arXiv:2304.01467v1 [math.OC] 4 Apr 2023

约束优化的部分精确罚函数方法

Nachuan Xiao * Xin Liu† 和 Kim-Chuan Toh ‡

2023年4月5日

摘要

本文聚焦于一类约束非线性优化问题（非线性规划，NLP），其中一些等式约束定义了 $\mathcal{M}$ 在 $\mathbb{R}^n$中的一个闭嵌入子流形。尽管可以采用多种现有的欧几里得空间约束优化方法直接求解 NLP，但这些方法通常无法识别 $\mathcal{M}$的流形结构。为了在直接应用这些现有优化方法时利用 $\mathcal{M}$ 的流形结构、提升效率，我们提出了一种适用于 NLP 的部分罚函数方法。在所提出的罚函数方法中，我们将 NLP 转化为欧几里得空间中的相应约束消解问题（CDP），并通过精确罚函数消除定义 $\mathcal{M}$ 的约束。我们建立了 NLP 与 CDP 的约束资格条件之间的关系，并证明在温和条件下，NLP 与 CDP 在可行域的某个邻域内具有相同的驻点和 KKT 点。因此，多种针对欧几里得空间中约束优化问题开发的现有优化方法，都可以通过 CDP 直接用于求解 NLP。初步数值实验表明，通过消除定义 $\mathcal{M}$的约束，与直接应用现有优化方法求解 NLP 相比，CDP 的计算效率更高，尤其是在高维情形下。

1 引言

本文考虑如下约束优化问题，

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & s(x)=0,\, v(x)\le 0, \end{aligned} \tag{1.1}$$

其中，等式约束的一个子集$s(x) = 0$ 满足线性无关约束资格条件 (线性无关约束资格条件)。因此，我们可以将 (1.1) 改写为以下约束非线性规划问题（非线性规划），

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & x\in\mathcal{M}:=\{x\in\mathbb{R}^n : c(x)=0\}, \\ & u_i(x)=0, \quad \forall i\in[N_E]:=\{1,...,N_E\}, \\ & v_j(x)\le 0, \quad \forall j\in[N_I]:=\{1,...,N_I\}. \end{aligned} \tag{NLP}$$

*新加坡国立大学运筹学与分析研究所，新加坡。（xnc@lsec.cc.ac.cn）。该作者的研究获得新加坡教育部学术研究基金第三层级项目（MOE-2019-T3-1-010）资助。

†中国科学院数学与系统科学研究院科学与工程计算国家重点实验室、中国科学院大学，中国。（liuxin@lsec.cc.ac.cn）。研究部分获得国家自然科学基金委员会资助（编号：12125108、11971466、11991021），以及中国科学院前沿科学重点研究计划资助（编号：ZDBS-LY-7022）。

‡新加坡国立大学数学系及运筹学与分析研究所，新加坡119076。（mattohkc@nus.edu.sg）。该作者的研究获得新加坡教育部学术研究基金第三层级项目（MOE-2019-T3-1-010）资助。

其中，$c(x) = 0$ 表示 (1.1) 中满足线性无关约束资格条件（LICQ）的等式约束子集$s(x) = 0$，而 $u(x) = 0$ 表示所有其余的等式约束。本文始终对非线性规划作如下假设，

假设1.1. 默认假设

1. f：$\mathbb{R}^n \to \mathbb{R}$ 在 $\mathbb{R}^n$ 上局部*Lipschitz*连续且为*Clarke*正则函数。*Clarke*正则函数的定义见本文的 *[12,* 定义 *2.3.4]*或定义 *2.2*。*2.* c：$\mathbb{R}^n \to \mathbb{R}^p$ 是局部利普希茨光滑映射，即 c 的转置雅可比矩阵记为 $J_c(x)$，在 $\mathbb{R}^n$ 上局部*Lipschitz*连续。*3.* 约束 $c(x) = 0$ 在 $\mathcal{M}$ 上满足线性无关约束资格条件。也就是说，对于任意 $x \in \mathcal{M}$，$J_c(x)$ 均为满秩矩阵。*4.* u：$\mathbb{R}^n \to \mathbb{R}^{N_E}$ 和 v：$\mathbb{R}^n \to \mathbb{R}^{N_I}$ 是局部利普希茨光滑映射。*5. NLP* 的可行域记为 $\mathcal{K} := \{x \in \mathcal{M} : u(x) = 0, v(x) \leq 0\}$，且非空。

非线性规划形式的优化问题在数据科学及许多其他相关领域中有广泛应用，包括计算黎曼流形上的几何平均数[20, 24, 6, 2, 7], 最大平衡割问题 [22], 二次指派问题 [10], 聚类问题 [32], 等。下面简要介绍一些非线性规划的例子。

例1.2 (黎曼质心问题**)**。 黎曼质心问题在纯数学 *[20, 6, 2],* 数据科学 *[24, 7],* 等领域有重要应用。给定黎曼流形 $\mathcal{M} := \{x \in \mathbb{R}^n : c(x) = 0\}$ 以及样本 $\{s_i : i \in [N]\}$ 来自 $\mathcal{M}$，黎曼质心可表示为

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{N} \sum_{i=1}^{N} \|x - s_i\|^2 \\ s.t. \quad & c(x) = 0, \ \|x - s^*\|^2 \leq r, \end{aligned} \tag{1.2}$$

其中，s^* 是 $\mathcal{M}$ 上的一个固定点，$r > 0$ 为常数。

例1.3 (图二分的最小平衡割**)**。 给定一个无向图，最小割问题旨在将顶点划分为两个大小相同的簇，同时确保两个簇之间的边数尽可能少。如 *[22],* 所述，这类问题可以松弛为以下带有额外线性约束的黎曼优化问题，

$$\begin{aligned} \min_{X \in \mathbb{R}^{m \times q}} \quad & -\frac{1}{4} \mathrm{tr}\left(X^\top L X\right) \\ s.t. \quad & \mathrm{Diag}(XX^\top) = I_m, \\ & X^\top e = 0, \end{aligned} \tag{1.3}$$

其中，L 是图的拉普拉斯矩阵，$e \in \mathbb{R}^m$ 是全1向量，约束 $\mathrm{Diag}(XX^\top) = I_m$在 $\mathbb{R}^{m \times q}$ 中定义了一个斜流形。

如 [1], 中所述，非线性规划中的等式约束 $c(x) = 0$ 定义了 $\mathbb{R}^n$ 中的嵌入子流形$\mathcal{M}$。因此，非线性规划可以被视为在 $\mathbb{R}^n$ 上带有等式约束$[c(x); u(x)] = 0$ 和不等式约束 $v(x) \leq 0$的《优化》问题；也可以被视为在嵌入子流形 $\mathcal{M}$ 上，附加约束为 $u(x) = 0$ 和 $v(x) \leq 0$的《优化》问题。

根据对约束 $c(x) = 0$ 的处理方式，可以将用于求解非线性规划的优化方法分为两大类。一类将非线性规划视为 $\mathbb{R}^n$ 中的标准非线性约束优化问题，其中约束 $c(x) = 0$ 作为一组额外的等式约束嵌入其中，而不对其底层流形结构进行任何特殊处理。于是，大量的欧几里得优化方法（即为 $\mathbb{R}^n$ 中的约束优化问题设计的方法）都可以直接用于求解非线性规划。这些方法包括增广拉格朗日法（ALM） [17, 27], 序列二次规划法（SQP） [14], 内点法

[8], 滤波方法 [15], 等。此外，得益于过去几十年来在 $\mathbb{R}^n$ 中求解约束优化问题所积累的丰富经验，开发出了各种高效的欧几里得求解器（即为 $\mathbb{R}^n$ 中的约束优化开发的求解器），包括 ALGENCAN [3, 4] 和 Ipopt [31], ，并广泛应用于求解非线性规划。

然而，由于约束$c(x) = 0$ 在 $\mathbb{R}^n$ 中定义了一个嵌入子流形，另一类优化方法 [22, 7] 将$u(x) = 0$ 和$v(x) \leq 0$ 视为嵌入子流形 $\mathcal{M}$上的附加约束，即，

$$\begin{aligned} \min_{x \in \mathcal{M}} \quad & f(x) \\ \text{s.t.} \quad & u(x) = 0,\ v(x) \leq 0. \end{aligned} \tag{1.4}$$

遵循文献 [1], 中提出的广为认可的框架，现有的黎曼优化方法（即为黎曼流形上的优化而设计的优化方法）通过将高效的欧几里得优化方法从 $\mathbb{R}^n$ 推广到 $\mathcal{M}$而发展起来，包括黎曼序列二次规划方法 [5, 28], 、黎曼增广拉格朗日方法 [22, 30], 、黎曼非光滑罚函数方法 [22], 、黎曼弗兰克–沃尔夫方法 [34], 等。与直接应用欧几里得优化方法求解非线性规划相比，这些黎曼优化方法通过将$c(x) = 0$ 识别为黎曼流形，并利用其几何结构$\mathcal{M}$ [22]，有望更高效地求解问题。

尽管这些黎曼优化方法是求解 (1.4) 的有力替代方案，但开发这些黎曼方法依赖于底层流形的几何工具，包括黎曼导数、回缩及其逆、向量传输等 [1, 9]。确定这些几何工具通常颇具挑战性，实例见 [1, 9, 29] 。因此，这些几何工具仅在若干知名黎曼流形上得到了深入研究，这使得现有黎曼优化方法的应用仅限于若干知名流形。

此外，即使已有适用于 $\mathcal{M}$ 的几何工具，要依据 [1] 中描述的框架，从欧几里得优化方法开发黎曼优化方法仍非易事。要将欧几里得优化方法转换为其黎曼版本，需要将f、u 和v 的导数替换为黎曼导数，引入回缩以确保迭代点保持在 $\mathcal{M}$ 中，并采用向量传输在 $\mathcal{M}$ 的不同切空间之间移动向量。因此，现有的黎曼优化方法数量有限，尤其是与日益增多的欧几里得优化方法相比更是如此。此外，与各种现有的高效欧几里得求解器（即基于欧几里得优化方法开发的求解器）形成对比的是，目前尚无公开发布的非线性规划黎曼求解器。考虑到开发黎曼优化方法及其相应求解器所需付出的巨大努力，这些方法难以跟上欧几里得空间中非凸约束优化的发展与进步。

简而言之，直接通过欧几里得优化方法求解非线性规划，可以受益于各种现有的高效算法和求解器，十分便利。另一方面，通过黎曼优化方法求解非线性规划，可以利用 $\mathcal{M}$的结构来实现更高的效率，但其中涉及的几何工具可能会给开发高效的黎曼优化求解器带来很大困难。因此，自然可以考虑如何结合欧几里得优化方法和黎曼优化方法各自的优势。更确切地说，我们受到以下问题的启发：

> 我们能否直接应用欧几里得优化求解器来求解非线性规划，同时通过利用由约束 $c(x) = 0$定义的嵌入子流形 $\mathcal{M}$ 的结构来实现更高的效率？

最近， [35] 表明，以下光滑优化问题

$$\min_{x \in \mathbb{R}^n} \quad f(x), \qquad \text{s. t. } c(x) = 0, \tag{1.5}$$

等价　　转化为以下约束消解函数的无约束最小化问题，

$$f(\mathcal{A}(x)) + \frac{\beta}{2} \|c(x)\|^2. \tag{1.6}$$

其中， $\beta \geq 0$是惩罚参数，约束消解映射 $\mathcal{A}$ 应满足以下假设。

假设 1.4. *$\mathcal{A}$的整体假设*

1. $\mathcal{A}$： $\mathbb{R}^n \to \mathbb{R}^n$ 在 $\mathbb{R}^n$中是局部利普希茨光滑的；
2. $\mathcal{A}(x) = x$ 对任意 $x \in \mathcal{M}$均成立；
3. 对任意 $x \in \mathcal{M}$， $c(\mathcal{A}(x))$ 的雅可比矩阵等于0 。也就是说，对任意 $x \in \mathcal{M}$， $J_A(x)J_c(x) = 0$ 成立，其中 $J_A(x) \in \mathbb{R}^{n\times n}$ 表示 $\mathcal{A}$ 在 x 处的转置雅可比矩阵。

各种重要流形的约束消解映射 $\mathcal{A}$ 的构造可参见 [35, 第 4]节。正如 [35], 所证明的，在适当选择惩罚参数 β后，(1.5) 和 (1.6) 在 $\mathcal{M}$ 附近具有相同的驻点。此外， [35] 表明，构造 $\mathcal{A}$ 完全无需借助 $\mathcal{M}$ 的任何几何工具。因此，可以直接应用各种无约束优化方法，通过求解 (1.6) 来求解 (1.5)。

