arXiv:2601.02134v1 [math.OC] 5 Jan 2026

约束满足 PL 条件时自适应精度二次罚函数法的复杂度

Florentin Goyens*[1] 和 Geovani N. Grapiglia[1]

[1]比利时新鲁汶，UCLouvain ICTEAM研究所

2026年1月6日

摘要

我们研究了用于带等式约束的光滑非凸优化问题的二次罚函数法（QPM）。假设约束违反在可行集附近满足 PL 条件，我们推导出求得近似一阶 KKT 点的更精确最坏情况复杂度界。当目标函数和约束二次连续可微时，我们证明，配备适当一阶内层求解器的 QPM 至多需要 $\mathcal{O}(\varepsilon_0^{-1}\varepsilon_1^{-2})$ 次一阶预言机调用，即可找到一个 $(\varepsilon_0,\ \varepsilon_1)$-近似 KKT 点——也就是说，该点是 ε_0-近似可行且 ε_1-近似驻定的。此外，当目标函数和约束三次连续可微时，我们证明，配备适当二阶内层求解器的 QPM 至多需要 $\mathcal{O}\left(\varepsilon_0^{-1/2}\varepsilon_1^{-3/2}\right)$ 次二阶预言机调用，即可找到一个 $(\varepsilon_0,\ \varepsilon_1)$-近似 KKT 点。我们还为子问题提出了一种自适应可行性感知停止准则，它会在远离可行状态时放宽驻定性容差。该规则在保留所有理论保证的同时，大幅降低了实际计算开销。

关键词： 二次罚函数法、最坏情况复杂度、二阶方法、自适应子问题容差。

AMS 2020 学科分类： 90C30,65 K05.

目录

1 引言 **2**

1.1 贡献 .3 1.2 相关文献 .4 1.3 内容.4

2 二次罚函数法5 3 无 PL 条件下的外层迭代复杂度6 4 PL 条件下的外层迭代复杂度7 5 使用一阶内层求解器时的总评估复杂度11 5.1 关于一阶内层求解器的引理.11 5.2 无PL条件的一阶内层求解器.13 5.3 PL 条件下一阶内层求解器.14

*通讯作者： florentin.goyens@uclouvain.be。F. Goyens 的资助由比利时国家科学研究基金会FNRS通过 T.0001.23号资助项目提供。

6 使用二阶内层求解器时的总评估复杂度14 6.1 关于二阶内层求解器的引理.15 6.2 无 PL 条件下的二阶内层求解器.16 6.3 PL 条件下的二阶内层求解器.17

7 说明性数值结果 **18**

1 引言

我们考虑如下等式约束优化问题

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x) \text{ subject to } c(x)=0, \tag{P}$$

其中 $f\colon\ \mathbb{R}^n\to\mathbb{R}$ 且 $c\colon\ \mathbb{R}^n\to\mathbb{R}^m$ 是光滑且可能非凸的函数，并且 $m\le n$。罚函数方法的思想是将（P）转化为一系列无约束优化问题。在其第 k次迭代中，二次罚函数法（QPM）通过近似最小化罚函数目标来计算 x_{k+1}。

$$Q_{\beta_k}(x)=f(x)+\frac{\beta_k}{2}\|c(x)\|^2, \tag{1.1}$$

其中 $\beta_k>0$ 是罚参数，用于确定赋予约束违反的权重。每次内层最小化（也称为子问题）都会采用更大的 β_k值。这会提高不可行解的代价，从而推动迭代点逐渐接近可行域。当 $\beta_k\to\infty$时，Q_{β_k} 的极小化点理想情况下会趋近于约束问题（P）的可行点。二次罚函数法历史悠久（Zangwill，1967；Bertsekas，1997），也是非凸优化中的一个重要主题，部分原因在于它是增广拉格朗日法的构建基础。

传统上，优化方法的理论分析侧重于渐近收敛——也就是说，确定迭代点收敛（或其聚点收敛）到问题 (P) 的 KKT 点所需满足的条件。近年来，研究重点转向了最坏情况复杂度，旨在回答以下问题：

> 最坏情况下，需要多少次迭代（以及多少次预言机调用）才能生成一个近似 *KKT* 点？

给定正容差 ε_0 和 ε_1，问题 (ε_0，ε_1)-近似 KKT 点是指满足以下条件的点 $x\in\mathbb{R}^n$，使得存在 $\lambda\in\mathbb{R}^m$ 满足

$$\|c(x)\|\le\varepsilon_0, \qquad \text{and} \qquad \|\nabla f(x)-\sum_{i=1}^m\lambda_i\nabla c_i(x)\|\le\varepsilon_1. \tag{($(\varepsilon_0,\varepsilon_1)$-KKT)}$$

最坏情况复杂度结果通常用预言机调用次数来表示。调用一阶预言机 可在给定点 $x\in\mathbb{R}^n$ 处返回 f、c 的值及其一阶导数。二阶预言机 还会提供二阶导数。也可以使用高阶预言机，但本文不予考虑。

在 QPM 的背景下，外层迭代复杂度指的是在达到一个(ε_0，ε_1)-KKT 点之前求解的子问题数量（或等价地，罚参数的更新次数）。在每次外层迭代中，无约束优化方法会根据可用的预言机，采用一阶或二阶方法近似最小化罚函数 Q_{β_k}。

内层评估复杂度是指近似最小化单个子问题所需的预言机调用次数。这两个概念共同构成总评估复杂度，即求得一个 $(\varepsilon_0,\ \varepsilon_1)$-KKT 点所需的预言机调用总数。

理解并改进这些界对于优化算法的设计至关重要，因为从复杂度分析中获得的洞见可能为设计具有更佳实际性能的方法提供线索。

1.1 贡献

我们的工作改进了 QPM 获取问题（P）近似一阶KKT点的复杂度界。我们分析了一个简单且实用的 QPM 版本（算法 1），未引入任何旨在促进理论分析的辅助机制。

我们的主要贡献如下：

1. 我们为子问题引入了自适应的可行性感知容差，当远离可行性时放宽驻点性要求，从而避免在早期迭代中浪费计算（定义 2.1）。这显著提升了实际性能。

2. 我们改进了 QPM 的外层迭代复杂度界。如果约束违反在可行集附近满足 Polyak–Łojasiewicz（PL）条件（假设 A4），我们证明 $\beta_k \geq \mathcal{O}(\varepsilon_0^{-1})$ 可确保 $\|c(x_{k+1})\| \leq \varepsilon_0$（引理 4.3），从而改进现有的 $\beta_k \geq \mathcal{O}(\varepsilon_0^{-2})$ 界（Grapiglia，2023）。PL 条件是约束优化中的一种主流假设，以各种名称出现在相关文献中。在大量文献中，该条件被假设为在全局或所有迭代点上成立；而在本文中，我们只要求它在可行集的一个邻域内成立。

3. 我们建立了总评估复杂度界。根据A4，当目标函数和约束二阶连续可微时，我们证明，采用适当的一阶内层求解器的 QPM 至多需要 $\tilde{\mathcal{O}}(\varepsilon_0^{-1}\varepsilon_1^{-2})$ 次一阶预言机调用，即可找到一个 $(\varepsilon_0,\varepsilon_1)$-KKT 点。此外，根据A4，当目标函数和约束三次连续可微时，我们证明，采用适当的二阶内层求解器的 QPM 至多通过$(\varepsilon_0,\varepsilon_1)$-KKT 点所需的 $\tilde{\mathcal{O}}\left(\varepsilon_0^{-1/2}\varepsilon_1^{-3/2}\right)$ 次二阶预言机调用即可达到该点。我们不假设 f 和 c的导数在包含迭代点和试探点的集合上 Lipschitz 连续，因为这可能难以事先保证。相反，我们只假设 f 具有有界次水平集（A5）。

下表总结了本文得出的复杂度结果。

QPM：目标 $(\varepsilon_0,\ \varepsilon_1)$-KKT	无 PL 条件	满足 PL 条件（A4）
罚参数 $\beta_{\max}$	$\mathcal{O}\left(\varepsilon_0^{-2}\right)^*$	$\mathcal{O}\left(\varepsilon_0^{-1}\right)$
外层迭代复杂度	$\mathcal{O}\left(\log\left(\beta_0^{-1}\varepsilon_0^{-2}\right)\right)^*$	$\mathcal{O}\left(\log\left(\beta_0^{-1}\varepsilon_0^{-1}\right)\right)$
总评估复杂度 采用一阶内层求解器（A6）	$\mathcal{O}\left(\log\left(\beta_0^{-1}\varepsilon_0^{-2}\right)\varepsilon_0^{-2}\,\varepsilon_1^{-2}\right)^*$ $\mathcal{O}\left(\log\left(\beta_0^{-1}\varepsilon_0^{-2}\right)\varepsilon_0^{-1}\,\varepsilon_1^{-3/2}\right)$	$\mathcal{O}\left(\log\left(\beta_0^{-1}\varepsilon_0^{-1}\right)\varepsilon_0^{-1}\,\varepsilon_1^{-2}\right)$ $\mathcal{O}\left(\log\left(\beta_0^{-1}\varepsilon_0^{-1}\right)\varepsilon_0^{-1/2}\,\varepsilon_1^{-3/2}\right)$
总评估复杂度 采用二阶内层求解器（A9）		

表1：二次罚函数法在约束违反满足和不满足 PL 条件时的复杂度结果汇总。 *对于子问题容差恒定的 QPM，Grapiglia于2023年已在（$\tau(x) \equiv \varepsilon_1$ 中的（2.4）中给出。

1.2 相关文献

近年来，约束非凸问题的方法的最坏情况评估复杂度得到了广泛研究（Cartis 等人，2022）。对于一阶方法，Cartis 等人（2011）分析了一种精确罚函数法，证明至多需要 $\mathcal{O}(\varepsilon^{-5})$次问题评估，即可达到一个 ε-KKT 点；这相当于在我们的定义中达到一个满足 $\varepsilon = \varepsilon_0 = \varepsilon_1$ 的$(\varepsilon_0, \varepsilon_1)$-KKT 点。假设罚参数保持有界，该复杂度可改进为 $\mathcal{O}(\varepsilon^{-2})$。两阶段方法（Cartis 等人，2014）、非精确恢复方法（Bueno 和 Martínez，2020）以及采用 Fletcher增广拉格朗日函数的方法（Goyens 等人，2024）也取得了相当的 $\mathcal{O}(\varepsilon^{-2})$界。对于序列二次规划方法，Facchinei 等人（2021）建立了 $\mathcal{O}(\varepsilon^{-4})$界，Curtis 等人（2024）随后证明了$\mathcal{O}(\varepsilon^{-2})$界。

