arXiv:2302.00363v2 [math.OC] 2 Aug 2023

隐式增广拉格朗日与广义优化

Alberto De Marchi*

2023年8月3日

摘要

本文研究不作任何凸性假设的广义非线性规划，涵盖多种问题，包括非光滑目标函数、组合结构以及集值非线性约束。我们将增广拉格朗日框架扩展至这一广泛的问题类别，同时保留隐式形式，并仅将松弛变量作为形式上的工具引入。然而，由于对松弛变量进行边缘化，所得广义增广拉格朗日函数缺乏正则性。基于参数优化，我们提出一种专门的驻点概念，以更好地刻画作为一系列子问题近似解而生成的迭代点。利用这一变分刻画和提升表示，我们推导出合适的乘子更新规则，并进一步为一种带保护机制的增广拉格朗日方案建立渐近性质和收敛保证。一个说明性数值示例展示了去除凸性假设所带来的建模灵活性，并表明所倡导的隐式方法具有实际优势。

关键词。非光滑非凸优化、广义非线性规划、乘子方法、广义增广拉格朗日函数、近端算法**AMS学科分类**。65K10，90C26，65K05

*慕尼黑联邦国防军大学，航空航天工程系，应用数学与科学计算研究所，85577 诺伊比贝格，德国。 EMAIL alberto.demarchi@unibw.de，ORCID 0000-0002-3545-6898。

1 引言

优化问题通常蕴含丰富的数学结构可供研究。本文探讨不考虑任何凸性的广义规划情形，即如下形式的有限维优化问题

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \varphi(x) := f(x) + g(c(x)), \tag{P}$$

其中 x 为决策变量， f 和 c 均为光滑映射，且 g 仅为下半连续函数。问题数据 f 和 g 允许为非凸映射；非光滑代价函数 g 不一定连续，也不一定取实数值；映射 c 则可能是非线性的。非光滑项 g 与光滑映射 c 的复合带来了极大的建模灵活性，涵盖了广泛的问题类型 [36]；正则化项和约束均可编码其中，也可对它们进行组合 [9, 7, 13]。**(P)** 的重要特例包括非线性 [4] 规划、析取 [2]规划、结构化优化 [32, 40]，以及约束结构化优化 [15]。抽象问题 **(P)** 也被称为“扩展非线性规划” [36]，，它通过引入复合项，突破了非线性规划的传统形式，从而更好地表示应用中出现的结构。

求解约束优化问题的一项基本技术是增广拉格朗日函数（AL）框架 [35, 4, 6]。引入一个辅助变量 $z \in \mathbb{R}^m$后，**(P)** 可等价地改写为以下形式

$$\underset{x\in\mathbb{R}^n,\ z\in\mathbb{R}^m}{\text{minimize}} \quad f(x) + g(z) \qquad \text{subject to} \quad c(x) - z = 0, \tag{P_S}$$

其目标函数更简单且可分离，但受一些（显式）约束限制。人们在 [15, 引理 31]中观察到，与原问题（P）相比，以这种方式引入松弛变量并不会影响极小化解和驻点——考虑到 [3]，这多少令人惊讶。诸如（P_S）这样的约束极小化问题适合采用增广拉格朗日方法来求解；我们采用这种方法，因为它能轻松处理非光滑性 [18, 19]。增广拉格朗日方法是一种框架，而非单一算法；过去几十年来，人们提出了多种方法，以不同方式体现其基本思想。

自诞生以来， [24, 33, 34], 增广拉格朗日方案已得到深入而广泛的研究， [10, 4, 20, 6, 21], 也包括无限维情形 [26]。在凸情形下，人们认识到，“乘子法” [24, 33] 对应于通过近端点算法（PPA）求解相关的对偶问题 [34, 35]，，而对偶理论在这一对应关系中发挥着关键作用。随后，Bertsekas 利用充分的最优性条件和局部对偶性，在非线性规划中得到了类似结果 [4]。依托某种局部对偶性（而这又由某种局部凸性促成），增广拉格朗日方案可追溯至近端点算法的一种应用，从而确立了其收敛性。

Rockafellar 最近沿用了这一思路 [37, 38]，将这些想法扩展到更广泛的情形（P），其中 g 为凸。与这条至少可追溯至 [36, 第 2], 节的研究脉络不同，我们在此关注的是研究广义非线性规划中的增广拉格朗日方法，且不作任何凸性假设。当 g可能为非凸时，抽象框架（P）不仅涵盖了多种问题，还提供了一种更灵活、更具表现力的建模范式，无需

进行“适当的重构”。然而，舍弃 g 的凸性是有代价的，因为利用近端点算法的视角似乎更加困难。尽管如此，奠定经典“乘子法”基础的移位惩罚方法仍然适用，也可以直接加以利用。不过，g 的非凸性会导致其近端映射可能具有集值性，进而造成正则性缺失，由此带来困难。在此，我们旨在更好地理解其对（P）数值方法的分析与设计所产生的影响，包括理论性质、收敛保证和实际方面。

以下全程考虑下列总体假设，不再另行说明。技术性定义见第2.1节。

假设 1。在(P)中满足以下条件：

(i) $f\colon \mathbb{R}^n \to \mathbb{R}$ 和 $c\colon \mathbb{R}^n \to \mathbb{R}^m$ 连续可微；*(ii)* $g\colon \mathbb{R}^m \to \mathbb{R}$ 是真、下半连续且近端有界；*(iii)* $\inf \varphi \in \mathbb{R}$。

假设 1(iii)与问题的适定性有关，因为代价函数$\varphi := f + g \circ c$ 的定义域非空且有下界。根据假设 1(ii) 中的近端有界性，对于某些合适的步长，g 的近端映射是良定义的。此外，我们采用一项实用假设，即（仅）可用或易于计算以下计算预言机：

- 给定 $x \in \mathbb{R}^n$、代价函数值 $f(x) \in \mathbb{R}$ 和梯度 $\nabla f(x) \in \mathbb{R}^n$；
- 给定 $v \in \mathbb{R}^m$ 和 $\gamma \in (0, \gamma_g)$，其中 γ_g 是 g的近端有界阈值，近端点 $z \in \operatorname{prox}_{\gamma g}(v)$以及其中的代价函数值 $g(z) \in \mathbb{R}$；
- 给定 $x \in \mathbb{R}^n$ 和 $v \in \mathbb{R}^m$，约束函数值 $c(x) \in \mathbb{R}^m$ 以及转置雅可比-向量乘积 $c'(x)^\top v \in \mathbb{R}^n$，其中 $c'\colon \mathbb{R}^n \to \mathbb{R}^{m\times n}$ 表示 c的雅可比矩阵。

以下介绍的数值方法仅依赖这些基本运算，因此从设计上就是一阶且无矩阵的；因此，它只涉及简单操作，内存占用较低。

概要 第1.1节第1.1节介绍相关工作和主题，随后在 第1.2节介绍我们的贡献。第2.1节给出记号、预备材料和有用工具，随后在 第2.2节介绍相关的最优性和驻点概念。第3节详细介绍并刻画所提出的增广拉格朗日方案。第4节开发并分析内层增广拉格朗日问题的求解器。最后，在 第5节报告一个说明性数值示例，并在 第6节给出一些总结性意见。

1.1 相关工作

该模板（P）涵盖许多经典问题类别，包括结构化的两项问题 $f + g$ [32] 和三项问题 $f + g + h$，以及带有

由线性算子作用于非光滑项的情形，包括两项问题 $g + h \circ L$、三项问题 $f + g + h \circ L$ [28]，或带有线性约束的情形，例如交替方向乘子法 [7]。这些问题类别大多已在凸情形下得到广泛研究，而非凸性主要只在结构化的两项问题中得到考察 $f + g$ [40]。

结合增广拉格朗日方法，近期已有研究考虑非利普希茨或仅下半连续的代价函数 [8, 19, 15]。文献第2.1节定义了注意收敛；进行极限分析似乎有必要满足这一性质，但 [8] 忽略了这一问题， [15]则对此予以关注。由于难以事先验证，若要确保注意收敛，可以进一步假设 g 在其定义域上连续。尽管这一假设远弱于凸性，但正如文献 [16]所指出的，许多实际问题都满足这一条件。

另一个问题是增广拉格朗日算法的良定义性。事实上，增广拉格朗日子问题可能是不适定的，因为由于约束松弛 [6]，无法事先保证增广拉格朗日函数有下界。鉴于本文关注的是 g的作用，我们不考虑这种可能性，并假设所有子问题都是适定的。虽然可能需要满足其他条件，但有几种方法或可用于放宽这一假设，例如利用不可松弛约束进行局部化最小化 [1] ，或在子问题中加入信赖域，以及采用不同的增广拉格朗日方法 [21, 19]。

约束结构化优化 [15] 旨在最小化总和 $f(x) + g(x)$（变量为 $x \in \mathbb{R}^n$ ，且满足集合隶属约束 $c(x) \in D$，其中 c 是光滑映射，而 D 是非空闭集。因此，它比非线性规划和析取规划更一般 [4, 2], ，但属于模板（P）。反过来，（P）总可以通过引入一些辅助变量，重构为约束结构化规划。因此，在 [15] 中研究的问题类别，在P）假设 1下同样具有一般性，但我们在此采用的解策略本质上不同。

1.2 方法与贡献

本研究的核心思想可追溯至作者的博士论文 [13, 第 1]章。为了改进分析并弥补推理中的薄弱之处，尤其是解决子求解器适用的终止准则问题，后续研究采用了不同的思路，最终形成了 [15]。其中，如（$\mathrm{P_S}$）所示，引入松弛变量，使约束显式化，然后应用传统增广拉格朗日方案。相反，我们在此坚持隐式方法，该方法在 [13], 中提出，并提供严格的收敛性分析。具体而言，我们仅将提升重构（$\mathrm{P_S}$）作为理论工具，以此构建求解（P）的策略，并借助更精细的变分工具研究其性质。

对（$\mathrm{P_S}$）的增广拉格朗日函数关于松弛变量进行边缘化 z ，可得到（P）的广义（或隐式）增广拉格朗日函数 [18, 13, 38]。然而，由于 g的非凸性，广义增广拉格朗日函数通常不具备连续可微性（或近端友好性）。为克服这种正则性缺失，

- 我们提出了 Υ-驻点性的概念，从而能够精确刻画增广拉格朗日子问题所寻求的（近似）解，

由此更好地体现预期的理念。ϒ-驻点性植根于参数优化，既足以支持外层 AL 方案，又便于计算和数值处理。事实上，我们证明，这一概念在隐式形式和显式形式的 M-驻点性之间提供了一种中间约束资格。

有了 ϒ-驻点性的概念，我们将重点放在求解增广拉格朗日子问题上，即最小化增广拉格朗日函数。这构成了增广拉格朗日算法的主要计算任务和关键步骤。我们探讨直接求解增广拉格朗日子问题而不引入松弛变量的可能性。具体而言，

- 我们设计并分析了一种子求解器，它利用 g 的近端预言机来最小化广义增广拉格朗日函数，并通过生成近似的 ϒ-驻点，满足增广拉格朗日方法的要求。

令人想起神奇步骤 [11, 第 10.4 节。1], 子求解器背后的理念是，仅将松弛变量作为算法的副产品，而非决策变量。随后，在迭代方案中采用这些变量，以生成能够带来充分改进的更新，并以适当的（近似）驻点性证书作为终止条件。

总的来说，如果没有收敛结果，较弱的假设和更强的驻点概念就没有吸引力。因此，

- 我们研究了带保护的隐式增广拉格朗日方法在（P）下的渐近性质，并在 假设 1 下建立了与经典非线性规划中相当的（全局）收敛保证。

此外，我们还说明了如何在不引入松弛变量的情况下采用隐式方法，从而提升建模灵活性和实际性能。坚持使用紧凑的问题表述不仅可以避免增大问题规模，还能更好地利用现有的计算预言机，并有助于避免伪造的（局部）极小值和驻点。

2 记号与基础知识

本节将介绍记号、预备定义和一些有用的结果。

2.1 预备知识

用 $\mathbb{R}$ 和 $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ 分别表示实数轴和扩展实数轴。 $\mathbb{R}^n$ 中所有元素均为 0 或 1 的向量分别记作 0_n 和 1_n；若上下文已明确 n，则分别简写为 0 和 1。用 $\mathbb{B}_r(x)$ 表示以 x 为中心、半径为 r的闭球。扩展实值函数 $h\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ 的有效定义域记为 $\operatorname{dom} h := \{x \in \mathbb{R}^n \mid h(x) < \infty\}$。若 $\operatorname{dom} h \neq \emptyset$，则称 h为真函数；若对所有$\bar{x} \in \mathbb{R}^n$， $h(\bar{x}) \leq \liminf_{x \to \bar{x}} h(x)$，则称其下半连续（lsc）。对于某个常数 $\tau \in \mathbb{R}$，$\operatorname{lev}_{\leq \tau} h := \{x \in \mathbb{R}^n \mid h(x) \leq \tau\}$ 表示与 h对应的 τ-子水平集。若对每个 $\tau \in \mathbb{R}$，函数 h 的 τ-子水平集都是有界的，则称该函数（下）水平有界。接下来，我们引入一致水平有界性的概念，该概念在参数极小化中发挥重要作用 [39, 定义 1.16]。函数$f\colon \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ 取值于 $f(x,u)$，若对每个 x，在 u的局部一致意义下水平有界，则称其在

$\alpha \in \mathbb{R}$和 $\bar{u} \in \mathbb{R}^m$存在 $\varepsilon > 0$，使得集合 $\{(x,\ u) \mid f(x,\ u) \le \alpha,\ \|u - \bar{u}\| \le \varepsilon\}$在 $\mathbb{R}^n \times \mathbb{R}^m$中有界。我们使用记号 $F\colon\ \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 表示点到集合函数 $F\colon\ \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^m)$，其中 $\mathcal{P}(\mathbb{R}^m)$ 是 $\mathbb{R}^m$ 的幂集（即$\mathbb{R}^m$的所有子集构成的集合）。

