \widetilde{\mathrm{H}}_{t+1} \widetilde{\mathrm{H}}_t^{-1/2} \right\| \leq \left(2 \cdot \frac{\mathcal{E}(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1})}{\mathcal{E}(\widetilde{\boldsymbol{H}}_{t+1}\boldsymbol{H}^{-1})} \right)^2$，则 $\|\mathrm{r}_{t+1}\|_{\widetilde{\boldsymbol{H}}_{t+1}^{-1}} \leq \|\mathrm{r}_{t+1}\|_{\widetilde{\boldsymbol{H}}_t^{-1}}$，但 $\left\| \widetilde{\mathrm{H}}_{t+1}^{1/2} \widetilde{\mathrm{H}}_t^{-1} \widetilde{\mathrm{H}}_{t+1}^{1/2} \right\| = \left\| \widetilde{\mathrm{H}}_t^{-1/2} \widetilde{\mathrm{H}}_{t+1} \widetilde{\mathrm{H}}_t^{-1/2} \right\|$，因此

$$\Phi_t \leq \Phi_{t-1} \cdot \left(2 \cdot \frac{\mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right)}{\mathcal{E}\left(\widetilde{\boldsymbol{H}}_{t+1}\boldsymbol{H}^{-1}\right)} \right)^2 .$$

If $1 \leq \left\| \widetilde{\boldsymbol{H}}_t^{1/2} \widetilde{\boldsymbol{H}}_{t+1}^{-1} \widetilde{\boldsymbol{H}}_t^{1/2} \right\| \leq \left(2 \cdot \frac{\mathcal{E}(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1})}{\mathcal{E}(\widetilde{\boldsymbol{H}}_{t+1}\boldsymbol{H}^{-1})} \right)^2$, then $\left\| \widetilde{\boldsymbol{H}}_{t+1}^{1/2} \widetilde{\boldsymbol{H}}_t^{-1} \widetilde{\boldsymbol{H}}_{t+1}^{1/2} \right\| = \frac{1}{\left\| \widetilde{\boldsymbol{H}}_t^{1/2} \widetilde{\boldsymbol{H}}_{t+1}^{-1} \widetilde{\boldsymbol{H}}_t^{1/2} \right\|} \leq 1$,

but $\frac{\|\boldsymbol{r}_{t+1}\|^2_{\widetilde{\boldsymbol{H}}_{t+1}^{-1}}}{\|\boldsymbol{r}_t\|^2_{\widetilde{\boldsymbol{H}}_t^{-1}}} = \frac{\left\| \widetilde{\boldsymbol{H}}_t^{-1/2} \boldsymbol{r}_t \right\|^2_{\widetilde{\boldsymbol{H}}_t^{1/2} \widetilde{\boldsymbol{H}}_{t+1}^{-1} \widetilde{\boldsymbol{H}}_t^{1/2}}}{\left\| \widetilde{\boldsymbol{H}}_t^{-1/2} \boldsymbol{r}_t \right\|_2^2} \leq \left\| \widetilde{\boldsymbol{H}}_t^{1/2} \widetilde{\boldsymbol{H}}_{t+1}^{-1} \widetilde{\boldsymbol{H}}_t^{1/2} \right\|$, and therefore

$$\Phi_t \leq \Phi_{t-1} \cdot \left(2 \cdot \frac{\mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right)}{\mathcal{E}\left(\widetilde{\boldsymbol{H}}_{t+1}\boldsymbol{H}^{-1}\right)} \right)^2 .$$

最后，由于每当预条件器发生变化时，偏心率最坏情况下都会缩小至原来的 $\frac{2}{\sqrt{1+\frac{1}{\sqrt{\beta}}}} \geq \beta^{1/4}$ 倍（引理 34），而且$\mathcal{E}(\mathrm{X}) \geq 1$ 始终成立，因此这种情况最多会发生 $T_{\text{prec}} = {}^4\log \mathcal{E}(\boldsymbol{X}_0) \log(1/\beta)$次。因此，令

$$T = T_{\text{prec}} + \log_{1-\beta} \varepsilon = \frac{4 \log \mathcal{E}\left(\boldsymbol{X}_0\right)}{\log\left(\frac{1}{\beta}\right)} + \frac{\log\left(1/\varepsilon\right)}{\log\left(1/\left(1-\beta\right)\right)}$$

我们可以保证

$$\Phi_T \leq \Phi_0 \left(1-\beta\right)^{T - T_{\text{prec}}} = \Phi_0 \cdot \varepsilon = \varepsilon\,,$$

从而得到 $\|\mathrm{r}_T\|_2^2 \leq \varepsilon \|\mathrm{r}_0\|_2^2$。最后，令 $\beta = \frac{1}{2}$，我们得到 $T = O\left(\log\, \mathcal{E}(\mathrm{X}_0) + \log \frac{1}{\varepsilon}\right)$，证明完毕。

□

C 来自第4 节的证明

C.1 鲁棒性证明

引理 35。 设有一个向量 $\boldsymbol{u}$及其范数的一个上估计 $n(\boldsymbol{u})$，满足 $\|\boldsymbol{u}\| \leq n(\boldsymbol{u}) \leq \|\boldsymbol{u}\| \cdot \alpha$，其中 $\alpha \geq 1$。如果 $\boldsymbol{u}$ 满足

1. $\boldsymbol{u}^{\top} \boldsymbol{X}^{-1} \boldsymbol{u} \geq \gamma \cdot \boldsymbol{u}^{\top} \boldsymbol{u}$, *then* $\frac{\mathcal{E}\left(\boldsymbol{X}+\frac{\boldsymbol{u}\boldsymbol{u}^{\top}}{n(\boldsymbol{u})^2}\right)}{\mathcal{E}(\boldsymbol{X})} \leq \frac{2}{\sqrt{1+\frac{\gamma}{\alpha^2}}}$,

2. $\boldsymbol{u}^{\top} \boldsymbol{X} \boldsymbol{u} \geq \gamma \cdot \boldsymbol{u}^{\top} \boldsymbol{u}$, *then* $\frac{\mathcal{E}\left(\boldsymbol{X}-\frac{\boldsymbol{X}\boldsymbol{u}\boldsymbol{u}^{\top}\boldsymbol{X}}{n(\boldsymbol{u})^2+\boldsymbol{u}^{\top}\boldsymbol{X}\boldsymbol{u}}\right)}{\mathcal{E}(\boldsymbol{X})} \leq \frac{2}{\sqrt{1+\frac{\gamma}{\alpha^2}}}$.

证明。 我们采用与引理 5 的证明类似的方法。在第一种情况下，利用引理 4，我们得到

$$\frac{\mathcal{E}\left(\boldsymbol{X}+\frac{\boldsymbol{u}\boldsymbol{u}^{\top}}{n(\boldsymbol{u})^2}\right)}{\mathcal{E}(\boldsymbol{X})} = \frac{1+\frac{\boldsymbol{u}^{\top}(\boldsymbol{I}+\boldsymbol{X})^{-1}\boldsymbol{u}}{n(\boldsymbol{u})^2}}{\sqrt{1+\frac{\boldsymbol{u}^{\top}\boldsymbol{X}^{-1}\boldsymbol{u}}{n(\boldsymbol{u})^2}}} \leq \frac{2}{\sqrt{1+\frac{\boldsymbol{u}^{\top}\boldsymbol{X}^{-1}\boldsymbol{u}}{n(\boldsymbol{u})^2}}} \leq \frac{2}{\sqrt{1+\frac{\boldsymbol{u}^{\top}\boldsymbol{X}^{-1}\boldsymbol{u}}{\alpha^2\|\boldsymbol{u}\|^2}}} \leq \frac{2}{\sqrt{1+\frac{\gamma}{\alpha^2}}} .$$