二阶方法可改善复杂度：自适应三次正则化和信赖域漏斗方案可达到最坏情况速率 $\mathcal{O}(\varepsilon^{-3/2})$ (Cartis等人，2013；Curtis等人，2018），而高阶模型可达到 $\mathcal{O}(\varepsilon^{-(p+1)/p})$，适用于 p 阶方法（$p \geq 2$），如（Birgin等人，2016；Martínez，2017；Cartis等人，2022）所示。

非精确罚函数法，例如二次罚函数法和增广拉格朗日法，也受到了广泛关注。在约束满足非奇异性条件时，Li等人(2021)证明了非精确增广拉格朗日方法的 $\tilde{\mathcal{O}}(\varepsilon^{-3})$ 界。

Birgin 和 Martínez（2020）给出了 Algencan 算法（Andreani等人，2008）的复杂度结果。该算法是一种旨在提高实际性能的增广拉格朗日方法。他们证明，在若干情形下，外层迭代次数呈对数级，并将其与具有任意复杂度的内层求解器相结合。Grapiglia（2023）证明，采用非自适应子问题精度 $\tau(x) \equiv \varepsilon_1$ 并在子问题中使用带 Armijo 线搜索的梯度下降法的二次罚函数法（算法 1），至多需要 $\tilde{\mathcal{O}}(\varepsilon^{-4})$次问题评估，即可找到一个 ε-KKT 点。对于 f 和 c 弱凸的情形，Lin等人（2020）证明了二次罚函数法的 $\tilde{\mathcal{O}}(\varepsilon^{-3})$界。

对于线性约束问题，Kong 等人（2019）证明了采用加速一阶内层求解器的二次罚函数法具有 $\mathcal{O}(\varepsilon^{-3})$ 界。在 Slater 型条件下，近端点法或增广拉格朗日法变体可达到 $\tilde{\mathcal{O}}(\varepsilon^{-5/2})$，例如（Li 和 Xu，2021；Lin 等人，2020；Melo 等人，2020）。Grapiglia 和 Yuan（2021）为采用 p 阶内层求解器的增广拉格朗日方法获得了 $\tilde{\mathcal{O}}(\varepsilon^{-(p+1)/p})$ 界，与 Xie 和 Wright（2021）的近端增广拉格朗日界的阶一致。随后，He 等人（2023）改进了这一界：在 LICQ 条件下得到 $\tilde{\mathcal{O}}(\varepsilon^{-7/2})$ 的结果，而在不满足 LICQ 时得到 $\tilde{\mathcal{O}}(\varepsilon^{-11/2})$。

在罚函数方法之外，黎曼优化方法专门用于处理（P）的实例，其中可行集是光滑流形，并且切空间和到可行集的投影具有便捷的表达式。在利普希茨连续性假设下，黎曼梯度下降至多经过 $(0, \varepsilon_1)$-KKT 点需迭代 $\mathcal{O}(\varepsilon_1^{-2})$ 次（Boumal 等人，2019）。二阶黎曼信赖域法也具有类似的界；对于一类具有严格鞍点的函数，该界可改进为 $\mathcal{O}(\log\log(\varepsilon^{-1}))$ （Goyens 和 Royer，2025）。

总体而言，比较这些结果并非易事，因为它们针对不同的问题类别，通常对 c 采用不同的正则性假设，或对初始点的质量提出不同要求。本文通过为一种允许非精确子问题解的二次罚函数法建立复杂度界，推进了现有研究。

1.3 内容

本文结构安排如下。在第 2 节中，我们定义二次罚函数法（QPM），并引入可行性感知容差。在第3 节中，我们证明，当子问题中的精度采用Grapiglia，2023 年提出的

自适应容差（2.4）时，（4 节中，我们在约束违反满足 PL 条件（A4）的情况下，推导出改进的外层迭代复杂度界。在第 5 节中，我们给出采用一阶方法执行内层最小化时 QPM 的总评估复杂度界。在第6 节中，我们给出采用二阶方法执行内层最小化时的总评估复杂度界。在第 7 节中，我们通过实验展示研究结果，并表明子问题中的自适应精度能够提升实际表现。

符号

令 $\|\cdot\|$ 表示 $\mathbb{R}^n$中的通常 2-范数。令 $J(x) \in \mathbb{R}^{m\times n}$ 表示函数c的雅可比矩阵，并令 $\mathbb{N}^*$ 表示正自然数集合。我们用 conv(A) 表示集合 A的凸包。

2 二次罚函数法

我们考虑算法1中描述的二次罚函数法。关于二次罚函数法的经典结果（Bertsekas，1997，命题 4.2.1）表明，在温和的假设下，Q_{β_k} 的全局极小点会收敛到（P）的全局极小点，当 $\beta_k \to \infty$时。这在实践中价值不大，因为不希望 β_k 无限增大，否则子问题会变得病态。此外，通常无法找到二次罚函数的全局极小点。这引出了两个重要问题：

(i) β_k 应以多快的速度增长?

(ii) 每个子问题应求解到何种精度?

在迭代 k时，我们将罚参数更新为 $\beta_{k+1} = \alpha\beta_k$，其中 $\alpha > 1$ 是 QPM 的一个参数。我们要求 Q_{β_k} 的近似极小点使 Q_{β_k}的值相较于当前点 x_k 和初始点 x_0 (2.3)有所下降；同时还要求该点近似满足一阶驻点条件（2.4）。

在（Grapiglia，2023）中，每次迭代都将子问题求解至目标精度 ε_1。然而，在优化软件中，通常会在初期迭代时采用大于 ε_1 的驻点性容差。实践中会使用各种启发式方法，例如从$\sqrt{\varepsilon_1}$开始，并在每次外层迭代时将容差除以10 （Algencan、Andreani等人（2008））；（Eckstein和Silva，2013）还提出了另一种策略。我们提出一种关于梯度范数的自适应实用容差。精度与当前可行性水平成正比；当迭代点远离可行集时，这是一种自然的节省计算资源的方式。

定义2.1（可行性感知容差）。给定容差 $\varepsilon_0 > 0$，$\varepsilon_1 > 0$，可行性感知容差 是一个函数 τ：$\mathbb{R}^n \to \mathbb{R}_+$，对于某个 $\tau_{\text{cap}} \geq \varepsilon_1$ 由下式给出

$$\tau(x) := \max\Big\{\varepsilon_1,\ \min\Big(\tau_{\text{cap}},\ \tfrac{\varepsilon_1}{\varepsilon_0}\ \|c(x)\|\Big)\Big\}. \tag{2.1}$$

这确保了

$$\tau(x) \geq \varepsilon_1 \quad \text{for all } x \in \mathbb{R}^n, \qquad \text{and} \qquad \tau(x) = \varepsilon_1 \text{ whenever } \|c(x)\| \leq \varepsilon_0. \tag{2.2}$$

因此，一旦 QPM 生成 x_{k+1} 满足 $\|c(x_{k+1})\| \leq \varepsilon_0$，便可自动推出 $\|\nabla Q_{\beta_k}(x_{k+1})\| \leq \varepsilon_1$，这使其成为一个 ($\varepsilon_0$，$\varepsilon_1$)-KKT 点。在我们的数值实验（第 7 节）中，我们比较了这一选择

$$\tau(x) := \max\left(\varepsilon_1, \frac{\varepsilon_1}{\varepsilon_0}\|c(x)\|\right),$$

corr对应于 $\tau_{\text{cap}} = +\infty$；采用非自适应选择 $\tau(x) \equiv \varepsilon_1$，其给定形式为 $\tau_{\text{cap}} = 0$.

算法 1二次罚函数法（QPM）

给定： $\varepsilon_0 > 0$、$\varepsilon_1 > 0$、$\alpha > 0$、$\beta_0 \geq 1$、可行性感知容差 τ：$\mathbb{R}^n \to \mathbb{R}_+$，以及 $x_0 \in \mathbb{R}^n$
$k \leftarrow 0$
当 真 时执行
　求取 $x_{k+1} \in \mathbb{R}^n$ 的近似极小点，使其满足

$$Q_{\beta_k}(x_{k+1}) \leq \min\{Q_{\beta_k}(x_k), Q_{\beta_k}(x_0)\} \tag{2.3}$$

and

$$\|\nabla Q_{\beta_k}(x_{k+1})\| \leq \tau(x_{k+1}). \tag{2.4}$$

　如果 $\|c(x_{k+1})\| \leq \varepsilon_0$，**则**
　　返回 x_{k+1} ▷ $(\varepsilon_0, \varepsilon_1)$-KKT 点（乘子为 $-\beta_k c(x_{k+1})$）
　结束 if
　$\beta_{k+1} = \alpha\beta_k$。
　$k \leftarrow k + 1$
结束 while

3 无 PL 条件下的外层迭代复杂度

本节为 QPM（算法 1）建立外层迭代复杂度界。本节的主要结果已在（Grapiglia，2023）中给出，其中采用了 $\tau(x) \equiv \varepsilon_1$ 的选择。我们将分析扩展到包含（2.4）中的自适应容差。

本节不对约束的雅可比矩阵作任何正则性假设。第4节将展示，在约束违反满足 PL 条件的假设下，可获得更优的外层迭代复杂度保证。

我们给出标准假设。

A1. 函数 f 和 c可微。

A2. 存在 $f_{\text{low}} \in \mathbb{R}$，使得 $f(x) \geq f_{low}$ 对所有 $x \in \mathbb{R}^n$均成立。

A3. 初始迭代点 $x_0 \in \mathbb{R}^n$ 满足 $\|c(x_0)\| \leq \dfrac{\varepsilon_0}{\sqrt{2}}$。 —=

以下引理以逆否形式表明：若在某次迭代 k时，罚参数 β_k 足够大，则 QPM 会终止。具体而言，若 $\beta_k \geq \mathcal{O}(\varepsilon_0^{-2})$，则$\|c(x_{k+1})\| \leq \varepsilon_0$。