我们的动机源于黎曼优化中的约束消解方法[35]。(1.6) 中约束消解函数的构造启发我们考虑在非线性规划中用相应的约束消解函数替换f、u 和v ，并通过精确罚函数化移除非线性规划中的约束$c(x) = 0$ 。由此得到以下约束消解问题（CDP）：

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & h(x) := f(\mathcal{A}(x)) + \frac{\beta}{2}\|c(x)\|^2 \\ \text{s.t.} \quad & \tilde{u}_i(x) := u_i(\mathcal{A}(x)) + \frac{\tau_i}{2}\|c(x)\|^2 = 0, \quad i \in [N_E], \\ & \tilde{v}_j(x) := v_j(\mathcal{A}(x)) + \frac{\gamma_j}{2}\|c(x)\|^2 \leq 0, \quad j \in [N_I], \end{aligned} \tag{CDP}$$

其中， β、 $\{\tau_i\}$ 和 $\{\gamma_j\}$ 均为非负罚参数。由于 CDP 中约束消解映射 $\mathcal{A}$ 的构造独立于 $\mathcal{M}$ [35]的几何工具，我们可以直接将各种现有的欧几里得求解器用于 CDP，而无需计算回缩、向量传输、黎曼微分等几何工具。此外，与非线性规划相比，CDP 消除了约束$c(x) = 0$。因此，使用欧几里得求解器求解由此得到的 CDP 有望提高效率。

然而，现有的约束消解方法仅适用于在 $\mathcal{M}$ 上最小化目标函数f，且不含任何其他约束的情形。如何建立非线性规划与 CDP 之间的等价关系，尤其是在存在非光滑目标函数和其他约束时，仍有待解决。更重要的是，由于非线性规划与 CDP 的约束不同，因此应仔细分析二者在约束资格条件、驻点和 KKT 点方面的关系。

贡献 本文提出一种针对非线性规划的约束消解方法，将其转化为 CDP，并消去约束 $c(x) = 0$ 。我们首先证明，非线性规划与 CDP 的约束资格条件具有等价性，具体而言，非线性规划的一大类约束资格条件可推出 CDP 的所有 KKT 点都是其一阶驻点。此外，我们证明，在温和条件下，非线性规划与 CDP 在可行域上具有相同的驻点和 Karush–Kuhn–Tucker（KKT）点。除此之外，我们还针对非线性规划中 $N_E = 0$ 的特殊情形，证明了更强的等价性结果。这些关于非线性规划与 CDP 等价性的结果表明，将非线性规划转化为 CDP 能够保留约束资格条件的有效性，同时保持非线性规划的驻点和 KKT 点不变。因此，我们可以直接采用各种现有的欧几里得方法求解非线性规划。

我们对黎曼质心问题和最小平衡割问题进行了数值实验，以展示所提出的约束消解方法的优势。作为基准，我们将所提出的约束消解方法与上述方法进行比较。

欧几里得优化方法只需将 $c(x) = 0$ 作为一组额外的等式约-束来考虑。初步数值实验表明，将现有的欧几里得优化方法应用于 CDP 可以实现更高的计算效率，尤其是在高维情况下。这些数值实验进一步表明，通过 CDP 求解非线性规划可以享有现有黎曼优化方法所带来的类似优势，同时避免确定底层流形的几何工具以及设计黎曼优化求解器时所面临的困难。

结构 本文的安排如下。第 2 节确定本文证明所需的记号、定义和常数。我们建立 CDP 的理论性质，并讨论仅含不等式约束的特殊情形。CDP 理论性质的证明见附录。第 4 节给出若干说明性数值示例，展示 CDP 如何让各种欧几里得优化方法得以直接实现。最后一节对本文作总结。

2 记号、定义和常数

2.1 记号

设 range(A) 为由矩阵 A 的列向量张成的子空间，$\|\cdot\|_1$ 和 $\|\cdot\|$分别表示向量或算子的 ℓ_1-范数和 ℓ_2-范数。记号 diag(A) 和 Diag(x) 分别表示由矩阵 A 的对角元素构成的向量，以及以 x 的元素 $\in \mathbb{R}^n$ 为对角元素的对角矩阵。我们分别用 $\lambda_{\min}$(A) 和 $\lambda_{\max}$(A) 表示 A 的最小和最大特征值。此外，σ_l(A) 表示 A 的第 l 大奇异值，$\sigma_{\min}$(A) 表示 A 的最小奇异值。进一步，对于任意矩阵 A $\in \mathbb{R}^{n\times p}$，我们用 $A^\dagger$表示 A 的伪逆。也就是说，$A^\dagger \in \mathbb{R}^{p\times n}$满足 $AA^\dagger A = A$、$A^\dagger AA^\dagger = A^\dagger$，且 $A^\dagger A$ 和 $AA^\dagger$ 均为对称矩阵 [16]。

对于子流形 $\mathcal{M}$，我们将 $\mathcal{T}_x$ 定义为 $\mathcal{M}$ 在 x 处的切空间，可表示为 $\mathcal{T}_x := \{d \in \mathbb{R}^n : J = \{ \in \mathbb{R}^n \top = 0 \ \forall \in \mathcal{T}_x\}$ $\mathcal{T}_x$ $\mathcal{N}_x$ $\mathcal{N}_x =$ $_c$ (x)$^\top$d $= 0\}$。此外，$\mathcal{N}_x$ 表示 $\mathcal{M}$ 在 x 处的法空间，即 $\mathcal{N}_x$: d : d u, u。根据 和 的定义，有 值域(J (x))。此外，对于任意 x $\in \mathbb{R}^n$，我们将到 $\mathcal{M}$ 的投影定义为 proj(x, $\mathcal{M}$): $= \arg\min_{y\in\mathcal{M}} \|x - y\|$。值得一提的是，对于任意 x $\in \mathbb{R}^n$，上述问题的最优性条件意味着，对于任意 y $\in$proj(x, $\mathcal{M}$)，有 x $-$ y $\in$值域(J$_c$(y))。此外，dist(x, $\mathcal{M}$)表示 x 与 $\mathcal{M}$之间的距离，即 dist(x, $\mathcal{M}$): $= \inf_{y\in\mathcal{M}} \|x - y\|$。

映射 $\mathcal{A}$ 和c 的转置雅可比矩阵分别记为$J_A(x) \in \mathbb{R}^{n\times n}$ 和$J_c(x) \in \mathbb{R}^{n\times p}$。也就是说，令$c_i$ 和 $\mathcal{A}_i$ 分别表示映射c 和 $\mathcal{A}$ 的第i个坐标，则J_c 和J_A 定义为

$$J_c(x) := \left[\nabla c_1(x), ..., \nabla c_p(x)\right], \quad \text{and} \quad J_A(x) := \left[\nabla \mathcal{A}_1(x), ..., \nabla \mathcal{A}_n(x)\right].$$

类似地，我们将J_u 和J_v 分别定义为u 和v的输运雅可比矩阵。此外，对于 $\mathbb{R}^n$的给定子集 $\mathcal{G}$，我们定义J_A $(x)\mathcal{G} := \{J_A(x)d : d \in \mathcal{G}\}$。令

$$\mathcal{A}^k(x) := \underbrace{\mathcal{A}(\mathcal{A}(\cdots \mathcal{A}(x)\cdots))}_{k \text{ times}},$$

对于 k ≥ 1，并定义 $\mathcal{A}^0$(x) := x, $\mathcal{A}^\infty$(x) := $\lim_{k\to+\infty}$ $\mathcal{A}^k$(x)。此外，在本文余下部分，我们记 g(x) := f ($\mathcal{A}$(x))。另外，将以 x 为中心、半径为 ρ 的闭球记为 $\mathcal{B}_{x,\rho} := \{y \in \mathbb{R}^n : \|y - x\| \leq \rho\}$。最后，对于任意正整数 n，我们使用记号 [n] := {1, 2,..., n}。我们将$\mathcal{K}_A$ 记为 CDP 的可行域，即 $\mathcal{K}_A := \{x \in \mathbb{R}^n : \tilde{}u(x) = 0, \tilde{}v(x) \leq 0\}$，并将 $\mathcal{F}$(x) 记为非线性规划中不等式约束的活跃指标集，即 $\mathcal{F}(x) := \{j \in [N_I] : v_j(x) = 0\}$。类似地，我们将 CDP 中不等式约束的活跃指标集记为 $\mathcal{F}_A(x) := \{j \in [N_I] : \tilde{}v_j(x) = 0\}$。注意，对于任意 x $\in \mathcal{M}$，有 $\mathcal{F}$(x) = $\mathcal{F}_A$(x)，因为对于任意 x $\in \mathcal{M}$ 且 j $\in [N_I]$，均有 $\tilde{}v_j(x) = v_j(x)$。此外，容易验证 $\mathcal{K} \subseteq \mathcal{K}_A$。

2.2 预备知识

2.2.1 次微分与正则性

定义 2.1. f 在 $x \in \mathbb{R}^n$ 沿方向 d 的广义方向导数，记为 $f^\star(x, d)$，定义如下：

$$f^\star(x,d) = \limsup_{y \to x, t \downarrow 0} \frac{f(y+td) - f(y)}{t}. \tag{2.1}$$

广义　梯度或 f 在 $x \in \mathbb{R}^{n\times p}$ 处的（*Clarke*）次微分，记作 $\partial f(x)$，定义为　s

$$\partial f(x) := \{w \in \mathbb{R}^n : \langle w, d\rangle \le f^\star(x,d), \textit{for all } d \in \mathbb{R}^n\}. \tag{2.2}$$

定义 2.2. 如果对于每个方向 d，单侧方向导数存在，则称 f 在 $x \in \mathbb{R}^n$ 处是（*Clarke*）正则的

$$f'(x,d) = \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} \tag{2.3}$$

存在，且 $f'(x,d) = f^\star(x,d)$。

2.2.2 约束资格条件和最优性条件

在本小节中，我们分别给出非线性规划（NLP）和（CDP）的约束优化问题的约束资格条件定义。值得一提的是，非线性规划（NLP）和（CDP）都是 $\mathbb{R}^n$ 中的约束优化问题，因此可以用相同的方式定义它们的约束资格条件。回顾一下，$\mathcal{K}$ 和 $\mathcal{K}_A$ 分别表示非线性规划（NLP）和（CDP）的可行域，我们给出（NLP）和（CDP）的约束资格条件定义如下。

定义 2.3. 对于 $\mathbb{R}^n$中的任意闭锥 $\mathcal{C}$，其极锥 $\mathcal{C}^\circ$ 定义为

$$\mathcal{C}^\circ := \{d \in \mathbb{R}^n : \langle d, w\rangle \le 0,\ \forall w \in \mathcal{C}\}.$$

定义 2.4. 对于任意 $x \in \mathcal{K}$，相对于 $\mathcal{K}$ 的（*Bouligand*）切锥定义为

$$\mathcal{T}_{\mathcal{K}}(x) := \left\{ d \in \mathbb{R}^n : \exists x_k \in \mathcal{K} \to x,\ t_k \downarrow 0,\ s.t.\ d = \lim_{k\to+\infty} \frac{x_k - x}{t_k} \right\}.$$

类似地，对于任意 $x \in \mathcal{K}_A$，相对于 $\mathcal{K}_A$ 的（*Bouligand*）切锥定义为

$$\mathcal{T}_{\mathcal{K}_A}(x) := \left\{ d \in \mathbb{R}^n : \exists x_k \in \mathcal{K}_A \to x,\ t_k \downarrow 0,\ s.t.\ d = \lim_{k\to+\infty} \frac{x_k - x}{t_k} \right\}.$$

定义 2.5。对于任意 $x \in \mathcal{K}$，相对于 $\mathcal{K}$ 的线性化锥定义为

$$\mathcal{T}^{lin}_{\mathcal{K}}(x) := \left\{ d \in \mathbb{R}^n : \langle d, \nabla c_l(x)\rangle = 0, \langle d, \nabla u_i(x)\rangle = 0, \left\langle d, \nabla v_j(x)\right\rangle \le 0, l \in [p], i \in [N_E], j \in \mathcal{F}(x) \right\}.$$