对于第二 第二种情况下，我们利用偏心率在反演下是不变量这一事实，并且 hen ce

$$\begin{aligned}\frac{\mathcal{E}\left(\boldsymbol{X}-\frac{\boldsymbol{X}\boldsymbol{u}\boldsymbol{u}^{\top}\boldsymbol{X}}{n(\boldsymbol{u})^2+\boldsymbol{u}^{\top}\boldsymbol{X}\boldsymbol{u}}\right)}{\mathcal{E}(\boldsymbol{X})} &= \frac{\mathcal{E}\left(\boldsymbol{X}-\frac{\frac{\boldsymbol{X}\boldsymbol{u}\boldsymbol{u}^{\top}\boldsymbol{X}}{n(\boldsymbol{u})^2}}{1+\frac{\boldsymbol{u}^{\top}\boldsymbol{X}\boldsymbol{u}}{n(\boldsymbol{u})^2}}\right)}{\mathcal{E}(\boldsymbol{X})} = \frac{\mathcal{E}\left(\left(\boldsymbol{X}^{-1}+\frac{\boldsymbol{u}\boldsymbol{u}^{\top}}{n(\boldsymbol{u})^2}\right)^{-1}\right)}{\mathcal{E}(\boldsymbol{X})} = \frac{\mathcal{E}\left(\boldsymbol{X}^{-1}+\frac{\boldsymbol{u}\boldsymbol{u}^{\top}}{n(\boldsymbol{u})^2}\right)}{\mathcal{E}\left(\boldsymbol{X}^{-1}\right)} \\ &\leq \frac{2}{\sqrt{1+\frac{\gamma}{\alpha^2}}} .\end{aligned}$$

□

C.2 用于估计海森矩阵-向量乘积的引理

以下两个估计引理利用了自协调函数的性质。

引理 36。 设 $g: \ K \to \mathbb{R}$ 是一个自协调函数，设 $\boldsymbol{y} \in$ 属于 (K)，且设 $\boldsymbol{H}_{\boldsymbol{y}} := \nabla^2 g(\boldsymbol{y})$是 $\boldsymbol{y}$ 处的海森矩阵，满足 $\|\boldsymbol{H}_{\boldsymbol{y}}\| \leq B$。那么，对于任意向量 $\boldsymbol{v} \in \mathbb{R}^n$，使用对 g 的梯度预言机进行两次调用，我们可以得到一个估计值

$$\begin{aligned} n_{\boldsymbol{y}}(\boldsymbol{v}) &= \left(\frac{1}{1-\frac{1}{1000}}\right) \cdot \sqrt{\frac{1}{\tau} \cdot \langle \boldsymbol{v}, \nabla g(\boldsymbol{y}+\tau \boldsymbol{v}) - \nabla g(\boldsymbol{y}) \rangle} \ , where \\ \tau &= \frac{1}{1000\|\boldsymbol{v}\| B} \ , \end{aligned}$$

使得

$$\|\boldsymbol{v}\|_{\boldsymbol{H}_{\boldsymbol{y}}} \leq n_{\boldsymbol{y}}(\boldsymbol{v}) \leq \|\boldsymbol{v}\|_{\boldsymbol{H}_{\boldsymbol{y}}} \cdot \left(\frac{1}{1-\frac{1}{1000}}\right)^2 .$$

证明。 我们接下来分析该近似的质量。由于

$$\nabla g\left(\boldsymbol{y}+\tau \boldsymbol{v}\right)-\nabla g\left(\boldsymbol{y}\right)=\left(\int_{0}^{1}\nabla^{2}g\left(\boldsymbol{y}+t\cdot\tau \boldsymbol{v}\right)dt\right)\cdot\left(\tau \boldsymbol{v}\right)\,,$$

我们有

$$\begin{aligned}n_{\boldsymbol{y}}\left(\boldsymbol{v}\right)&=\frac{1}{1-\frac{1}{1000}}\cdot\sqrt{\frac{1}{\tau}\cdot\left\langle \boldsymbol{v},\left(\int_{0}^{1}\nabla^{2}g\left(\boldsymbol{y}+t\cdot\tau \boldsymbol{v}\right)dt\right)\tau \boldsymbol{v}\right\rangle}\\&=\frac{1}{1-\frac{1}{1000}}\cdot\left\langle \boldsymbol{v},\left(\int_{0}^{1}\nabla^{2}g\left(\boldsymbol{y}+t\cdot\tau \boldsymbol{v}\right)dt\right)\boldsymbol{v}\right\rangle^{1/2}\,.\end{aligned}$$

利用该界

$$\left\|\tau \boldsymbol{v}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\leq\tau\cdot\left\|\boldsymbol{v}\right\|\cdot\left\|\boldsymbol{H}_{\boldsymbol{y}}\right\|\leq\tau\cdot\left\|\boldsymbol{v}\right\|\cdot B=\frac{1}{1000}\,,$$

我们将自协调性应用于夹逼估计。

$$\left(1-\left\|\tau \boldsymbol{v}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\right)^{2}\cdot\boldsymbol{H}_{\boldsymbol{y}}\preceq\nabla^{2}g\left(\boldsymbol{y}+t\cdot\tau \boldsymbol{v}\right)\preceq\left(\frac{1}{1-\left\|\tau \boldsymbol{v}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}}\right)^{2}\cdot\boldsymbol{H}_{\boldsymbol{y}}\,,$$

这也意味着我们的查询点是可行的，因为沿着 y 到 y $+\tau$v 的整条路径，海森矩阵始终有界。

$$\left(1-\frac{1}{1000}\right)^{2}\cdot\boldsymbol{H}_{\boldsymbol{y}}\preceq\nabla^{2}g\left(\boldsymbol{y}+t\cdot\tau \boldsymbol{v}\right)\preceq\left(\frac{1}{1-\frac{1}{1000}}\right)^{2}\cdot\boldsymbol{H}_{\boldsymbol{y}}\,.$$

将其代回我们的 $n_{\boldsymbol{y}}$ (v) 恒等式中，可得

$$\frac{1}{1-\frac{1}{1000}}\cdot\left(1-\frac{1}{1000}\right)\cdot\left\|\boldsymbol{v}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\leq n_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\leq\left(\frac{1}{1-\frac{1}{1000}}\right)^{2}\cdot\left\|\boldsymbol{v}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\,,$$

从而完成证明。 □

引理 37. 设 $g\colon\ \mathbb{R}^{n}\to\mathbb{R}$ 是一个自协调函数，设 $\boldsymbol{y}\in\mathbb{R}^{n}$，并设 $\boldsymbol{H}_{\boldsymbol{y}}:=\nabla^{2}g(\boldsymbol{y})$ 是 $\boldsymbol{y}$ 处的海森矩阵，且满足 $\|\boldsymbol{H}_{\boldsymbol{y}}\|\leq B$，其中 $B\geq1$。那么，对于任意向量 $\boldsymbol{v}\in\mathbb{R}^{n}$，我们可以通过对 g 调用两次梯度预言机，获得海森矩阵与向量的乘积的估计：

$$\begin{aligned}p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)&=\frac{1}{\tau}\left(\nabla g\left(\boldsymbol{y}+\tau \boldsymbol{v}\right)-\nabla g\left(\boldsymbol{y}\right)\right)\ ,where\\\tau&=\frac{1}{1000\left\|\boldsymbol{v}\right\|B^{21}}\ ,\end{aligned}$$

使得

$$\left\|p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}\leq\frac{1}{400\cdot B^{20}}\left\|\boldsymbol{v}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\ .$$