引理 3.1 (Grapiglia (2023),推论 3.3)。 在 *A1*、*A2*、*A3*成立时，若 $\|c(x_k)\| > \varepsilon_0$ 对 k =1、.....、$T-1$，则

$$\beta_{T-2} < 4(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}. \tag{3.1}$$

回想一下，$\mathbb{N}^*$ 表示正自然数集。现在我们证明，QPM 的外层迭代次数至多为引理 3.1中给出的罚参数上界的对数因子。

定理 理 3.2（不满足 PL 条件时 QPM 的外层迭代）。 在 *A1*、*A2*、*A3*, 定义 e

$$T(\varepsilon_0) := \inf\{k \in \mathbb{N}^* \colon \|c(x_k)\| \le \varepsilon_0\}. \tag{3.2}$$

于是,

$$T(\varepsilon_0) < \hat{T}(\varepsilon_0) := 2 + \log_\alpha \left(4(f(x_0) - f_{\text{low}})\beta_0^{-1}\varepsilon_0^{-2}\right), \tag{3.3}$$

且由算法 $x_{T(\varepsilon_0)}$ 生成的点 *1* 是一个 $(\varepsilon_0,\ \varepsilon_1)$-*KKT* 点，属于 (P)，并满足

$$\left\|c\left(x_{T(\varepsilon_0)}\right)\right\| \le \varepsilon_0 \qquad \textit{and} \qquad \left\|\nabla f\left(x_{T(\varepsilon_0)}\right) + \beta_{T(\varepsilon_0)-1}\sum_{i=1}^{m} c_i\left(x_{T(\varepsilon_0)}\right)\nabla c_i\left(x_{T(\varepsilon_0)}\right)\right\| \le \varepsilon_1. \tag{3.4}$$

证明。 根据 $T(\varepsilon_0)$的定义， $\|c(x_k)\| > \varepsilon_0$对 $k = 1$，...， $T(\varepsilon_0) - 1$。由引理 3.1，可得

$$\alpha^{T(\varepsilon_0)-2}\beta_0 = \beta_{T(\varepsilon_0)-2} < 4(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}.$$

因此， $T(\varepsilon_0)$ 是有限的，并且

$$T(\varepsilon_0) < 2 + \log_\alpha \left(4(f(x_0) - f_{\text{low}})\beta_0^{-1}\varepsilon_0^{-2}\right) =: \hat{T}(\varepsilon_0).$$

最后，由构造可知 $\left\|c\left(x_{T(\varepsilon_0)}\right)\right\| \le \varepsilon_0$ ，由 (2.4) 可知

$$\left\|\nabla Q_{\beta_{T(\varepsilon_0)-1}}\left(x_{T(\varepsilon_0)}\right)\right\| \le \tau\left(x_{T(\varepsilon_0)}\right) = \varepsilon_1.$$

总之，恒等式

$$\left\|\nabla Q_{\beta_{T(\varepsilon_0)-1}}\left(x_{T(\varepsilon_0)}\right)\right\| = \left\|\nabla f\left(x_{T(\varepsilon_0)}\right) + \beta_{T(\varepsilon_0)-1}\sum_{i=1}^{m} c_i\left(x_{T(\varepsilon_0)}\right)\nabla c_i\left(x_{T(\varepsilon_0)}\right)\right\|,$$

意味着 $x_{T(\varepsilon_0)}$ 是一个$(\varepsilon_0, \varepsilon_1)$-KKT 点，其拉格朗日乘子为 $-\beta_{T(\varepsilon_0)-1}c\left(x_{T(\varepsilon_0)}\right)$。 □

在下一节中，我们将作出一项额外假设：在可行集附近，约束违反满足 PL 条件。这使我们能够改进复杂度结果。

4 PL 条件下的外层迭代复杂度

本节将给出 QPM 的一个新复杂度结果，改进定理 3.2 中的界。上一节的分析仅利用了每个子问题中罚函数值的下降（2.3）。为得到更精确的界，我们利用近似驻点条件（2.4）。最自然的做法是，在约束雅可比矩阵满足额外正则性条件的情况下进行分析。

A4 (约束违反的 PL 条件)。 存在正常数 $R > \varepsilon_0$ 和 $\sigma_{\min}$，使得对集合中的所有 x

$$\mathcal{C}_R := \{x \in \mathbb{R}^n \colon \|c(x)\| \le R\}, \tag{4.1}$$

我们有

$$\left\|J(x)^\top c(x)\right\| \ge \sigma_{\min}\|c(x)\|, \tag{4.2}$$

其中 $J(x) \in \mathbb{R}^{m\times n}$ 表示 c 的雅可比矩阵。

注意，(4.2) 是约束违反的 Polyak–Łojasiewicz（PL）条件

$$\phi(x) = \tfrac{1}{2}\|c(x)\|^2.$$

事实上，(4.2) 等价于

$$\phi(x) \le \frac{1}{2\sigma_{\min}^2}\,\|\nabla\phi(x)\|^2 \quad \text{for all } x \in \mathcal{C}_R. \tag{4.3}$$

它比点 x 处通常的 LICQ 条件更弱；该条件要求 $J(x)$ 满秩，但我们要求在可行集的管状邻域内存在一个正的下界 $\sigma_{\min} > 0$。正则性条件 (4.2) 在文献中相当常见，并有多种称呼。大多数分析都假设该条件在 $\mathbb{R}^n$上成立，这排除了许多常见的应用场景。值得注意的是，我们从一个局部条件推导出了一个全局界。

我们还要求 f 的一个子水平集有界。

A5. 集合 $\mathcal{L}_f\,(2f(x_0) - f_{\text{low}})$ 有界。

首先，我们给出几个满足假设的例子。对于以下任一例子，令 f 具有强制性（即 $\lim_{\|x\|\to\infty} f(x) = +\infty$）是满足 A5的充分条件。

例 1 （仿射约束）。 问题

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} && f(x) \\ &\textit{subject to} && Ax = b, \end{aligned} \tag{4.4}$$

满足 $\operatorname{rank}(A) = m$ 的条件 *A4*。

例 2 (正交约束)。 问题

$$\begin{aligned} &\min_{X\in\mathbb{R}^{n\times p}} && f(X) \\ &\textit{subject to} && X^\top X = \mathrm{I}_p, \end{aligned} \tag{4.5}$$

满足 *A4* 对于任意 $\varepsilon_0 < R < 1$ 和 $\sigma_{\min} \le 2\sqrt{1-R}$。

证明。 参见（Goyens 等人，2024，第 4 页）。 □

例 3 (二元约束)。 问题

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} && f(x) \\ &\textit{subject to} && x_i(1-x_i) = 0, \quad \textit{for } i = 1,\dots,n, \end{aligned} \tag{4.6}$$

编码二元约束 $x_i \in \{0,1\}$，满足 *A4* 对于任意 R 使得 $\varepsilon_0 < R < \frac{1}{4}$ 和 $\sigma_{\min} = \sqrt{1-4R} > 0$。

证明。 c的雅可比矩阵是对角矩阵， $J(x) = \operatorname{diag}\big(1-2x_1,\dots,1-2x_n\big)$. 利用恒等式

$$(1-2x_i)^2 = 1 - 4c_i(x),$$

对于任意满足 $\|c(x)\| \le R$ 的 x，我们得到

$$\begin{aligned}\|J(x)^\top c(x)\|^2 &= \sum_{i=1}^{n}(1-2x_i)^2\, c_i(x)^2 \\ &= \sum_{i=1}^{n}(1-4c_i)\, c_i(x)^2 \\ &\ge (1-4R)\sum_{i=1}^{n} c_i(x)^2 \\ &= (1-4R)\|c(x)\|^2.\end{aligned}$$

因此，

$$\|J(x)^\top c(x)\| \ge \sqrt{1-4R}\,\|c(x)\| \qquad \text{for all } x \in \mathcal{C}_R. \qquad \square$$

此外，结合多个非奇异约束可得到一个满足A4的约束函数。

命题 4.1（Goyens 等人（2024），命题 1.1）。对于 $i = 1、2、\ldots\ldots、k$，设 k 个函数 $c_i: \mathbb{R}^n \to \mathbb{R}^{m_i}$ 满足 A4，其常数为 R_i 和 σ_i。那么，函数 $c: \mathbb{R}^n \to \mathbb{R}^m$，其中 $m = m_1 + \cdots + m_k$ 由 $c(x) = (c_1(x)\ldots\ldots, c_k(x))^\top$ 定义，满足 A4，其常数为 $R = \min(R_1, \ldots\ldots, R_k)$ 和 $\sigma = \min(\sigma_1, \ldots\ldots, \sigma_k)$。

现在我们转而分析A4条件下的复杂度，并首先给出一个引理，说明 QPM 的迭代点都包含在函数 f的某个子水平集中。

引理 4.2。在A1、A2、A3条件下，若 $\|c(x_k)\| > \varepsilon_0$ 对于 $k = 1, \ldots\ldots, T-1$ 均成立，则

$$x_k \in \mathcal{L}_f(2f(x_0) - f_{\text{low}}) := \{x \in \mathbb{R}^n : f(x) \le 2f(x_0) - f_{\text{low}}\}, \tag{4.7}$$

对于 $k = 0, \ldots, T-1$。

证明。显然 $x_0 \in \mathcal{L}_f(2f(x_0) - f_{\text{low}})$。令 $k \in \{1, \ldots, T-1\}$。下降条件$Q_{\beta_{k-1}}(x_k) \le Q_{\beta_{k-1}}(x_0)$ (2.3)，结合引理 3.1 和 $\|c(x_0)\| \le \frac{\varepsilon_0}{\sqrt{2}}$ 可得

$$\begin{aligned}f(x_k) &\le f(x_k) + \frac{\beta_{k-1}}{2}\|c(x_k)\|^2 \\ &\le f(x_0) + \frac{\beta_{k-1}}{2}\|c(x_0)\|^2 \\ &\le f(x_0) + 2(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}\frac{\varepsilon_0^2}{2} \\ &= 2f(x_0) - f_{\text{low}}.\end{aligned}$$

$\square$

接下一个引理表明，若 $\beta_k \ge \mathcal{O}(\varepsilon_0^{-1})$，则 QPM 会终止，即 $\|c(x_{k+1})\| \le \varepsilon_0$。