给定一个真且下半连续函数 $h\colon\ \mathbb{R}^n \to \mathbb{R}$，以及一点 $\bar{x}$，使得 $h(\bar{x})$ 有限，我们可以不必假设 h 连续，而改为借助序列$\{x^k\}$的 h-注意收敛，记作 $x^k \xrightarrow{h} \bar{x}$，其中该序列由 $x^k \to \bar{x}$ 给出，并满足 $h(x^k) \to h(\bar{x})$。根据 [39, 定义 ˆ 8.3], ，我们将 h 的正则次微分记为 $\partial h\colon\ \mathbb{R}^n \rightrightarrows \mathbb{R}^n$，其中

$$v \in \hat{\partial} h(\bar{x}) \quad :\Leftrightarrow \quad \liminf_{\substack{x \to \bar{x} \\ x \neq \bar{x}}} \frac{h(x) - h(\bar{x}) - \langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \ge 0.$$

h 的（极限）次微分 为 $\partial h\colon\ \mathbb{R}^n \rightrightarrows \mathbb{R}^n$，其中 $v \in \partial h(\bar{x})$ 当且仅当$\bar{x} \in \operatorname{dom} h$，且存在序列 $\{x^k\}$ 和 $\{v^k\}$，使得 $x^k \xrightarrow{h} \bar{x}$ 且 $v^k \in \hat{}\partial h(x^k)$，其中 $v^k \to v$。对于任意在 $\bar{x}$ [39, 附近连续可微的 $h_0\colon\ \mathbb{R}^n \to \mathbb{R}$，$h$ 在 $\bar{x}$ 处的次微分满足 $\partial(h + h_0)(\bar{x}) = \partial h(\bar{x}) + \nabla h_0(\bar{x})$。练习 8.8]。

给定参数值 $\gamma > 0$，近端映射$\operatorname{prox}_{\gamma h}$ 定义为

$$\operatorname{prox}_{\gamma h}(x) := \operatorname*{arg\,min}_z \left\{ h(z) + \frac{1}{2\gamma}\|z - x\|^2 \right\}. \tag{1}$$

若 h 是近端有界的，则称其满足：它是真函数，且 $h + \|\cdot\|^2/(2\gamma)$ 在$\mathbb{R}^n$ 上有下界，其中某个 $\gamma > 0$ [39, 满足定义 1.23]。所有此类 γ 的上确界称为 h 的近端有界阈值γ_h。特别地，如果 h 有一个仿射函数作为下界，则 $\gamma_h = \infty$。当 h 下半连续时，对于任意 $\gamma \in (0,\ \gamma_h)$，近端映射 $\operatorname{prox}_{\gamma h}$ 是局部有界、非空且紧值的 [39, 定理 1.25]。定义近端映射的最小化问题的值函数称为参数为 $\gamma > 0$ 的*Moreau*包络，记作 $h^\gamma\colon\ \mathbb{R}^n \to \mathbb{R}$，即

$$h^\gamma(x) := \inf_z \left\{ h(z) + \frac{1}{2\gamma}\|z - x\|^2 \right\}.$$

为了描述某个非空、闭但不一定是凸集 $D \subset \mathbb{R}^m$ 的几何性质，我们将运用变分分析中的一些工具。投影映射$\Pi_D\colon\ \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ 以及距离函数 $\operatorname{dist}_D\colon\ \mathbb{R}^m \to \mathbb{R}$ 定义为

$$\Pi_D(v) := \operatorname*{arg\,min}_{z \in D} \|z - v\| \qquad \text{and} \qquad \operatorname{dist}_D(v) \equiv \operatorname{dist}(v, D) := \inf_{z \in D} \|z - v\|.$$

当 D 非凸时，前者是集值映射，而后者始终是单值的。D 的示性函数 记作 $\iota_D\colon\ \mathbb{R}^m \to \mathbb{R}$，定义为：当$v \in D$时，$\iota_D(v) = 0$；否则，$\iota_D(v) = \infty$。若 D 非空且闭，则 ι_D 是真且下半连续的。ι_D 的近端映射是投影 Π_D；因此，Π_D 局部有界。给定 $z \in D$，D 在 z 处的极限法锥 是闭锥

$$\mathcal{N}_D^{\lim}(z) := \limsup_{v \to z} \operatorname{cone}(v - \Pi_D(v)).$$

极限法锥具有稳健性，即对所有 $z \in D$，它满足 $\mathcal{N}_D^{\lim}(z) = \limsup_{v \to z} \mathcal{N}_{\lim D}(v)$。对于任意真且下半连续函数 $h\colon\ \mathbb{R}^n \to \mathbb{R}$ 以及点 $\bar{x} \in \operatorname{dom} h$，我们有

$$\partial h(\bar{x}) = \left\{ v \in \mathbb{R}^n \mid (v, -1) \in \mathcal{N}^{\lim}_{\operatorname{epi} h}(\bar{x}, h(\bar{x})) \right\}, \tag{2}$$

其中 $\operatorname{epi} h := \{(x,\ \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid h(x) \le \alpha\}$ 表示 h [39, 定理 8 的上图。9] .

2.2 驻点概念

现在我们定义一些基本概念，并讨论 (P) 的最优性和驻点性条件。由于代价函数 φ 可能取扩展实数值，因此点的可行性概念是指该代价函数的定义域。若点 $x^* \in \mathbb{R}^n$ 满足 $x^* \in \operatorname{dom} \varphi$，即 $c(x^*) \in \operatorname{dom}\ g$，则称其为 (P) 的可行点。对于 (P)，我们要寻找使 φ 取得最小值的点，这些点必然是可行的。最优性的概念可以是全局的，也可以是局部的。若存在 $\delta > 0$，使得对于所有可行的$x \in \mathbb{B}_\delta(x^*)$，均有 $\varphi(x^*) \leq \varphi(x)$，则称可行点 $x^* \in \mathbb{R}^n$ 是 (P) 的局部最优点，或称局部极小点。此外，如果该不等式对所有可行的 $x \in \mathbb{R}^n$都成立，则称 x^* 是 (P) 的（全局）最优点。

着眼于开发一种“经济实用”的算法，我们的方法旨在寻找候选极小化解，并以必要最优性条件为指导。在不作任何凸性假设的情况下，一种合理的驻点概念是基于极限次微分的概念；参见 [29, 第 3]节和 [30, 定理 5。48]对于适当下半连续函数 h: $\mathbb{R}^n \to \mathbb{R}$，若 $x^* \in \operatorname{dom} h$ 是*M*-驻点，则有$0 \in \partial h$（x^*），这是 x^* [39, 定理 10。1]中最优性的必要条件。对于（P），复合结构产生了用拉格朗日形式表示的一阶最优性条件 [37, 38], ，这可从（$\mathrm{P_S}$）直观地看出。尽管它与 M-驻点性密切相关，我们仍倾向于使用KKT-驻点性这一术语，以强调其拉格朗日结构。

定义 2.1（KKT 驻点性）。相对于（P），点 $x^* \in \mathbb{R}^n$ 称为 ***KKT*** 驻点，如果存在乘子 $y^* \in \mathbb{R}^m$ 使得

$$0 = \nabla f(x^*) + c'(x^*)^\top y^*, \tag{3a}$$
$$y^* \in \partial g(c(x^*)). \tag{3b}$$

注意，（3b）隐含要求 x^*满足可行性，即 c（x^*） $\in$ 属于 g，否则次微分 ∂g（c（x^*））为空。此外，注意到，如果定义 2.1 中的非线性规划（g （P）的示性函数对应于非空闭凸集 [4, 38, 15]，那么定义 2.1 与著名的卡鲁什-库恩-塔克（KKT）条件一致。

尽管在非凸情形下可能存在歧义，但可以将拉格朗日乘子 y 称为 对偶 变量，并将（x, y）称为 原始-对偶 对。在非凸情形下研究增广拉格朗日方法的收敛性时，事实上出现了某种局部对偶性，Bertsekas 指出了这一点 [4]，而 Rockafellar 最近也在更广泛的背景下揭示了这一点 [38, 37]——但对于满足P的（假设 1）而言，这似乎为时尚早。

给定容差 $\varepsilon \geq 0$，对于最小化问题 h，近似 M-驻点性概念指 $\mathit{dist}(0, \partial h(x^*)) \leq \varepsilon$。我们也可以考虑放宽条件，以定义 定义 2.1 的近似对应概念。

定义 2.2 (近似 KKT 驻点性)。给定 $\varepsilon \geq 0$。相对于（P），若存在 $x^* \in \mathbb{R}^n$，且 ε-*KKT* 驻点，则称该点为此类点；其中存在 $z^* \in \operatorname{dom} g$ 和一个乘子$y^* \in \mathbb{R}^m$ 使得

$$\|\nabla f(x^*) + c'(x^*)^\top y^*\| \leq \varepsilon, \tag{4a}$$
$$\operatorname{dist}(y^*, \partial g(z^*)) \leq \varepsilon, \tag{4b}$$
$$\|c(x^*) - z^*\| \leq \varepsilon. \tag{4c}$$

应注意，ε-KKT 驻点性与 $\varepsilon = 0$ 恢复了 KKT 驻点性的概念，这与 M-驻点性类似。该事实源于 $\partial g(z)$ 在任意点 $z \in \operatorname{dom} g$ [39, 定理 8。6], 这意味着，对于任意 $v \in \mathbb{R}^m$，$\operatorname{dist}(v, \partial g(z)) = 0$ 等价于包含关系 $v \in \partial g(z)$。

定义 2.2 所允许的近似不仅涉及驻点性 (4a) 和互补性 (4b)，也会影响可行性：由于 (4c)，可能有 $c(x^*) \notin \operatorname{dom} g$。当 c 是非线性映射时，由于处理复合 $g \circ c$ 可能较为困难，因此在实践中，涵盖这种不精确性尤为重要。相反，当 c 是恒等映射时，处理 (P) 则容易得多。在这一特殊情形下，与定义 2.2 相比，我们采用了更严格的定义，因为我们认为近似 M-驻点的可行性是一项重要性质。因此，对于带有 $c := \mathrm{Id}$ 的 (P)，若点 $x^* \in \mathbb{R}^n$ 满足 $\operatorname{dist}(-\nabla f(x^*), \partial g(x^*)) \leq \varepsilon$，则称其为 ε-M-驻点。除了可行性，即 $x^* \in \operatorname{dom} g$，这还要求比 (4) 更严格的近似。由于只需调用 g 的近端映射即可协调“约束” $x \in \operatorname{dom} g$，因此针对这类双项问题 $f + g$ 存在许多可行的方法；例如参见 [9, 32, 40, 17] 及其中的参考文献。

我们还将 KKT 驻点性的渐近对应概念纳入理论工具箱。下文与 [6, 定义 3.1], 中的近似 KKT 条件十分相似，原始迭代点 x^k 和拉格朗日乘子 y^k的含义也相同。

定义 2.3 （AKKT 驻点性）。对于（P），若可行点 $x^* \in \mathbb{R}^n$ 满足以下条件，则称其为渐近 *KKT* 驻点：存在序列 $\{x^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^n$ 和$\{y^k\}_{k\in\mathbb{N}}$，$\{z^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^m$ 使得 $z^k \xrightarrow{g} c(x^*)$ 并且

$$\nabla f(x^k) + c'(x^k)^\top y^k \to 0, \tag{5a}$$
$$y^k \in \partial g(z^k), \tag{5b}$$
$$c(x)^k - z^k \to 0. \tag{5c}$$

注意到，定义 2.3 并不要求序列 $\{y^k\}_{k\in\mathbb{N}}$ 收敛。（原始）不可行性的一种度量由（5c）给出，其中引入了 z^k，以便在（5b）中考虑到条件 $c(x^k) \in \operatorname{dom} g$ 可能在迭代过程中不成立，尽管它渐近地应当成立。类似地，通过（5a）可以监测驻点条件中的误差，或对偶不可行性。最后，在对（$z^k \xrightarrow{g} c(x^*)$）取极限时，注意收敛5）中起到作用，因为我们旨在恢复 KKT 驻点性所需的极限次微分；参见定义 2.1。

(P的局部极小点在满足适当的约束资格条件时是 KKT 驻点；由于 g的非利普希茨性，该条件也取决于函数 g，参见 [22]。

然而，我们可以证明，(P的每个局部极小点始终都是 AKKT-驻点。相关结果见 [27, 定理 6。2], [29, 第5节。1] 和 [15, 命题 2。5]。

命题 2.4. 设 $x^* \in \mathbb{R}^n$ 是(P)的局部极小值点。那么， x^* 是 *AKKT*-驻点。

证明。 由 x^* 是 (P) 的局部最优性可知，存在 $\delta > 0$ 使得 $\varphi(x^*) \leq \varphi(x)$ 对所有可行的 $x \in \mathbb{B}_\delta(x^*)$ 均成立。因此， x^* 是局部化问题的唯一全局极小点

$$\underset{x\in\mathbb{B}_\delta(x^*)}{\text{minimize}} \quad \varphi(x) + \frac{1}{2}\|x - x^*\|^2. \tag{6}$$

现在考虑惩罚代理问题

$$\underset{\substack{x\in\mathbb{B}_\delta(x^*)\\ z\in\mathbb{B}_1(c(x^*))}}{\text{minimize}} \quad f(x) + g(z) + \frac{\varrho_k}{2}\|c(x) - z\|^2 + \frac{1}{2}\|x - x^*\|^2 \tag{P(k)}$$

其中， $k \in \mathbb{N}$ 是任意的， $\varrho_k > 0$，且序列 $\{\varrho_k\}_{k\in\mathbb{N}}$ 满足当 $k \to \infty$时 $\varrho_k \to \infty$。注意到该优化问题的目标函数是下半连续的，且其可行集非空（由 x^* 的可行性可知）并且紧，因此，对每个 $k \in \mathbb{N}$，它都存在一个全局极小点 (x^k, z^k)。因此，不失一般性，在抽取一个子序列后，我们假设 $x^k \to \widehat{x}$ 且 $z^k \to \widehat{z}$，其中某个 $\widehat{x} \in \mathbb{B}_\delta(x^*)$ 满足 $\widehat{z} \in \operatorname{dom} g \cap \mathbb{B}_1(c(x^*))$.