同样，对于任意 $x \in \mathcal{K}_A$，关于 $\mathcal{K}_A$ 的线性化锥定义为

$$\mathcal{T}^{lin}_{\mathcal{K}_A}(x) := \left\{ d \in \mathbb{R}^n : \langle d, \nabla \tilde{u}_i(x)\rangle = 0, \left\langle d, \nabla \tilde{v}_j(x)\right\rangle \le 0, i \in [N_E], j \in \mathcal{F}_A(x) \right\}$$

由定义 2.5 可知，对于任意 $x \in \mathcal{K}$，$\mathcal{T}_{\mathcal{K}}(x) \subset \mathcal{T}^{lin}_{\mathcal{K}}(x)$ 以及

$$(\mathcal{T}^{lin}_{\mathcal{K}}(x))^\circ = \left\{ \sum_{i\in[p]} \rho_l \nabla c_l(x) + \sum_{i\in[N_E]} \lambda_i \nabla u_i(x) + \sum_{j\in[N_I]} \mu_j \nabla v_j(x) \;\middle|\; \begin{array}{l} \rho \in \mathbb{R}^p,\ \lambda \in \mathbb{R}^{N_E}, \\ \mu \in \mathbb{R}^{N_I}_+, \mu_j = 0\ \forall\ j \notin \mathcal{F}(x) \end{array} \right\}.$$

同样可以证明，对于任意$x \in \mathcal{K}_A$, $\mathcal{T}_{\mathcal{K}_A}(x) \subset \mathcal{T}^{lin}_{\mathcal{K}_A}(x)$ 以及

$$(\mathcal{T}^{lin}_{\mathcal{K}_A}(x))^\circ = \Big\{ \sum_{i\in[N_E]} \lambda_i \nabla \tilde{u}_i(x) + \sum_{j\in[N_I]} \mu_j \nabla \tilde{v}_j(x) \mid \lambda \in \mathbb{R}^{N_E}, \mu \in \mathbb{R}^{N_I}_+, \mu_j = 0\ \forall\ j \notin \mathcal{F}_A(x) \Big\}.$$

定义义 2.6。对于给定的 $x \in \mathcal{K}$，我们在 x 处为非线性规划定义以下约束资格条件 .

- 若 $\{\nabla cl(x) : l \in [p]\} \cup \{\nabla ui(x) : i \in [NE]\} \cup \{\nabla vj(x) : j \in \mathcal{F}(x)\}$ 在 $\mathbb{R}n$ 中构成线性无关集，则非线性规划关于 x 满足线性无关约束资格条件（*LICQ*）。
- 若 $\{\nabla cl(x) : l \in [p]\} \cup \{\nabla ui(x) : i \in [NE]\}$ 线性无关，且存在 $d \in \mathbb{R}n$，使得非线性规划关于 x 满足 *Mangasarian-Fromovitz* 约束资格条件（*MFCQ*）。

$$\begin{aligned}&\langle d, \nabla c_l(x)\rangle = 0,\ \forall l \in [p],\\ &\langle d, \nabla u_i(x)\rangle = 0,\ \forall i \in [N_E],\\ &\langle d, \nabla v_j(x)\rangle \le 0,\ \forall j \in \mathcal{F}(x).\end{aligned}$$

- 若 $\mathcal{T}_{\mathcal{K}}(x)^\circ = \mathcal{T}lin_{\mathcal{K}}(x)^\circ$，则非线性规划关于 x 满足 *GCQ*约束资格条件。

类似地，对于任意给定的 $x \in \mathcal{K}_A$，我们定义 *CDP* 在 x 处满足的以下约束资格条件。

- 若 $\{\nabla \tilde{} ui(x) : i \in [NE]\} \cup \{\nabla \tilde{} vj(x) : j \in \mathcal{F}(x)\}$ 在 $\mathbb{R}n$ 中构成线性无关集，则 *CDP* 关于 x 满足线性无关约束资格条件（*LICQ*）。
- 若 x 满足关于 *CDP* 的 *Mangasarian-Fromovitz*约束规范（*MFCQ*），则 $\{\nabla \tilde{} ui(x) : i \in [NE]\}$ 线性无关，且存在 $d \in \mathbb{R}n$，使得

$$\begin{aligned}&\langle d, \nabla \tilde{u}_i(x)\rangle = 0,\ \forall i \in [N_E],\\ &\langle d, \nabla \tilde{v}_j(x)\rangle \le 0,\ \forall j \in \mathcal{F}_A(x).\end{aligned}$$

- 若 x 满足关于 *CDP* 的 *GCQ*约束资格条件，则 $\mathcal{T}_{\mathcal{K}^A}(x)^\circ = \mathcal{T}lin^{\mathcal{K}_A}(x)^\circ$。

值得一提的是，定义 2.6 中的约束资格条件等价于 [7, 定义 3.12],中定义的黎曼约束资格条件，正如 [7, 注 3.13]中所讨论的。此外，根据定义 2.4 中切锥的定义，我们将 NLP 和 CDP 的驻点定义如下。

定义 2.7. 若 $x \in \mathcal{K}$ 满足以下条件，则称其为 *NLP* 的一阶驻点

$$0 \in \partial f(x) + \mathcal{T}_{\mathcal{K}}(x)^\circ. \tag{2.4}$$

此外，若 $x \in \mathcal{K}_A$ 满足

$$0 \in \partial h(x) + \mathcal{T}_{\mathcal{K}_A}(x)^\circ. \tag{2.5}$$

2.2.3 KKT条件

在本小节中，我们介绍非线性规划和CDP的KKT条件与最优性条件的定义。

定义 2.8。对于任意给定的 $x \in \mathbb{R}^n$，非线性规划的拉格朗日函数定义为

$$\mathcal{L}_{NLP}(x, \lambda, \rho, \mu) := f(x) + \rho^\top c(x) + \lambda^\top u(x) + \mu^\top v(x), \tag{2.6}$$

其关于变量 x 的偏次微分定义为

$$\partial_x \mathcal{L}_{NLP}(x, \rho, \lambda, \mu) := \partial f(x) + \sum_{l\in[p]} \rho_l \nabla c_l(x) + \sum_{i\in[N_E]} \lambda_i \nabla u_i(x) + \sum_{j\in[N_I]} \mu_j \nabla v_j(x). \tag{2.7}$$

此外，对于任意给定的 $x \in \mathbb{R}^n$，*CDP*的拉格朗日函数定义为

$$\mathcal{L}_{CDP}(x, \lambda, \mu) := h(x) + \lambda^\top \tilde{u}(x) + \mu^\top \tilde{v}(x), \tag{2.8}$$

其关于变量 x 的偏次微分定义为

$$\partial_x \mathcal{L}_{CDP}(x, \lambda, \mu) := \partial h(x) + \sum_{i\in[N_E]} \lambda_i \nabla \tilde{u}_i(x) + \sum_{j\in[N_I]} \mu_j \nabla \tilde{v}_j(x). \tag{2.9}$$

基于非线性规划和 CDP 的拉格朗日函数，我们给出如下 KKT 点的定义。**定义 2.9**。对于任意给定的 $x \in \mathcal{M}$，若 x 是非线性规划的 *KKT* 点（乘子为 ρ，λ 且 μ），则

$$\begin{aligned} &0 \in \partial_x \mathcal{L}_{NLP}(x, \rho, \lambda, \mu), \\ &c(x) = 0,\ u(x) = 0,\ v(x) \le 0,\ \mu \ge 0,\ \mu^\top v(x) = 0. \end{aligned} \tag{2.10}$$

定义 2.10 式中，对于任意给定的 $x \in \mathbb{R}^n$，若 x 是 *CDP* 的 *KKT* 点（乘子为 λ 和 μ），则

$$\begin{aligned} &0 \in \partial_x \mathcal{L}_{CDP}(x, \lambda, \mu), \\ &\tilde{u}(x) = 0,\ \tilde{v}(x) \le 0,\ \mu \ge 0,\ \mu^\top \tilde{v}(x) = 0. \end{aligned} \tag{2.11}$$

2.3 常数

本文中，对于任意给定的$x \in \mathcal{M}$，我们选择一个正标量 ρ_x，使得

$$\rho_x := \underset{0<\rho\le 1}{\arg\max} \left\{ \rho : \inf\left\{ \sigma_{\min}(J_c(y)) : y \in \mathcal{B}_{x,\rho} \right\} \ge \frac{1}{2}\sigma_{\min}(J_c(x)) \right\}.$$

那么对于任意给定的 g 定的$x \in \mathcal{M}$，基于 ρ_x 的定义，我们定义 $\Theta_x := \{y \in \mathcal{X}: \|y - x\| \le \rho_x$ $\}$并设定若干常数如下：

- $\sigma_{x,c} := \sigma_{\min}(Jc(x))$;
- $M_{x,f} := \sup_{y\in\Theta_x} \sup_{w\in\partial f(\mathcal{A}(y))} \|w\|$;
- $M_{x,u} := \sup_{y\in\Theta_x} \|Ju(y)\|$;
- $M_{x,v} := \sup_{y\in\Theta_x} \|Jv(y)\|$;
- $M_{x,c} := \sup_{y\in\Theta_x} \|Jc(y)\|$;
- $M_{x,A} := \sup_{y\in\Theta_x} \|JA(y)\|$;
- $L_{x,c}$: =上确界y，$z\in\Theta_x$，$y\neq z \|Jc(y)$
  - $\|Jc(z)y z$;
- $L_{x,A}$:=上确界$y,z\in\Theta_x, y\neq z \|JA(y)$
  - $\|JA(z)y z$;
- $L_{x,b}$:=上确界$y,z\in\Theta_x, y\neq z \|JA(y)Jc(\mathcal{A}(y))$
  - $\mathcal{A}$ $\|JA(z)Jc((z))yz$.

值得一提的是，上述所有常数均与惩罚参数β无关。此外，对于任意给定的 x $\in \mathbb{R}^n$ λ μ $_{x\lambda\mu}$ = $_{xf} + \|\lambda\|_1$ $_{xu}$ + 及其乘子(,)，我们设定 M, , : $\|\mu\|_1$ $_{xv}$ $M,M,M,$. 此外，基于这些常数，我们设定

- $\varepsilon_x := \min\left\{ \frac{\rho_x}{2}, \frac{\sigma_{x,c}}{32L_{x,c}(M_{x,A}+1)}, \frac{\sigma_{x,c}^2}{8L_{x,b}M_{x,c}} \right\}$;
- $\Omega_x := \{y \in \mathbb{R}^n : \|y - x\| \le \varepsilon_x\}$;
- $\overline{\Omega}_x := \left\{ y \in \mathbb{R}^n : \|y - x\| \le \frac{\sigma_{x,c}\varepsilon_x}{4M_{x,c}(M_{x,A}+1)+\sigma_{x,c}} \right\}$;
- $\Omega := \bigcup_{x\in\mathcal{M}} \Omega_x$;
- $\overline{\Omega} := \bigcup_{x\in\mathcal{M}} \overline{\Omega}_x$.