证明。 我们继续分析误差。出于与引理 36 的证明中相同的原因，查询点是可行的。令 $\mathrm{H}\,\mathrm{H}\,\mathrm{y}_{+t\tau}$ $_{v}dt$，我们可以写成 $^{p_{y}}(\mathrm{v}\,)\;=\;\frac{1}{\tau}\mathrm{H}\;_{\tau}\cdot(\tau\mathrm{v})\;=\mathrm{H}\,_{\tau}\mathrm{v}$。因此，利用自协调性，我们得到界：

$$\begin{aligned}\left\|p_{\boldsymbol{y}}(\boldsymbol{v})-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}&=\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\left(\overline{\boldsymbol{H}}_{\tau}-\boldsymbol{H}_{\boldsymbol{y}}\right)\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\cdot\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\boldsymbol{v}\right\|_{2}\leq\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\left(\overline{\boldsymbol{H}}_{\tau}-\boldsymbol{H}_{\boldsymbol{y}}\right)\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right\|\|\boldsymbol{v}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\\&\leq\left(\left(\frac{1}{1-\tau\|\boldsymbol{v}\|_{\boldsymbol{H}_{\boldsymbol{y}}}}\right)^{2}-1\right)\|\boldsymbol{v}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\leq\left(\left(\frac{1}{1-\tau\cdot\|\boldsymbol{H}_{\boldsymbol{y}}\|\,\|\boldsymbol{v}\|}\right)^{2}-1\right)\|\boldsymbol{v}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\\&\leq\left(\left(\frac{1}{1-\frac{1}{1000\cdot B^{20}}}\right)^{2}-1\right)\|\boldsymbol{v}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\leq\frac{1}{400\cdot B^{20}}\|\boldsymbol{v}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\,.\end{aligned}$$

□

C.3 引理的证明 18

证明。 我们通过将使用估计量 p_{y} 和 n_{y} 得到的残差与使用对矩阵 H_{y} 的精确访问所得到的残差进行比较，来执行适当的误差分析。首先，我们利用以下事实，即

$$\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\leq n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\leq\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\cdot\left(\frac{1}{1-\frac{1}{1000}}\right)^{2}$$

and

$$\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}}=\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}+\left(\boldsymbol{H}_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)-p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}},$$

我们给出上界

$$\begin{aligned}&\left\|\boldsymbol{H}_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)-p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}}\qquad(10)\\&=\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\cdot\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\left(\boldsymbol{H}_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)-p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right)\right\|_{2}\\&\leq\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\right\|\left\|\boldsymbol{H}_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)-p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}\\&\leq\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\right\|\cdot\frac{1}{400\cdot B^{20}}\cdot\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\\&\leq\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\right\|\cdot\frac{1}{400\cdot B^{20}}\cdot\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\widetilde{\boldsymbol{H}}^{1/2}\right\|\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\cdot\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{2}\\&=\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\right\|\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\widetilde{\boldsymbol{H}}^{1/2}\right\|\cdot\frac{1}{400\cdot B^{20}}\cdot\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}\\&\leq\frac{1}{400\cdot B^{18}}\cdot\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}},\qquad(11)\end{aligned}$$

由此可得

$$\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}\cdot\left(1-\frac{1}{400B^{18}}\right)\leq\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}}\leq\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}\cdot\left(1+\frac{1}{400B^{18}}\right).\quad(12)$$

因此，我们可以夹住

$$\frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}}{\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}}\cdot\frac{1}{\left(1+\frac{1}{400B^{18}}\right)}\leq\frac{n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)}{\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}}}\leq\frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}}{\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}}\cdot\frac{\left(\frac{1}{1-\frac{1}{1000}}\right)^{2}}{\left(1-\frac{1}{400B^{18}}\right)}. \tag{13}$$

此外，我们可以测量残差误差。给定任意向量 v，我们有 e

$$\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\|_{\widetilde{\boldsymbol{H}}^{-1}}=\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}+\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\|_{\widetilde{\boldsymbol{H}}^{-1}},$$

并且我们得到上界

$$\begin{aligned}
\|\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\|_{\widetilde{\boldsymbol{H}}^{-1}} &= \left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\cdot\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\left(\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\right)\right\|_{2}\\
&\leq \left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\right\|\cdot\|\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}\\
&\leq \left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\right\|\cdot\frac{1}{400\cdot B^{20}}\cdot\|\boldsymbol{v}\|_{\boldsymbol{H}_{\boldsymbol{y}}}\\
&\leq \left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\right\|\cdot\frac{1}{400\cdot B^{20}}\cdot\left\|\widetilde{\boldsymbol{H}}^{1/2}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right\|\left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\cdot\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\boldsymbol{v}\right\|_{2}\\
&= \left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\right\|\left\|\widetilde{\boldsymbol{H}}^{1/2}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right\|\cdot\frac{1}{400\cdot B^{20}}\cdot\|\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}\\
&\leq \frac{1}{400\cdot B^{18}}\cdot\|\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}\\
&\leq \frac{1}{400\cdot B^{18}}\cdot\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}+\frac{1}{400\cdot B^{18}}\|\boldsymbol{b}\|_{\widetilde{\boldsymbol{H}}^{-1}}\\
&\leq \frac{1}{400\cdot B^{18}}\cdot\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}+\frac{1}{400\cdot B^{16}}\cdot,
\end{aligned} \tag{14}$$

由此我们得出

$$\begin{aligned}
\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}\left(1-\frac{1}{400\cdot B^{18}}\right)-\frac{1}{400\cdot B^{16}}&\leq\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\|_{\widetilde{\boldsymbol{H}}^{-1}}\\
&\leq\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}\left(1+\frac{1}{400\cdot B^{18}}\right)+\frac{1}{400\cdot B^{16}}.
\end{aligned} \tag{15}$$

因此，只要

$$\frac{1}{B}\leq\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\|_{\widetilde{\boldsymbol{H}}^{-1}},$$

正如.....中所保证的 该陈述中，我们还额外有 $\frac{1}{400\cdot B^{16}}\leq\frac{1}{400\cdot B^{15}}\cdot\|\mathfrak{b}-p_{\boldsymbol{y}}(\mathrm{v})\|_{\widetilde{\boldsymbol{H}}^{-1}}$，因此 s

$$\begin{aligned}
&\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}\left(1-\frac{1}{400\cdot B^{18}}\right)\left(1+\frac{1}{400\cdot B^{15}}\right)^{-1}\\
&\leq\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\|_{\widetilde{\boldsymbol{H}}^{-1}}\leq\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}\left(1+\frac{1}{400\cdot B^{18}}\right)\left(1+\frac{1}{400\cdot B^{15}}\right),
\end{aligned}$$

这意味着一个更简单的条件：

$$\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}\left(1-\frac{1}{200\cdot B^{15}}\right)\leq\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\|_{\widetilde{\boldsymbol{H}}^{-1}}\leq\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\|_{\widetilde{\boldsymbol{H}}^{-1}}\left(1+\frac{1}{200\cdot B^{15}}\right). \tag{16}$$

通过类似的分析，我们可以证明

$$\begin{aligned}
&\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}}\left(1-\frac{1}{200\cdot B^{9}}\right)\\
&\leq\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}}\leq\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}}\left(1+\frac{1}{200\cdot B^{9}}\right), \qquad (17)\\
&\left\|\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\left(1-\frac{1}{200\cdot B^{9}}\right)\\
&\leq\left\|\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{v}\right)\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\leq\left\|\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{v}\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\left(1+\frac{1}{200\cdot B^{9}}\right). \qquad (18)
\end{aligned}$$