我们现在论证 $\widehat{x} = x^*$ 和 $\widehat{z} = c(\widehat{x}) = c(x^*)$。为此，我们注意到，$(x^*, c(x^*))$ 对 (P(k)) 是可行的，由此得到如下估计

$$f(x^k) + g(z^k) + \frac{\varrho_k}{2}\|c(x^k) - z^k\|^2 + \frac{1}{2}\|x^k - x^*\|^2 \leq \varphi(x^*) \tag{7}$$

对每个 $k \in \mathbb{N}$。利用 g 的下半连续性、 $\varphi(x^*)$ 的有限性，以及收敛关系 $c(x^k) \to c(\widehat{x})$ 和 $z^k \to \widehat{z}$，对 (7) 中的 $k \to \infty$ 取极限可得 $c(\widehat{x}) = \widehat{z} \in \operatorname{dom} g$。因此， $\widehat{x}$ 对 (6) 是可行的，而 x^* 在 (P) 上的局部最优性意味着 $\varphi(x^*) \leq \varphi(\widehat{x})$。此外，利用 (7) 以及每个 $(x^k, z^k) \in \mathbb{R}^n \times \operatorname{dom} g$ 的最优性，我们得到 $\widehat{x} = x^*$，因为

$$\varphi(\widehat{x}) + \frac{1}{2}\|\widehat{x} - x^*\|^2 \leq \liminf_{k\to\infty} \left(f(x^k) + g(z^k) + \frac{\varrho_k}{2}\|c(x^k) - z^k\|^2 + \frac{1}{2}\|x^k - x^*\|^2 \right) \leq \varphi(x^*) \leq \varphi(\widehat{x}).$$

此外，注意到 (7) 也可得 $f(x^k) + g(z^k) \leq \varphi(x^*)$ 对每个 $k \in \mathbb{N}$，由

$$\varphi(x^*) \leq \liminf_{k\to\infty} f(x^k) + g(z^k) \leq \limsup_{k\to\infty} f(x^k) + g(z^k) \leq \varphi(x^*)$$

$\{z^k\}_{k\in\mathbb{N}}$ 注意收敛到 $\widehat{z} = c(x^*)$，即$z^k \xrightarrow{g} c(x^*)$。

现在，由于 $x^k \to x^*$ 和 $z^k \to c(x^*)$，不失一般性，我们可以假设$\{x^k\}_{k\in\mathbb{N}}$ 和 $\{z^k\}_{k\in\mathbb{N}}$ 分别取自 $\mathbb{B}_\delta(x^*)$ 和 $\mathbb{B}_1(c(x^*))$ 的内部，因为最终情况确实如此。因此，对于每个 $k \in \mathbb{N}$，(x^k, z^k) 是一个无约束（即约束不起作用的）极小值点，其

$$(x, z) \mapsto f(x) + g(z) + \frac{\varrho_k}{2}\|c(x) - z\|^2 + \frac{1}{2}\|x - x^*\|^2,$$

相关的最优性必要条件为

$$0 = \nabla f(x^k) + \varrho_k c'(x^k)^\top [c(x^k) - z^k] + x^k - x^*,$$
$$0 \in \partial g(z^k) - \varrho_k [c(x^k) - z^k].$$

因此，对每个 $y^k := \varrho_k [c(x^k) - z^k]$，令 $k \in \mathbb{N}$，则 (5) 中的条件可由 $x^k \to x^*$ 和 $z^k \to c(x^*)$ 推出。总之，这表明 x^* 是 ((P)) 的 AKKT-驻点。 □

若要保证 (P) 的局部极小点不仅是 AKKT-驻点，也是 KKT 驻点，就必须具备约束资格条件。类似于[15, 定义 2.6], ，以下正则性概念将第3节中的（相对较弱的）约束资格条件推广到 ([29, 第3节。2]P) 的非利普希茨情形，并且与中引入的一致约束资格条件密切相关 [27, 定义 6。8]。

定义 2.5（AKKT-正则性）。设 $\bar{x} \in \mathbb{R}^n$ 是 (P) 的一个可行点，即 $c(\bar{x}) \in \operatorname{dom} g$，并定义集值映射 $\mathcal{M}\colon \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n$，其定义为 $\mathcal{M}(x,z) := c'(x)^\top \partial g(z)$。那么，若满足

$$\limsup_{\substack{x \to \bar{x} \\ z \xrightarrow{g} c(\bar{x})}} \mathcal{M}(x,z) \subset \mathcal{M}(\bar{x}, c(\bar{x})).$$

关于定义 2.5中更一般的约束资格条件成立的充分条件，可以按照 [27]中的论证得到。由命题 2.4可直接得出以下结果，其思路与 [27, 命题 6。9]和 [15,推论 2。7]类似。

推论 2.6。 设 $x^* \in \mathbb{R}^n$ 是（P）的一个 *AKK*-正则 *AKKT*-驻点。那么，x^* 是（P）的一个 *KKT* 驻点。特别地，（P）的每个 *AKK*-正则局部极小点都是 *KKT* 驻点。

3 增广拉格朗日方法

增广拉格朗日框架中自然会出现具有嵌套结构的算法，该框架建立在一系列子问题之上。我们将在第 3.1 节介绍隐式方法并推导广义增广拉格朗日函数。随后，在第 3.2 节 中，我们将讨论与增广拉格朗日子问题相关的驻点概念，以便精确界定子问题“解”的资格，并由此明确子求解器应当输出的结果。接下来的第 3.3 节将详细阐述我们的算法框架，其收敛性分析见第 3.4 节。合适的终止准则将在第 3.5 节中讨论。随后第 4 节 将专门讨论增广拉格朗日子问题的数值求解。

3.1 隐式增广拉格朗日

引入一个辅助变量 $z \in \mathbb{R}^m$，（P）可等价地改写为形式（P_S），其目标函数具有结构化形式，并包含一些（显式）等式约束。

对于某个罚参数 $\mu > 0$，定义与（$\mathcal{L}_\mu^{\mathrm{S}}$: $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$相关联的增广拉格朗日函数 P_{S}）如下：

$$\mathcal{L}_\mu^{\mathrm{S}}(x,z,y) := f(x) + g(z) + \langle y, c(x) - z\rangle + \frac{1}{2\mu}\|c(x) - z\|^2 \\ = f(x) + g(z) + \frac{1}{2\mu}\|c(x) + \mu y - z\|^2 - \frac{\mu}{2}\|y\|^2. \quad (8)$$

在增广拉格朗日方案 [4, 10, 6], 中，要求对增广拉格朗日函数进行最小化（至少达到近似驻点性）。由于该函数是一个光滑函数与一个近端友好函数之和，因此可以利用其结构和计算预言机，通过例如近端梯度方法 [9, 32, 40]来求解 $\mathcal{L}_\mu^{\mathrm{S}}$ 的最小化问题。得益于 $\mathcal{L}_\mu^{\mathrm{S}}$的便利结构， [15] 中提出的方法依赖于这种“显式”方法。这里，我们不对 $\mathcal{L}_\mu^{\mathrm{S}}(\cdot,\cdot,y)$ 关于 (x,z) 进行联合最小化，而是利用原问题 (P) 所产生的结构，按照近端增广拉格朗日方法 [18, 13, 23, 14]的思路，正式消去松弛变量 z。给定某个 $\mu > 0$ 和 $y \in \mathbb{R}^m$，我们将预言机 $\mathbf{O}_\mu(\cdot,y)$ 定义为与 $\mathcal{L}_\mu^{\mathrm{S}}(\cdot,\cdot,y)$ 关于 z的逐点最小化相关联的映射。根据 (1) 和 (8) 中的定义，集值映射 $\mathbf{O}_\mu(\cdot,y)$: $\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 为

$$\mathbf{O}_\mu(x,y) := \operatorname*{arg\,min}_{z\in\mathbb{R}^m} \mathcal{L}_\mu^{\mathrm{S}}(x,z,y) = \operatorname*{arg\,min}_{z\in\mathbb{R}^m} \left\{ g(z) + \frac{1}{2\mu}\|c(x) + \mu y - z\|^2 \right\} \\ = \operatorname{prox}_{\mu g}\left(c(x) + \mu y\right). \quad (9)$$

将任意元素代回 $\mathcal{L}_\mu^{\mathrm{S}}(x,\cdot,y)$ 中的$\mathbf{O}_\mu(x,y)$，亦即在对应于对 z进行最小化的集合上计算增广拉格朗日函数，可得（单值的）增广拉格朗日函数 $\mathcal{L}_\mu$: $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 与 (P) 相关联：

$$\mathcal{L}_\mu(x,y) := \inf_{z\in\mathbb{R}^m} \mathcal{L}_\mu^{\mathrm{S}}(x,z,y) = \mathcal{L}_\mu^{\mathrm{S}}(x,z,y)\Big|_{z\in\mathbf{O}_\mu(x,y)} \\ = f(x) + g^\mu(c(x) + \mu y) - \frac{\mu}{2}\|y\|^2. \quad (10)$$

我们强调， g^μ: $\mathbb{R}^m \to \mathbb{R}$ 的 g Moreau包络是实值且严格连续的 [31], [39, 例 10。32], 但一般而言并非连续可微，因为 g 的近端映射可能是集值的。将其视为 (P) 的增广拉格朗日函数时， $\mathcal{L}_\mu$ 被称为广义增广拉格朗日函数，见 [38] (其中 g 为凸函数)。与采用显式松弛变量 $\mathcal{L}_\mu^{\mathrm{S}}$ 的增广拉格朗日函数相比，后者用于提升后的“约束”形式 (显式松弛变量 z 的形式 (P_{S})，我们也可将 $\mathcal{L}_\mu$ 称为隐式增广拉格朗日函数（对应于 (P)）。

我们将在第3节.3 介绍的算法属于增广拉格朗日框架 [4, 6, 38]，因为它生成满足以下递推关系的迭代点序列 x^k 和拉格朗日乘子 y^k ：

$$x^{k+1} \quad \operatorname*{arg\,min}_{x\in\mathbb{R}^n} \mathcal{L}_{\mu_k}(x,\hat{y}^k), \qquad y^{k+1} \quad \hat{y}^k + \frac{1}{\mu_k}\partial_y \mathcal{L}_{\mu_k}(x^{k+1},\hat{y}^k), \quad (11)$$

关于适当的参数值序列 $\mu_k > 0$ 和乘子估计 $\hat{}y^k \in \mathbb{R}^m$。在深入探讨细节之前，我们希望先让读者对算法有一个大致了解；

通过在（11）中使用符号“ ”，我们强调这里的记号不必过于较真。尤其是，对于（11）中的原始变量更新，我们不要求其为 全局或精确的局部极小点 $\mathcal{L}_{\mu_k}(\cdot, \hat{}y^k)$，只要求它是一个近似驻点（定义见第 3.2 节）。对于乘子，当（11）中的更新规则应用于非线性规划 [4, 6], 时，会化为传统形式；但如果 g 是非凸的，则可能需要从相关的集值次微分中选取特定元素。这种可能的选择也与第 3.2 节中引入的驻点概念密切相关。

在增广拉格朗日方法的第 k 次迭代中，需要近似求解如下形式的子问题

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \mathcal{L}_{\mu_k}(x, \hat{y}^k), \tag{12}$$

其中给定了 $\mu_k > 0$ 和 $\hat{}y^k \in \mathbb{R}^m$。由于增广拉格朗日函数可能是 g 非凸的，因此可能缺乏正则性。第 3.2 节将 引入一种驻点概念，它比 *(12)* 的 M-驻点性弱，但足以保证第 3.3 节详述的增广拉格朗日方法所生成的迭代点具有适当的渐近行为和性质，这一点将在后续第 3.4 节中展示。

3.2 参数优化的驻点性

鉴于推导 ($\mathcal{L}_\mu$ 中的边缘化步骤 10)，我们现在从参数优化的角度研究增广拉格朗日子问题 (12), 参见 [39, 第 1.F 节、第 10.C 节]。考虑一个目标函数 $p\colon\ \mathbb{R}^n \to \mathbb{R}$ 以及一个预言机 $\mathbf{O}\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 定义如下

$$p(x) := \inf_{z\in\mathbb{R}^m} P(x,z) \qquad \text{and} \qquad \mathbf{O}(x) := \operatorname*{arg\,min}_{z\in\mathbb{R}^m} P(x,z) \tag{13}$$

对于真下半连续函数 $P\colon\ \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_{\infty}$ 我们关注的是最小化 p，但首先要设计合适的驻点概念来刻画“解”。这一设定显然对应于子问题 (12)，其中显式和隐式增广拉格朗日函数 $\mathcal{L}^{\mathrm{S}}$ 和 $\mathcal{L}$ 分别扮演 P 和 p 的角色。此外，这也凸显了我们的意图：通过利用预言机 $z\mathbf{O}$，**使求解过程保持隐式，从而有效隐藏辅助变量**。

点 $\bar{x} \in \operatorname{dom} p$ 对于给定的 $\varepsilon \geq 0$ 是近似M-驻点性，这意味着要求 $\operatorname{dist}(0, \partial p(\bar{x})) \leq \varepsilon$。然而，由于 p 具有参数性质，次微分 ∂p 通常并非简单对象。下面考察参数优化中次微分运算的一些结果。回顾第2.1节，一致水平有界性的概念对应于水平有界性的参数扩展[39, 定义1.16], 我们假设 $P(x, z)$ 关于 z 在 x 局部一致地水平有界。那么，根据 [39, 定理10.13]，对于每个 $\bar{x} \in \operatorname{dom} p$，都有如下包含关系

$$\partial p(\bar{x}) \subseteq \Upsilon(\bar{x}), \qquad \text{where} \qquad \Upsilon(\bar{x}) := \bigcup_{\bar{z}\in\mathbf{O}(\bar{x})} \{v \,|\, (v,0) \in \partial P(\bar{x}, \bar{z})\}, \tag{14}$$

这意味着不等式

$$\operatorname{dist}(0, \partial p(\bar{x})) \geq \operatorname{dist}(0, \Upsilon(\bar{x})). \tag{15}$$

这为 M-驻点性度量提供了一个下界，但通常无法提供适合作为终止条件的上界。尽管如此，在（13）的参数优化设置中，我们可以采用以下驻点概念，此处称为 Υ-驻点性。

定义 3.1（Υ-驻点性）。令 $\varepsilon \geq 0$为固定值，并考虑真下半连续函数$P\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$，其中 $P(x,z)$ 在 z中关于 x局部一致地水平有界。分别按（13）和（14）定义$p\colon \mathbb{R}^n \to \mathbb{R}$和 $\Upsilon\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n$。那么，相对于 p，若点 $x^* \in$属于$\operatorname{dom} p$且 $\operatorname{dist}(0, \Upsilon(x^*)) \leq \varepsilon$，则称其为 ε-Υ-驻点。在 $\varepsilon = 0$的情形下，这样的点 x^* 称为 Υ-驻点。

注意，隐含地要求包含关系 $x^* \in \operatorname{dom} p$ 成立，以确保 $\Upsilon(x^*)$非空。此外，在精确情形下，即 $\varepsilon = 0$，根据定理 10.13]的闭性， Υ-驻点性的要求可简化为 $0 \in \Upsilon(x^*)$。

根据 Υ 在（14）中的定义，似乎可以将一个 ε-Υ-驻点 $x^* \in \mathbb{R}^n$ 与一个（可能不唯一的）证书 $z^* \in \mathbf{O}(x^*)$ 以及一对 （（x^*， z^*）且满足

$$\begin{aligned}\operatorname{dist}(0, \Upsilon(x^*)) = \min_{v \in \Upsilon(x^*)} \|v\| = \min_{\substack{v \in \mathbb{R}^n \\ z \in \mathbf{O}(x^*)}} \{\|v\| \,|\, (v,0) \in \partial P(x^*, z)\} \\ \leq \min_{v \in \mathbb{R}^n} \{\|v\| \,|\, (v,0) \in \partial P(x^*, z^*)\} \leq \varepsilon. \quad (16)\end{aligned}$$