注意，$\varepsilon_x \le \frac{\sigma_x^2}{8} L_{x,b} M_{x,c}$ 会自动推出 $\varepsilon_x \le \frac{\sigma_x}{4} L_{x,b}$，因为 $\sigma_{x,c} \le M_{x,c}$。此外，值得一提的是，对于任意给定的$x \in \mathcal{M}$，假设 1.1 保证 $\sigma_{x,c} > 0$ 和 $\varepsilon_x > 0$。进一步地，根据 $\overline{\Omega}$ 的定义，我们可以得出，对于任意给定的$x \in \mathcal{M}$，$\overline{\Omega}_x \subset \Omega_x \subset \Theta_x$ 成立，且 $\mathcal{M}$ 位于 $\overline{\Omega}$ 的内部。

3 理论性质

本节中，我们介绍 CDP 的理论性质。首先，我们在下述命题中给出 $\partial h(x)$ 的显式表达式。由于通过直接计算即可证明，我们省略其证明。

命题 3.1. 对于任意 $x \in \mathbb{R}^n$，均有

$$\partial h(x) = J_A(x)\partial f(\mathcal{A}(x)) + \beta J_c(x)c(x), \tag{3.1}$$
$$\nabla \tilde{u}_i(x) = J_A(x)\nabla u_i(\mathcal{A}(x)) + \tau_i J_c(x)c(x), \tag{3.2}$$
$$\nabla \tilde{v}_j(x) = J_A(x)\nabla v_j(\mathcal{A}(x)) + \gamma_j J_c(x)c(x). \tag{3.3}$$

此外，对于任意 $x \in \mathcal{M}$、$i \in [N_E]$ 和 $j \in [N_I]$，均有

$$\partial h(x) = J_A(x)\partial f(x),\ \nabla \tilde{u}_i(x) = J_A(x)\nabla u_i(x),\ \textit{and}\ \nabla \tilde{v}_j(x) = J_A(x)\nabla v_j(x).$$

3.1 约束资格条件

本小节分析非线性规划与 CDP 的约束资格条件之间的关系。我们证明，非线性规划中的线性无关约束资格条件和 Mangasarian-Fromovitz约束规范，分别蕴含 CDP 中相应约束资格条件成立；而非线性规划中的 GCQ约束资格条件则意味着，CDP 的 KKT 点均为 CDP 的一阶驻点。图 1 总结了本小节的主要结果。

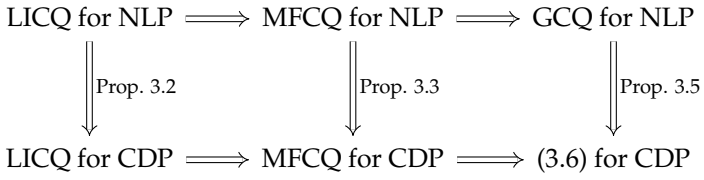

图 1：非线性规划与 CDP 之间约束资格条件的关系。定理 3.6 表明，式 (3.6) 可视为 CDP 的一种约束资格条件。此处“Prop.”是“Proposition”（命题）的缩写。

命题 3.2 至命题 3.5 的证明见附录 A.2 节。

命题 3.2。 对于任意 $x \in \mathcal{K}$，假设 *NLP* 在 x 处满足线性无关约束资格条件，则 *CDP* 在 x 处也满足线性无关约束资格条件。

命题3.3。 对于任意 $x \in \mathcal{K}$，假设在 x 处相对于非线性规划满足 *Mangasarian-Fromovitz*约束规范，则在 x 处相对于 *CDP* 也满足 *Mangasarian-Fromovitz*约束规范。

引理 3.4. 对于任意 $x \in \mathcal{K}$，有

$$\mathcal{T}^{lin}_{\mathcal{K}}(x) = J_A(x)^\top \mathcal{T}^{lin}_{\mathcal{K}_A}(x), \quad \mathcal{T}^{lin}_{\mathcal{K}_A}(x)^\circ \cap \mathrm{range}(J_A(x)) = J_A(x)\mathcal{T}^{lin}_{\mathcal{K}}(x)^\circ, \tag{3.4}$$
$$\mathcal{T}_{\mathcal{K}}(x) \subseteq J_A(x)^\top \mathcal{T}_{\mathcal{K}_A}(x), \quad \mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \mathrm{range}(J_A(x)) \subseteq J_A(x)\mathcal{T}_{\mathcal{K}}(x)^\circ. \tag{3.5}$$

在本小节的其余部分，我们旨在说明，非线性规划的 GCQ约束资格条件意味着 CDP 的所有 KKT点都是 CDP 的一阶驻点。我们首先展示当非线性规划的 GCQ约束资格条件在 $x \in \mathcal{M}$ 处成立时，$\mathcal{T}_{\mathcal{K}_A}(x)$ 与 $\mathcal{T}^{lin}_{\mathcal{K}_A}(x)$ 之间的关系。

命题 3.5。 对于任意 $x \in \mathcal{K}$，假设 *NLP* 在 x 处满足 *GCQ*约束资格条件，即 $\mathcal{T}_{\mathcal{K}}(x)^\circ = \mathcal{T}^{lin}_{\mathcal{K}}(x)^\circ$。则有

$$\mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \mathrm{range}(J_A(x)) = \mathcal{T}^{lin}_{\mathcal{K}_A}(x)^\circ \cap \mathrm{range}(J_A(x)). \tag{3.6}$$

正如许多已有研究 [25], 所指出的，GCQ通常被认为是最弱的约束资格条件之一，因为大量 NLP 的约束资格条件都能推出 NLP 满足 GCQ。关于各种约束资格条件之间关系的更多细节，感兴趣的读者可参阅 [33]。在下述定理中，我们说明命题 3.5 保证 CDP 的任意一阶驻点都满足 KKT条件。定理 3.6 的证明见 A.2 节。

定理 3.6。 对于任意 $x \in \mathcal{K}$，假设 $\mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap 值域(J_A(x)) = \mathcal{T}lin^{\mathcal{K}_A}(x)^\circ \cap 值域(J_A(x))$，那么 x 是 *CDP* 的一阶驻点，当且仅当 x 是 *CDP* 的 *KKT* 点。

3.2 驻点性

本小节旨在分析非线性规划与 CDP 在一阶驻点和 KKT 点方面的关系。图2总结了本小节的主要结果。此外，定理 3.7、定理 3.8 和定理3.10 的证明见 A.3 节。

$$\begin{array}{ccc} \text{FOSPs of NLP} & \xleftarrow{\text{Theo. 3.7}} & \text{FOSPs of CDP} \\ \Downarrow & & \Downarrow \\ \text{KKTs of NLP} & \underset{\text{Theo. 3.8}}{\overset{\text{Theo. 3.7}}{\Longleftrightarrow}} & \text{KKTs of CDP} \end{array}$$

图2：非线性规划与 CDP 的一阶驻点（FOSP）和 KKT 点（KKT）的关系。这里，“Theo.”是“Theorem”的缩写。

定理 3.7。 对于任意给定的 $x \in \mathcal{K}$，有

1. 如果 x 是 *CDP* 的一阶驻点，那么 x 就是非线性规划的一阶驻点；

2. x 是非线性规划的 *KKT* 点，当且仅当 x 是 *CDP* 的 *KKT* 点。

采用欧几里得优化方法求解 CDP，通常会得到一个收敛到非线性规划的 KKT 点的不可行序列。因此，分析非线性规划的 KKT 点与 CDP 的 KKT 点之间的关系至关重要。下述定理说明，在温和条件下，当 β 足够大时，CDP 在 $\overline{\Omega}_x$ 中针对某个 $x \in \mathcal{M}$ 的任意 KKT 点 y，都是非线性规划的 KKT 点。请注意，此处的结果与定理 3.7 的区别在于，我们不对定理 3.8 中的 KKT 点 $y \in \mathcal{K}$ 作出假设。

定理 3.8。 对于任意给定的 $x \in \mathcal{M}$，假设 $y \in \overline{\Omega}_x$ 是 *CDP* 的 *KKT* 点，其乘子为 λ 和 μ，那么有

$$\begin{aligned} &\operatorname{dist}(0, \partial_x \mathcal{L}_{CDP}(y, \lambda, \mu)) \\ &\geq \left(\frac{\sigma_{x,c}}{4(M_{x,A}+1)} \Big(\beta + \sum_{i \in [N_E]} \lambda_i \tau_i + \sum_{j \in [N_I]} \mu_j \gamma_j \Big) - \frac{4 L_{x,A} M_{x,\lambda,\mu}}{\sigma_{x,c}} \right) \|c(y)\| \end{aligned} \tag{3.7}$$

此外，假设

$$\beta + \sum_{i \in [N_E]} \lambda_i \tau_i + \sum_{j \in [N_I]} \mu_j \gamma_j \geq \frac{32 L_{x,A}(M_{x,A}+1) M_{x,\lambda,\mu}}{\sigma_{x,c}^2} \tag{3.8}$$

在 y 处成立，则 y 是非线性规划的 *KKT* 点。

注3.9。 定理*3.8*为 β 提供了一个取决于乘子 λ 和 μ的理论下界。在现有的各种欧几里得方法中，乘子 $\{\lambda_i\}_{i \in [N_E]}$ 被假设为或被证明是一致有界的。在这些情况下，我们可以立即为 β获得一个理论下界。

另一方面，当无法保证乘子 $\{\lambda_i\}_{i \in [N_E]}$ 一致有界时，一些现有研究可以估计乘子 *[23]* 的上界。因此，我们可以自适应地调整 β *to*

使其满足定理*3.8*中的理论下界，从而确保所采用的优化方法能够找到非线性规划的 *KKT* 点。

此外，对于 $N_E = 0$ 的特殊情形（即非线性规划中没有额外的等式约束），我们可以证明，*CDP* 中 β 的理论下界与乘子无关。详细讨论见第*3.3*节。

如 [35], 所示，我们知道，约束消解算子 $\mathcal{A}$ 能够将任意不可行点 $y \in \overline{\Omega}_x$ 的可行性违反量按二次速度降低，因此 $\mathcal{A}^\infty(y)$ 是可行点。在以下定理中，我们将证明 y 与 $\mathcal{A}^\infty(y)$ 在函数值和导数方面的关系，这对于刻画 $\overline{\Omega}_x$ 中这些不可行点处 CDP 的性质非常重要。

定理3.10。 对于任意给定的 $x \in \mathcal{K}_A$，假设 $y \in \overline{\Omega}_x$ 是 *CDP* 的一个 *KKT* 点，其乘子为 λ 和 μ，且

$$\beta + \sum_{i\in[N_E]} \lambda_i \tau_i + \sum_{j\in[N_I]} \mu_j \gamma_j \geq \frac{8M_{x,\lambda,\mu}(M_{x,A}+1)L_{x,b}}{\sigma_{x,c}^3}. \tag{3.9}$$

则有

$$\begin{aligned} &\mathcal{L}_{CDP}(y,\lambda,\mu) \geq \mathcal{L}_{CDP}(\mathcal{A}(y),\lambda,\mu), \\ &\mathcal{L}_{CDP}(y,\lambda,\mu) \geq \mathcal{L}_{CDP}(\mathcal{A}^\infty(y),\lambda,\mu). \end{aligned} \tag{3.10}$$

3.3 不等式约束情形

在本小节中，我们考虑非线性规划的一个特殊情形，即不存在附加约束$u(x) = 0$。更准确地说，我们在以下假设下分析非线性规划与 CDP 之间的等价性。

假设3.11。$N_E = 0$ 在非线性规划中。

由于不存在附加等式约束$u(x) = 0$，为简便起见，本小节将 CDP 的拉格朗日函数记为 $\mathcal{L}_{CDP}(x,\mu)$。此外，对于给定的$x \in \mathcal{M}$，我们对 β 和 $\{\gamma_j\}$ 的阈值作如下条件要求，且该条件独立于相应的乘子 μ。

条件3.12。给定任意 $x \in \mathcal{M}$，*CDP* 中的参数 β 和 $\{\gamma_j\}_{j\in[N_I]}$ 满足以下不等式，

$$\beta \geq \frac{64M_{x,f}(M_{x,A}+1)(L_{x,b}+\sigma_{x,c}L_{x,A})}{\sigma_{x,c}^3}, \quad \inf_{j\in[N_I]} \gamma_j \geq \frac{32L_{x,A}M_{x,v}(M_{x,A}+1)}{\sigma_{x,c}^2}. \tag{3.11}$$

注记3.13。当 $\mathcal{M}$ 为紧集时，存在有限集合 $\mathcal{I} \subset \mathcal{M}$，使得 $\mathcal{M} \subseteq \bigcup_{x\in\mathcal{I}} \Theta_x$。因此，我们可以为 $\sigma_{x,c}$ 和 ε_x 选取一致的正下界，同时为前述所有其他常数选取一致的上界。具体而言，当 $\mathcal{M}$ 为紧集时，在假设*3.11*下，我们可以为 *CDP* 的罚参数 β 和 $\{\gamma_j\}$ 选取一致的上界。

我们首先研究约束资格条件的等价性。以下引理说明，对于任意满足条件3.12的$x \in \mathcal{K}$，其中 β 和 $\{\gamma_j\}$ 成立，我们可以得到关于 $\mathcal{T}_\mathcal{K}(x)$ 与 $\mathcal{T}_{\mathcal{K}_A}(x)$之间关系的更精确结果。本小节的主要结果总结于图3中。引理 3.14、定理3.15、定理3.16和推论3.17的证明见附录A.4节。

引理3.14. 假设假设*3.11*成立，且*CDP*中的罚参数 β 和 $\{\gamma_j\}$ 满足条件*3.12*，则对任意$x \in \mathcal{K}$，均有

$$\mathcal{T}_\mathcal{K}(x) = J_A(x)^\top \mathcal{T}_{\mathcal{K}_A}(x), \tag{3.12}$$

$$\mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \mathrm{range}(J_A(x)) = J_A(x)\mathcal{T}_\mathcal{K}(x)^\circ. \tag{3.13}$$