L理 4.3。在 A1、A2、A3、A4、A5的条件下，若 $\|c(x_k)\| > \varepsilon_0$ 对于 $k = 1, \ldots, T-1$，则 n

$$\beta_{T-2} < \max\left\{\frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\}\varepsilon_0^{-1}, \tag{4.8}$$

其中 $L_{f,\,0} \ge 0$ 满足

$$\|\nabla f(x)\| \le L_{f,0} \quad \text{for all } x \in \mathcal{L}_f(2f(x_0) - f_{\text{low}}). \tag{4.9}$$

证明。注意到 $L_{f,\,0} \geq 0$ 定义良好，因为 $\mathcal{L}_f(2f(x_0) - f_{\text{low}})$ 有界（A5）。反设存在 $k \in \{0, \ldots, T-2\}$ 使得 $\|c(x_{k+1})\| > \varepsilon_0$ 且

$$\beta_k \geq \max\left\{\frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\} \varepsilon_0^{-1}. \tag{4.10}$$

条件 $Q_{\beta_k}(x_{k+1}) \leq Q_{\beta_k}(x_0)$ 得出

$$f(x_{k+1}) + \frac{\beta_k}{2}\|c(x_{k+1})\|^2 \leq f(x_0) + \frac{\beta_k}{2}\|c(x_0)\|^2. \tag{4.11}$$

由于 $\varepsilon_0 < R$，条件（4.10）意味着 $\beta_k \geq 4(f(x_0) - f_{\text{low}})/R^2$。利用 $\|c(x_0)\|^2 \leq \frac{\varepsilon_0^2}{2}$，可得

$$\begin{aligned}\|c(x_{k+1})\|^2 &\leq \frac{2(f(x_0) - f_{\text{low}})}{\beta_k} + \|c(x_0)\|^2 \\ &\leq \frac{R^2}{2} + \frac{R^2}{2} \leq R^2.\end{aligned}$$

因此，x_{k+1} 属于正则区域 $\mathcal{C}_R$，这保证了

$$\left\|J(x_{k+1})^\top c(x_{k+1})\right\| \geq \sigma_{\min}\|c(x_{k+1})\|.$$

注意到 $\nabla Q_\beta(x) = \nabla f(x) + \beta J(x)^\top c(x)$。因此，一阶条件（2.4）给出

$$\begin{aligned}\tau(x_{k+1}) &\geq \|\nabla Q_{\beta_k}(x_{k+1})\| \\ &= \left\|\nabla f(x_{k+1}) + \beta_k J(x_{k+1})^\top c(x_{k+1})\right\| \\ &\geq \beta_k \left\|J(x_{k+1})^\top c(x_{k+1})\right\| - \|\nabla f(x_{k+1})\| \\ &\geq \beta_k \sigma_{\min}\|c(x_{k+1})\| - \|\nabla f(x_{k+1})\|,\end{aligned} \tag{4.12}$$

这给出

$$\sigma_{\min}\beta_k\|c(x_{k+1})\| \leq \tau(x_{k+1}) + \|\nabla f(x_{k+1})\|. \tag{4.13}$$

由引理 4.2，$x_{k+1} \in \mathcal{L}_f(2f(x_0) - f_{\text{low}})$，因此 $\|\nabla f(x_{k+1})\| \leq L_{f,\,0}$。由于 $\|c(x_{k+1})\| > \varepsilon_0$，我们有

$$\tau(x_{k+1}) = \max\left\{\varepsilon_1, \min\left(\tau_{\text{cap}}, \frac{\varepsilon_1}{\varepsilon_0}\|c(x_{k+1})\|\right)\right\} \leq \max\left\{\varepsilon_1, \frac{\varepsilon_1}{\varepsilon_0}\|c(x_{k+1})\|\right\} = \frac{\varepsilon_1}{\varepsilon_0}\|c(x_{k+1})\|.$$

由此可得

$$\sigma_{\min}\beta_k\|c(x_{k+1})\| \leq \frac{\varepsilon_1}{\varepsilon_0}\|c(x_{k+1})\| + L_{f,0},$$

也就是说，

$$\|c(x_{k+1})\| \leq \frac{L_{f,0}}{\sigma_{\min}\beta_k - \frac{\varepsilon_1}{\varepsilon_0}}.$$

由于 $\beta_k \geq \frac{L_{f,\,0}+\varepsilon_1}{\sigma_{\min}\varepsilon_0}$，因此有 $\|c(x_{k+1})\| \leq \varepsilon_0$，这就产生了矛盾。 □

现在，我们给出 QPM 在约束违反满足 PL 条件时改进后的外层迭代次数上界。

定理 4.4 (PL 条件下 QPM 的外层迭代次数)。 在 *A1*、*A2*、*A3*、*A4*、*A5*成立时，令

$$T(\varepsilon_0) := \inf\{k \in \mathbb{N}^* : \ \|c(x_k)\| \le \varepsilon_0\}. \tag{4.14}$$

那么，

$$T(\varepsilon_0) \le \tilde{T}(\varepsilon_0) := 2 + \log_\alpha \left(\max\left\{ \frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R} \right\} \beta_0^{-1} \varepsilon_0^{-1} \right), \tag{4.15}$$

且由算法 $x_{T(\varepsilon_0)}$ 生成的点 *1* 是 $(\varepsilon_0, \varepsilon_1)$-*KKT* 点，满足 (P)，

$$\left\|c\left(x_{T(\varepsilon_0)}\right)\right\| \le \varepsilon_0 \quad \textit{and} \quad \left\| \nabla f\left(x_{T(\varepsilon_0)}\right) + \beta_{T(\varepsilon_0)-1} \sum_{i=1}^{m} c_i\left(x_{T(\varepsilon_0)}\right) \nabla c_i\left(x_{T(\varepsilon_0)}\right) \right\| \le \varepsilon_1. \tag{4.16}$$

证明。 根据 $T(\varepsilon_0)$的定义，有 $\|c(x_k)\| > \varepsilon_0$ 对于 $k = 1, \ldots, T(\varepsilon_0) - 1$。引理 4.3 给出

$$\alpha^{T(\varepsilon_0)-2}\beta_0 = \beta_{T(\varepsilon_0)-2} < \beta_{\max} := \max\left\{ \frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R} \right\} \varepsilon_0^{-1}.$$

因此， $T(\varepsilon_0)$ 有限，特别是

$$\alpha^{T(\varepsilon_0)-2}\beta_0 = \beta_{T(\varepsilon_0)-2} < \beta_{\max}. \tag{4.17}$$

这就得到

$$T(\varepsilon_0) \le 2 + \log_\alpha \left(\frac{\beta_{\max}}{\beta_0} \right) =: \tilde{T}(\varepsilon_0). \tag{4.18}$$

最后，根据构造有 $\left\|c(x_{T(\varepsilon_0)})\right\| \le \varepsilon_0$，这保证 $x_{T(\varepsilon_0)}$是一个 $(\varepsilon_0, \varepsilon_1)$-KKT点，其乘子为 $-\beta_{T(\varepsilon_0)-1}c(x_{T(\varepsilon_0)})$ (参见定理 3.2的证明）。 □

注 4.1。 正如引理 *4.3* 的证明所示，我们可以将要求 $\|c(x_0)\| \le \varepsilon_0/\sqrt{2}$ 替换为 $x_0 \in \mathcal{C}_R$，并保持收敛性，前提是 ∇f 在各迭代点上的范数保持有界。

5 使用一阶内层求解器时的总评估复杂度

本节及下一节将推导 QPM（算法 1）的总评估复杂度界，从而给出预言机调用次数（对 f、c 及其导数进行函数评估）的上界。内层评估复杂度通过对 Q_{β_k}、 ∇Q_{β_k}和$\nabla^2 Q_{\beta_k}$的函数评估次数来衡量，其中每次此类评估分别对应一次一阶或二阶预言机调用。

本节给出 QPM 的内层求解器为一阶方法时的结果。我们将详细介绍在约束违反（A4）满足和不满足 PL 条件时的结果。

5.1 引理：一阶内层求解器

我们假设，在 QPM 的每个子问题中，名为 $\mathcal{M}_1$ 的一阶方法都会对 Q_{β_k} 进行最小化。以下关于（内层）评估复杂度的假设，是无约束非凸优化中一阶方法的典型假设。

A6. 给定一个连续可微函数 $F:\mathbb{R}^n \to \mathbb{R}$，其满足

$$\mathcal{L}_F(\tilde{x}) = \{z \in \mathbb{R}^n \mid F(z) \le F(\tilde{x})\}$$

对某个 $\tilde{x} \in \mathbb{R}^n$有界。从 $\tilde{x}$ 开始的方法 $\mathcal{A}_1$ 至多需要

$$C_{\mathcal{A}_1} L_1 \big(F(\tilde{x}) - F_{low}\big)\, \varepsilon^{-2}$$

F 和 ∇F 的函数评估次数，以生成满足 $\|\nabla F(x)\| \le \varepsilon$的点，其中 F_{low} 是 F的下界，正常数 $C_{\mathcal{A}_1}$ 仅取决于方法 $\mathcal{A}_1$，且 ∇F 在$\mathrm{conv}(\mathcal{L}_F(F(\tilde{x})))$上是 L_1-*Lipschitz*连续的。

传统的带 Armijo线搜索的梯度下降法满足 A6 (Grapiglia，2023，附录1）。

我们作出以下光滑性假设。

A7. 函数 f 和 c均为二阶连续可微。

Recall that $\phi(x) = \frac{1}{2}\|c(x)\|^2$.