如果找到任意这样一对（x^*， z^*），则（16）中的上界就为 ε-Υ-驻点性提供证书： x^*。

以下结果表明，（近似）M-驻点性比（近似） Υ-驻点性具有更强的约束资格，这意味着前者是后者的充分条件，而一般而言并非必要条件。所需的蕴含关系可直接由（15）中关于 Υ-驻点性度量的上界推出。

命题 3.2。 令 $\varepsilon \geq 0$ 固定。则，相对于 p 的最小化，任何ε-*M*-驻点也都是 ε-Υ-驻点。

反向蕴含通常不成立，但在某些（可能是局部的）凸性假设下成立；例如参见 [39, 定理 2.26]。对于下面的结果也有类似的考虑，该结果表明，相对于 p 的（近似）Υ-驻点性，比相对于 P 的（近似）M-驻点性具有更强的约束资格。

命题 3.3。 设 $\varepsilon \geq 0$ 固定，且(x^*, z^*) 是与 p 相关的任意 ε-Υ-驻点对。那么，相对于 P 的最小化，(x^*, z^*) 是 ε-*M*-驻点。

证明。 根据 ε-Υ-驻点性相对于 p，该对 (x^*, z^*) 必然满足 $x^* \in \operatorname{dom} p$和 $z^* \in \mathbf{O}(x^*)$。接着，利用 (16) 中的上界，可以推得

$$\begin{aligned}\operatorname{dist}(0, \partial P(x^*, z^*)) = \min_{\substack{v \in \mathbb{R}^n \\ u \in \mathbb{R}^m}} \{\|(v,u)\| \,|\, (v,u) \in \partial P(x^*, z^*)\} \\ \leq \min_{v \in \mathbb{R}^n} \{\|v\| \,|\, (v,0) \in \partial P(x^*, z^*)\} \leq \varepsilon,\end{aligned}$$

其中，第一个不等式是由限制条件 $u := 0$ 得出的。证明 (x^*, z^*) 满足 $\mathrm{dist}(0, \partial P(x^*, z^*)) \leq \varepsilon$，即可证得所述结论。 □

然而请注意，一般而言，命题 3.3 中的结论并不适用于所有$z^* \in \mathbf{O}(x^*$，即使 x^* 是 ε-Υ-驻点，也不例外。 p事实上，有些点对（x^*， z^*）可能不满足（16），而且它们未必是 ε-M-驻点。 P

此时显然，对于（8）和（10），预言机$\mathbf{O}$ 指的是$\mathbf{O}^{\mu}(\cdot, \hat{y})$，其中给定了某个 $\mu > 0$ 和 $\hat{y} \in \mathbb{R}^m$。此外，我们将重点关注具有以下结构的特殊情形： P 具有如下结构

$$P(x, z) := F(x, z) + g(z), \tag{17a}$$

$$F(x, z) := f(x) + \frac{1}{2\mu}\|c(x) + \mu\hat{y} - z\|^2, \tag{17b}$$

其中 F： $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 是连续可微的。

3.3 算法

本节详细介绍一种用于求解形式为（P）的广义规划的增广拉格朗日方法。本文着重研究一种受 [6, 算法 4.1] 启发的带保护机制的增广拉格朗日方案，并考察其收敛性质。与用于求解非线性规划的经典增广拉格朗日方法或乘子罚函数方法相比， [10, 4], 该变体对拉格朗日乘子采用带保护机制的更新规则，且具有更强的全局收敛性质[26]。虽然我们的分析仅限于这一特定算法，但稍作修改即可得到其他算法的类似结果，例如 [19, 算法 1]。

整体方法见 算法 3.1。虽然没有明确要求起始点，但在 步骤 4 求解子问题时，应以当前原始变量估计为热启动 x^{k-1}，从而利用初始猜测。增广拉格朗日子问题出现在 步骤 4，其中包含罚参数为当前值 μ_k 、估计值为 $\hat{y}^k$ 的隐式增广拉格朗日函数。关键是，步骤 4 要求计算一个近似 Υ-驻点对(x^k, z^k)，而不只是点 x^k，针对 $\mathcal{L}_{\mu_k}(\cdot, \hat{y}^k)$。由于它可能用于验证近似驻点性，因此很可能无需额外工作即可获得 z^k 关于 x^k 的证书——第 4 节开发的子求解器便是如此。

算法 3.1： 针对 (P) 的带保护机制的隐式增广拉格朗日方法

数据： $\epsilon > 0$

结果： ϵ-KKT 驻点 x^*

1 选择 $\mu_0 > 0$, $\theta, \kappa \in (0,1)$, $Y \subset \mathbb{R}^m$ 非空有界

2 **对于** $k = 0, 1, 2 \ldots$ **执行**

3 选择 $\hat{y}^k \in Y$ 和 $\varepsilon_k \geq 0$

4 计算 $\mathcal{L}$的一个 ε_k-Υ-驻点对 (x^k, z^k) $_{\mu_k}(\cdot, \hat{y}^k)$

5 令 $y^k \leftarrow \hat{y}^k + \mu_k^{-1}[c(x^k) - z^k]$ 和 $V^k \leftarrow \|c(x^k) - z^k\|$

6 **如果** 最大值$\{\varepsilon_k, V^k\} \leq \epsilon$，**则返回** $x^* \leftarrow x^k$

7 **如果** $k = 0$ **或** $V^k \leq \theta V^{k-1}$，**则** $\mu_{k+1} \leftarrow \mu_k$，**否则** $\mu_{k+1} \leftarrow \kappa\mu_k$

保护拉格朗日乘子估计 $\hat{y}^k$ 取自一个有界集 $Y \subset \mathbb{R}^m$，在 第 3 步中。虽然并非必要，但 $\hat{y}^k$ 的选择也应取决于当前的

估计 y^{k-1}。此外， Y 的选择可以利用 先验 知识 ∂g及其结构，以生成更好的拉格朗日乘子估计。在实践中，建议将保护估计 $\hat{}y^k$ 设为拉格朗日乘子 y^{k-1} 在 Y上的投影，从而只要 y^{k-1}仍处于 Y之内，就有效地采用经典方法；更深入的讨论请参见 [6, 第 4 节。1] 以及 [26, 第 3] 节。

受经典一阶拉格朗日乘子估计的启发，步骤 5 中的更新规则围绕（5a）及以下恒等式进行设计

$$\nabla_x \mathcal{L}_\mu^{\mathrm{S}}(x, z, \hat{y}) = \nabla f(x) + \frac{1}{\mu} c'(x)^\top [c(x) + \mu\hat{y} - z].$$

给定一个 在x^k， z^k处通过步骤 4得到的 y^k ，对其作这一特定选择便可构造出

$$\|\nabla f(x^k) + c'(x^k)^\top y^k\| \le \varepsilon_k,$$

从而有效连接内循环和外循环，并为（x^k， y^k）处的对偶不可行性提供上界。步骤 6 概括了返回近似 KKT 驻点所需的终止条件；相关内容将在第3.5节中讨论。在步骤 7 处采用单调性测试，以监测迭代点的原始不可行性，并相应更新罚参数。为促使 V^k : $= \|c$（x^k） $- z^k\|$ 趋于零，当下降不足时，罚参数 μ_k 将在步骤 7 处减小。

3.4 收敛性分析

在整个收敛性分析中，我们假设算法 3.1 定义良好，即要求步骤 4 中的每个子问题都存在一个近似驻点。此外，以下论述假设由算法 3.1生成的序列 $\{x^k\}_{k\in\mathbb{N}}$ 存在某个聚点 x^* 。一般而言，应采用强制性或（水平集）有界性论证来验证这些前提条件。

鉴于其实际重要性，我们着重关注局部或称经济型求解器，它们在步骤 4中为子问题返回作为候选局部极小点的驻点。相反，我们不探讨将子问题求解至全局最优性的情形，因为其分析将与 [6, 第 5] 章和 [26]中的经典结果类似。

与非凸情形下的罚函数型方法通常出现的情况一样，算法 3.1可能生成对于（P）不可行的聚点。遵循标准论证，下面的结果给出了保证极限点可行的条件；参见 [5, 问题 4。12], [15, 命题 3。2]。

命题 3.4。设假设 *1*成立，并考虑由 $\{x^k\}_{k\in\mathbb{N}}$ 生成的迭代点序列算法 *3.1* ，其中 $\epsilon = 0$。那么， x^* 的每个聚点 $\{x^k\}_{k\in\mathbb{N}}$都是（(P）的可行点，只要以下条件中至少有一个成立：

(i) 对某个子序列 $\{x^k\}_{k\in K}$， $x^k \xrightarrow{\varphi}_K x^*$ ；

(ii) 存在某个 $B \in \mathbb{R}$ ，使得 $\mathcal{L}_{\mu_k}(x^k, \hat{}y^k) \le B$ 对所有 $k \in \mathbb{N}$均成立；

(iii) $\{\mu_k\}_{k\in\mathbb{N}}$ 远离零，且$\operatorname{dom} g$ 是闭的；

(iv) $\{\mu_k\}_{k\in\mathbb{N}}$ 远离零，且对所有 $k \in \mathbb{N}$， $\mathcal{L}_{\mu_k}(x^k, \hat{}y^k) \le \mathcal{L}_{\mu_k}(x^{k-1}, \hat{}y^k)$。

证明。设 $x^* \in \mathbb{R}^n$ 是 $\{x^k\}$ 的任意一个聚点，并设 $\{x^k\}_{k\in K}$ 是满足 $x^k \to_K x^*$ 的一个子序列。我们将证明 x^* 对 (P) 可行，即 $c(x^*) \in \mathbf{dom}\, g$。在讨论各项断言之前，我们指出，若 $\{\mu_k\}$ 远离零，则算法 3.1 的步骤 7 中的条件意味着，对于 $k \to \infty$，有 $V^k := \|c(x^k) - z^k\| \to 0$。因此，由于 $x^k \to_K x^*$，根据连续性可知，我们还有 $z^k \to_K c(x^*)$。最后，方便起见，记算法 3.1 中初始化的有界集 $Y \subset \mathbb{R}^m$ 的有限大小为 $\Delta_Y \in \mathbb{R}$，使得对所有 $\hat{y} \in Y$，均有 $\|\hat{y}\| \leq \Delta_Y$。

(i) 对于 φ-注意收敛至 $\{x^k\}_{k\in K}$ 需要满足 $x^* \in \mathbf{dom}\, x^*$。因此，根据 φ: φ 以及 $= g \circ c$ 的连续性，必有 c （ c） $x^* \in \mathbf{dom}$。

(ii) 鉴于关于 $\{\mu_k\}$ 远离零的初步观察，只需考虑 $\mu_k \to 0$ 的情形，以证明 $z^k \to_K c$（x^*）。利用（10）和$z^k \in \mathbf{O}_{\mu_k}$（$x^k$, $\hat{y}^k$），由上界可得

$$B \geq \mathcal{L}_{\mu_k}(x^k, \hat{y}^k) = f(x^k) + g(z^k) + \frac{1}{2\mu_k}\|c(x^k) + \mu_k \hat{y}^k - z^k\|^2 - \frac{\mu_k}{2}\|\hat{y}^k\|^2$$

对所有 $k \in \mathbb{N}$ 均成立。重新整理各项并考虑 $\mu_k \to 0$，可得

$$\|c(x^k) + \mu_k \hat{y}^k - z^k\|^2 \leq 2\mu_k \left[B + \frac{\mu_k}{2}\|\hat{y}^k\|^2 - f(x^k) - g(z^k)\right] \to_K 0,$$

其中，该极限由 $\{\hat{y}^k\}_{k\in\mathbb{N}}$ 和 $\{z^k\}_{k\in K}$ 的有界性推出；后者由于 $x^k \to_K x^*$ 以及近端映射的局部有界性而保持有界。再次利用 $\mu_k \hat{y}^k \to 0$，该极限意味着 $c(x^k) - z^k \to_K 0$。因此，由于 $x^k \to_K x^*$，根据连续性可知 $z^k \to_K c(x^*)$。现在，鉴于 $x^k \to_K x^*$、$\{\hat{y}^k\}_{k\in\mathbb{N}}$ 的有界性以及 $\mu_k \leq \mu_0$，根据上界条件还可得到 $\{g(z^k)\}_{k\in K}$ 有上界：

$$\limsup_{k\to_K\infty} g(z^k) \leq B + \limsup_{k\to_K\infty} \frac{\mu_k}{2}\|\hat{y}^k\|^2 - f(x^k) \leq B + \frac{\mu_0}{2}\Delta_Y^2 - f(x^*) < \infty.$$

结合 $z^k \to_K c(x^*)$ 以及 g 的下半连续性，可知 $c(x^*) \in \mathbf{dom}\ g$。

(iii) 由于 $\{\mu_k\}_{k\in\mathbb{N}}$ 始终远离零，因此有 $z^k \to_K c(x^*)$。接着，由于 $z^k \in \mathbf{O}_{\mu_k}(x^k, \hat{y}^k) \subseteq \mathbf{dom}\ g$ 对所有 $k \in \mathbb{N}$ 均成立，根据步骤 4，由 $\mathbf{dom}\ g$ 是闭集可知 $c(x^*) \in \mathbf{dom}\, g$。

(iv) 由于 $\{\mu_k\}_{k\in\mathbb{N}}$ 始终远离零点，因此有 $z^k \to_K c(x^*)$，并且由于 步骤 7中的条件，$\{\mu_k\}_{k\in\mathbb{N}}$ 渐近地保持不变，即存在某个有限索引 $k_* \in \mathbb{N}$，使得 $\mu_k = \mu_* > 0$ 对所有 $k \geq k_*$均成立。进而，原始不可行性度量 V^k 呈线性下降，因为必有 $V^k \leq \theta V^{k-1}$ 对所有 $k \geq k_*$均成立。现在，重新整理（10），并利用关于 z的显式最小化，可得到上界

$$\begin{aligned}
\mathcal{L}_{\mu_*}(x^{k-1}, \hat{y}^k) &= f(x^{k-1}) + \min_z \left\{ g(z) + \langle \hat{y}^k, c(x^{k-1}) - z\rangle + \frac{1}{2\mu_*}\|c(x^{k-1}) - z\|^2 \right\} \\
&\leq f(x^{k-1}) + g(z^{k-1}) + \langle \hat{y}^k, c(x^{k-1}) - z^{k-1}\rangle + \frac{1}{2\mu_*}\|c(x^{k-1}) - z^{k-1}\|^2 \\
&= \mathcal{L}_{\mu_*}(x^{k-1}, \hat{y}^{k-1}) + \langle \hat{y}^k - \hat{y}^{k-1}, c(x^{k-1}) - z^{k-1}\rangle
\end{aligned}$$

for 所有 $k > k_*$。利用柯西-施瓦茨不等式，下降条件给出 s

$$\mathcal{L}_{\mu_*}(x^k, \hat{y}^k) \le \mathcal{L}_{\mu_*}(x^{k-1}, \hat{y}^k) \le \mathcal{L}_{\mu_*}(x^{k-1}, \hat{y}^{k-1}) + 2\Delta_Y V^{k-1}$$