在下面的定理中，我们使用引理 3.14 得到关于非线性规划与 CDP 之间一阶驻点等价性的更精确结果，即 β 和 $\{\gamma_j\}$ 的下界与乘子 μ 无关。

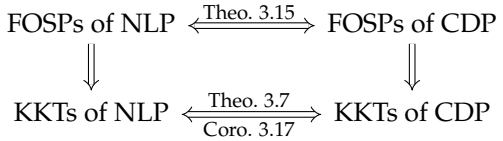

图3：在假设3.11下，非线性规划与CDP之间一阶驻点（一阶驻点）和KKT点（KKT点）的关系。此处“Theo.”是“定理”的缩写，而“Coro.”是“推论”的缩写。

定理3.15。 对于任意给定的 $x \in \mathcal{K}$，假设假设*3.11*成立，且 *CDP* 中的罚参数 β 和 $\{\gamma_j\}$满足条件*3.12*。当且仅当 x 是 *CDP* 的一阶驻点时，x 才是非线性规划的一阶驻点。

根据定理3.15，我们证明定理3.8中 KKT 点的等价性可以进一步提升为一阶驻点的等价性。

定理3.16。 对于任意给定的 $x \in \mathcal{K}$，假设假设*3.11*成立，且 *CDP* 中的罚参数 β 和 $\{\gamma_j\}$ 满足条件*3.12*。如果 $y \in \overline{\Omega}_x$ 是 *CDP* 的一阶驻点，那么 y 是非线性规划的一阶驻点。

推论3.17表明，对于任意 $x \in \mathcal{K}$，非线性规划和 CDP 在 $\overline{\Omega}_x$中具有相同的 KKT 点，其中 β 和 $\{\gamma_j\}$ 的下界与乘子 μ无关。

推论3.17. 对于任意给定的 $x \in \mathcal{K}$，假设假设*3.11*成立，*CDP* 中的罚参数 β 和 $\{\gamma_j\}$ 满足条件*3.12*。那么，$y \in \overline{\Omega}_x$ 是 *CDP* 的 *KKT* 点，当且仅当 y 是非线性规划的 *KKT* 点。

下一个推论可直接由定理3.10得出，因此为简洁起见，我们省略其证明。

推论3.18. 对于任意给定的 $x \in \mathcal{K}_A$，假设假设*3.11*成立，*CDP* 中的罚参数 β 和 $\{\gamma_j\}$满足条件*3.12*。若 $y \in \overline{\Omega}_x$ 是 *CDP* 的 *KKT* 点，且其乘子为 μ，则有

$$\begin{aligned} &\mathcal{L}_{CDP}(y,\mu) \geq \mathcal{L}_{CDP}(\mathcal{A}(y),\mu), \\ &\mathcal{L}_{CDP}(y,\mu) \geq \mathcal{L}_{CDP}(\mathcal{A}^{\infty}(y),\mu). \end{aligned} \tag{3.14}$$

4 数值实验

本节介绍通过约束消解方法 CDP 将各种现有高效求解器应用于非线性规划的初步数值结果。

4.1 基本设置

所有数值实验均在一台配备双路 Intel(R) Xeon(R) Gold 6242R CPU @ 3.10GHz×20的服务器上进行，运行 Ubuntu 20.04 下的 Python 3.8。此外，我们选择了多种基于不同方法开发的现有求解器，包括内点求解器 TRCON[11, 13] 和 IPOPT [31]、基于增广拉格朗日法的求解器 ALGENCAN [3]，以及求解器 PSQP（http://www.cs.cas.cz/~luksan/ 和基于序列规划方法开发的 SLSQP [21]。我们为所有求解器选择相同的初始点，并在运行时间超过 1200 秒时终止这些求解器。下面介绍各个求解器的实现细节。

- ALGENCAN（版本 3.1.1）：ALGENCAN 是一个以 GENCAN 为子程序的 FORTRAN 软件包。我们通过 pyOpt 软件包提供的 Python 接口调用 ALGENCAN 求解器 [26]。我们设置“epsopt $= 10^{-6}$”，其余参数均保留默认值。

• IPOPT（版本 3.14.5）：Ipopt 求解器以 C++ 编写，我们通过 CyIpopt 软件包提供的 Python 接口运行该求解器。我们将“海森矩阵近似 = 有限内存”设为采用 (L)BFGS 方法计算近似海森矩阵。此外，我们设置“tol $= 10^{-6}$”，其余参数均保留默认值。

• PSQP：PSQP 求解器以 FORTRAN 编写，我们通过 pyOpt 软件包提供的 Python 接口运行它。我们设置“TOLG $= 10^{-6}$”，其余参数均保持默认值。

• SLSQP：SLSQP 是 SciPy 软件包提供的 FORTRAN 求解器，其 Python 接口集成在 SciPy-optimize 软件包中。我们设置“tol $= 10^{-6}$”和“ftol $= 10^{-5}$”，其余参数均保持默认值。

• TRCON：求解器 TRCON 完全使用 Python 编写，基于信赖域方法。我们设置“hess = None”，以采用 BFGS 策略计算目标函数的近似海森矩阵，固定“gtol $= 10^{-6}$”，其余参数均保持默认值。

值得一提的是，在我们的设置下，所有受测求解器仅使用目标函数和约束的一阶导数。此外，我们使用自动微分软件包 autograd (https://github.com/HIPS/autograd) 自动计算导数。在数值结果中，我们报告这些受测求解器得到的函数值，以及结果 $\tilde{x}$ 的可行性和驻点性。这里，可行性通过 $\|u(\tilde{x})\| + \|c(\tilde{x})\| + \|\max\{v(\tilde{x}), 0\}\|$ 衡量，驻点性则通过 $\|\nabla_x \mathcal{L}_{NLP}(\tilde{x}, \tilde{\rho}, \tilde{\lambda}, \tilde{\mu})\|$ 衡量，其中 $\tilde{\rho}, \tilde{\lambda}, \tilde{\mu} = \arg\min \|\mathcal{L}_{NLP}(\tilde{x}, \rho, \lambda, \mu)\|^2, \rho \in \mathbb{R}^p, \lambda \in \mathbb{R}^{N_E}, \mu \in \mathbb{R}^{N_I}_+$

4.2 黎曼质心

在本小节中，我们将所有测试过的求解器用于求解 CDP，以测试约束消解方法在例 1.2 上的数值性能，并将结果与基线方法（即直接将测试过的求解器用于求解非线性规划）进行比较。在例 1.2 中，我们选择子流形 $\mathcal{M}$ 为辛 Stiefel 流形 $\mathcal{S}_{m,q} := \{X \in \mathbb{R}^{m\times q} : X^\top Q_m X = Q_q\}$。此外，我们选取 $N = 1000$，并在辛流形上随机生成 s^* 和 $\{s_i\}_{1\le i\le N}$。对于例 1.2 中的黎曼质心问题，相应的 CDP 可表示为

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \frac{1}{N}\sum_{i=1}^{N} \|\mathcal{A}(x) - s_i\|^2 + \frac{\beta}{2}\|c(x)\|^2 \\ \text{s.t.} \quad & \|\mathcal{A}(x) - s^*\|^2 \le r, \end{aligned} \tag{4.1}$$

其中我们固定 $\beta = 1$。此外，对于所有比较的求解器，我们在每个测试实例中均将初始点设为 s^*。

数值结果见表1。从表中可以得出，将现有求解器用于求解 CDP 比直接用于求解非线性规划高效得多。对于某些大规模实例，部分求解器的效率提升可达 10–40 倍。此外，我们注意到，SLSQP求解器直接用于求解非线性规划时，会因病态最小二乘子问题而出错。进一步地，我们的结果表明，ALGECAN 直接用于求解非线性规划时，在几乎所有测试实例上都不稳定。

4.3 图二分的最小平衡割

在本小节中，我们对例1.3进行数值实验，其中选择矩阵L 作为给定图 $\mathcal{G}$的拉普拉斯矩阵，即$L = D - A$，其中$D \in \mathbb{R}^{m\times m}$ 是顶点度的对角矩阵，而$A \in \mathbb{R}^{m\times m}$ 是邻接矩阵。在数值实验中，我们通过内置函数$\mathcal{G} = \text{Erdos Renyi}(m, \rho)$，使用 Erd¨os-R˜enyi 模型随机生成图 $\mathcal{G}$，其中任意两个节点 $\mathcal{G}$ 独立地以概率 ρ 连接。

表1：非线性规划与 CDP 在黎曼质心问题上的比较。

测试问题		求解器	函数值		次驻点性		可行性		CPU时间 (s)	
			CDP	NLP	CDP	NLP	CDP	NLP	CDP	NLP
$(q,\ r) = (10,\ 0.01)$	$m = 50$	ALGENCAN	5.44e+00	5.44e+00	4.07e-08	3.75e-07	4.77e-11	1.20e-09	0.37	3.39
		IPOPT	5.44e+00	5.44e+00	9.95e-07	1.68e-07	1.00e-08	1.00e-06	0.61	9.77
		PSQP	5.44e+00	5.44e+00	7.01e-07	3.46e-07	7.42e-11	3.55e-12	0.42	1.49
		SLSQP	5.44e+00	-	3.80e-07	-	5.87e-06	-	3.68	-
		TRCON	5.44e+00	5.46e+00	5.99e-07	7.06e-04	2.16e-12	1.11e-07	0.59	30.39
	米 $= 100$	ALGENCAN	5.60e+00	-	3.95e-07	-	6.88e-11	-	1.03	-
		IPOPT	5.60e+00	5.60e+00	4.88e-07	2.14e-07	1.08e-08	1.00e-06	1.02	30.11
		PSQP	5.60e+00	5.60e+00	8.46e-07	2.11e-07	2.20e-09	4.10e-12	0.56	4.43
		SLSQP	5.60e+00	-	4.51e-07	-	8.22e-06	-	30.93	-
		TRCON	5.60e+00	5.74e+00	9.32e-07	5.81e-02	1.33e-12	6.45e-06	1.11	45.36
	$m = 150$	ALGENCAN	5.84e+00	-	5.71e-07	-	3.77e-09	-	1.13	-
		IPOPT	5.84e+00	5.84e+00	5.48e-07	3.51e-07	9.98e-09	1.00e-06	1.57	75.25
		PSQP	5.84e+00	5.84e+00	5.32e-07	7.18e-07	5.30e-11	4.18e-11	1.58	10.07
		SLSQP	5.84e+00	-	6.07e-07	-	6.21e-06	-	112.83	-
		TRCON	5.84e+00	6.01e+00	9.83e-07	6.73e-02	1.62e-12	1.85e-05	2.53	63.93
	$m = 200$	ALGENCAN	5.88e+00	-	4.69e-07	-	4.99e-11	-	1.47	-
		IPOPT	5.88e+00	5.88e+00	2.64e-07	3.87e-08	1.14e-08	1.00e-06	1.91	57.91
		PSQP	5.88e+00	5.88e+00	6.70e-07	4.08e-07	1.44e-10	6.97e-13	2.61	17.54
		SLSQP	5.88e+00	-	6.18e-07	-	7.40e-06	-	256.97	-
		TRCON	5.88e+00	6.06e+00	9.54e-07	7.12e-02	2.06e-11	2.32e-05	4.89	104.14
$(m,r) = (100,\ 0.01)$	$q = 6$	ALGENCAN	3.42e+00	-	1.72e-07	-	2.65e-09	-	0.60	-
		IPOPT	3.42e+00	3.42e+00	2.42e-07	1.06e-07	1.00e-08	1.00e-06	0.65	6.33
		PSQP	3.42e+00	3.42e+00	2.05e-07	6.85e-07	6.24e-10	9.60e-12	0.21	0.66
		SLSQP	3.42e+00	-	4.05e-04	-	7.90e-06	-	3.98	-
		TRCON	3.42e+00	3.42e+00	7.54e-07	7.46e-04	1.42e-12	1.11e-15	0.76	2.74
	$q = 10$	ALGENCAN	5.60e+00	-	3.95e-07	-	6.88e-11	-	1.03	-
		IPOPT	5.60e+00	5.60e+00	4.88e-07	2.14e-07	1.08e-08	1.00e-06	1.02	30.11
		PSQP	5.60e+00	5.60e+00	8.46e-07	2.11e-07	2.20e-09	4.10e-12	0.56	4.43
		SLSQP	5.60e+00	-	4.51e-07	-	8.22e-06	-	30.93	-
		TRCON	5.60e+00	5.74e+00	9.32e-07	5.81e-02	1.33e-12	6.45e-06	1.11	45.36
	$q = 14$	ALGENCAN	8.08e+00	-	4.09e-07	-	1.90e-10	-	1.82	-
		IPOPT	8.08e+00	8.08e+00	3.42e-07	4.52e-08	1.13e-08	1.00e-06	1.92	56.56
		PSQP	8.08e+00	8.08e+00	7.33e-07	3.03e-07	3.18e-11	1.11e-11	1.80	18.54
		SLSQP	8.08e+00	-	4.12e-07	-	6.21e-06	-	110.97	-
		TRCON	8.08e+00	8.27e+00	4.87e-07	7.78e-02	1.98e-12	4.92e-06	2.21	114.87
	$q = 18$	ALGENCAN	9.98e+00	-	3.76e-07	-	4.77e-10	-	2.04	-
		IPOPT	9.98e+00	9.98e+00	6.74e-07	3.63e-08	1.03e-08	1.00e-06	2.33	86.40
		PSQP	9.98e+00	9.98e+00	8.64e-07	2.73e-07	1.59e-10	8.89e-13	2.34	52.86
		SLSQP	9.98e+00	-	1.84e-07	-	6.28e-06	-	219.27	-
		TRCON	9.98e+00	1.02e+01	9.72e-07	7.83e-02	1.27e-11	4.48e-05	4.13	293.02
$(m,q) = (100,\ 10)$	$r = 0.001$	ALGENCAN	5.78e+00	-	4.75e-08	-	1.13e-10	-	1.78	-
		IPOPT	5.78e+00	5.78e+00	9.99e-07	7.01e-10	1.04e-08	1.00e-05	1.68	38.36
		PSQP	5.78e+00	5.78e+00	9.90e-07	6.45e-07	2.81e-10	2.70e-12	1.62	8.04
		SLSQP	5.78e+00	-	5.52e-07	-	9.97e-06	-	56.10	-
		TRCON	5.78e+00	5.83e+00	6.83e-07	1.11e-01	3.97e-12	2.03e-05	6.65	45.06
	$r = 0.01$	ALGENCAN	5.60e+00	-	3.95e-07	-	6.88e-11	-	1.03	-
		IPOPT	5.60e+00	5.60e+00	4.88e-07	2.14e-07	1.08e-08	1.00e-06	1.02	30.11
		PSQP	5.60e+00	5.60e+00	8.46e-07	2.11e-07	2.20e-09	4.10e-12	0.56	4.43
		SLSQP	5.60e+00	-	4.51e-07	-	8.22e-06	-	30.93	-
		TRCON	5.60e+00	5.74e+00	9.32e-07	5.81e-02	1.33e-12	6.45e-06	1.11	45.36
	$r = 0.1$	ALGENCAN	5.35e+00	-	1.10e-06	-	9.34e-10	-	0.52	-
		IPOPT	5.35e+00	5.35e+00	8.12e-07	7.60e-08	1.00e-08	1.00e-07	0.57	25.64
		PSQP	5.35e+00	5.35e+00	3.69e-07	3.18e-07	2.00e-11	2.06e-10	0.18	2.19
		SLSQP	5.35e+00	-	4.92e-05	-	8.65e-06	-	7.40	-
		TRCON	5.35e+00	5.51e+00	8.62e-07	6.38e-03	5.49e-12	1.23e-05	0.46	43.00
	$r = 1$	ALGENCAN	4.85e+00	-	6.35e-07	-	3.45e-09	-	0.82	-
		IPOPT	4.85e+00	4.85e+00	8.07e-07	1.36e-07	9.96e-09	1.06e-08	0.94	49.84
		PSQP	4.85e+00	4.85e+00	9.32e-07	6.70e-07	5.93e-11	3.51e-10	0.20	2.67
		SLSQP	4.85e+00	-	4.57e-04	-	6.61e-06	-	6.01	-
		TRCON	4.85e+00	4.90e+00	6.62e-07	2.52e-02	8.24e-12	1.06e-11	0.46	11.87