引理5.1。在*A5* 和*A7*的条件下，存在常数 $L_{f,\,1}$ 和 $L_{\phi,\,1} > 0$，使得

$$\left\|\nabla^2 f(x)\right\| \le L_{f,1} \qquad \textit{and} \qquad \left\|\nabla^2 \phi(x)\right\| \le L_{\phi,1} \tag{5.1}$$

对所有 $x \in \mathrm{conv}\,(\mathcal{L}_f(2f(x_0) - f_{\mathrm{low}}))$。

此外，对于任意 $\beta > 0$，我们有 ∇Q_β 在$(L_{f,1} + \beta L_{\phi,1})$-*Lipschitz*连续于$\mathrm{conv}(\mathcal{L}_f(2f(x_0) - f_{\mathrm{low}}))$ 上。

证明。根据 A5，子水平集 $\mathcal{L}_f(2f(x_0)-f_{\mathrm{low}})$ 有界，因此其凸包有界。根据 A7，海森矩阵 $\nabla^2 f$ 和 $\nabla^2 P$ 连续，因此存在常数 L_f,1 和 $L_{\phi,1} > 0$，使得 (5.1) 成立。因此，对于所有属于 $\mathrm{conv}(\mathcal{L}_f(2f(x_0) - f_{\mathrm{low}}))$ 的 x 以及任意 $\beta > 0$,

$$\begin{aligned}\left\|\nabla^2 Q_\beta(x)\right\| &\le \left\|\nabla^2 f(x)\right\| + \beta\left\|\nabla^2 \phi(x)\right\| \\ &\le L_{f,1} + \beta L_{\phi,1}.\end{aligned}$$

因此，∇Q_β 在 $\mathrm{conv}\,(\mathcal{L}_f(2f(x_0) - f_{\mathrm{low}}))$ 上是 $(L_{f,1} + \beta L_{\phi,1})$-Lipschitz连续的。 □

下一个引理给出了采用一阶求解器的 QPM 的内层评估复杂度。

引理5.2（$\mathcal{M}_1$的内层评估复杂度）。在 *A2*、*A5* 和 *A7*下，考虑算法 *k1*的第 $\mathcal{M}_1$ 次迭代，其中采用单调一阶方法 Q_{β_k} 从

$$\tilde{x}_{k,0} := \operatorname*{argmin}_{x \in \{x_k, x_0\}} Q_{\beta_k}(x). \tag{5.2}$$

如果 $\mathcal{M}_1$ 满足*A6*，则方法 $\mathcal{M}_1$ 至多经过 x_{k+1} 次迭代即可生成满足子问题条件及（2.4）的 (2.3)

$$2C_{\mathcal{M}_1}(L_{f,1} + \beta_k L_{\phi,1})(f(x_0) - f_{\mathrm{low}})\varepsilon_1^{-2} \tag{5.3}$$

一阶预言机调用，其中$L_{f,\,1}$ 和$L_{\phi,\,1}$ 在引理*5.1*中定义。

证明。 根据A2， Q_{β_k} 有下界：

$$Q_{\beta_k}(x) = f(x) + \frac{\beta_k}{2}\|c(x)\|_2^2 \geq f(x) \geq f_{\text{low}}, \quad \text{for all } x \in \mathbb{R}^n. \tag{5.4}$$

令 $x \in \mathcal{L}_{Q_{\beta_k}}(\tilde{x}_{k,0})$，由于 $\|c(x_k)\| > \varepsilon_0$，引理 3.1 给出

$$f(x) \leq Q_{\beta_k}(x) \leq Q_{\beta_k}(\tilde{x}_{k,0}) \leq Q_{\beta_k}(x_0) = f(x_0) + \frac{\beta_k}{2}\|c(x_0)\|^2 \leq 2f(x_0) - f_{\text{low}}, \tag{5.5}$$

这表明

$$\mathcal{L}_{Q_{\beta_k}}(\tilde{x}_{k,0}) \subset \mathcal{L}_f(2f(x_0) - f_{\text{low}}).$$

因此，根据引理 5.1， ∇Q_{β_k} 在 $\mathcal{L}_{Q_{\beta_k}}(\tilde{x}_{k,0})$ 上是 $(L_{f,1} + \beta_k L_{\phi,1})$-Lipschitz连续的。根据 A6，由于对于所有 $x \in \mathbb{R}^n$， $\varepsilon_1 \leq \tau(x)$ 会生成满足 $\|\nabla Q_{\beta_k}(x_{k+1})\| \leq \tau(x_{k+1})$ 且 $Q_{\beta_k}(x_{k+1}) \leq Q_{\beta_k}(\tilde{x}_{k,0})$ 的 $\mathcal{M}_1$ x_{k+1}，最多需要

$$C_{\mathcal{M}_1}(L_{f,1} + \beta_k L_{\phi,1})(Q_{\beta_k}(\tilde{x}_{k,0}) - f_{\text{low}})\varepsilon_1^{-2} \tag{5.6}$$

次一阶预言机调用。结论由 (5.5) 得出。 □

5.2 不满足 PL 条件的内层一阶求解器

根据引理 5.2的内层评估复杂度，我们得到算法 1在约束违反不满足 PL 条件时的总评估复杂度如下。

定理5.3 （无PL条件下一阶内层求解器的总评估复杂度）。 在 *A2*、*A3*、*A5*、*A7*成立的条件下，假设在*QPM*（算法*1*）的每次迭代中，点 x_{k+1} 由一个单调一阶方法计算 $\mathcal{M}_1$ ，其初始点为

$$\tilde{x}_{k,0} = \operatorname{argmin}_{x \in \{x_k, x_0\}} Q_{\beta_k}(x).$$

若方法 $\mathcal{M}_1$ 满足 *A6*，则*QPM*至多经过 $(\varepsilon_0, \varepsilon_1)$ *-KKT* 点，所需

$$8\alpha\hat{T}(\varepsilon_0)C_{\mathcal{M}_1}(f(x_0) - f_{\text{low}})^2(L_{f,1} + L_{\phi,1})\varepsilon_0^{-2}\varepsilon_1^{-2} \tag{5.7}$$

ˆ次一阶预言机调用，其中 $L_{f、1}$ 和 $L_{\phi、1}$ 定义见引理 *5.1* ，而 $T(\varepsilon_0)$ 定义见（3.3）。

ˆ证明。 根据定理 3.2，外层迭代次数 $T(\varepsilon_0)$ 的上界为 $T(\varepsilon_0)$ 。由于$\|c(x_k)\| > \varepsilon_0$对于 $k = 1, \ldots, T(\varepsilon_0) - 1$，引理 5.2 表明，算法 **1**的第 k次迭代至多需要

$$2C_{\mathcal{M}_1}(L_{f,1} + \beta_k L_{\phi,1})(f(x_0) - f_{\text{low}})\varepsilon_1^{-2}$$

一阶预言机调用。引理3.1意味着 $1 \leq \beta_{T(\varepsilon_0)-1} \leq 4\alpha(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}$。由此可知，一阶预言机调用的总次数有如下界：

$$\begin{aligned}
&\sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{M}_1}(L_{f,1} + \beta_k L_{\phi,1})(f(x_0) - f_{\text{low}})\varepsilon_1^{-2} \\
&\leq \sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{M}_1}\beta_k(L_{f,1} + L_{\phi,1})(f(x_0) - f_{\text{low}})\varepsilon_1^{-2} \\
&\leq \hat{T}(\varepsilon_0)8\alpha C_{\mathcal{M}_1}(L_{f,1} + L_{\phi,1})(f(x_0) - f_{\text{low}})^2\varepsilon_0^{-2}\varepsilon_1^{-2}.
\end{aligned}$$

□

根据定理 5.3，若约束违反不满足 PL 条件，采用一阶内层求解器的 QPM 至多需要 $\mathcal{O}\big(|\log(\beta_0^{-1}\varepsilon_0^{-2})|\varepsilon_0^{-2}\varepsilon_1^{-2}\big)$ 次一阶预言机调用，才能找到一个 $(\varepsilon_0, \varepsilon_1)$-KKT 点。特别地，当 $\beta_0 = \mathcal{O}(\varepsilon_0^{-2})$时，复杂度界可简化为 $\mathcal{O}(\varepsilon_0^{-2}\varepsilon_1^{-2})$。

5.3 PL 条件下的一阶内层求解器

在约束违反满足 PL 条件时，总评估复杂度可降低 ε_0^{-1}倍（A4）。

定理 5.4（PL 条件下一阶内层求解器的总评估复杂度）。 在 *A2*、*A3*、 *A4*、 *A5*、 *A7*下，假设在 *QPM*（算法 *1*）的每次迭代中，点 x_{k+1} 均使用单调一阶方法 $\mathcal{M}_1$ 并以

$$\tilde{x}_{k,0} = \operatorname{argmin}_{x\in\{x_k,x_0\}} Q_{\beta_k}(x).$$

*I*若该方法 $\mathcal{M}_1$ 满足 *A6*，则 *QPM* 至多生成一个 $(\varepsilon_0,\ \varepsilon_1)$ *-KKT* 点，所需迭代次数不超过 t

$$2\alpha\tilde{T}(\varepsilon_0)C_{\mathcal{M}_1}(L_{f,1}+L_{\phi,1})(f(x_0)-f_{\text{low}})\max\left\{\frac{L_{f,0}+\varepsilon_1}{\sigma_{\min}},\frac{4(f(x_0)-f_{\text{low}})}{R}\right\}\varepsilon_0^{-1}\varepsilon_1^{-2} \tag{5.8}$$

一阶预言机调用，其中 $L_{f,1}$ 和 $L_{\phi,1}$ 定义见引理*5.1*， $\sigma_{\min}$ 和 R 定义见˜于 *A4*， $T(\varepsilon_0)$ 定义见 （4.15） ，而 $L_{f,0}$ 定义见引理 *4.3*。

˜证明。 根据定理 4.4，外层迭代次数 $T(\varepsilon_0)$ 上界为 $T(\varepsilon_0)$。由于$\|c(x_k)\|>\varepsilon_0$对于 $k=1,\ldots,\ T(\varepsilon_0)-1$，根据引理 5.2，算法 **1** 的第 k次迭代至多需要

$$2C_{\mathcal{M}_1}(L_{f,1}+\beta_k L_{\phi,1})(f(x_0)-f_{\text{low}})\varepsilon_1^{-2}$$

一阶预言机调用。引理 4.3表明

$$1\le\beta_{T(\varepsilon_0)-1}<\alpha\max\left\{\frac{L_{f,0}+\varepsilon_1}{\sigma_{\min}},\frac{4(f(x_0)-f_{\text{low}})}{R}\right\}\varepsilon_0^{-1}. \tag{5.9}$$