对所有 $k > k_*$，其中我们用到了 $\|\hat{y}^k - \hat{y}^{k-1}\| \le 2\Delta_Y$。然后，对这一论证进行望远镜求和，$V^k$ 的线性递减给出

$$\mathcal{L}_{\mu_*}(x^{k+\ell}, \hat{y}^{k+\ell}) - \mathcal{L}_{\mu_*}(x^{k-1}, \hat{y}^{k-1}) \le 2\Delta_Y \sum_{m=0}^{\ell} V^{k+m-1} \le 2\Delta_Y V^{k-1} \sum_{m=0}^{\ell} \theta^{\ell}$$

对所有 $k > k_*$ 和 $\ell \in \mathbb{N}$ 均成立。因此，由于 $\theta \in (0,1)$，对任意固定的 $k > k_*$ 取极限可得

$$\limsup_{\ell \to \infty} \mathcal{L}_{\mu_*}(x^{k+\ell}, \hat{y}^{k+\ell}) \le \mathcal{L}_{\mu_*}(x^{k-1}, \hat{y}^{k-1}) + \frac{2\Delta_Y}{1-\theta} V^{k-1} < \infty,$$

这表明，只要舍弃前期的迭代点，$\mathcal{L}_{\mu_k}(x^k, \hat{y}^k)$ 就存在一个上界。然后，x^* 的可行性可按(ii)中的方式得出。 □

以下结果讨论拉格朗日乘子估计。考虑 g 是某个集合 D 的示性函数且该集合内点非空这一特殊情形，命题 3.5表明，对于在极限处（严格）不活跃的约束，拉格朗日乘子会趋于零。这说明算法 3.1 的步骤 5 中的更新规则符合其预期行为，因为乘子会按 [6, 第 4]节中所述发生偏移。

命题 3.5。 设 $g \equiv \iota_D$ 对某个非空闭集 $D \subseteq \mathbb{R}^m$成立。令 假设 *1* 成立，并考虑由 $\{x^k\}_{k\in\mathbb{N}}$ 生成的迭代点序列，其中 $\epsilon = 0$。设 x^* 是 $\{x^k\}_{k\in\mathbb{N}}$ 的一个聚点，且 $\{x^k\}_{k\in K}$ 是满足 $x^k \to_K x^*$的子序列。若 $c(x^*) \in \operatorname{int} D$，则 $y^k = 0$ 对所有足够 $k \in K$ 大的情形均成立。

证明。 令 $c(x^*) \in \operatorname{int} D \neq \emptyset$，并注意到 ι_D 的近端映射与到 D 的投影一致，即对所有 $\gamma > 0$，$\operatorname{prox}_{\gamma\iota_D} = \Pi_D$。分两种情况讨论。

- 若 $\mu_k \to 0$，由 $\{\hat{y}^k\}_{k\in\mathbb{N}}$ 的有界性及 c 的连续性，存在 $k_1 \in K$ 使得对所有 $k \in K$，$k \ge k_1$，$c(x^k) + \mu_k \hat{y}^k \in \operatorname{int} D$。
- 如果 $\{\mu_k\}_{k\in\mathbb{N}}$有正的下界，则由 步骤 7 可知 $c(x^k) - z^k \to_K 0$。然后，根据 c 和 $x^k \to_K x^*$ 的连续性，$z^k \to_K c(x^*) \in \operatorname{int} D$ 成立。因此，存在 $k_2 \in K$ 使得 $z^k \in \operatorname{int} D$ 对所有 $k \in K$，$k \ge k_2$均成立。

在这两种情况下，由于 $z^k \in \mathbf{O}_{\mu_k}(x^k, \hat{y}^k) = \Pi_D(c(x^k) + \mu_k \hat{y}^k)$ 且 $z^k \in \operatorname{int} D$对所有 $k \in K$足够大的值都成立，因此必有 $z^k = c(x^k) + \mu_k \hat{y}^k$，进而 根据 步骤 5可得。 □

请注意，如果 D 具有可分结构，则可以利用这一点轻松地改进该断言。例如，如果 D 是一个超盒，则证明中的每一步都可以逐分量应用，从而得到 [6,定理 4.1]。

以下收敛结果为算法 3.1提供了基础理论支持。它表明，在子序列注意收敛的条件下，聚点是（P）的 AKKT-驻点。对于任意聚点 x^* 以及任意满足 $\{x^k\}_{k\in K}$ 的子序列 $x^k \to_K x^*$，命题 3.4(i) 保证了 x^* 的可行性

在 $x^k \overset{\varphi}{\to}_K x^*$ 的假设下。然而， φ-注意收敛还意味着$g(c^k) \to_K g(c(x^*))$ 当 $x^k \to_K x^*$ 时成立；例如，如果（x^* 是可行的且） g相对于 dom g连续，则该条件可直接满足。具体而言，对于所有 $x \in \operatorname{dom} g$和 $\{x^\ell\} \subseteq \operatorname{dom} g$，只要 $x^\ell \to x$ ，就有 $g(x^\ell) \to g(x)$。我们在 [15, 例 3.4]中说明了 φ-注意收敛的重要性。

定理 3.6。 设 假设 *1*成立，并考虑由 $\{x^k\}_{k\in\mathbb{N}}$ 生成的迭代点序列 算法 *3.1* ，其中 $\epsilon = 0$ 且 $\varepsilon_k \to 0$。设 $x^* \in \mathbb{R}^n$ 是 $\{x^k\}_{k\in\mathbb{N}}$ 的一个聚点，且 $\{x^k\}_{k\in K}$ 是满足 $x^k \overset{\varphi}{\to}_K x^*$的子序列。那么， x^* 是问题 (P)的 *AKKT*-驻点。

证明。 我们声称，由 $\{x^k\}_{k\in K}$、 $\{y^k\}_{k\in K}$ 和 $\{z^k\}_{k\in K}$ 生成的子序列算法 3.1 满足 定义 2.3 中的性质，因此可以证明 x^* 是问题 (P) 的 AKKT-驻点。首先，根据命题 3.4(i)， φ-注意收敛 $\{x^k\}_{k\in K}$ 意味着 x^* 对于问题 (P) 是可行的。由 算法 3.1的第 5 步 和 第 4 步，根据 ε_k-Υ-驻点性可知， x^k 带有证书 z^k 时，

$$\eta^k := \nabla f(x^k) + c'(x^k)^\top y^k = \nabla_x \mathcal{L}^{\mathrm{S}}_{\mu_k}(x^k, z^k, \hat{y}^k)$$

满足 $\|\eta^k\| \le \varepsilon_k$。因此，根据构造可得 $\eta^k \to_K 0$。此外，由 步骤 5 可知，对于所有 $k \in \mathbb{N}$,

$$y^k = \hat{y}^k + \frac{c(x^k) - z^k}{\mu_k} \in \partial g(z^k),$$

其中，该包含关系由 $z^k \in \mathrm{O}_{\mu_k}(x^k, \hat{}y^k)$ 得出（事实上，它来自 z^k 的最优性必要条件）。接下来只需证明 $c(x^k) - z^k \to_K 0$。为此，我们考虑以下两种情况。

- 若 $\{\mu_k\}_{k\in\mathbb{N}}$ 有正下界，则 步骤 7 中的条件意味着 $V^k :=\|c(x^k) - z^k\| \to 0$，因此特别有 $c(x^k) - z^k \to_K 0$。
- 反之，若 $\mu_k \to 0$，则利用 c 的连续性以及 $\{\hat{}y^k\}_{k\in\mathbb{N}}$ 的有界性。这些性质推出 $c(x^k) + \mu_k \hat{}y^k \to_K c(x^*) \in \operatorname{dom} g$，其中该包含关系来自 x^* 的可行性。然后，由于对所有 $k \in \mathbb{N}$ 都有 $z^k \in \mathrm{O}_{\mu_k}(x^k, \hat{}y^k)$，根据式 (9) 和 [16,引理 3.3] ，可得

$$z^k \in \operatorname{prox}_{\mu_k g}(c(x^k) + \mu_k \hat{y}^k) \to_K c(x^*)$$

  当 $\mu_k \to 0$ 时。因此，极限 $c(x^k) - z^k \to_K 0$ 成立。

综上，这证明 x^* 是 (P) 的 AKKT-驻点。 □

约束优化算法旨在寻找可行点，并在约束条件下最小化目标函数。采用成本可承受的局部优化技术时，不能指望找到不可行性度量的全局极小点。然而，下一个结果证明，算法 3.1 在 $\{\varepsilon_k\}_{k\in\mathbb{N}}$ 有界的情况下，至少能找到不可行性度量的驻点。注意，该性质不要求 $\varepsilon_k \to 0$，而只要求有界性；参见 [6, 定理 6。3]。

命题 3.7。 设 假设 *1* 成立，并假设 $\operatorname{dom} g$ 是闭集。考虑由 $\{x^k\}_{k\in\mathbb{N}}$ 算法 *3.1* 生成的迭代点序列 $\epsilon = 0$，其中 $\{\varepsilon_k\}_{k\in\mathbb{N}}$ 有界。设 $x^* \in \mathbb{R}^n$ 是 $\{x^k\}_{k\in\mathbb{N}}$ 的一个聚点，并且 $\{x^k\}_{k\in K}$ 是满足$x^k \to_K x^*$的一个子序列。那么， x^* 是

$$\operatorname*{minimize}_{x\in\mathbb{R}^n} \operatorname{dist}^2(c(x), \operatorname{dom} g).$$

证明。 根据 命题 3.4(iii) 以及 $\operatorname{dom} g$ 为闭集，若 $\{\mu_k\}_{k\in\mathbb{N}}$ 与零保持正距离，则每个聚点 x^* 都是（P）的可行点，即它是该可行性问题的全局极小点，因此根据目标函数 [30, 命题 5.3] 的严格连续性，它是该问题的一个 M- 驻点。于是，只需考虑以下情形 $\mu_k \to 0$。

我们用等价的提升表示重写该可行性问题

$$\operatorname*{minimize}_{\substack{x\in\mathbb{R}^n \\ (z,\alpha)\in\operatorname{epi} g}} \quad \frac{1}{2}\|c(x) - z\|^2 \tag{18}$$

通过上图 reformulation 得到。包含关系 $(z, \alpha) \in \operatorname{epi} g$ 要求 $g(z) \leq \alpha < \infty$，从而编码了 $z \in \operatorname{dom} g$。对于（$x^*$ （18）的 M-驻点性要求存在某个 $z^* \in \mathbb{R}^m$ 使得

$$c'(x^*)^\top [c(x^*) - z^*] = 0, \tag{19a}$$

$$(c(x^*) - z^*, 0) \in \mathcal{N}^{\lim}_{\operatorname{epi} g}(z^*, g(z^*)). \tag{19b}$$

由于 算法 3.1 的 步骤 4，有 $\|\eta^k\| \leq \varepsilon_k$，其中

$$\eta^k := \nabla f(x^k) + \frac{1}{\mu_k} c'(x^k)^\top \left[c(x^k) + \mu_k \hat{y}^k - z^k\right],$$

对所有 $k \in K$ 均成立。令 $\mu_k \eta^k$ 在 $k \to_K \infty$ 时取极限，并利用 $\mu_k \to 0$ 和 $\{\eta^k\}_{k\in\mathbb{N}}$ 的有界性以及 $\{\nabla f(x^k)\}_{k\in K}$，可得（19a）成立。此外，根据 $z^k \in \operatorname{O}_{\mu_k}(x^k, \hat{y}^k) = \operatorname{prox}_{\mu_k g}(c(x^k) + \mu_k \hat{y}^k)$ 的最小化性质，有 $z^k \in \operatorname{dom} g$，且

$$\frac{1}{\mu_k}\left[c(x^k) + \mu_k \hat{y}^k - z^k\right] \in \partial g(z^k)$$

对所有 $k \in \mathbb{N}$ 都成立。由 $\{c(x^k) + \mu_k \hat{y}^k\}_{k\in K}$ 的有界性以及近端映射的局部有界性可知，$\{z^k\}_{k\in K}$ 也保持有界。因此，由于对所有 $k \in \mathbb{N}$ 都有 $z^k \in \operatorname{dom} g$，且 $\operatorname{dom} g$ 是闭的，我们可以不失一般性地假设，对于某个$z^* \in \operatorname{dom} g$， $z^k \to_K z^*$。利用式（2）中给出的关系，两边乘以 $\mu_k > 0$，并利用 $\mathcal{N}_{\operatorname{limepi} g}(z^k, g(z^k))$ 是一个锥这一事实，可得

$$(c(x^k) + \mu_k \hat{y}^k - z^k, -\mu_k) \in \mathcal{N}^{\lim}_{\operatorname{epi} g}(z^k, g(z^k))$$

对所有 $k \in \mathbb{N}$取 $k \to_K \infty$的极限，利用极限法锥的稳健性[30, 命题 1。20] 以及 $z^k \to_K z^* \in \operatorname{dom} g$，得到（19b），证明完毕。

□

3.5 终止准则

算法 3.1 的 步骤 4 涉及最小化式（10）中定义的隐式增广拉格朗日函数。随后，步骤 5中的乘子更新使我们能够对原问题（P得出结论，如定理 3.6所示。此外，回顾 AKKT 驻点条件（5），结合步骤 3，可以构造一个序列 $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ 使得 $\varepsilon_k \to 0$。给定一些用户指定的容差$\varepsilon^{\text{dual}}$、$\varepsilon^{\text{prim}} > 0$，当满足以下条件时，合理的做法是宣布成功收敛：$\varepsilon_k \leq \varepsilon^{\text{dual}}$ 以及 $V^k := \|c(x^k) - z^k\| \leq \varepsilon^{\text{prim}}$。 定理 3.6 表明，如果 $\{x^k\}_{k\in K}$ 的任一子序列收敛到一个可行点 $x^* \in \mathbb{R}^n$，那么这些终止准则（尤其是后者）会在有限次迭代内得到满足。由于这种情况未必发生，因此需要一种（局部）不可行性检测机制，实际实现中通常也会包含这种机制。

给定一些容差，算法 3.1 可以在步骤 7 的下降要求以及步骤 4 [5]的最优性方面采用放宽条件。在步骤 3 中，只要对足够大的ε_k 满足 $\varepsilon_k \leq \varepsilon^{\text{dual}}$，内部容差 $k \in \mathbb{N}$就可以保持远离零值的有界状态。类似地，步骤 7 的条件也可以放宽，允许（包含）$V^k \leq \varepsilon^{\text{prim}}$的情形，从而避免不必要的过强惩罚。