此外，我们在斜流形上随机生成一个初始点。在我们提出的方法中，我们将 (1.3) 转化为以下带有 q 个等式约束的优化问题，其中$\mathbb{R}^{m\times q}$。

$$\begin{aligned} \min_{X\in\mathbb{R}^{m\times q}} \quad & -\frac{1}{4}\mathrm{tr}\left(\mathcal{A}(X)^\top L\mathcal{A}(X)\right)+\frac{\beta}{4}\left\|\mathrm{Diag}(XX^\top)-I_m\right\|^2 \\ \text{s.t.} \quad & \mathcal{A}(X)^\top e=0, \end{aligned} \tag{4.2}$$

其中 $\mathcal{A}(X) = 2X\left(\mathrm{Diag}(XX^\top) + I_q\right)^{-1} \in \mathbb{R}^{m\times q}$，且所有测试实例的 β 均固定为 0.05。

数值结果如表2所示。我们观察到，所有求解器都成功为非线性规划和 CDP 求得了一个解。从表2可以得出，将现有求解器用于 CDP，效率显著高于直接用于非线性规划。对于部分参与比较的求解器，效率提升甚至超过10倍。此外，在所有求解器中，PSQP 的整体表现似乎最佳。因此，我们可以得出结论：CDP 使现有求解器能够直接应用，同时借助非线性规划的部分流形结构，保持较高的效率。此外，值得一提的是，构造 CDP 不依赖于黎曼流形 $\mathcal{M}$ 的几何性质。通过 CDP 求解非线性规划，我们利用了约束的黎曼结构，同时避免了处理黎曼流形 $\mathcal{M}$ 的几何工具。

5 结论

本文聚焦于一般的约束优化问题 NLP，其中部分等式约束在 $\mathbb{R}^n$ 中定义了一个嵌入子流形。我们提出了一种适用于 NLP 的约束消解方法，通过直接应用各种欧几里得优化求解器，同时利用 (NLP) 中的黎曼结构，以提升便利性和效率。在我们提出的约束消解方法中，求解 NLP 被转化为求解约束消解问题 CDP；CDP 是 $\mathbb{R}^n$ 中的一个约束优化问题，已从 NLP 中消除了流形约束 $c(x) = 0$。在温和条件下，我们证明了 NLP 与 CDP 在约束资格条件、驻点和 KKT 点方面的等价性。此外，对于 NLP 除 $c(x) = 0$ 外没有其他等式约束这一特殊情形，我们证明了 NLP 与 CDP 等价性的更强结论。因此，我们可以直接应用现有的欧几里得优化方法求解 CDP，并充分利用过去数年在 $\mathbb{R}^n$ 中求解约束优化问题所积累的丰富经验。

我们对黎曼质心问题和平衡图二分问题进行了初步数值实验，其中应用了包括 ALGENCAN、IPOPT、PSQP、SLSQP 和 TRCON 在内的各种现有数值求解器来求解非线性规划和 CDP。数值结果表明，将这些求解器应用于 CDP，在效率和稳定性方面能够取得优于直接求解非线性规划方法的性能。因此，我们可以得出结论：我们的约束消解方法能够方便地直接应用各种欧几里得优化求解器，并且通过利用非线性规划中 $\mathcal{M}$ 的结构，可以进一步提升其数值性能。

参考文献

[1] P-A Absil、Robert Mahony 和 Rodolphe Sepulchre。《矩阵流形上的优化算法》。普林斯顿大学出版社，2009年。[2] Bijan Afsari。黎曼质心：存在性、唯一性和凸性。《美国数学会会刊》，139(2):655–673，2011年。[3] Roberto Andreani、Ernesto G Birgin、Jos´e Mario Mart´ınez 和 Mar´ıa Laura Schuverdt。关于具有一般下层约束的增广拉格朗日函数方法。《SIAM Journal on Optimization》，18(4):1286–1309，2008年。

表2：非线性规划与 CDP 在最小平衡割问题上的比较。

测试问题		求解器	函数值		次驻点性		可行性		CPU时间（秒）	
			CDP	NLP	CDP	NLP	CDP	NLP	CDP	NLP
$(q,\rho)=(2,0.1)$	米 $=50$	ALGENCAN	-1.62e+01	-1.62e+01	4.05e-07	5.80e-07	3.12e-09	8.06e-09	0.85	2.21
		IPOPT	-1.62e+01	-1.62e+01	5.66e-07	5.42e-07	1.16e-10	3.52e-12	0.63	0.62
		PSQP	-1.62e+01	-1.62e+01	6.11e-07	5.22e-07	4.49e-11	5.39e-11	0.09	0.61
		SLSQP	-1.62e+01	-1.62e+01	4.74e-07	2.38e-07	4.08e-11	8.68e-13	0.16	0.72
		TRCON	-1.62e+01	-1.62e+01	9.41e-07	8.80e-06	2.24e-12	1.38e-10	0.70	6.02
	$m=100$	ALGENCAN	-3.43e+01	-3.43e+01	5.27e-07	6.25e-07	2.81e-09	3.51e-09	1.12	6.37
		IPOPT	-3.43e+01	-3.43e+01	9.88e-07	5.15e-07	1.95e-11	2.35e-11	0.79	7.73
		PSQP	-3.43e+01	-3.43e+01	7.71e-07	9.95e-07	4.44e-11	5.50e-11	0.17	2.16
		SLSQP	-3.43e+01	-3.43e+01	6.57e-07	1.23e-07	8.26e-11	3.42e-13	0.70	2.04
		TRCON	-3.43e+01	-3.43e+01	9.26e-07	9.92e-07	4.93e-12	8.01e-13	0.80	17.63
	$m=200$	ALGENCAN	-7.44e+01	-7.44e+01	1.64e-07	2.92e-07	2.09e-09	9.09e-10	5.84	34.68
		IPOPT	-7.44e+01	-7.44e+01	6.20e-07	8.92e-07	4.74e-10	5.48e-11	1.83	85.03
		PSQP	-7.44e+01	-7.44e+01	9.31e-07	9.34e-07	2.31e-10	1.51e-10	0.38	24.35
		SLSQP	-7.44e+01	-7.44e+01	9.89e-07	1.21e-07	3.30e-11	5.18e-13	7.45	23.67
		TRCON	-7.44e+01	-7.42e+01	9.96e-07	1.65e-02	1.17e-11	3.18e-04	2.21	49.28
	米 $=400$	ALGENCAN	-1.63e+02	-	7.15e-07	-	1.37e-10	-	60.40	-
		IPOPT	-1.63e+02	-1.63e+02	9.17e-07	9.36e-07	7.83e-11	2.82e-10	8.41	739.39
		PSQP	-1.63e+02	-1.63e+02	8.14e-07	7.08e-07	2.12e-10	1.22e-10	1.19	251.36
		SLSQP	-1.63e+02	-1.63e+02	5.22e-07	7.00e-08	4.45e-11	2.68e-13	75.74	181.46
		TRCON	-1.63e+02	-1.62e+02	7.81e-05	6.09e-03	9.61e-08	1.86e-05	6.92	136.06
$(米,\ \rho)=(100,0.1)$	$q=2$	ALGENCAN	-3.43e+01	-3.43e+01	5.27e-07	6.25e-07	2.81e-09	3.51e-09	1.12	6.37
		IPOPT	-3.43e+01	-3.43e+01	9.88e-07	5.15e-07	1.95e-11	2.35e-11	0.79	7.73
		PSQP	-3.43e+01	-3.43e+01	7.71e-07	9.95e-07	4.44e-11	5.50e-11	0.17	2.16
		SLSQP	-3.43e+01	-3.43e+01	6.57e-07	1.23e-07	8.26e-11	3.42e-13	0.70	2.04
		TRCON	-3.43e+01	-3.43e+01	9.26e-07	9.92e-07	4.93e-12	8.01e-13	0.80	17.63
	$q=4$	ALGENCAN	-3.60e+01	-3.60e+01	3.52e-07	2.61e-07	8.78e-09	3.22e-09	1.24	10.57
		IPOPT	-3.60e+01	-3.60e+01	7.42e-07	3.82e-07	1.42e-11	7.74e-12	1.27	40.78
		PSQP	-3.60e+01	-3.60e+01	1.07e-06	7.60e-07	9.36e-11	1.04e-10	0.34	5.54
		SLSQP	-3.60e+01	-3.60e+01	8.38e-07	1.20e-07	1.83e-12	9.12e-13	4.06	7.64
		TRCON	-3.60e+01	-3.60e+01	9.09e-07	1.78e-04	1.75e-10	2.83e-08	1.57	24.09
	$q=6$	ALGENCAN	-3.81e+01	-3.81e+01	2.50e-07	1.79e-07	6.34e-10	2.46e-09	1.90	8.76
		IPOPT	-3.81e+01	-3.81e+01	5.76e-07	7.24e-07	2.45e-10	1.83e-10	1.83	83.78
		PSQP	-3.81e+01	-3.81e+01	8.55e-07	9.49e-07	9.75e-11	1.47e-10	0.91	10.96
		SLSQP	-3.81e+01	-3.81e+01	6.33e-07	7.37e-08	2.42e-11	5.82e-13	14.35	22.08
		TRCON	-3.81e+01	-3.81e+01	3.85e-04	4.02e-05	2.38e-08	1.47e-09	6.54	30.70
	$q=8$	ALGENCAN	-4.10e+01	-4.10e+01	7.32e-08	8.80e-08	4.17e-09	2.07e-09	1.55	10.11
		IPOPT	-4.10e+01	-4.10e+01	8.89e-07	6.54e-07	2.87e-11	7.69e-12	1.93	111.52
		PSQP	-4.10e+01	-4.10e+01	9.74e-07	8.92e-07	6.41e-11	8.06e-13	1.33	13.38
		SLSQP	-4.10e+01	-4.10e+01	8.13e-07	7.13e-08	2.61e-11	3.78e-13	25.97	41.20
		TRCON	-4.10e+01	-4.10e+01	2.63e-05	1.49e-05	5.83e-11	1.75e-10	7.59	37.19
$(米,\ 夸脱)=(100,2)$	$\rho=0.05$	ALGENCAN	-3.28e+01	-3.30e+01	7.08e-07	8.31e-08	1.91e-09	2.44e-09	0.99	7.99
		IPOPT	-3.28e+01	-3.27e+01	8.61e-07	6.02e-07	1.83e-09	1.13e-11	1.77	11.97
		PSQP	-3.28e+01	-3.30e+01	8.63e-07	8.36e-07	9.60e-11	1.08e-10	0.26	2.82
		SLSQP	-3.28e+01	-3.30e+01	9.32e-07	1.34e-07	2.84e-12	4.58e-13	1.08	2.33
		TRCON	-3.28e+01	-3.30e+01	9.88e-07	3.30e-05	4.22e-12	1.80e-09	1.44	19.61
	$\rho=0.1$	ALGENCAN	-3.43e+01	-3.43e+01	5.27e-07	6.25e-07	2.81e-09	3.51e-09	1.12	6.37
		IPOPT	-3.43e+01	-3.43e+01	9.88e-07	5.15e-07	1.95e-11	2.35e-11	0.79	7.73
		PSQP	-3.43e+01	-3.43e+01	7.71e-07	9.95e-07	4.44e-11	5.50e-11	0.17	2.16
		SLSQP	-3.43e+01	-3.43e+01	6.57e-07	1.23e-07	8.26e-11	3.42e-13	0.70	2.04
		TRCON	-3.43e+01	-3.43e+01	9.26e-07	9.92e-07	4.93e-12	8.01e-13	0.80	17.63
	$\rho=0.2$	ALGENCAN	-4.21e+01	-4.21e+01	9.60e-08	3.21e-07	1.22e-09	5.36e-09	1.68	7.66
		IPOPT	-4.21e+01	-4.21e+01	9.95e-07	9.88e-07	1.67e-09	1.55e-10	0.73	8.55
		PSQP	-4.21e+01	-4.21e+01	6.60e-07	7.11e-07	6.00e-11	2.82e-10	0.16	3.22
		SLSQP	-4.21e+01	-4.21e+01	9.57e-07	1.21e-07	8.63e-12	6.58e-13	0.61	2.75
		TRCON	-4.21e+01	-4.20e+01	9.98e-07	2.18e-02	3.60e-12	2.89e-04	1.27	19.15
	$\rho=0.3$	ALGENCAN	-4.37e+01	-4.37e+01	2.52e-07	1.77e-08	1.46e-09	1.98e-10	7.51	10.21
		IPOPT	-4.37e+01	-4.37e+01	3.61e-07	4.90e-07	1.46e-11	4.18e-11	1.83	10.92
		PSQP	-4.37e+01	-4.37e+01	9.88e-07	6.41e-07	6.01e-11	3.81e-11	0.21	4.78
		SLSQP	-4.37e+01	-4.37e+01	8.90e-07	1.30e-07	7.63e-11	6.61e-13	0.97	3.99
		TRCON	-4.37e+01	-4.37e+01	6.80e-03	1.95e-03	4.21e-04	1.68e-06	3.87	19.91