现在我们考虑新的残差 $\|\mathbf{r}'\|^2_{\widetilde{\boldsymbol{H}}^{-1}}$。根据假设中所述，若 $\|\mathbf{r}'\|^2_{\widetilde{\boldsymbol{H}}^{-1}} \geq \frac{1}{B}$，则可以对其进行如下界定：

$$\begin{aligned}
&\left\|\boldsymbol{r}'\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}=\left\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}'\right)\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}\leq\left\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}'\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}\left(1+\frac{1}{200\cdot B^{15}}\right)^2\\
&=\left\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}-\boldsymbol{H}_{\boldsymbol{y}}\left(\frac{n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2}{\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}}\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right)\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}\cdot\left(1+\frac{1}{200\cdot B^{15}}\right)^2\\
&=\left(\left\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}+\frac{n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^4}{\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|^4_{\widetilde{\boldsymbol{H}}^{-1}}}\cdot\left\|\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}\right.\\
&\left.-2\cdot\frac{n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2}{\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|^2_{\widetilde{\boldsymbol{H}}^{-1}}}\left\langle\widetilde{\boldsymbol{H}}^{-1/2}\left(\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right),\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\rangle\right)\cdot\left(1+\frac{1}{200\cdot B^{15}}\right)^2.
\end{aligned}$$

利用 $\|\mathbf{b}-p_{\boldsymbol{y}}(\mathbf{x})\|_{\widetilde{\boldsymbol{H}}^{-1}} \geq 1/B$，我们给出下界：

$$
\begin{aligned}
&\left\langle \widetilde{\boldsymbol{H}}^{-1/2}\left(\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right), \widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\rangle \\
&= \left\|\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}^{2} - \left\langle \widetilde{\boldsymbol{H}}^{-1/2}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right), \widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1/2}\cdot\widetilde{\boldsymbol{H}}^{-1/2}\left(\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\rangle \\
&\geq \left\|\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}^{2} - \left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1/2}\right\| \left\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \left\|p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \\
&\geq \left\|\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}^{2} \\
&\quad - \left\|\widetilde{\boldsymbol{H}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1/2}\right\| \left\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \left(\frac{1}{400\cdot B^{18}}\cdot\left\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}+\frac{1}{400\cdot B^{16}}\right) \\
&\geq \left\|\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}^{2} - \frac{1}{400\cdot B^{16}}\left\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}}\left\|\boldsymbol{b}-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \\
&\quad - \frac{1}{400\cdot B^{14}}\left\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \\
&\geq \left\|\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}^{2} - \frac{1}{200\cdot B^{15}}\left\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^{2} - \frac{1}{400\cdot B^{14}}\left\|\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right\|_{\widetilde{\boldsymbol{H}}^{-1}} \\
&\geq \left\|\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}^{2}\left(1-\frac{1}{200\cdot B^{15}}-\frac{1}{400\cdot B^{14}}\right) \\
&\geq \left\|\widetilde{\boldsymbol{H}}^{-1}\left(\boldsymbol{b}-p_{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}^{2}\left(1-\frac{1}{200\cdot B^{14}}\right). \qquad (19)
\end{aligned}
$$

回到关于新残差的不等式，利用（16），（13），（19）我们得到：

$$
\begin{aligned}
&\left\|\boldsymbol{r}'\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2 \\
&\leq \|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2 \cdot \left(1+\frac{1}{200\cdot B^{15}}\right)^4 \\
&+ \frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y}^4}{\left\|\boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^4}\left\|\boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2 \cdot \frac{\left(\frac{1}{1-\frac{1}{1000}}\right)^8}{\left(1-\frac{1}{400B^{18}}\right)^4}\cdot\left(1+\frac{1}{200\cdot B^{15}}\right)^2 \\
&- 2\cdot\frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y}^2}{\left\|\boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y}^2\left(1-\frac{1}{200\cdot B^{14}}\right)^2\cdot\frac{1}{\left(1+\frac{1}{400B^{18}}\right)}\left(1+\frac{1}{200\cdot B^{15}}\right)^2 \\
&\leq \|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2 \cdot \left(1+\frac{1}{200\cdot B^{15}}\right)^4 \\
&- \frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y}^4}{\left\|\boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2} \\
&\cdot\left(2\left(1-\frac{1}{200\cdot B^{14}}\right)^2\cdot\frac{1}{\left(1+\frac{1}{400B^{18}}\right)}\left(1+\frac{1}{200\cdot B^{15}}\right)^2-\frac{\left(\frac{1}{1-\frac{1}{1000}}\right)^8}{\left(1-\frac{1}{400B^{18}}\right)^4}\cdot\left(1+\frac{1}{200\cdot B^{15}}\right)^2\right) \\
&\leq \|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2 \cdot \left(1+\frac{1}{200\cdot B^{15}}\right)^4 - \frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y}^4}{\left\|\boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\cdot\frac{9}{10}.
\end{aligned}
$$

因此，失败条件

$$
\frac{\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2-\|\boldsymbol{r}'\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}{\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\leq\beta
$$

这意味着

$$
1-\left(1+\frac{1}{200\cdot B^{15}}\right)^4+\frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y}^4}{\left\|\boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\cdot\frac{9}{10}\leq 1-\frac{\|\boldsymbol{r}'\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}{\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\leq\beta
$$

因此

$$
\frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y}^4}{\left\|\boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\leq\frac{10}{9}\left(\beta+\left(1+\frac{1}{200\cdot B^{15}}\right)^4-1\right)
$$

$$
\frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y}^4}{\left\|\boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\widetilde{\boldsymbol{H}}^{-1}}^2\|\boldsymbol{r}\|_{\widetilde{\boldsymbol{H}}^{-1}}^2}\leq\frac{20}{9}\beta,
$$

这与引理 15 一样表明 $\boldsymbol{r}$ 是一个偏心率证书，并完成证明。 □

C.4 引理的证明19

证明。 只需证明，算法 3 执行的秩一更新仍足以降低偏心率。为此，我们分析两类证书。

类型 1. 令 $\mathrm{X} = \mathrm{H}^{-1/2}{}_{y}\widetilde{\ }\mathrm{H}\,\mathrm{H}^{-1/2}{}_{y}$ 且 $\mathrm{u} = \mathrm{H}^{1/2}{}_{y}\widetilde{\ }\mathrm{H}_{-1}\mathrm{r}$，我们有 $\mathrm{u}^{\top}\mathrm{X}\,\mathrm{u} \geq \frac{1}{\sqrt{\frac{20}{9}\beta}}\mathrm{u}^{\top}\mathrm{u}$。由于该证书要求我们进行的更新，为了应用引理 5，需要精确获取范数 $\left\|\widetilde{\ }\mathrm{H}_{-1}\mathrm{r}\right\|_{\boldsymbol{H}^{y}}$，而我们无法获取，因此我们改用一个鲁棒版本（附录 C.1中的引理 35）。由于 $\|\mathrm{u}\|_2 = \left\|\widetilde{\ }\mathrm{H}_{-1}\mathrm{r}\right\|_{\boldsymbol{H}^{y}}$，根据引理 36，有 $\|\mathrm{u}\|_2 \leq n_y\left(\widetilde{\ }\mathrm{H}_{-1}\mathrm{r}\right) \leq \|\mathrm{u}\|_2 \cdot \left(\frac{1}{1-\frac{1}{1000}}\right)^2$，因此利用引理 35，我们得到，将预条件器更改为 $\widetilde{\ }\mathrm{H}\prime$，使得

$$
\begin{aligned}
&\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1/2} \\
&= \boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2} - \frac{\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}\left(\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\left(\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^{\top}\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}}{n_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2 + \left\|\boldsymbol{H}^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|^2_{\boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2}}} \\
&= \boldsymbol{H}^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1/2} - \frac{\boldsymbol{H}^{-1/2}\boldsymbol{r}\boldsymbol{r}^{\top}\boldsymbol{H}^{-1/2}}{n_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2 + \|\boldsymbol{r}\|^2_{\widetilde{\boldsymbol{H}}^{-1}}},
\end{aligned}
$$

等价地，令

$$
\widetilde{\boldsymbol{H}}' = \widetilde{\boldsymbol{H}} - \frac{\boldsymbol{r}\boldsymbol{r}^{\top}}{n_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^2 + \|\boldsymbol{r}\|^2_{\widetilde{\boldsymbol{H}}^{-1}}},
$$