4 内层问题与求解器

本节将详细讨论隐式增广拉格朗日函数（定义见（10））的最小化，重点关注在某些固定的 $\mu_k > 0$和 $\hat{y}^k \in \mathbb{R}^m$下对子问题（12）进行数值求解。如第 3.1 节所述，可以通过联合最小化显式增广拉格朗日函数 $\mathcal{L}^{\text{S}}_{\mu_k}(\cdot, \cdot, \hat{y}^k)$（8）中的形式来完成此任务，即同时针对原变量和松弛变量进行最小化。由于其具有复合结构，可以采用任何适用于复合优化的方法来完成此任务；正如 [15, 第 3 节4],所指出的一个典型例子是近端梯度算法。

我们采用一种不同的方法，旨在直接最小化 $\mathcal{L}_{\mu_k}(\cdot, \hat{y}^k)$，以便在没有辅助变量的情况下求解内层问题。为此，我们像（13）中那样，灵活运用由预言机联系起来的隐式形式与显式形式。在 第 4.1 节中，讨论了一种用于最小化 p 的算法；该算法利用现有预言机以及 P 在（17）中的结构。令人想起 [13, 第 1.5 节。4], 该方案以预言机最优性为支点，弥补正则性的不足，宛如一个可靠的神奇步骤[11, 6]。事实上，基于 F 在预言机响应处的光滑部分 P的梯度，可以构造出 p 的代理梯度。随后，只需稍加抽象便可发现，任何现成的无约束光滑优化算法都能以极其直观的方式直接用于最小化隐式增广拉格朗日函数。

在本节中，令 $\mu_k > 0$ 和 $\hat{y}^k \in \mathbb{R}^m$ 为给定的固定参数。记 $p\colon \mathbb{R}^n \to \mathbb{R}$ 和 $F\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 分别为隐式增广拉格朗日函数及其光滑部分。此外，令 $P\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 为显式增广拉格朗日函数；参见第 3.2 节。仿照从式 (8) 中的 $\mathcal{L}^{\text{S}}_{\mu}(\cdot, \cdot, y)$ 到式 (10) 中的 $\mathcal{L}_{\mu}(\cdot, y)$ 的转换，我们假设存在一个预言机 O，可对 P 关于 $z \in \mathbb{R}^m$ 执行显式（精确的全局）最小化，这显然是指式 (9)。对于我们的增广拉格朗日方法的第 k 次迭代，记 $\mathsf{O} = \mathsf{O}_{\mu_k}(\cdot, \hat{y}^k)$。注意，对于所有 $x \in \mathbb{R}^n$，都有 $\mathsf{O}(x) \subseteq \operatorname{dom} g$，因此 $p(x) < \infty$。

我们在假设 1下归纳一些基本性质。

命题 4.1。 考虑 (13)、(17)，并设假设 *1* 成立。那么，

(i) F: $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 连续可微；*(ii)* g: $\mathbb{R}^m \to \mathbb{R}$ 是真、下半连续且近端有界；*(iii)* P: $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $P(x, z)$ 关于 x 局部一致地水平有界于 z；*(iv)* $\mathbf{O}$: $\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 局部有界、非空且取紧值；*(v)* p: $\mathbb{R}^n \to \mathbb{R}$ 严格连续（即局部利普希茨）。

证明。上述断言可直接由假设 1(i)–(ii) 以及 (13) 和 (17) 中的定义得出。特别地，(iii) 源于 g 的近端有界性，以及 $\mu_k \leq \gamma_g$、固定的 $\hat{}y^k$，还有 f 和 c 的光滑性。断言 (iv) 是假设 1(i) 中 c 的光滑性以及近端映射 [39,定理 1.25] 的性质的结果。断言 (v) 由恒等式 $p(\cdot) = f(\cdot) + g^{\mu_k}(c(\cdot) + \mu_k \hat{}y^k)$ 以及 Moreau包络严格连续这一事实得出 [39, 例 10.32]。 □

由于 假设 1 和 命题 4.1(v)，目标函数 p 是真且下半连续的。然而，由于无法事先保证 p 存在极小值点，我们也考虑以下假设。一般而言，该假设可在某些强制性或（水平集）有界性条件下得到验证。

假设 2. 考虑 (13)。则，$\inf p \in \mathbb{R}$。

注 4.1. 问题形式 (P) 也可以包含一个抽象约束$x \in X$，其中集合 $X \subseteq \mathbb{R}^n$ 非空、闭且简单。尽管为使阐述更简洁而未将其纳入，本文中的结果很容易扩展到包含一个凸 X 的情形。如 [37]，所述，一种做法是将 $x \in X$ 嵌入模板 (P)，方法是将 $c(x)$ 替换为 $\bar{c}(x) := (c(x)\ x)$，并将 $g(z)$ 替换为 $\bar{g}(z, z') := g(z) + \iota_X(z')$。然而，这种方法会引入一个额外的松弛分量 z'及一个额外的乘子分量 y'，同时松弛简单约束 $x \in X$。相反，按照 [1] 的做法，更精确的策略是将 $x \in X$ 直接视为增广拉格朗日子问题中的不可松弛约束，而不将其纳入隐式增广拉格朗日函数。利用到X的投影，内层求解器可以通过设计生成满足该抽象约束的迭代点。然而，当 X 可能是非凸时，要制定适当的终止准则以判定近似驻点性，似乎并非易事。

将增广拉格朗日子问题局部化到一个紧集中的可能性 X 有助于确保假设 2。结合 p 的适当性和下半连续性，$X \subset \mathbb{R}^n$ 的紧致性可确保局部化后，p 的最小化是适定的。不过，秉承信赖域方法 [11], 的思路，有必要调整允许区域X，以免干扰外层增广拉格朗日循环的收敛。这些问题留待未来研究。

4.1 带预言机的下降方法

本节将详细介绍并讨论一种用于最小化 p的算法，该算法沿着 p的下降方向逐步推进。配备预言机的下降方法见算法 4.1，其模式与 [6, 算法 81]相近；忽略变量 z 时，可恢复出一种用于无约束光滑优化的下降方法。我们证明了该算法的良定义性，并在第 4.2 节考察其收敛性质。

算法 4.1： 带预言机的非单调下降方法

数据： $x^0 \in \mathbb{R}^n$, $\varepsilon > 0$

结果： ε-Υ-驻点 $x^* \in \mathbb{R}^n$，其证书为 $z^* \in \mathbf{O}(x^*)$

1 选择参数 $\alpha, \beta, \theta \in (0,1)$, $\omega, \nu \in (0, 1]$ 并设置 $\Phi_0 \leftarrow p(x^0)$

2 **对于** $j = 0, 1, \ldots$ **执行**

3 选择 $z^j \in \mathbf{O}(x^j)$，并设置 $\nabla p^j \leftarrow \nabla_x F(x^j, z^j)$

4 **如果** $\|\nabla p^j\| \le \varepsilon$，**则返回** $(x^*, z^*) \leftarrow (x^j, z^j)$

5 计算 $d^j \in \mathbb{R}^n$，使得

$$\langle \nabla p^j, d^j \rangle \le -\theta \|\nabla p^j\| \|d^j\| \quad \text{and} \quad \|d^j\| \ge \omega \|\nabla p^j\|$$

6 计算 $x^{j+1} \in \mathbb{R}^n$ 和最大的 $\gamma_j = \beta^{\ell_j}$, $\ell_j \in \mathbb{N}$，使得

$$p(x^{j+1}) \le p(x^j + \gamma_j d^j) \le \Phi_j + \alpha \gamma_j \langle \nabla p^j, d^j \rangle$$

7 设置 $\Phi_{j+1} \leftarrow (1 - \nu)\Phi_j + \nu p(x^{j+1})$

这里有几点说明。首先，该方法只要求为变量 x 提供一个初始点 $x^0 \in \mathbb{R}^n$，而不需要为 z 提供。然后，在步骤 3 中，会计算 p 在 x^j 处的代理梯度 $\nabla p^j \in \mathbb{R}^n$：预言机返回一个任意元素 $z^j \in \mathbf{O}(x^j)$，根据结构 (17)，我们令 ∇p^j 为在 (x^j, z^j) 处的偏导数 $\partial_x P$。在步骤 5 中，必须选择一个合适的搜索方向 $d^j \in \mathbb{R}^n$，可能有多种选择。最简单的选择是由负梯度给出的最速方向，即 $d^j := -\nabla p^j$；对于任意的 $\theta \in (0, 1)$ 和 $\omega \in (0, 1]$，该方向始终满足步骤 5 中的条件。另一种可能是考虑基于局部二次近似的类牛顿方向。在这一领域中，Barzilai-Borwein 步长和（有限记忆）BFGS 方向是广受欢迎的策略；概述请见 [40, 第 5.1 节。2] 随后，步骤 6 的内循环会寻找合适的步长 $\gamma_j > 0$，使得更新$x^j + \gamma_j d^j$ 能够带来某种改进。具体而言，此处将这一回溯线搜索过程作为全局化机制。最后，当 $\nu \in (0, 1)$ 时，步骤 7 的更新规则允许放宽目标值必须单调递减的要求，从而降低线搜索的保守性，并可能鼓励采用更长的步长，提升实际表现 [6, 第 8 节。2], [40]不过，通过 $\nu := 1$，可以确保$\{p(x^j)\}_{j \in \mathbb{N}}$ 呈现单调变化。

关键在于，步骤 6中用于调整步长 γ_k 的试探性更新和回溯条件仅涉及变量 x。此外，步骤 3中的预言机评估也可以从算法中形式上剔除，并纳入代理梯度

对 p 在 x^j处的评估。这得益于预言机所选元素 z^j 的任意性 **O**，因此我们可以保持变量 z 为隐式。这些观察表明，算法 4.1等价于直接应用于最小化 p 的下降法（使用代理梯度），因此，配备预言机的算法仅是一种形式化工具，对用户而言是透明的。

4.2 收敛性分析

本节将研究 算法 4.1 的收敛性质，利用 F 的连续可微性以及预言机最优性，证明关于 x的充分下降。我们首先将证明 算法 4.1 定义良好，即每次迭代都能在有限时间内终止。这是一个可实现算法必须满足的重要理论和实践条件。

引理 4.2 （定义良好性）。 设 假设 *1* 成立，并考虑由算法 *4.1*生成的迭代点。则，

（*i*）在每次迭代中，步骤 *6* 中的回溯次数均为有限；

（*ii*）在第 j 次迭代结束时，有

$$p(x^{j+1}) \leq \Phi_j + \delta_j \qquad \textit{and} \qquad p(x^{j+1}) \leq \Phi_{j+1} \leq \Phi_j + \nu\delta_j$$

where $\delta_j := \alpha\gamma_j\langle\nabla p^j, d^j\rangle < 0$;

（*iii*）每个迭代点 x^j 均保持在子水平集 $\{x \in \mathbb{R}^n \,|\, p\ (x) \quad \leq \Phi_0 < \infty \qquad \}$.

证明。 回顾 命题 4.1 中汇总的基本性质。

(i) 假设 $\|\nabla p^j\| > \varepsilon$，以排除平凡情形。由于 $\theta \in (0,1)$ 且$\omega \in (0,1]$, 选择 $d^j \leftarrow -\nabla p^j$ 始终可行，因此 步骤 5 定义良好。记 $x^{j+} := x^j + \gamma_j d^j$ 以及 $z^{j+} \in \mathbf{O}(x^{j+})$。于是，

$$p(x^{j+}) = F(x^{j+}, z^{j+}) + g(z^{j+}) \leq F(x^{j+}, z) + g(z)$$

对所有 $z \in \mathbb{R}^m$均成立。特别地， $p(x^{j+}) \leq F(x^{j+}, z^j) + g(z^j)$。由 F 的连续可微性，结合 步骤 3 和 5 可得

$$\lim_{\gamma\to 0} \frac{F(x^j + \gamma d^j, z^j) - F(x^j, z^j)}{\gamma} = \langle\nabla_x F(x^j, z^j), d^j\rangle = \langle\nabla p^j, d^j\rangle < 0.$$

因此，

$$\lim_{\gamma\to 0} \frac{F(x^j + \gamma d^j, z^j) - F(x^j, z^j)}{\gamma\langle\nabla p^j, d^j\rangle} = 1.$$

因此，由于 $\alpha \in (0,1)$，对于足够小的某个 $\gamma_j = \beta^{\ell_j} > 0$ ，我们有

$$\frac{F(x^j + \gamma_j d^j, z^j) - F(x^j, z^j)}{\gamma_j\langle\nabla p^j, d^j\rangle} \geq \alpha.$$

由于 $\langle \nabla p^j,\ d^j \rangle < 0$ 根据 步骤 5 和 $\gamma_j > 0$，我们推导出

$$F(x^{j+}, z^j) \le F(x^j, z^j) + \alpha\gamma_j \langle \nabla p^j, d^j \rangle.$$

现在，调用预言机以获得 $z^{j+} \in \mathbf{O}(x^{j+})$，我们得到

$$\begin{aligned} p(x^{j+}) = F(x^{j+}, z^{j+}) + g(z^{j+}) &\le F(x^{j+}, z^j) + g(z^j) \\ &\le F(x^j, z^j) + g(z^j) + \alpha\gamma_j \langle \nabla p^j, d^j \rangle = p(x^j) + \alpha\gamma_j \langle \nabla p^j, d^j \rangle. \end{aligned}$$

然后，由于对所有 $j \in \mathbb{N}$，界 $\Phi_j \ge p(x^j)$ 均成立，因此内循环有限步终止。该界通过归纳法得到：若 $j = 0$，则根据构造有 $\Phi_j = p(x^j)$；若 $j > 0$，则该界由 步骤 7 得出。

(ii) 第一个不等式由结合 (i) 与条件在步骤 6处不成立这一事实得到。再结合 步骤 7，可得下界

$$\begin{aligned} \Phi_{j+1} = (1-\nu)\Phi_j + \nu p(x^{j+1}) &\ge (1-\nu)[p(x^{j+1}) - \delta_j] + \nu p(x^{j+1}) \\ &= p(x^{j+1}) - (1-\nu)\delta_j \ge p(x^{j+1}) \end{aligned}$$

以及上界

$$\Phi_{j+1} = (1-\nu)\Phi_j + \nu p(x^{j+1}) \le (1-\nu)\Phi_j + \nu(\Phi_j + \delta_j) = \Phi_j + \nu\delta_j.$$

(iii) 由以下事实可得，$\Phi_0 = p(x^0) < \infty$： $z^0 \in \mathbf{O}(x^0) \in \operatorname{dom} g \ne \emptyset$。然后，通过对 $p(x^j) \le \Phi_j \le \ldots \le \Phi_0$ 进行望远镜求和，断言直接由 (ii) 得出。 □