[4] Roberto Andreani、Ernesto G Birgin、Jos´e Mario Mart´ınez 和 Maria Laura Schuverdt。在常数正线性相关约束资格条件下的增广拉格朗日方法。Mathematical Programming，111(1):5–32，2008年。[5] Yu Bai 和 Song Mei。从黎曼优化的视角分析序列二次规划。arXiv 预印本 arXiv:1805.08756，2018年。[6] Marcel Berger。《黎曼几何全景》。2003年。[7] Ronny Bergmann 和 Roland Herzog。光滑流形上 KKT条件与约束资格条件的内在表述。SIAM Journal on Optimization，29(4):2423–2444，2019年。[8] Wei Bian、Xiaojun Chen 和 Yinyu Ye。非 Lipschitz、非凸最小化问题的内点算法复杂度分析。Mathematical Programming，149(1-2):301–327，2015年。[9] Nicolas Boumal。《光滑流形上的优化导论》。在线可获取，2020年5月。[10] Rainer E Burkard、Stefan E Karisch 和 Franz Rendl。QAPLIB——二次指派问题库。Journal of Global Optimization，10:391–403，1997年。[11] Richard H Byrd、Mary E Hribar 和 Jorge Nocedal。大规模非线性规划的内点算法。SIAM Journal on Optimization，9(4):877–900，1999年。[12] Frank H Clarke。《优化与非光滑分析》，第5卷。SIAM，1990年。[13] Andrew R Conn、Nicholas IM Gould 和 Philippe L Toint。《信赖域方法》。SIAM，2000年。[14] Frank E Curtis、Tim Mitchell 和 Michael L Overton。一种用于非光滑、非凸约束优化的 BFGS-SQP 方法及其相对最小化剖面评估。Optimization Methods and Software，32(1):148–181，2017年。[15] Roger Fletcher 和 Sven Leyffer。不使用罚函数的非线性规划。Mathematical programming，91(2):239–269，2002年。[16] Gene H Golub 和 Charles F Van Loan。《矩阵计算》，第3卷。JHU Press，2013年。[17] Magnus R Hestenes。乘子法与梯度法。Journal of Optimization Theory and Applications，4(5):303–320，1969年。[18] Xiaoyin Hu、Nachuan Xiao、Xin Liu 和 Kim-Chuan Toh。斯蒂费尔流形上非光滑优化的约束消解方法。arXiv 预印本 arXiv:2205.10500，2022年。[19] Xiaoyin Hu、Nachuan Xiao、Xin Liu 和 Kim-Chuan Toh。一种改进的双层优化无约束方法。arXiv 预印本 arXiv:2208.00732，2022年。[20] Hermann Karcher。黎曼质心与磨光函数平滑。Communications on Pure and Applied Mathematics，30(5):509–541，1977年。[21] Dieter Kraft 等。序列二次规划软件包。1988年。[22] Changshuo Liu 和 Nicolas Boumal。带约束黎曼流形优化的简单算法。Applied Mathematics & Optimization，第1–33页，2019年。[23] Olvi L Mangasarian。非凸可微非线性规划驻点拉格朗日乘子的可计算数值界。Operations Research Letters，4(2):47–48，1985年。[24] Maher Moakher。旋转群中的均值与平均。SIAM Journal on Matrix Analysis and Applications，24(1):1–16，2002年。[25] Jorge Nocedal 和 Stephen J. Wright。《数值优化》。1999年。

[26] Ruben E. Perez、Peter W. Jansen 和 Joaquim R. R. A. Martins。pyOpt：一种基于 Python、面向非线性约束优化的对象化框架。结构与多学科优化，45(1):101–118，2012年。doi: 10.1007/s00158-011-0666-3。[27] Michael JD Powell。最小化问题中处理非线性约束的一种方法。《优化》，第283–298页，1969年。[28] Anton Schiela 和 Julian Ortiz。流形上等式约束优化的序列二次规划方法。*arXiv* 预印本 *arXiv:2005.06844*，2020年。[29] Nguyen Thanh Son、P-A Absil、Bin Gao 和 Tatjana Stykel。通过迹最小化与黎曼优化求解辛特征值问题。*arXiv* 预印本 *arXiv:2101.02618*，2021年。[30] Tianyun Tang 和 Kim-Chuan Toh。在代数簇上求解图均衡划分半正定规划。*arXiv* 预印本 *arXiv:2112.04256*，2021年。[31] Andreas W¨achter 和 Lorenz T Biegler。大规模非线性规划内点滤波线搜索算法的实现。*MathematicalProgramming*，106(1):25–57，2006年。[32] Shuai Wang、Tsung-Hui Chang、Ying Cui 和 Jong-Shi Pang。基于正交非负矩阵分解模型与非凸惩罚优化的聚类。《*IEEE*信号处理汇刊》，69:5273–5288，2021年。[33] Ziteng Wang、Shu-Cherng Fang 和 Wenxun Xing。论约束资格条件：动机、设计与相互关系。《工业与管理优化杂志》，9(4):983，2013年。[34] Melanie Weber 和 Suvrit Sra。通过黎曼弗兰克-沃尔夫方法进行流形上的非凸随机优化。*arXiv* 预印本 *arXiv:1910.04194*，2019年。[35] Nachuan Xiao、Xin Liu 和 Kim-Chuan Toh。黎曼优化中的约束消解。《运筹学数学》，即将刊印，2023年。

A 理论结果的证明

A.1 预备引理

在本小节中，我们给出 [35, 第 3 节1]中的若干预备引理。

引理 A.1. 对于任意给定的 $x \in \mathcal{M}$，对任意 $y \in \Omega_x$，以下不等式成立，

$$\frac{1}{M_{x,c}} \|c(y)\| \leq \operatorname{dist}(y, \mathcal{M}) \leq \frac{2}{\sigma_{x,c}} \|c(y)\|. \tag{A.1}$$

引理 A.2. 对于任意给定的 $x \in \mathcal{M}$，有

$$\|\mathcal{A}(y) - y\| \leq \frac{2(M_{x,A}+1)}{\sigma_{x,c}} \|c(y)\|, \qquad \textit{for any } y \in \Omega_x. \tag{A.2}$$

引理 A.3. 对于任意给定的 $x \in \mathcal{M}$，有

$$\|c(\mathcal{A}(y))\| \leq \frac{4L_{x,b}}{\sigma_{x,c}^2} \|c(y)\|^2, \qquad \textit{for any } y \in \Omega_x. \tag{A.3}$$

引理 A.4。对于任意给定的 $x \in \mathcal{M}$，包含关系 $J_A(x)^\top d \in \mathcal{T}_x$ 对任意 $d \in \mathbb{R}^n$ 都成立。此外，当 $d \in \mathcal{T}_x$ 时，有 $J_A(x)^\top d = d$。

引理 A.5。对于任意 $x \in \mathcal{M}$，$J_A(x)d = 0$ 当且仅当 $d \in \mathcal{N}_x =$ 值域$(J_c(x))$。

引理 A.6. 对于任意 $x \in \mathcal{M}$，有 $J_A(x)^2 = J_A(x)$.