我们得到

$$
\mathcal{E}\left(\widetilde{\boldsymbol{H}}'\boldsymbol{H}^{-1}\right) \leq \mathcal{E}\left(\widetilde{\boldsymbol{H}}\boldsymbol{H}^{-1}\right) \cdot \frac{2}{\sqrt{1 + \frac{1}{\left(\frac{1}{1-\frac{1}{1000}}\right)^4\sqrt{\frac{20}{9}\beta}}}} \leq \frac{2}{\sqrt{1 + \frac{99}{100}\cdot\frac{1}{\sqrt{\frac{20}{9}\beta}}}}.
$$

类型 2。令 $\mathrm{X} = \mathrm{H}^{-1/2}\mathrm{y}\widetilde{\ }\mathrm{H}\mathrm{H}^{-1/2}\mathrm{y}$ 且 $\mathrm{u} = \mathrm{H}^{1/2}\mathrm{y}\widetilde{\ }\mathrm{H}_{-1}\mathrm{r}$，我们有 $\mathrm{u}^{\top}\mathrm{X}^{-1}\mathrm{u} \geq \frac{1}{\sqrt{\frac{20}{9}\beta}}\mathrm{u}^{\top}\mathrm{u}$。由于该证书所要求的更新将需要我们为了应用引理5，精确获取范数 $\left\|\widetilde{\ }\mathrm{H}_{-1}\mathrm{r}\right\|_{\boldsymbol{H}}\mathrm{y}$ 和向量 $\mathrm{H}\widetilde{\ }\mathrm{H}_{-1}\mathrm{r}$，而这是我们无法获得的。使用引理37，我们得到一个满足以下条件的近似值 $p_y\left(\widetilde{\ }\mathrm{H}_{-1}\mathrm{r}\right)$：

$$
\left\|p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right) - \boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y^{-1}} \leq \frac{1}{400\cdot B^{20}}\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y} = \frac{1}{400\cdot B^{20}}\|\boldsymbol{u}\|_2 .
$$

因此令 $\mathrm{u}' = \mathrm{H}^{-1/2}{}_y p_y\left(\widetilde{\ }\mathrm{H}_{-1}\mathrm{r}\right)$，我们有 $\mathrm{u}\prime^{\top}\mathrm{X}^{-1}\mathrm{u}' \geq \left(\frac{1-\frac{1}{100}}{1+\frac{1}{100}}\right)^2 \cdot \mathrm{u}'^{\top}\mathrm{u}'$.这是因为

$$
\begin{aligned}
\frac{\boldsymbol{u}'^{\top}\boldsymbol{X}^{-1}\boldsymbol{u}'}{\boldsymbol{u}'^{\top}\boldsymbol{u}'} &= \frac{\|\boldsymbol{u}'\|_{\boldsymbol{X}^{-1}}^2}{\|\boldsymbol{u}'\|_2^2} = \frac{\|\boldsymbol{u}'-\boldsymbol{u}+\boldsymbol{u}\|_{\boldsymbol{X}^{-1}}^2}{\|\boldsymbol{u}'-\boldsymbol{u}+\boldsymbol{u}\|_2^2} \geq \frac{\left(\|\boldsymbol{u}\|_{\boldsymbol{X}^{-1}} - \|\boldsymbol{u}'-\boldsymbol{u}\|_{\boldsymbol{X}^{-1}}\right)^2}{\left(\|\boldsymbol{u}\|_2 + \|\boldsymbol{u}'-\boldsymbol{u}\|_2\right)^2} \\
&= \frac{\left(\|\boldsymbol{u}\|_{\boldsymbol{X}^{-1}} - \left\|\boldsymbol{H}_y^{-1/2}p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right) - \boldsymbol{H}_y^{-1/2}\boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{X}^{-1}}\right)^2}{\left(\|\boldsymbol{u}\|_2 + \left\|\boldsymbol{H}_y^{-1/2}p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right) - \boldsymbol{H}_y^{-1/2}\boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2\right)^2} \\
&\geq \frac{\left(\|\boldsymbol{u}\|_{\boldsymbol{X}^{-1}} - \left\|\boldsymbol{X}^{-1}\right\|\left\|p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right) - \boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y^{-1}}\right)^2}{\left(\|\boldsymbol{u}\|_2 + \left\|p_y\left(\widetilde{\boldsymbol{H}}^{-1}\nabla g\left(\boldsymbol{y}\right)\right) - \boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y^{-1}}\right)^2} \\
&\geq \frac{\left(\|\boldsymbol{u}\|_{\boldsymbol{X}^{-1}} - \frac{1}{400\cdot B^{20}}\|\boldsymbol{u}\|_2\right)^2}{\left(\|\boldsymbol{u}\|_2 + \frac{1}{400\cdot B^{20}}\|\boldsymbol{u}\|_2\right)^2} \\
&\geq \frac{\left(\|\boldsymbol{u}\|_{\boldsymbol{X}^{-1}} - B^2\cdot\frac{1}{400\cdot B^{20}}\|\boldsymbol{u}\|_{\boldsymbol{X}^{-1}}\right)^2}{\left(\|\boldsymbol{u}\|_2 + \frac{1}{400\cdot B^{20}}\|\boldsymbol{u}\|_2\right)^2} \\
&\geq \left(1-\frac{1}{B^{17}}\right)\cdot\frac{\|\boldsymbol{u}\|_{\boldsymbol{X}^{-1}}^2}{\|\boldsymbol{u}\|_2^2} \geq \left(1-\frac{1}{B^{17}}\right)\cdot\frac{1}{\sqrt{\frac{20}{9}\beta}}\cdot
\end{aligned}
$$

此外，根据引理36，我们得到一个近似值 $\|\mathrm{u}\|_2 \leq n_y(\mathrm{u}) \leq \left(\frac{1}{1-\frac{1}{1000}}\right)^2\|\mathrm{u}\|_2$。现在我们证明 n_y (u) 是 $\|\mathrm{u}\prime\|_2$ 的良好近似。事实上，沿用之前的方法，我们有

$$
\left|\left\|\boldsymbol{u}'\right\|_2 - \|\boldsymbol{u}\|_2\right| \leq \left\|\boldsymbol{u}'-\boldsymbol{u}\right\|_2 \leq \left\|p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right) - \boldsymbol{H}_y\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y^{-1}} \leq \frac{1}{400\cdot B^{20}}\|\boldsymbol{u}\|_2\ ,
$$

因此，

$$
\|\boldsymbol{u}\|_2\left(1-\frac{1}{400\cdot B^{20}}\right) \leq \left\|\boldsymbol{u}'\right\|_2 \leq \|\boldsymbol{u}\|_2\left(1+\frac{1}{400\cdot B^{20}}\right)\ ,
$$

由此可得：

$$
\begin{aligned}
\left\|\boldsymbol{u}'\right\|_2 \leq \left(1+\frac{1}{400\cdot B^{20}}\right)n_y\left(\boldsymbol{u}\right) &\leq \left(\frac{1}{1-\frac{1}{1000}}\right)^2\cdot\frac{1+\frac{1}{400\cdot B^{20}}}{1-\frac{1}{400\cdot B^{20}}}\cdot\left\|\boldsymbol{u}'\right\|_2 \\
&\leq \left(\frac{1}{1-\frac{1}{1000}}\right)^2\cdot\left(1+\frac{1}{100\cdot B^{20}}\right)\cdot\left\|\boldsymbol{u}'\right\|_2\ .
\end{aligned}
$$