接下来，我们对该算法进行渐近分析。

引理 4.3。 设 假设 *1* 和 *2* 成立，并考虑由 $\{x^j\}_{j\in\mathbb{N}}$通过 算法 *4.1* 在 $\varepsilon = 0$下生成的迭代序列 $\{\Phi_j\}_{j\in\mathbb{N}}$ 和 $\{p(x^j)\}_{j\in\mathbb{N}}$均收敛到有限值 $p_* \ge \inf p$，其中前者从上方收敛。

证明。 关于 $\{\Phi_j\}_{j\in\mathbb{N}}$，结论可由 引理 4.2(ii) 和 假设 2推出。接着，重新整理 步骤 7 中的更新规则可得 $p(x^{j+1}) = \nu^{-1}(\Phi_{j+1} - \Phi_j) + \Phi_j$成立。令 $j \to \infty$取极限，由 $\nu > 0$ 的收敛性可知 $\{\Phi_j\}_{j\in\mathbb{N}}$也收敛到 $\{p(x^j)\}_{j\in\mathbb{N}}$，且二者收敛到相同的有限值 p_*。 □

下面的 定理 4.4 保证全局收敛至 Υ-驻点，即由 算法 4.1生成的迭代点的任意聚点均是相对于最小化 Υ而言的 p-驻点。

定理 4.4。 设 假设 *1*和 *2* 成立，并考虑由下述方法生成的迭代序列$\{x^j\}_{j\in\mathbb{N}}$： 算法 *4.1*，其中 $\varepsilon = 0$。设 $x^* \in \mathbb{R}^n$ 是 $\{x^j\}_{j\in\mathbb{N}}$的一个聚点。那么， x^* 是 Υ-驻点，满足 p，即存在 $z^* \in \mathbf{O}(x^*)$， 使得$\nabla_x F(x^*,\ z^*) = 0$。

证明。 如果 算法 4.1 在有限次迭代后终止，并返回 $(x^j,\ z^j)$，那么必有 $\|\nabla p^j\| = 0$，因此 $\nabla_x F(x^j,\ z^j) = 0$。由于 $z^j \in \mathbf{O}(x^j)$ 对所有 $j \in \mathbb{N}$均成立，因此x^j 将是 Υ-驻点。

现在反过来假设算法 4.1 进行了无限多次迭代。令 $\{x^j\}_{j\in J}, J := \{j_0, j_1, j_2 \ldots\} \subseteq \mathbb{N}$ 为一个子序列，使得 $x^j \to_J x^*$。由 p的连续性，根据命题 4.1(v) 可得 $p(x^j) \to_J p(x^*)$。由于步骤 6处的线搜索条件，且 $j_{i+1} \geq j_i + 1$，我们有

$$p(x^{j_{i+1}}) \leq p(x^{j_i+1}) \leq \Phi_{j_i} + \alpha\gamma_{j_i}\langle \nabla p^{j_i}, d^{j_i}\rangle < \Phi_{j_i}.$$

于是，引理 4.3 蕴含 $\gamma_j\langle \nabla p^j, d^j\rangle \to_J 0$，因为 $\alpha > 0$，并且由于步骤 5，必有 $\gamma_j\|\nabla p^j\|\|d^j\| \to_J 0$。由于极限中的任一项都可能趋于零，因此存在三种（并非互斥的）情形。现在我们说明，经过抽取子序列后，在每种情形下都有 $\nabla p^j \to_J 0$ 。首先，若 ∇p^j 趋于零，则结论显然成立。其次，若 $d^j \to_J 0$，则步骤 5 处的条件蕴含 ∇p^j 也趋于零，从而得到结论。第三，若 $\gamma_j \to_J 0$，我们将证明存在一个更大的步长 $\hat{}\gamma_j$，它也趋于零，而沿该步长函数的下降幅度不足；这只有在梯度趋于零时才可能发生。

我们先来注意一下尝试步骤 6失败的原因。此前（因而已被否决的）步长 $\hat{}\gamma_j := \beta^{-1}\gamma_j = \beta^{\ell_j - 1}$ 给出

$$p(x^j + \hat{\gamma}_j d^j) > \Phi_j + \alpha\hat{\gamma}_i\langle \nabla p^j, d^j\rangle \geq p(x^j) + \alpha\hat{\gamma}_j\langle \nabla p^j, d^j\rangle, \tag{20}$$

其中，最后一个不等式由界 $\Phi_j \geq p(x^j)$ 得出，该界见引理 4.2(ii)。在 $\gamma_j \to_J 0$的情形下，$\hat{}\gamma_j \to_J 0$ 也必然如此，因为 $\ell_j \to_J 0$。重新整理各项，并利用（ P ）中17的结构，任意 $z^j \in \mathbf{O}(x^j)$ 的最小化性质可得

$$\begin{aligned}
\alpha\langle \nabla p^j, d^j\rangle &< \frac{p(x^j + \frac{\gamma_j}{\beta}d^j) - p(x^j)}{\frac{\gamma_j}{\beta}} \\
&\leq \frac{F(x^j + \frac{\gamma_j}{\beta}d^j, z^j) + g(z^j) - F(x^j, z^j) - g(z^j)}{\frac{\gamma_j}{\beta}} \\
&= \frac{F(x^j + \frac{\gamma_j}{\beta}d^j, z^j) - F(x^j, z^j)}{\frac{\gamma_j}{\beta}}.
\end{aligned}$$

对所有 $j \in \mathbb{N}$ 定义 $s^j := \hat{}\gamma_j d^j$ ，则有 $\|s^j\| \to_J 0$。可能需要提取一个子序列并重新定义 J，令 $s^* \in \mathbb{R}^n$ 满足 $s^j/\|s^j\| \to_J s^*$。现在，对所有$j \in J$，由中值定理可知，存在 $\zeta_j \in [0, 1]$ 使得

$$\begin{aligned}
\langle \nabla_x F(x^j + \zeta_j s^j, z^j), s^j\rangle &= F(x^j + s^j, z^j) - F(x^j, z^j) \\
&= F(x^j + s^j, z^j) + g(z^j) - F(x^j, z^j) - g(z^j) \\
&\geq p(x^j + s^j) - p(x^j) \\
&> \alpha\langle \nabla p^j, s^j\rangle = \alpha\langle \nabla_x F(x^j, z^j), s^j\rangle,
\end{aligned}$$

其中，不等式源于预言机最优性以及 (20)。两边除以 $\|s^j\|$ 并重新整理可得

$$\frac{\langle \nabla_x F(x^j + \zeta_j s^j, z^j) - \alpha\nabla_x F(x^j, z^j), s^j\rangle}{\|s^j\|} > 0.$$

由于 $\zeta_j s^j \to_J 0$ 以及 ∇F的连续性，取极限可得

$$\limsup_{j\to_J\infty} \frac{\langle \nabla_x F(x^j+\zeta_j s^j, z^j) - \alpha \nabla_x F(x^j, z^j), s^j\rangle}{\|s^j\|} = \limsup_{j\to_J\infty} (1-\alpha)\frac{\langle \nabla_x F(x^j, z^j), s^j\rangle}{\|s^j\|} \le 0,$$

其中该不等式由 $\langle \nabla_x F(x^j, z^j)\ s^j\rangle = \hat{}\gamma_j \langle \nabla p^j, d^j\rangle < 0$ 对所有 $j \in \mathbb{N}$，根据 步骤 5，以及 $\alpha < 1$ 得出。回顾 $s^j/\|s^j\| \to_J s^*$，上述不等式表明必有 $\langle \nabla_x F(x^j, z^j)\ s^*\rangle = \langle \nabla p^j,\ s^*\rangle \to_J 0$。然后，根据 步骤 5 中的条件，必有 $\nabla p^j \to_J 0$。

最后，由于 $\nabla p^j := \nabla_x F(x^j, z^j) \to_J 0$ 在所有情况下都成立，且对所有 $j \in J, x^j \to_J x^*$，$z^j \in \mathbf{O}(x^j)$，并且 ∇F 连续，因此我们得出 $\nabla_x F(x^*, z^*) = 0$ 对某个 $z^* \in \mathbf{O}(x^*)$ 成立，从而证明了 x^* 对于 p 是 Υ-驻点。 □

4.3 终止准则

算法 4.1若不加入 步骤 4，便会无限运行，生成无限的迭代点序列。正如引理 4.2(ii) 和引理 4.3分别证明的，随着迭代进行，优值和目标值趋于下降，并收敛到某个有限值。因此，需要加入适当的终止准则，以停止算法并返回一个在某种意义上相对于最小化 p 而言令人满意的迭代点。

由于对所有 $j \in \mathbb{N}$，$z^j \in \mathbf{O}(x^j)$，因此只需检查x^j的（近似）驻点性。为此，我们为算法 4.1 添加终止条件 $\|\nabla_x F(x^j, z^j)\| \le \varepsilon$。根据式 (17) 中 P 的特定结构，该条件刻画了 x^j 相对于 p的 ε-Υ-驻点性。借助定理 4.4，我们现在可以轻松推知，对于任意容差 $\varepsilon > 0$，算法 4.1 都会在有限步内终止；这表明其输出符合算法 3.1 步骤 4 中外层增广拉格朗日框架的要求。

推论 4.5. 设 假设 *1* 成立，并考虑由 $\{x^j, z^j\}_{j\in\mathbb{N}}$生成的迭代序列 算法 *4.1* 在 $\varepsilon > 0$ 条件下得到。假设序列 $\{x^j\}_{j\in\mathbb{N}}$ 存在一个聚点。那么，经过有限步后，将返回一个 ε-Υ-驻点 $x^* \in \operatorname{dom} p$，其证书为 $z^* \in \mathbf{O}(x^*)$。

证明。根据定理 4.4，算法在有限个迭代点后（设为 j_* 个）终止，因为沿任意收敛子序列都有 $\|\nabla_x F(x^j, z^j)\| = \|\nabla p^j\| \to_J 0\{x^j\}_{j\in J}$。由于对于所有 $j \in \mathbb{N}$，$z^j \in \mathbf{O}(x^j)$，因此输出 $x^* := x^{j_*}$ 的证书 $z^* := z^{j_*}$ 满足 $z^* \in \mathbf{O}(x^*)$。根据终止准则，$\nabla_x F(x^*, z^*)$ 的大小不超过 ε，因此对于 p，数对 (x^*, z^*) 是 ε-Υ-驻点。 □

请注意，推论 4.5中假设极限点存在，但并未保证其存在。一般而言，可通过采用（水平集）有界性或强制性论证来满足这一假设。

5 数值示例与比较

本节展示了（非凸）广义规划框架所提供的灵活性，并给出了一些数值结果，以支持第3节提出的隐式方法。我们以一个著名的带消失约束问题作为说明性示例。

我们已在增广拉格朗日近端求解器（ALPS）中实现了所提出的方法，该求解器是开源软件包 Bazinga.jl [12], 的一部分，其中包含下文讨论的示例。由于我们的目标仅是为隐式方法提供动机和数值支持，因此不介绍实现细节。尽管如此，我们要指出，该方法在很大程度上沿用了 ALS 的模式 [15], ；ALS 是一种基于显式处理松弛变量的增广拉格朗日求解器。具体而言，算法 3.1的步骤 4默认使用加速近端梯度求解器 PANOC 求解$^+$ [17]。在我们的数值测试中，这一选择便于对不同方法和形式进行比较并获得可比结果。作为一种近端梯度法，PANOC$^+$ 能够处理显式增广拉格朗日函数的双项结构；事实上，它是 ALS 中默认的内层求解器 [15]。不过，PANOC$^+$ 也可以作为基于梯度的下降法，因此可根据第 3.1 节，将其用作第 4.1 节中的内层求解器。

5.1 含消失约束的问题

我们考虑以下二维问题，该问题源自桁架拓扑优化，是一个众所周知的学术问题；相关讨论十分详尽，见[25,第 9.5 节1]。

$$\begin{aligned} \underset{x\in\mathbb{R}^2}{\text{minimize}} \quad & 4x_1 + 2x_2 \\ \text{subject to} \quad & x_1 \geq 0, \quad x_2 \geq 0 \\ & x_1 > 0 \quad \Rightarrow \quad x_1 + x_2 \geq 5\sqrt{2} \\ & x_2 > 0 \quad \Rightarrow \quad x_1 + x_2 \geq 5 \end{aligned} \tag{21}$$

由于其结构，这些逻辑蕴含会产生所谓的消失约束，通常通过要求 $x_1(x_1 + x_2 - 5\sqrt{2}) \geq 0$ 且 $x_2(x_1 + x_2 - 5) \geq 0$，并满足 $x_1, x_2 \geq 0$ 来重新表述。问题 (21) 的可行集由无界多面体 $\{(x_1, x_2) \in \mathbb{R}^2_+ \,|\, x_1 + x_2 \geq 5\sqrt{2}\}$、与之相连的线段 $\{(0, x_2) \in \mathbb{R}^2 \,|\, 5 \leq x_2 \leq 5\sqrt{2}\}$以及孤立点 (0, 0) 组成。该问题在 $x^\star = (0, 0)$ 处取得全局最小值，并在 $x^\diamond = (0, 5)$ 处取得局部最小值；见图1。

可以通过多种方式将 (21) 转换为 (P) 的形式。我们将考虑三种不同的表述，其处理约束的显式程度依次递增：隐式、中间形式（或部分显式形式，不为平凡约束引入松弛变量分量）以及（完全）显式形式。定义代价函数 $f(x) := 4x_1 + 2x_2$。此外，方便起见，将 $D_{\mathrm{VC}} \subset \mathbb{R}^2$ 记为与一个消失约束相关联的集合：$D_{\mathrm{VC}} := \{(a, b) \in \mathbb{R}^2 \,|\, a \geq 0, \quad ab \geq 0\}$；该集合非空、闭且非凸，其（集值）投影 Π_D 可以很容易地计算。

- An 将 (21) 隐式表示为 (P) 的形式，可由数据函数给出

$$c_{\mathrm{impl}}(x) := (x_1,\, x_1 + x_2 - 5\sqrt{2},\, x_2,\, x_1 + x_2 - 5)^\top$$
$$g_{\mathrm{impl}}(c) := \iota_{D_{\mathrm{VC}}}(c_1, c_2) + \iota_{D_{\mathrm{VC}}}(c_3, c_4).$$

对于尺寸（n，m）＝（2，4），这似乎是（21）最紧凑的形式。

- 通过引入松弛变量，如（Ps）所示，可将（21）显式重构为（P）。由此得到一个尺寸为（n，m）＝（6，8）的优化问题。请注意，向 ALS 输入隐式形式时，ALS 会在内部将原问题显式重构为提升表示 [15,第 3] 节。