因此，一阶预言机调用的总次数有界于

$$\sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{M}_1}(L_{f,1}+\beta_k L_{\phi,1})(f(x_0)-f_{\text{low}})\varepsilon_1^{-2} \tag{5.10}$$

$$\le \sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{M}_1}(L_{f,1}+L_{\phi,1})\beta_k(f(x_0)-f_{\text{low}})\varepsilon_1^{-2} \tag{5.11}$$

$$\le T(\varepsilon_0)2C_{\mathcal{M}_1}(L_{f,1}+L_{\phi,1})(f(x_0)-f_{\text{low}})\alpha\max\left\{\frac{L_{f,0}+\varepsilon_1}{\sigma_{\min}},\frac{4(f(x_0)-f_{\text{low}})}{R}\right\}\varepsilon_0^{-1}\varepsilon_1^{-2} \tag{5.12}$$

$$\le 2\alpha\tilde{T}(\varepsilon_0)C_{\mathcal{M}_1}(L_{f,1}+L_{\phi,1})(f(x_0)-f_{\text{low}})\max\left\{\frac{L_{f,0}+\varepsilon_1}{\sigma_{\min}},\frac{4(f(x_0)-f_{\text{low}})}{R}\right\}\varepsilon_0^{-1}\varepsilon_1^{-2}.\ \square$$

根据定理 5.4，在约束违反满足 PL 条件（A4）时，采用一阶内层求解器的 QPM 至多需要 $\mathcal{O}\left(\left|\log(\beta_0^{-1}\varepsilon_0^{-1})\right|\varepsilon_0^{-1}\varepsilon_1^{-2}\right)$ 次一阶预言机调用，即可找到一个 $(\varepsilon_0,\ \varepsilon_1)$-KKT 点。特别地，当 $\beta_0=\mathcal{O}\left(\varepsilon_0^{-1}\right)$时，复杂度界简化为 $\mathcal{O}\left(\varepsilon_0^{-1}\varepsilon_1^{-2}\right)$。

6 二阶内层求解器的总评估复杂度

我们考虑由二阶方法（称为 $\mathcal{M}_2$）最小化子问题时，QPM 的总评估复杂度。我们详细介绍了在约束违反满足和不满足 PL 条件两种情况下的结果。

6.1 二阶内层求解器的引理

我们考虑关于问题以及二阶求解器（内部）评估复杂度的以下假设。

A8. 函数 f 和 c是三次连续可微的。

A9. 给定一个三次连续可微函数 F: $\mathbb{R}^n \to \mathbb{R}$，考虑采用方法 $\mathcal{M}_2$ 从 $\tilde{x} \in \mathbb{R}^n$开始最小化 F，其中

$$\mathcal{L}_F(\tilde{x}) = \{z \in \mathbb{R}^n | F(z) \leq F(\tilde{x})\}$$

有界。存在一个称为 $\Delta_{\max}$的算法常数 $\mathcal{M}_2$，使得 $\mathcal{M}_2$ 访问的所有试探点和迭代点都属于

$$\Omega_F = \{x + d \in \mathbb{R}^n | x \in \mathcal{L}_F(\tilde{x}) \ and \ \|d\| \leq \Delta_{\max}\},$$

并且方法 $\mathcal{M}_2$ 至多需要

$$C_{\mathcal{M}_2}(F(\tilde{x}) - F_{\text{low}})L^{1/2}\varepsilon^{-3/2}$$

次对 F、 ∇F和$\nabla^2 F$ 进行函数评估，以生成一个点 $x \in \mathbb{R}^n$ ，满足 $\|\nabla F(x)\| \leq \varepsilon$，其中 $\nabla^2 F$ 在Ω_F上是L-*Lipschitz*连续的。

具有最优最坏情况复杂度的无约束最小化二阶算法的开发，是一个非常活跃的研究领域。例如，Hamad 和 Hinder（2025）提出的信赖域方法，在加入上界 $\Delta_{\max}$ 以限制信赖域半径后，满足 A9。在 A5 和 A9的条件下，我们定义有界集

$$\Omega_* := \text{conv}\{x + d \in \mathbb{R}^n : x \in \mathcal{L}_f(2f(x_0) - f_{\text{low}}) \text{ and } \|d\| \leq \Delta_{\max}\}. \tag{6.1}$$

此外，在A8条件下，存在有限常数

$$L_{f,2} := \sup_{x \in \Omega_*} \|\mathrm{D}^3 f(x)\| \qquad \text{and} \qquad L_{P,2} := \sup_{x \in \Omega_*} \|\mathrm{D}^3 \phi(x)\|. \tag{6.2}$$

在下面的引理中，我们证明对所有 $\beta \geq 0$，海森矩阵 $\nabla^2 Q_\beta$ 在 Ω_* 上 Lipschitz连续。我们强调，我们并未假设 f 和 c 的某个导数具有 Lipschitz 连续性；我们仅假设 f 具有有界次水平集（A5）。

引理6.1。 在 *A2*、*A5*、*A8*、*A9*条件下，对任意 $\beta > 0$，海森矩阵 $\nabla^2 Q_\beta$ 在集合 $(L_{f,\,2} + \beta L_{\phi,\,2})$- *Lipschitz* 连续$\Omega_*$ (6.1)上。

证明。 集合 Ω_* 在 (6.1) 中定义，且由 A5 界定；($L_{f,2}$ 和 $L_{\phi,2}$) 中的常数 6.2) 均定义良好且有限。我们发现，对于所有 $x \in \Omega_*$ 和 $\beta > 0$,

$$\begin{aligned} \left\|\mathrm{D}^3 Q_\beta(x)\right\| &\leq \left\|\mathrm{D}^3 f(x) + \beta \mathrm{D}^3 \phi(x)\right\| \\ &\leq \left\|\mathrm{D}^3 f(x)\right\| + \beta \left\|\mathrm{D}^3 \phi(x)\right\| \\ &\leq L_{f,2} + \beta L_{\phi,2}. \end{aligned}$$

因此，对于所有 $x \in \Omega_*$， $\left\|\mathrm{D}^3 Q_\beta(x)\right\| \leq L_{f,2} + \beta L_{\phi,2}$ 成立，因此 $\nabla^2 Q_\beta$ 在 Ω_* 上是 $(L_{f,2} + \beta L_{\phi,2}$)-Lipschitz连续的。 □

由此可以推导出 QPM 的总评估复杂度界。我们首先给出二阶子问题求解器的内层评估复杂度。

引理 6.2（$\mathcal{M}_2$的内层评估复杂度）。 在 *A2*、*A5* 和 *A8*条件下，考虑算法 k *1* 中的迭代，其中单调二阶方法 $\mathcal{M}_2$ 最小化 Q_{β_k}，初始点为

$$\tilde{x}_{k,0} = \operatorname{argmin}_{x\in\{x_k,x_0\}} Q_{\beta_k}(x).$$

若 $\mathcal{M}_2$ 满足 *A9*，该方法 $\mathcal{M}_2$ 会生成 x_{k+1} 满足子问题条件及 (2.4)，且至多需要 (2.3)

$$2C_{\mathcal{M}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + \beta_k L_{\phi,2})^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \tag{6.3}$$

次二阶预言机调用，其中 $L_{f,\ 2}$ 和 $L_{\phi,\ 2}$ 定义见 (6.2)。

证明。 根据 A 2, $Q_{\beta_k}\ x\ \geq\ f_{\text{low}}\quad x\ \in\ \mathbb{R}^n\quad \mathcal{L}_{Q_{\beta_k}}\ \tilde{x}_{k\,0}\ \subset$, $\mathcal{L}_f\ 2f\ x_0\ - f_{\text{low}}\quad 5\ 5\quad \mathcal{M}_2\quad Q_{\beta_k}$ 我们有 () 对所有 都成立。我们还有 (,) (()) (。)。因此，从 $\tilde{x}_{k,0}$ 开始应用 的所有迭代点和试探点均留在集合 Ω_* (6.1)。此外，根据引理 $\nabla^2 Q_{\beta_k}$，海森矩阵 $(L_{f,2}+\beta_k L_{\phi,2})$在 Ω_* 上是 6.1-Lipschitz连续的。

因此，由于对所有 $x \in \mathbb{R}^n$, $\mathcal{M}_2\ \varepsilon_1 \leq \tau(x)$, x_{k+1}会生成满足 $\|\nabla Q_{\beta_k}(x_{k+1})\| \leq\tau(x_{k+1})$ 和 $Q_{\beta_k}(x_{k+1}) \leq Q_{\beta_k}(\{\text{v14},\ 0)$ 的结果，且至多需要

$$C_{\mathcal{M}_2}(Q_{\beta_k}(\tilde{x}_{k,0}) - f_{\text{low}})(L_{f,2} + \beta_k L_{\phi,2})^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \tag{6.4}$$

二阶预言机调用。结论由 $Q_{\beta_k}(\tilde{x}_{k,\ 0}) \leq 2f(x_0) - f_{\text{low}}$ (5.5) 得出。 □

根据这一结果，我们推导出算法 1 的总评估复杂度界。

6.2 不满足 PL 条件时的二阶内层求解器

本节给出在约束违反不满足 PL 条件时，采用二阶内层求解器的 QPM 的总评估复杂度界。

定理 6.3（采用二阶内层求解器且不满足 PL 条件时的总评估复杂度）。 在 *A2*、*A3*、*A5*、*A8* 条件下，假设 *QPM*（算法 *1*）的每次迭代中，点 x_{k+1} 由一个单调二阶方法 $\mathcal{M}_2$ 从以下初始点计算得到

$$\tilde{x}_{k,0} = \operatorname{argmin}_{x\in\{x_k,x_0\}} Q_{\beta_k}(x).$$

如果方法 $\mathcal{M}_2$ 满足 *A9*，那么 *QPM* 至多经过以下次数的迭代即可生成一个 $(\varepsilon_0,\ \varepsilon_1)$-*KKT* 点：

$$4\hat{T}(\varepsilon_0)C_{\mathcal{M}_2}\sqrt{\alpha}(f(x_0) - f_{\text{low}})^{\frac{3}{2}}(L_{f,2} + L_{\phi,2})^{\frac{1}{2}}\varepsilon_0^{-1}\varepsilon_1^{-\frac{3}{2}} \tag{6.5}$$