~ ~因此，根据引理 35的第一种情况，将 $\boldsymbol{H}$ 更新为 $\boldsymbol{H}'$ 使得

$$
\boldsymbol{X}' := \boldsymbol{H}_y^{-1/2}\widetilde{\boldsymbol{H}}'\boldsymbol{H}_y^{-1/2} = \boldsymbol{H}_y^{-1/2}\widetilde{\boldsymbol{H}}\boldsymbol{H}_y^{-1/2} + \frac{\boldsymbol{H}_y^{-1/2}p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)p_y\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^{\top}\boldsymbol{H}_y^{-1/2}}{\left(1+\frac{1}{400\cdot B^{20}}\right)^2\cdot n_y\left(\widetilde{\boldsymbol{H}}^{-1}\nabla g\left(\boldsymbol{y}\right)\right)^2}
$$

并等价地设定

$$\widetilde{\boldsymbol{H}}'=\widetilde{\boldsymbol{H}}+\frac{p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^{\top}}{\left(1+\frac{1}{400\cdot B^{20}}\right)^{2}n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\nabla g\left(\boldsymbol{y}\right)\right)^{2}}$$

改善偏心率。其含义是，

$$\frac{\mathcal{E}\left(\boldsymbol{X}'\right)}{\mathcal{E}\left(\boldsymbol{X}\right)}\leq\frac{2}{\sqrt{1+\left(\frac{1}{\left(\frac{1}{1-\frac{1}{1000}}\right)^{2}\cdot\left(1+\frac{1}{100\cdot B^{20}}\right)}\right)^{2}\frac{1}{\sqrt{\frac{20}{9}\beta}}}}\leq\frac{2}{\sqrt{1+\frac{99}{100}\cdot\frac{1}{\sqrt{\frac{20}{9}\beta}}}}.$$

再次注意到，实现此更新所需的时间主要用于进行常数次梯度查询，以及执行涉及 $\widetilde{\mathrm{H}}$和 $\widetilde{\mathrm{H}^{-1}}$ 的常数次矩阵-向量乘法，这需要 $O\left(n^{2}\right)$ 时间。

预条件器范数的变化。 为了证明的目的，我们只需要对 $\left\|\widetilde{\mathrm{H}}\right\|$ 或 $\left\|\widetilde{\mathrm{H}^{-1}}\right\|$所经历的范数增幅提供稍弱的保证。每当预条件器范数增加时，都是由于类型 2 证书，在这种情况下我们有：

$$\widetilde{\boldsymbol{H}}'=\widetilde{\boldsymbol{H}}+\frac{p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^{\top}}{\left(1+\frac{1}{400\cdot B^{20}}\right)^{2}n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^{2}},$$

so

$$\left\|\widetilde{\boldsymbol{H}}'-\widetilde{\boldsymbol{H}}\right\|\leq\frac{\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|_{2}^{2}}{n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^{2}}$$

并利用引理 37，我们知道

$$\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)-\boldsymbol{H}_{\boldsymbol{y}}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}\leq\frac{1}{400\cdot B^{20}}\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}},$$

$$\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}\leq\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\left(1+\frac{1}{400\cdot B^{20}}\right),$$

$$\left\|p_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)\right\|_{2}\leq\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1}\right\|\cdot\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\left(1+\frac{1}{400\cdot B^{20}}\right),$$

此外，根据引理 36，$\left\|\widetilde{\mathrm{H}}^{-1}\mathbf{r}\right\|_{\boldsymbol{H}^{y}}\leq n_{\boldsymbol{y}}\left(\widetilde{\mathrm{H}}^{-1}\mathbf{r}\right)\leq\|\mathbf{u}\|_{2}\cdot\left(\frac{1}{1-\frac{1}{1000}}\right)^{2}$，我们得出

$$\left\|\widetilde{\boldsymbol{H}}'-\widetilde{\boldsymbol{H}}\right\|\leq\frac{\left(\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1}\right\|\cdot\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\left(1+\frac{1}{400\cdot B^{20}}\right)\right)^{2}}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}^{2}}=\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1}\right\|^{2}\left(1+\frac{1}{400\cdot B^{20}}\right)^{2}.$$

同样地，在 $\widetilde{\mathrm{H}'}^{-1}$ 增加的情况下，这是由于存在一个 1 型证书，此时我们有：

$$\widetilde{\boldsymbol{H}'}^{-1}=\widetilde{\boldsymbol{H}}^{-1}+\frac{\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\boldsymbol{r}^{\top}\widetilde{\boldsymbol{H}}^{-1}}{n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^{2}},$$

因此

$$\left\|\widetilde{\boldsymbol{H}'}^{-1}-\widetilde{\boldsymbol{H}}^{-1}\right\|\leq\frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2^2}{n_{\boldsymbol{y}}\left(\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right)^{2}}\leq\frac{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2^2}{\left\|\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y}^2}=\frac{\left\|\boldsymbol{H}_y^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_{\boldsymbol{H}_y^{-1}}^2}{\left\|\boldsymbol{H}_y^{1/2}\widetilde{\boldsymbol{H}}^{-1}\boldsymbol{r}\right\|_2^2}\leq\left\|\boldsymbol{H}_y^{-1}\right\|.$$

□

D 第 5 节的证明

引理 38. 设 g_μ 为自协调的。假设 $\|\nabla g_\mu(\boldsymbol{y})\|_{\boldsymbol{H}^{-1}\boldsymbol{y}}\leq\ 1/20$ 且 $\|\nabla g_\mu(\boldsymbol{y})-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\|_{\boldsymbol{H}^{-1}\boldsymbol{y}}\leq\varepsilon\leq 1/20$。则令 $\boldsymbol{y}'=\boldsymbol{y}-\boldsymbol{x}$,

$$\left\|\nabla g_\mu\left(\boldsymbol{y}'\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}'}}\leq 2\varepsilon+7\cdot\left\|\nabla g_\mu\left(\boldsymbol{y}\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}^2.$$

此外，

$$\left\|\boldsymbol{y}'-\boldsymbol{y}\right\|_{\boldsymbol{H}_{\boldsymbol{y}}}\leq\varepsilon+\left\|\nabla g_\mu\left(\boldsymbol{y}\right)\right\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}.$$

证明。令 $\mathrm{H}=\int_0^1\mathrm{H}_{\boldsymbol{y}+(\boldsymbol{y}'-\boldsymbol{y})t}dt$，由自协调性可知

$$\max\left\{\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\boldsymbol{H}_{\boldsymbol{y}'}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right\|,\left\|\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\overline{\boldsymbol{H}}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right\|\right\}\leq\left(\frac{1}{1-\|\boldsymbol{y}'-\boldsymbol{y}\|_{\boldsymbol{H}_{\boldsymbol{y}}}}\right)^2=\left(\frac{1}{1-\|\boldsymbol{x}\|_{\boldsymbol{H}_{\boldsymbol{y}}}}\right)^2.$$

此外，我们记作

$$\|\boldsymbol{x}\|_{\boldsymbol{H}_{\boldsymbol{y}}}=\|\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}\leq\|\nabla g_\mu\left(\boldsymbol{y}\right)-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}+\|\nabla g_\mu\left(\boldsymbol{y}\right)\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}\leq\varepsilon+\|\nabla g_\mu\left(\boldsymbol{y}\right)\|_{\boldsymbol{H}_{\boldsymbol{y}}^{-1}}\leq\frac{1}{10}.\quad(20)$$