- 通过避免为平凡约束引入松弛变量分量，可将（21）转换为（P）的中间表示。这种中间形式的尺寸为（n，m）＝（4，6），其数据函数为

$$c_{\mathrm{inte}}(x) := (x_1,\, x_1 + x_2 - 5\sqrt{2} - x_3,\, x_2,\, x_1 + x_2 - 5 - x_4,\, x_3,\, x_4)^\top$$
$$g_{\mathrm{inte}}(c) := \iota_{D_{\mathrm{VC}}}(c_1, c_5) + \iota_{D_{\mathrm{VC}}}(c_3, c_6) + \iota_0(c_2) + \iota_0(c_4).$$

我们对所有问题表示均运行 ALPS。我们也记录 ALS [15] 在隐式形式上的结果，以便与 ALPS 上的显式形式结果进行比较。我们在 $51^2 = 2601$ 中取一组初始点 x^0，并令初始拉格朗日乘子估计为 $y^0 = 0$；初始化松弛变量，使其满足重构后得到的线性约束。对于所有实例，两个求解器均返回成功状态，且总是指示全局或局部极小值点（欧几里得距离在 10^{-6} 以内）。我们在图1中描绘（21）的可行域，并更详细地展示这些结果。每个初始点 x^0 均以一个标记表示，该标记指示从 x^0出发所到达的终止点。由于集合 D_{VC} 是非凸的，算法 3.1的迭代点取决于在投影到 D_{VC}时投影结果不唯一所导致的任意选择。通过选取不同的投影，我们观察到了类似但并不完全相同的数值结果。

表1汇总了概括数值结果的统计数据。考虑隐式形式时，ALPS 在大多数情况下（93%）返回全局极小点 $x^\star = (0,\, 0)$。采用中间形式时，ALPS 似乎只有在从附近位置开始时才会稳定地找到 $x^\star$（28%）；采用完全显式形式时，这一比例更低（19%），使用 ALS 时则为 16%。尽管提升 [15, 引理 3。1], 并不会影响解和驻点，但实际达到的目标值却存在显著差距。然而，优势可能并不止于此，因为隐式方法在减少（中位数）迭代次数和运行时间方面似乎也极为有效。我们不禁怀疑，这正是预言机所起的作用。

6 总结

在完全非凸情形下对广义优化进行的理论研究，促成了对增广拉格朗日（AL）框架的重要扩展。借助提升重构，研究者推导出了KKT型驻点条件

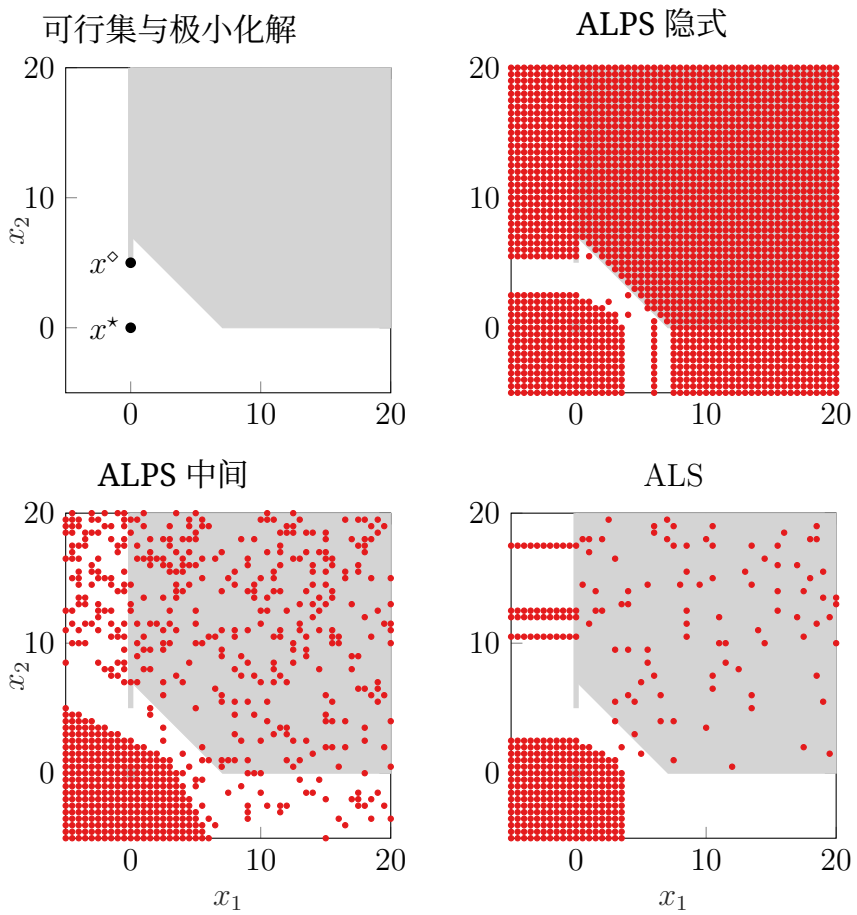

图1：使用不同方法和表述形式得到的解的比较。可行集（灰色背景）以及起始点网格（2601 个均匀分布的点），并根据返回的解标记相应的终止状态：全局极小点$x^\star = (0,0)$（红色圆圈）或局部极小点 $x^\diamond = (0,5)$（不标记）。

并以拉格朗日形式表述，同时补充了渐近性质和正则性概念。研究者定义了隐式增广拉格朗日函数及其子问题，通过松弛变量将其显式对应形式边缘化；由于集值性，这似乎牺牲了一定的正则性。随后，我们基于参数化解释提出了 Υ-驻点性的概念，用于刻画子问题的解。尽管对于隐式增广拉格朗日函数而言，它弱于M-驻点性，但 Υ-驻点性已被证明足以有效研究渐近性质。我们考虑了一种带保护机制的增广拉格朗日算法，并建立了适时的收敛保证，其水平与非线性规划中的保证相当。

不过，采用隐式理念并非只是出于理论上的好奇。我们开发并刻画了一种求解隐式增广拉格朗日子问题的求解器，该求解器确实能够生成近似 Υ-平稳迭代点。借助近端预言机，它处理的决策变量更少，同时在内部保留松弛变量。与更显式的方法相比，这似乎能构建更紧凑的模型，并提升实际性能；一个示例问题的数值结果印证了这一点。

严峻 增广拉格朗日方面的问题和疑问仍未解决。通常会假设 e

表1：比较不同方法和公式所获得的解及其性能。统计（外层和累计内层）迭代次数、耗时，以及返回全局极小点 $x^\star$ 的频率。

		ALPS 隐式	ALPS 中间	ALPS 显式	ALS
外层迭代次数	中位数	10	10	10	10
	max	10	10	10	10
总内层迭代次数	中位数	10	77	117	117
	max	802	598	217	205
运行时间 [ms]	中位数	0.17	2.1	1.9	3.0
	max	155	130	46	36
全局极小值点	abs	2413	721	496	405
	rel	92.77%	27.72%	19.07%	15.57%

增广拉格朗日子问题是适定的，这一假设可能基于强制性或有界性论证。然而，一般而言，为了克服这一顾虑，可以研究增广拉格朗日方法与信赖域策略的结合，正如注4.1中所提示的。另一个值得关注的主题是注意收敛，其作用应得到仔细考察。目前尚不清楚，关于渐近驻点条件，注意收敛是否可以省略。最后，未来的研究应进一步阐明如何将驻点刻画为隐式增广拉格朗日函数的鞍点，建立广义优化的局部对偶理论，并探讨其与近端点算法之间的关系。

致谢

我谨向Andreas Themelis（九州大学）和Patrick Mehlitz（科特布斯-森夫滕贝格勃兰登堡工业大学）致谢，感谢他们对本文早期稿件提出批判性且富有建设性的意见。

参考文献

[1] Roberto Andreani、Ernesto G. Birgin、Jos´e M. Mart´ınez 和 Maria Laura Schuverdt. 关于具有一般下层约束的增广拉格朗日方法。*SIAM*优化期刊，18(4):1286–1309，2008年。[2] Egon Balas. 析取规划。施普林格，查姆，2018年。[3] Mat´u˘s Benko 和 Patrick Mehlitz. 关于优化理论中的隐变量。非光滑分析与优化期刊，2:7215，2021年。[4] Dimitri P. Bertsekas. 约束优化与拉格朗日乘子方法。Athena Scientific，1996年。[5] Ernesto G. Birgin 和 Jos´e M. Mart´ınez. 用于约束优化的非单调罚参数增广拉格朗日方法。计算优化与应用，51(3):941–965，2012年。

[6] Ernesto G. Birgin 和 Jos´e M. Mart´ınez. <style>约束优化的实用增广拉格朗日方法。工业与应用数学学会，费城，宾夕法尼亚州，2014年。[7] Stephen Boyd、Neal Parikh、Eric Chu、Borja Peleato 和 Jonathan Eckstein. 通过交替方向乘子法实现分布式优化与统计学习。《机器学习基础与趋势》，3(1):1–122，2011年。[8] Xiaojun Chen、Lei Guo、Zhaosong Lu 和 Jane J. Ye. 一种用于非利普希茨非凸规划的增广拉格朗日方法。SIAM 数值分析杂志，55:168–193，2017年。[9] Patrick L. Combettes 和 Jean-Christophe Pesquet. 信号处理中的近端分裂方法，第185–212页。施普林格，纽约，2011年。[10] Andrew R. Conn、Nicholas I. M. Gould 和 Philippe L. Toint. 一种适用于一般约束和简单界限优化问题的全局收敛增广拉格朗日算法。SIAM 数值分析杂志，28(2):545–572，1991年。[11] Andrew R. Conn、Nicholas I. M. Gould 和 Philippe L. Toint. <style>信赖域方法。工业与应用数学学会，2000年。[12] Alberto De Marchi. Bazinga.jl：一个用于 Julia 中约束结构化优化的工具箱。在线资源：https://github.com/aldma/Bazinga.jl，2020年。访问日期：2023年1月21日。[13] Alberto De Marchi. 约束结构化优化的增广拉格朗日方法与近端方法。博士论文，慕尼黑联邦国防军大学，2021年。[14] Alberto De Marchi. 一种用于凸二次规划的原始-对偶牛顿近端方法。计算优化与应用，81(2):369–395，2022年3月。[15] Alberto De Marchi、Xiaoxi Jia、Christian Kanzow 和 Patrick Mehlitz. 约束复合优化与增广拉格朗日方法。数学规划，201(1):863–896，2023年。[16] Alberto De Marchi 和 Andreas Themelis. 一种用于非凸优化的内点近端梯度法，2022年。arXiv:2208.00799v1。[17] Alberto De Marchi 和 Andreas Themelis. 局部利普希茨梯度连续性条件下的近端梯度算法。优化理论与应用期刊，194(3):771–794，2022年。[18] Neil K. Dhingra、Sei Zhen Khong 和 Mihailo R. Jovanovi´c. 用于非光滑复合优化的近端增广拉格朗日方法。IEEE 自动控制汇刊，64(7):2861–2868，2019年。[19] Brecht Evens、Puya Latafat、Andreas Themelis、Johan Suykens 和 Panagiotis Patrinos. 将神经网络训练视为最优控制问题：一种增广拉格朗日方法。载于《2021年第60届IEEE决策与控制会议（CDC）》，第5136–5143页。IEEE，2021年。[20] Philip E. Gill 和 Daniel P. Robinson. 一种原始-对偶增广拉格朗日函数。计算优化与应用，51(1):1–25，2012年。

[21]Geovani Nunes Grapiglia 和 Ya-xiang Yuan. 非凸优化增广拉格朗日方法的复杂度. IMA 数值分析期刊, 41(2):1546–1568, 2020年.[22] Lei Guo 和 Jane J. Ye. 非利普希茨非线性规划的必要最优性条件与精确罚. 数学规划, 168(1):571–598, 2018.[23] Ben Hermans、Andreas Themelis 和 Panagiotis Patrinos. QPALM：用于非凸二次规划的近端增广拉格朗日方法. 数学规划计算, 14(3):497–541, 2022年.[24] Magnus R. Hestenes. 乘子与梯度方法. 优化理论与应用期刊, 4(5):303–320, 1969.[25] Tim Hoheisel. 带消失约束的数学规划. 博士论文, 维尔茨堡大学，德国维尔茨堡, 2009.[26] Christian Kanzow、Daniel Steck 和 Daniel Wachsmuth. 巴拿赫空间中优化问题的一种增广拉格朗日方法. SIAM 控制与优化期刊, 56(1):272–291, 2018.[27] Alexander Y. Kruger 和 Patrick Mehlitz. 最优性条件、近似驻点性及其应用——一个超越利普希茨性的故事. ESAIM：控制、优化与变分法, 28:42, 2022年.[28] Puya Latafat 和 Panagiotis Patrinos. 用于求解涉及三个算子的单调包含的非对称前向-后向-伴随分裂. 计算优化与应用, 68(1):57–93, 2017年.[29] Patrick Mehlitz. 非光滑优化问题的渐近驻点性与正则性. 非光滑分析与优化期刊, 1:6575, 2020年.[30] Boris S. Mordukhovich. 变分分析与广义微分，第一部分：基础理论，第二部分：应用. 施普林格，柏林, 2006.[31] Jean Jacques Moreau. 希尔伯特空间中的邻近性与对偶性. 法国数学会公报, 93:273–299, 1965.[32] Neal Parikh 和 Stephen Boyd. 近端算法. 优化基础与趋势, 1(3):127–239, 2014年.[33] Michael J. D. Powell. 一种处理极小化问题中非线性约束的方法，第283–298页. 学术出版社, 1969.[34] R. Tyrrell Rockafellar. 通过无约束优化求解非线性规划问题的对偶方法. 数学规划, 5(1):354–373, 1973.[35] R. Tyrrell Rockafellar. 凸规划中增广拉格朗日函数及近端点算法的应用. 运筹研究数学, 1(2):97–116, 1976.[36] R. Tyrrell Rockafellar. 扩展非线性规划. 载于 Gianni Di Pillo 和 Franco Giannessi（编），《非线性优化及相关主题》，应用优化第36卷, 第381–399页. 克鲁维尔, 2000年。

[37] R. Tyrrell Rockafellar. 增广拉格朗日函数与局部最优性充分条件中的隐藏凸性. 数学规划, 198:159–194, 2023 年 .[38] R. Tyrrell Rockafellar. 增广拉格朗日方法在非线性规划之外的推广中的收敛性. 数学规划, 199:375–420, 2023年.[39] R. Tyrrell Rockafellar 和 Roger J. B. Wets. 变分分析, 第317卷. 施普林格, 1998.[40] Andreas Themelis、Lorenzo Stella 和 Panagiotis Patrinos. 两个非凸函数之和的前向-后向包络：进一步性质与非单调线搜索算法. *SIAM*优化期刊, 28(3):2274–2303, 2018.