ˆ二阶预言机调用，其中 $L_{f,\ 2}$ 和 $L_{\phi,\ 2}$ 定义于 （6.2），而 T（ε_0）定义于 （3.3）。

ˆ证明。 根据定理 3.2，外层迭代次数 $T(\varepsilon_0)$ 上界为 $T(\varepsilon_0)$。由于$\|c(x_k)\| > \varepsilon_0$ 对 $k = 1, \ldots,$ $T(\varepsilon_0) - 1$成立，引理 6.2 给出，算法 k 1 的第 1 次迭代最多需要

$$2C_{\mathcal{M}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + \beta_k L_{\phi,2})^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \tag{6.6}$$

次二阶预言机调用。引理 3.1 给出 $1 \le \beta_{T(\varepsilon_0)-1} \le 4\alpha(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}$。因此，二阶预言机调用总数的上界为

$$\begin{aligned}
&\sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{M}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + \beta_k L_{\phi,2})^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \\
&\le \sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{M}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + L_{\phi,2})^{\frac{1}{2}}\beta_k^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \\
&\le 2T(\varepsilon_0)C_{\mathcal{M}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + L_{\phi,2})^{\frac{1}{2}}(4\alpha(f(x_0) - f_{\text{low}})\varepsilon_0^{-2})^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \\
&\le 4\hat{T}(\varepsilon_0)C_{\mathcal{M}_2}\sqrt{\alpha}(f(x_0) - f_{\text{low}})^{\frac{3}{2}}(L_{f,2} + L_{\phi,2})^{\frac{1}{2}}\varepsilon_0^{-1}\varepsilon_1^{-\frac{3}{2}}.
\end{aligned}$$

□

根据定理 6.3，若约束违反不满足 PL 条件，则采用二阶内层求解器的 QPM 至多需要 $\mathcal{O}\left(|\log(\beta_0^{-1}\varepsilon_0^{-2})|\varepsilon_0^{-1}\varepsilon_1^{-3/2}\right)$ 次二阶预言机调用，才能找到一个 $(\varepsilon_0, \varepsilon_1)$-KKT 点。特别地，当 $\beta_0 = \mathcal{O}(\varepsilon_0^{-2})$时，复杂度界简化为 $\mathcal{O}\left(\varepsilon_0^{-1}\varepsilon_1^{-3/2}\right)$。

6.3 PL 条件下的二阶内层求解器

现在，我们给出约束违反满足 PL 条件且子问题采用二阶求解器时的总评估复杂度界。

定理 6.4（采用 PL 条件下二阶内层求解器时的总评估复杂度）。 在 *A2*、*A3*、*A4*、*A5*、*A8*条件下，假设在 *QPM* 的每次迭代（算法 *1*）中，点 x_{k+1}由单调二阶方法计算得到 $\mathcal{M}_2$，其初始点为

$$\tilde{x}_{k,0} = \text{argmin}_{x\in\{x_k,x_0\}} Q_{\beta_k}(x).$$

如果方法 $\mathcal{M}_2$ 满足 *A9*，则 *QPM* 至多生成一个 $(\varepsilon_0, \varepsilon_1)$-*KKT*，所需的

$$2\tilde{T}(\varepsilon_0)C_{\mathcal{M}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + L_{\phi,2})^{\frac{1}{2}}\sqrt{\alpha}\max\left\{\frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\}^{\frac{1}{2}}\varepsilon_0^{-\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \quad (6.7)$$

次二阶预言机调用，其中 $L_{f,2}$ 和 $L_{\phi,2}$ 定义见 (6.2)，$\sigma_{\min}$ 和 R 定义见*A4*，

~ $T(\varepsilon_0)$ 定义见 (4.15)，且 $L_{f,0}$ 定义见引理 *4.3*。

~证明。 根据定理 4.4，外层迭代次数 $T(\varepsilon_0)$ 的上界为 $T(\varepsilon_0)$。引理4.3 给出

$$1 \le \beta_{T(\varepsilon_0)-1} < \alpha\max\left\{\frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\}\varepsilon_0^{-1}. \quad (6.8)$$

引理6.2 表明，算法 k 的第 1 次迭代至多需要

$$2C_{\mathcal{M}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + \beta_k L_{\phi,2})^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \quad (6.9)$$

二阶　　次预言机调用。因此，二阶预言机调用总数的上界为　　y

$$\begin{aligned}
&\sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{M}_2}(f(x_0)-f_{\text{low}})(L_{f,2}+\beta_k L_{\phi,2})^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \\
&\leq \sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{M}_2}(f(x_0)-f_{\text{low}})(L_{f,2}+L_{\phi,2})^{\frac{1}{2}}\beta_k^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \\
&\leq 2\tilde{T}(\varepsilon_0)C_{\mathcal{M}_2}(f(x_0)-f_{\text{low}})(L_{f,2}+L_{\phi,2})^{\frac{1}{2}}\sqrt{\alpha}\max\left\{\frac{L_{f,0}+\varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0)-f_{\text{low}})}{R}\right\}^{\frac{1}{2}} \varepsilon_0^{-\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}}.\square
\end{aligned}$$

根据定理 6.4，在约束违反满足 PL 条件（A4）的情况下，配备二阶内层求解器的 QPM 至多需要 $\mathcal{O}\left(\left|\log(\beta_0^{-1}\varepsilon_0^{-1})\right|\varepsilon_0^{-1/2}\varepsilon_1^{-3/2}\right)$ 次二阶预言机调用，即可找到一个 $(\varepsilon_0,\ \varepsilon_1)$-KKT 点。特别地，当 $\beta_0 = \mathcal{O}(\varepsilon_0^{-1})$时，复杂度界可简化为 $\mathcal{O}\left(\varepsilon_0^{-1/2}\varepsilon_1^{-3/2}\right)$。

7 示例数值结果

本节将展示我们的理论研究结果。我们通过数值实验展示子问题中采用可行性感知容差带来的性能提升，并比较子问题中一阶求解器与二阶求解器的性能。

我们的实验使用 Julia 实现的算法1。复现实验结果的代码可在 www.github.com/flgoyens/QPM获取。我们的实现包含用于求解子问题的一阶方法和二阶方法。一阶方法称为 QPM-GD，使用梯度下降法，并通过 Armijo 线搜索进行内层最小化。二阶方法称为 QPM-TR，使用信赖域法和精确海森矩阵进行内层最小化。信赖域子问题通过截断共轭梯度法求解（Conn 等人，2000）。在实验中，除非另有说明，我们使用 $\alpha = 1.2$ 和 $\beta_0 = 1$ 的取值。

在下面的实验中，我们比较两种子问题容差设定的表现：非自适应规则

$$\|\nabla Q_{\beta_k}(x_{k+1})\| \leq \varepsilon_1, \tag{7.1}$$

与自适应容差

$$\|\nabla Q_{\beta_k}(x_{k+1})\| \leq \tau_k := \max\left(\varepsilon_1, \frac{\varepsilon_1}{\varepsilon_0}\|c(x_{k+1})\|\right), \tag{7.2}$$

其中，τ_k 用作图中的标签，表示第 k 次迭代时的自适应容差　　.

我们的测试问题是在单位球面上最小化扩展 Rosenbrock 函数：对于 n 为偶数的情形，

$$\begin{aligned}
\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & \sum_{i=1}^{n/2}\left[100(x_{2i}-x_{2i-1}^2)^2+(1-x_{2i-1})^2\right] \\
\text{subject to} \quad & \|x\|^2 = 1.
\end{aligned} \tag{7.3}$$

问题是光滑的 A8，代价函数有下界 A2，具有强制性 A5，且约束函数满足 A4 作为例 2 的特例，其中 $p = 1$。初始点的选取

as

$$x_0 = \sqrt{\frac{1}{n}\left(1 + \frac{\varepsilon_0}{\sqrt{2}}\right)}\,\mathbf{1}_n,$$

是为了确保 $\|c(x_0)\| = \varepsilon_0/\sqrt{2}$ (A3)。

一阶内层最小化：QPM-GD

图1 展示了 QPM-GD 在维度为 $n = 1000$的问题 (7.2) 中的两种变体。蓝色曲线表示子问题中的自适应容差（标记为 τ_k），橙色曲线表示常数容差（7.1）（标记为 ε_1）。显然，自适应容差在每个子问题中所需的内层迭代次数更少。要达到 10^{-3}的精度，自适应 QPM 使用了 1570 个梯度步数，而非自适应 QPM 使用了3259个梯度步数。表2 显示，两种方法在容差设为 10^{-6} 时均以相同精度终止，并报告了自适应版本显著更少的预言机调用次数。

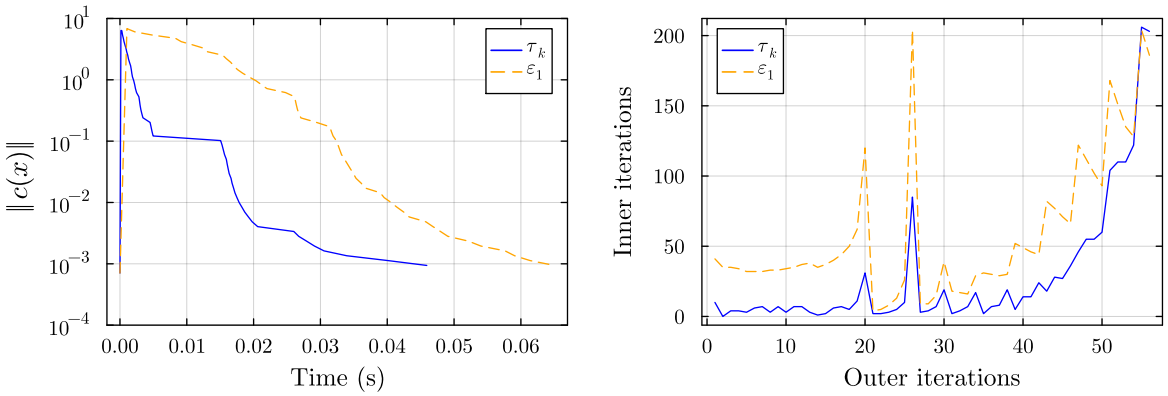


图 1：比较 QPM-GD 在问题 (7.3) 上采用自适应和固定子问题容差的表现，其中 $\beta_0 = 1$, $\alpha = 1.2$, $n = 10^3$, $\varepsilon_0 = \varepsilon_1 = 10^{-3}$。内层迭代总数：1570 对应 τ_k，以及3259 对应 ε_1。

方法	$\|c(x_{\text{final}})\|$	$f(x_{\text{final}})$	Q_β 评估	∇Q_β 评估
QPM-GD (τ_k)	7.1×10^{-7}	515.76	8441	4583
QPM-GD (ε_1)	7.1×10^{-7}	515.76	12079	7771

表 2：QPM-GD 与 τ_k 对比 ε_1 在问题（7.3）中，$n = 1000$, $\varepsilon_0 = \varepsilon_1 = 10^{-6}$.