利用这些结果，我们可以将新的梯度界定为：

$$\begin{aligned}
\left\|\nabla g_\mu\left(\boldsymbol{y}'\right)\right\|_{\boldsymbol{H}_y^{-1}} &= \left\|\nabla g_\mu\left(\boldsymbol{y}\right)+\overline{\boldsymbol{H}}\left(\boldsymbol{y}'-\boldsymbol{y}\right)dt\right\|_{\boldsymbol{H}_y^{-1}} \\
&= \left\|\nabla g_\mu\left(\boldsymbol{y}\right)-\overline{\boldsymbol{H}}\boldsymbol{x}\right\|_{\boldsymbol{H}_y^{-1}} \\
&= \left\|\nabla g_\mu\left(\boldsymbol{y}\right)-\boldsymbol{H}_{\boldsymbol{y}}\boldsymbol{x}+\left(\boldsymbol{H}_{\boldsymbol{y}}-\overline{\boldsymbol{H}}\right)\boldsymbol{x}\right\|_{\boldsymbol{H}_y^{-1}} \\
&\leq \varepsilon+\left\|\left(\boldsymbol{H}_{\boldsymbol{y}}-\overline{\boldsymbol{H}}\right)\boldsymbol{x}\right\|_{\boldsymbol{H}_y^{-1}} \\
&= \varepsilon+\left\|\left(\boldsymbol{I}-\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\overline{\boldsymbol{H}}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right)\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\boldsymbol{x}\right\|_2 \\
&\leq \varepsilon+\left\|\boldsymbol{I}-\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\overline{\boldsymbol{H}}\boldsymbol{H}_{\boldsymbol{y}}^{-1/2}\right\|\left\|\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\boldsymbol{x}\right\|_2 \\
&\leq \varepsilon+\left(1-\left(\frac{1}{1-\|\boldsymbol{x}\|_{\boldsymbol{H}_y}}\right)^2\right)\|\boldsymbol{x}\|_{\boldsymbol{H}_y} \\
&\leq \varepsilon+\frac{2\|\boldsymbol{x}\|_{\boldsymbol{H}_y}^2}{\left(1-\|\boldsymbol{x}\|_{\boldsymbol{H}_y}\right)^2} \\
&\leq \varepsilon+\left(\frac{10}{9}\right)^2\cdot 2\|\boldsymbol{x}\|_{\boldsymbol{H}_y}^2 .
\end{aligned}$$

利用（20），我们得到

$$\begin{aligned}
\left\|\nabla g_\mu\left(\boldsymbol{y}'\right)\right\|_{\boldsymbol{H}_y^{-1}} &\leq \varepsilon+\left(\frac{10}{9}\right)^2\cdot 2\left(\varepsilon+\|\nabla g_\mu\left(\boldsymbol{y}\right)\|_{\boldsymbol{H}_y^{-1}}\right)^2 \\
&\leq \varepsilon+\left(\frac{10}{9}\right)^2\cdot 4\left(\varepsilon^2+\|\nabla g_\mu\left(\boldsymbol{y}\right)\|_{\boldsymbol{H}_y^{-1}}^2\right) \\
&\leq \varepsilon\left(1+\left(\frac{10}{9}\right)^2\cdot\frac{1}{5}\right)+\left(\frac{10}{9}\right)^2\cdot 4\cdot\|\nabla g_\mu\left(\boldsymbol{y}\right)\|_{\boldsymbol{H}_y^{-1}}^2 \\
&\leq \varepsilon\left(1+\frac{1}{4}\right)+5\cdot\|\nabla g_\mu\left(\boldsymbol{y}\right)\|_{\boldsymbol{H}_y^{-1}}^2 .
\end{aligned}$$

最后，改变范数后，我们有

$$\begin{aligned}
\left\|\nabla g_\mu\left(\boldsymbol{y}'\right)\right\|_{\boldsymbol{H}_{y'}^{-1}} &\leq \left\|\nabla g_\mu\left(\boldsymbol{y}'\right)\right\|_{\boldsymbol{H}_y^{-1}}\cdot\left\|\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\boldsymbol{H}_{\boldsymbol{y}'}^{-1}\boldsymbol{H}_{\boldsymbol{y}}^{1/2}\right\| \\
&\leq \left\|\nabla g_\mu\left(\boldsymbol{y}'\right)\right\|_{\boldsymbol{H}_y^{-1}}\cdot\left(\frac{1}{1-\|\boldsymbol{x}\|_{\boldsymbol{H}_y}}\right)^2 \\
&\leq \left(\varepsilon\left(1+\frac{1}{4}\right)+5\cdot\|\nabla g_\mu\left(\boldsymbol{y}\right)\|_{\boldsymbol{H}_y^{-1}}^2\right)\cdot\frac{100}{81} \\
&\leq 2\varepsilon+7\cdot\|\nabla g_\mu\left(\boldsymbol{y}\right)\|_{\boldsymbol{H}_y^{-1}}^2 .
\end{aligned}$$

□

引理 39. 设 $g_\mu(\mathrm{y}) = \frac{\langle c,\ \rangle}{\mu} + \phi(\mathrm{y})$。则 $\nabla g_{\mu/(1-\delta)}(\mathrm{y}) = (1+\delta)\nabla g_\mu(\mathrm{y}) - \delta\nabla\phi(\mathrm{y})$。

证明。 我们如下证明恒等式：

$$\begin{aligned}\nabla g_{\mu(1-\delta)}(\boldsymbol{y}) &= \frac{\boldsymbol{c}}{\mu/(1+\delta)} + \nabla\phi(\boldsymbol{y}) = (1+\delta)\left(\frac{\boldsymbol{c}}{\mu} + \nabla\phi(\boldsymbol{y})\right) - \delta\nabla\phi(\boldsymbol{y}) \\ &= (1+\delta)\nabla g_{\mu}(\boldsymbol{y}) - \delta\nabla\phi(\boldsymbol{y}) .\end{aligned}$$

□

参考文献

[And94] Tsuyoshi Ando.矩阵理论中的主序与不等式。《线性代数及其应用》, 199:17–67, 1994.27[Ans00] Kurt MAnstreicher. 半定规划的体积障碍函数。《运筹学数学》,25(3):365–380, 2000. 8[AV95] David SAtkinson andPravin MVaidya.一种使用解析中心的凸规划割平面算法。《数学规划》,69(1-3):1–43, 1995. 8[BBE+22] J. Blikstad, J. Van Den Brand, Y. Efron, S. Mukhopadhyay, andD.Nanongkai. 二分图匹配近乎最优的通信与查询复杂度。载于2022年IEEE第63届计算机科学基础年会（FOCS），第1174–1185页，美国加利福尼亚州洛斯阿拉米托斯, 2022年11月。IEEE计算机学会。 3[BE15] Sébastien Bubeckand RonenEldan. 熵障碍函数：一种简单且最优的通用自协调障碍函数。载于PeterGrünwald、Elad Hazan和Satyen Kale（编），《第28届学习理论会议论文集》，《机器学习研究论文集》第40卷，第279–279页，法国巴黎, 2015年7月3–6日。PMLR。 3[BNS94] Richard H Byrd, Jorge Nocedal,and Robert B Schnabel. 拟牛顿矩阵的表示及其在有限内存方法中的应用。《数学规划》, 63(1):129–156, 1994. 2, 8[Bro70] CharlesGeorge Broyden.一类双秩最小化算法的收敛性 1. 一般考虑。《IMA应用数学杂志》, 6(1):76– 90, 1970年3月。2, 8[BTN01] Aharon Ben-Tal and ArkadiNemirovski.《现代凸优化》。《工业与应用数学学会》, 2001. 2[Bub15] Sébastien Bubeck. 凸优化：算法与复杂度。《Found. Trends Mach. Learn.》, 8(3-4):231–357, 2015年11月。2[BV04] Dimitris Bertsimas andSantosh Vempala.通过随机游走求解凸规划。《J.ACM》, 51(4):540–556, 2004年7月。8[Che21] YuansiChen. KLS猜想中等周系数的一个近似常数下界。《几何与泛函分析》, 31(1):34–61, 2021. 3, 25

[CLS19] Michael B. Cohen、Yin Tat Lee 和 Zhao Song。在当前矩阵乘法时间内求解线性规划。载于《第51届ACM SIGACT计算理论研讨会论文集》，STOC 2019，第938–942页，美国纽约州纽约，2019年。计算机协会。3, 8[DH07] John Dunagan 和 Nicholas J.A. Harvey。通过共轭梯度方法迭代构造预条件器。载于《第三十九届ACM计算理论研讨会论文集》，STOC '07，第207–216页，美国纽约州纽约，2007年。计算机协会。2, 5, 9, 13, 14[Fle70] Roger Fletcher。变度量算法的新方法。《计算机杂志》，13(3):317–322，1970年。2, 8[FM90] Anthony V Fiacco 和 Garth P McCormick。非线性规划：序贯无约束最小化技术。SIAM，1990年。8[Fri55] KR Frisch。凸规划的对数势函数方法。奥斯陆大学经济学研究所备忘录，5(6)，1955年。8[Gol70] Donald Goldfarb。通过变分方法推导出的一族变度量方法。《计算数学》，24(109):23–26，1970年。2, 8[GS19] Jacek Gondzio 和 FNC Sobral。二次问题内点法的拟牛顿方法。《计算优化与应用》，74(1):93–120，2019年。2, 8[GS22] Jacek Gondzio 和 Francisco NC Sobral。线性规划拟牛顿原始-对偶内点算法的多项式最坏情况迭代复杂度。arXiv预印本 arXiv:2208.08771，2022年。2, 8[Gül96] Osman Güler。内点法中的障碍函数。《《运筹学数学》》，21(4):860–885，1996年。3[HJS+22] B. Huang、S. Jiang、Z. Song、R. Tao 和 R. Zhang。更快地求解半正定规划：一种稳健的内点法框架及高效实现。载于《2022年IEEE第63届计算机科学基础年会（FOCS）》，第233–244页，美国加利福尼亚州洛斯阿拉米托斯，2022年11月。IEEE计算机学会。3, 4, 8[Hua18] Jianyu Huang。实用的快速矩阵乘法算法。博士论文，2018年。4[JKL+20] Haotian Jiang、Tarun Kathuria、Yin Tat Lee、Swati Padmanabhan 和 Zhao Song。半定规划的更快内点法。载于《2020年IEEE第61届计算机科学基础年会（FOCS）》，第910–918页，2020年。2, 3, 4, 7, 8, 28, 29[JLSW20] Haotian Jiang、Yin Tat Lee、Zhao Song 和 Sam Chiu-wai Wong。用于凸优化、凸-凹博弈及其应用的改进切割平面方法。载于《第52届ACM SIGACT计算理论研讨会论文集》，STOC 2020，第944–953页，美国纽约州纽约，2020年。计算机协会。3, 4, 8

[JLV22] Arun Jambulapati、Yin Tat Lee 和 Santosh S Vempala。《关于 kls 常数的略微改进界》。arXiv 预印本，arXiv:2208.11644，2022 年。3，25[Kha79] Leonid Genrikhovich Khachiyan。《线性规划中的多项式算法》。载于《科学院学报》，第 244 卷，第 1093–1096 页。俄罗斯科学院，1979 年。3[Kha80] Leonid G Khachiyan。《线性规划中的多项式算法》。《苏联计算数学与数理物理》，20(1)：53–72，1980 年。8[KMY89] Masakazu Kojima、Shinji Mizuno 和 Akiko Yoshise。《线性规划的原始-对偶内点法》。载于《数学规划进展》，第 29–47 页。Springer，1989 年。8[KOSZ13] Jonathan A. Kelner、Lorenzo Orecchia、Aaron Sidford 和 Zeyuan Allen Zhu。《一种用于在近线性时间内求解 sdd 系统的简单组合算法》。载于第四十五届 ACM 计算理论年会论文集，STOC '13，第 911–920 页，美国纽约州纽约市，2013 年。计算机协会。2[KTE88] Leonid G Khachiyan、Sergei Pavlovich Tarasov 和 I. I. Erlikh。《内接椭球方法》。载于《苏联数学文献汇编》，第 37 卷，第 226–230 页，1988 年。8[LLV20] Aditi Laddha、Yin Tat Lee 和 Santosh Vempala。《强自协调性与采样》。载于第五十二届 ACM SIGACT 计算理论研讨会论文集，STOC 2020，第 1212–1222 页，美国纽约州纽约市，2020 年。计算机协会。3，11，25[LS14] Yin Tat Lee 和 Aaron Sidford。《线性规划的路径搜索方法：在 $\widetilde{O}(\sqrt{rank})$ 次迭代内求解线性规划，以及更快的最大流算法》。载于 2014 年 IEEE 第 55 届计算机科学基础年会，第 424–433 页，2014 年。8[LSW15] Yin Tat Lee、Aaron Sidford 和 Sam Chiu-Wai Wong。《一种更快的切割平面方法及其对组合优化和凸优化的影响》。载于 2015 年 IEEE 第 56 届计算机科学基础年会，第 1049–1065 页，2015 年。3，4，8[LSZ19] Yin Tat Lee、Zhao Song 和 Qiuyi Zhang。《在当前矩阵乘法时间内求解经验风险最小化问题》。载于《学习理论会议》，第 2140–2157 页。PMLR，2019 年。3，8[Meg89] Nimrod Megiddo。《线性规划中通往最优集的路径》。载于《数学规划进展》，第 131–158 页。Springer，1989 年。8[Nes98] Yurii Nesterov。《凸规划导论讲义·第一卷：基础课程》。讲义，3(4)：5，1998 年。24[Nes18] Yurii Nesterov。《凸优化讲义》，第 137 卷。Springer，2018 年。2，3

[NN89] Yurii Nesterov 和 Arkadi Nemirovski。《凸规划中的自协调函数与多项式时间方法》。预印本，苏联科学院中央经济与数学研究所，苏联莫斯科，1989 年。8[NN92] Yu. Nesterov 和 A. Nemirovsky。《凸规划问题的锥形式与对偶性》。《优化方法与软件》，1(2)：95–115，1992 年。2，3，4，8[NN94] Yurii Nesterov 和 Arkadii Nemirovskii。《凸规划中的内点多项式算法》。工业与应用数学学会，1994 年。8，11[Ren01] James Renegar。《凸优化中内点法的数学视角》。SIAM，2001 年。22，33[Sha70] David F Shanno。《函数极小化中拟牛顿方法的条件性》。《计算数学》，24(111)：647–656，1970 年。2，8[Sho77] Naum Z Shor。《凸规划问题中的带空间扩展的截断方法》。《控制论》，13(1)：94–96，1977 年。8[Tun01] Levent Tunçel。《将原始-对偶内点法推广到锥形式的凸优化问题》。《计算数学基础》，1(3)：229–254，2001 年。2，8[Vai89] Pravin M Vaidya。《一种在凸集上极小化凸函数的新算法》。载于第三十届计算机科学基础年会，第 338–343 页，1989 年。8[VB96] Lieven Vandenberghe 和 Stephen Boyd。《半定规划》。《SIAM 评论》，38(1)：49–95，1996 年。2[vdB20] Jan van den Brand。《在当前矩阵乘法时间内运行的确定性线性规划求解器》。载于第三十一届 ACM-SIAM 离散算法研讨会论文集，SODA ’20，第 259–278 页，美国，2020 年。工业与应用数学学会。3，8[vdBLSS20] Jan van den Brand、Yin Tat Lee、Aaron Sidford 和 Zhao Song。《在近线性时间内求解高而稠密的线性规划》。载于第五十二届 ACM SIGACT 计算理论研讨会论文集，STOC 2020，第 775–788 页，美国纽约州纽约市，2020 年。计算机协会。3，8[VR73] Dale E Varberg 和 A Wayne Roberts。《凸函数》。纽约—伦敦，1973 年。27[YN76] David B Yudin 和 Arkadi S Nemirovski。《数学规划问题信息复杂度的评估》。《经济学与数学方法》，12：128–142，1976 年。8