二阶内层最小化：QPM-TR

图 2 展示了 QPM-GD 和 QPM-TR 在同一个问题实例（7.3）上的表现，其中 $n = 1000$，二者均在子问题中采用自适应容差。显然，QPM-TR 求解该问题所需的时间少得多。此外，表 3 显示，二阶方法终止时的函数值小于一阶方法。

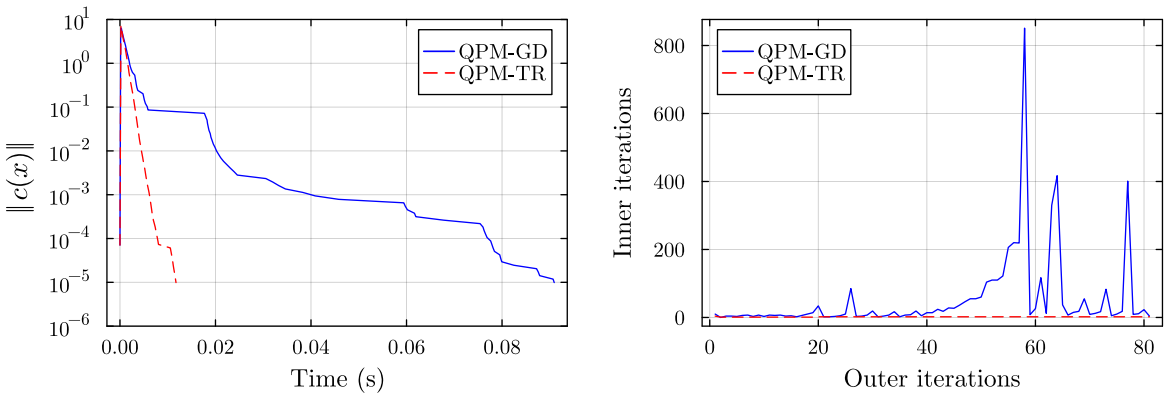


图 2：比较 QPM-GD 和 QPM-TR 在问题 (7.3) 上采用自适应子问题容差时的表现，其中 $\beta_0 = 1$，$\alpha = 1.2$，$n = 10^3$，$\varepsilon_0 = \varepsilon_1 = 10^{-5}$。内层迭代总数：4316 次（QPM-GD），以及146 次（QPM-TR）。

方法	$\|c(x_{\text{final}})\|$	$f(x_{\text{final}})$	Q_β 求值	∇Q_β 求值	$\nabla^2 Q_\beta$ 评估
QPM-GD（τ_k）	7.1×10^{-7}	515.76	8441	4583	0
QPM-TR（τ_k）	9.2×10^{-7}	456.31	548	265	262

表 3：QPM-GD 与 QPM-TR 在问题（7.3）上的表现，其中 $n = 1000$，$\varepsilon_0 = \varepsilon_1 = 10^{-6}$ 和 $\tau(x) = \tau_k$。

结论

本文分析了二次罚函数法（QPM）在光滑非凸等式约束优化问题上的最坏情况预言机复杂度，分别考察了约束违反满足和不满足 PL 条件的情形。在不满足 PL 条件时，对于采用适当一阶和二阶内层求解器的 QPM，我们分别建立了复杂度界 $\tilde{\mathcal{O}}(\varepsilon_0^{-2}\varepsilon_1^{-2})$ 和 $\tilde{\mathcal{O}}(\varepsilon_0^{-1}\varepsilon_1^{-3/2})$，以求得 $(\varepsilon_0, \varepsilon_1)$-KKT 点。在约束违反满足 PL 条件时，这些界分别改进为 $\tilde{\mathcal{O}}(\varepsilon_0^{-1}\varepsilon_1^{-2})$和 $\tilde{\mathcal{O}}(\varepsilon_0^{-1/2}\varepsilon_1^{-3/2})$，这体现了正则性假设所带来的可行性精度依赖性的改善。在这两种情形下，我们还证明，当初始罚参数选为与 ε_0 的适当次幂成比例时，对数依赖关系 ε_0^{-1}会消失。我们的分析适用于采用放宽子问题停止准则的 QPM 变体。利用这一灵活性，我们提出了一种可行性感知停止规则：当远离可行性时，自适应地放宽驻点精度要求。该准则保留了所有理论保证，并可显著提升实际求解速度，初步数值实验验证了这一点。

参考文献

Andreani, R., Birgin, E. G., Martínez, J. M., 和 Schuverdt, M. L. (2008). 关于具有一般下层约束的增广拉格朗日法。*SIAM* 优化期刊, 18(4):1286–1309。

Bertsekas, D. P. (1997). 非线性规划。运筹学会会刊, 48(3):334–334。

Birgin, E. G., Gardenghi, J. L., Martínez, J. M., Santos, S. A. 和 Toint, Ph. L.（2016）。基于未缩放KKT条件和高阶模型的非线性约束优化评估复杂度-

析。*SIAM* 优化期刊，26(2)：951–967。Birgin, E. G. 和 Martínez, J. M.（2020）。增广拉格朗日算法的复杂度与性能。最优化方法与软件，35(5)：885–920。Boumal, N.、Absil, P.-A. 和 Cartis, C.（2019）。流形上非凸优化的全局收敛率。*IMA*数值分析期刊，39(1)：1–33。Bueno, L. F. 和 Martínez, J. M.（2020）。约束优化中非精确恢复法的复杂度。*SIAM* 优化期刊，30(1)：80–101。Cartis, C.、Gould, N. I. 和 Toint, P. L.（2011）。复合函数极小化的评估复杂度及其在非凸非线性规划中的应用。*SIAM* 优化期刊，21(4)：1721–1739。Cartis, C.、Gould, N. I. 和 Toint, P. L.（2013）。潜在秩亏非线性最小二乘问题的三次正则化方法的评估复杂度及其与约束非线性优化的相关性。*SIAM* 优化期刊，23(3)：1553–1574。Cartis, C.、Gould, N. I. M. 和 Toint, P. L.（2014）。约束非线性优化中一阶临界点的求解复杂度。数学规划，144(1)：93–106。Cartis, C.、Gould, N. I. M. 和 Toint, P. L.（2022）。非凸优化算法的评估复杂度。*MOS-SIAM*最优化丛书。Conn, A. R.、Gould, N. I. 和 Toint, P. L.（2000）。信赖域方法，第1卷。SIAM。Curtis, F. E.、O’Neill, M. J. 和 Robinson, D. P.（2024）。非线性等式约束随机优化中SQP方法的最坏情况复杂度。数学规划，205(1)：431–483。Curtis, F. E.、Robinson, D. P. 和 Samadi, M.（2018）。等式约束优化中信赖漏斗算法的复杂度分析。*SIAM* 优化期刊，28(2)：1533–1563。Eckstein, J. 和 Silva, P. J. S.（2013）。增广拉格朗日函数的实用相对误差准则。数学规划，141(1-2)：319–348。Facchinei, F.、Kungurtsev, V.、Lampariello, L. 和 Scutari, G.（2021）。非凸约束优化中的幽灵罚项：递减步长与迭代复杂度。运筹数学，46(2)：595–627。Goyens, F.、Eftekhari, A. 和 Boumal, N.（2024）。等式约束下二阶点的计算：重新审视Fletcher增广拉格朗日函数。最优化理论与应用期刊，201(3)：1198–1228。Goyens, F. 和 Royer, C. W.（2025）。严格鞍函数的黎曼信赖域方法及其复杂度保证。数学规划，213(1)：863–905。Grapiglia, G. N.（2023）。非凸优化中二次罚函数法的最坏情况评估复杂度。最优化方法与软件，38(4)：781–803。Grapiglia, G. N. 和 袁亚湘（2021）。非凸优化中增广拉格朗日方法的复杂度。*IMA*数值分析期刊，41(2)：1508–1530。

Hamad, F. 和 Hinder, O.（2025）。一种简单实用的自适应信赖域方法。http://arxiv.org/abs/2412.02079。He, C.、Lu, Z. 和 Pong, T. K.（2023）。一种基于牛顿共轭梯度的增广拉格朗日方法，用于求解具有复杂度保证的非凸等式约束优化中的二阶驻点。*SIAM* 优化期刊，33(3):1734–1766。Kong, W.、Melo, J. G. 和 Monteiro, R. D. C.（2019）。用于求解线性约束非凸复合规划的二次罚函数加速非精确近端点方法的复杂度。*SIAM* 优化期刊，29(4):2566–2593。Li, Z.、Chen, P.-Y.、Liu, S.、Lu, S. 和 Xu, Y.（2021）。约束非凸优化的速率改进型非精确增广拉格朗日方法。载于第二十四届国际人工智能与统计学会议论文集，第2170–2178页。PMLR。Li, Z. 和 Xu, Y.（2021）。面向弱凸目标函数的凸约束规划的基于增广拉格朗日的 一阶方法。*INFORMS*优化期刊，3(4):373–397。Lin, Q.、Ma, R. 和 Xu, Y.（2020）。用于约束非凸优化的非精确近端点罚方法。http://arxiv.org/abs/1908.11518。Martínez, J. M.（2017）。约束优化中的高阶模型正则化。*SIAM* 优化期刊，27(4):2447–2458。Melo, J. G.、Monteiro, R. D. C. 和 Wang, H.（2020）。用于求解线性约束光滑非凸复合优化问题的非精确近端加速增广拉格朗日方法的迭代复杂度。http://arxiv.org/abs/2006.08048。Xie, Y. 和 Wright, S. J.（2021）。具有非线性等式约束的非凸优化中近端增广拉格朗日方法的复杂度。科学计算杂志，86(3):1–30。Zangwill, W. I.（1967）。通过罚函数进行非线性规划。管理科学，13(5):344–358。