A.2 第3.1节的证明

引理 A.7。 对于任意 $x \in \mathcal{K}$，都有

$$\mathcal{T}_{\mathcal{K}}(x) \subseteq \mathcal{T}_{\mathcal{K}_A}(x), \tag{A.4}$$

$$\mathcal{T}^{lin}_{\mathcal{K}}(x) \subseteq \mathcal{T}^{lin}_{\mathcal{K}_A}(x). \tag{A.5}$$

该结果可直接由以下事实推出： $\mathcal{K} \subseteq \mathcal{K}_A$。

引理 A.8. 对于任意闭锥 $\mathcal{X}_1 \subset \mathcal{R}^n$，令 $\mathcal{X}_2 = J_A(x)^\top \mathcal{X}_1$，则

$$\mathcal{X}_1^\circ \cap \text{range}(J_A(x)) = J_A(x)\mathcal{X}_2^\circ. \tag{A.6}$$

证明. 对于任意 $d_3 \in J_A(x)\mathcal{X}_2^\circ$，存在 $d_2 \in \mathcal{X}_2^\circ$使得 $d_3 = J_A(x)d_2$。因此，根据极锥的定义，对于任意 $d_1 \in \mathcal{X}_1$，有

$$0 \geq \left\langle d_2, J_A(x)^\top d_1 \right\rangle = \langle J_A(x)d_2, d_1 \rangle = \langle d_3, d_1 \rangle . \tag{A.7}$$

因此，由于 $d_1 \in \mathcal{X}_1$ 的任意性，我们可以得出 $d_3 \in \mathcal{X}_1^\circ$，从而 $J_A(x)\mathcal{X}_2^\circ \subseteq \mathcal{X}_1^\circ$。结合 $J_A(x)\mathcal{X}_2^\circ \subset$ 是 (J_A (x)) 的值域这一事实，可得

$$J_A(x)\mathcal{X}_2^\circ \subseteq \mathcal{X}_1^\circ \cap \text{range}(J_A(x)). \tag{A.8}$$

另一方面，对于任意 $d_3 \in \mathcal{X}_1^\circ \cap$ 属于 $(J_A(x))$ 的值域， $\langle d_3, d_1 \rangle \leq 0$ 对任意 $d_1 \in \mathcal{X}_1$ 均成立。注意到 $d_3 \in$ 属于 $(J_A(x))$ 的值域，我们有 $J_A(x)J_A(x)^\dagger d_3 = d_3$。因此，对于任意 $d_1 \in \mathcal{X}_1$，有

$$0 \geq \langle d_3, d_1 \rangle = \left\langle J_A(x)J_A(x)^\dagger d_3, d_1 \right\rangle = \left\langle J_A(x)^\dagger d_3, J_A(x)^\top d_1 \right\rangle, \tag{A.9}$$

这说明 $J_A(x)^\dagger d_3 \in \mathcal{X}_2^\circ$。因此，我们得出结论：

$$d_3 = J_A(x)J_A(x)^\dagger d_3 \in J_A(x)\mathcal{X}_2^\circ. \tag{A.10}$$

因此，结合 (A.8)，我们得到 $\mathcal{X}_1^\circ \cap$ 值域$(J_A(x)) = J_A(x)\mathcal{X}_2^\circ$。证明完毕。

□

引理 A.9。 对于任意 $x \in \mathcal{K}$，有 $\mathcal{T}_{\mathcal{K}}(x) \subseteq \mathcal{T}_x$。

证明。 对于任意$d \in \mathcal{T}_{\mathcal{K}}(x)$，存在序列 $\{x_k\} \subset \mathcal{K} \subset \mathcal{M}$ 和 $\{t_k\} \to 0$，使得$d = \lim_{k\to+\infty} \frac{x_k - x}{t_k}$。根据 ~~$\mathcal{T}_x$的~~定义，我们可以得出$d \in \mathcal{T}_x$，从而完成证明。 □

命题 3.2 的证明

证明。 对于任意系数 $\hat{}\lambda \in \mathbb{R}^{N_E}$ 和 $\hat{}\mu \in \mathbb{R}^{N_I}$，使得

$$0 = \sum_{i\in[N_E]} \hat{\lambda}_i \nabla \tilde{u}_i(x) + \sum_{j\in\mathcal{F}_A(x)} \hat{\mu}_j \nabla \tilde{v}_j(x), \tag{A.11}$$

成立

$$\begin{aligned} 0 &= \sum_{i\in[N_E]} \hat{\lambda}_i \nabla \tilde{u}_i(x) + \sum_{j\in\mathcal{F}_A(x)} \hat{\mu}_j \nabla \tilde{v}_j(x) \\ &= J_A(x) \left(\sum_{i\in[N_E]} \hat{\lambda}_i \nabla u_i(x) + \sum_{j\in\mathcal{F}_A(x)} \hat{\mu}_j \nabla v_j(x) \right). \end{aligned} \tag{A.12}$$

注意，对于任意$x \in \mathcal{M}$，都有 $\tilde{}v(x) = v(x)$，因此 $\mathcal{F}_A(x) = \mathcal{F}(x)$ 对任意$x \in \mathcal{M}$都成立。由引理 A.5 可知,

$$\sum_{i\in[N_E]} \hat{\lambda}_i \nabla u_i(x) + \sum_{j\in\mathcal{F}(x)} \hat{\mu}_j \nabla v_j(x) \in \text{range}(J_c(x)). \tag{A.13}$$

也就是说，存在 $\hat{}\rho \in \mathbb{R}^p$ 使得

$$\sum_{l\in[p]} \hat{\rho}_l \nabla c_l(x) + \sum_{i\in[N_E]} \hat{\lambda}_i \nabla u_i(x) + \sum_{j\in\mathcal{F}(x)} \hat{\mu}_j \nabla v_j(x) = 0. \tag{A.14}$$

由于在x处相对于非线性规划满足线性无关约束资格条件，因此
$\{\nabla_{i} \in [E]\} \cup \{\nabla_{l} \in [\,]\} \cup u(x): iNc(x): lpv(x): j(x)$是一个线性无关
$\{\nabla_{j} \in \mathcal{F}\}$ $\rho = 0\ \lambda = 0\ \mu = 0$
集，因此我们得到 $\hat{}$、$\hat{}$ 和 $\hat{}$。因此，我们得出结论：在x处，相对于 CDP 的线性无关约束资格条件成立。

□

命题3.3的证明

证明。 根据相对于非线性规划的 Mangasarian-Fromovitz约束规范的定义，有 $\{\nabla u_i(x) : i \in [N_E]\} \cup \{\nabla c_l(x) : l \in [p]\}$ 是一个线性无关集，并且存在$d \in \mathbb{R}^n$ 使得

$$\begin{aligned} &\langle d, \nabla v_j(x)\rangle \le -1,\ \forall j \in \mathcal{F}(x);\\ &\langle d, \nabla u_i(x)\rangle = 0,\ \forall i \in [N_E];\\ &\langle d, \nabla c_i(x)\rangle = 0,\ \forall i \in [p]. \end{aligned} \tag{A.15}$$

Th因此，对于任意满足 $\sum_{i\in[N_E]} \hat{}\lambda_i J_A(x)\nabla u_i(x) = 0$的 $\hat{}\lambda \in \mathbb{R}^{N_E}$，引理 A.5 推出 tha t

$$\sum_{i\in[N_E]} \hat{\lambda}_i \nabla u_i(x) \in \mathrm{range}(J_c(x)).$$

由 $\{\nabla u_i(x) : i \in [N_E]\} \cup \{\nabla c_l(x) : l \in [p]\}$ 的线性无关性，可得 $\hat{}\lambda = 0$。因此，我们可以得出 $\{J_A(x)\nabla u_i(x) : i \in [N_E]\}$ 是一个线性无关集。此外，(A.15) 表明 $d \in \mathcal{T}_x$。结合引理 A.4，对于任意 $j \in [N_I]$，有

$$\langle d, J_A(x)\nabla v_j(x)\rangle = \left\langle J_A(x)^\top d, \nabla v_j(x)\right\rangle = \langle d, \nabla v_j(x)\rangle \le -1. \tag{A.16}$$

因此，我们得出，关于 CDP 的 Mangasarian-Fromovitz约束规范在 x 处成立。 □

引理 3.4 的证明

证明。 **证明 (3.4)：** 对于任意 $d_1 \in \mathcal{T}^{lin}_{\mathcal{K}_A}(x)$，根据 $\mathcal{T}^{lin}_{\mathcal{K}_A}(x)$ 的定义，有 t

$$\langle d_1, J_A(x)\nabla u_i(x)\rangle = 0,\ \langle d_1, J_A(x)\nabla v_j(x)\rangle \le 0,\ \forall\, i \in [N_E],\ j \in \mathcal{F}_A(x). \tag{A.17}$$

No注意到 $\mathcal{F}_A(x) = \mathcal{F}(x)$ 对任意 $x \in \mathcal{K}$ 均成立。于是，对于任意 $i \in [N_E]$、$j \in \mathcal{F}(x)$，我们得到 in

$$\left\langle J_A(x)^\top d_1, \nabla u_i(x)\right\rangle = 0, \text{ and } \left\langle J_A(x)^\top d_1, \nabla v_j(x)\right\rangle \le 0. \tag{A.18}$$

由于 $x \in \mathcal{K} \subset \mathcal{M}$，假设 1.4 表明，对于任意 $l \in [p]$，$\langle J_A(x)^\top d1, \nabla c_l(x)\rangle = 0$ 成立。因此，根据定义 2.5，有 $J_A(x)^\top d_1 \in \mathcal{T}^{lin}_{\mathcal{K}}(x)$。由于 $d_1 \in \mathcal{T}^{lin}_{\mathcal{K}_A}(x)$ 的任意性，我们得到 $J_A(x)^\top \mathcal{T}^{lin}_{\mathcal{K}_A}(x) \subseteq \mathcal{T}^{lin}_{\mathcal{K}}(x)$。

另一方面，引理 A.7 表明 $\mathcal{T}^{lin}_{\mathcal{K}}(x) \subseteq \mathcal{T}^{lin}_{\mathcal{K}_A}(x)$。注意到 $\mathcal{T}^{lin}_{\mathcal{K}}(x) \subset \mathcal{T}_x$，由引理 A.4 可知

$$\mathcal{T}^{lin}_{\mathcal{K}}(x) = J_A(x)^\top \mathcal{T}^{lin}_{\mathcal{K}}(x) \subseteq J_A(x)^\top \mathcal{T}^{lin}_{\mathcal{K}_A}(x). \tag{A.19}$$

因此，有以下结论：

$$J_A(x)^\top \mathcal{T}^{lin}_{\mathcal{K}_A}(x) = \mathcal{T}^{lin}_{\mathcal{K}}(x). \tag{A.20}$$

结合引理 A.8，可得：

$$\mathcal{T}^{lin}_{\mathcal{K}_A}(x)^\circ \cap \mathrm{range}(J_A(x)) = J_A(x)\mathcal{T}^{lin}_{\mathcal{K}}(x)^\circ. \tag{A.21}$$

证明 (3.5)： 引理 A.7 表明 $\mathcal{T}_{\mathcal{K}}(x) \subseteq \mathcal{T}_{\mathcal{K}_A}(x)$。此外，根据引理 A.9，有 $\mathcal{T}_{\mathcal{K}}(x) \subset \mathcal{T}_x$。因此，根据引理 A.4，我们有

$$\mathcal{T}_{\mathcal{K}}(x) = J_A(x)^\top \mathcal{T}_{\mathcal{K}}(x) \subseteq J_A(x)^\top \mathcal{T}_{\mathcal{K}_A}(x). \tag{A.22}$$

因此 $\left(J_A(x)^\top \mathcal{T}_{\mathcal{K}_A}(x)\right)^\circ \subset \mathcal{T}_{\mathcal{K}}(x)^\circ$。结合引理 A.8，可得

$$\mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \text{range}(J_A(x)) = J_A(x)\left(J_A(x)^\top \mathcal{T}_{\mathcal{K}_A}(x)\right)^\circ \subseteq J_A(x)\mathcal{T}_{\mathcal{K}}(x)^\circ. \tag{A.23}$$

证明完毕。 □

命题 3.5 的证明

证明。 由引理 3.4，可得

$$\begin{aligned} &\mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \text{range}(J_A(x)) \subseteq J_A(x)\mathcal{T}_{\mathcal{K}}(x)^\circ \\ &= J_A(x)\mathcal{T}_{\mathcal{K}}^{lin}(x)^\circ = \mathcal{T}_{\mathcal{K}_A}^{lin}(x)^\circ \cap \text{range}(J_A(x)) \\ &\subseteq \mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \text{range}(J_A(x)). \end{aligned} \tag{A.24}$$

这里，第一个包含关系由（3.5）得出，第二个等式由（3.4）推出，最后一个包含关系利用了以下事实：对于任意 $x \in \mathcal{K}_A$，$\mathcal{T}^{lin}{}_{\mathcal{K}_A}(x)^\circ \subseteq \mathcal{T}_{\mathcal{K}_A}(x)^\circ$ 成立。因此，我们得到

$$\mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \text{range}(J_A(x)) = \mathcal{T}_{\mathcal{K}_A}^{lin}(x)^\circ \cap \text{range}(J_A(x)), \tag{A.25}$$

从而证明完毕。 □

定理 3.6 的证明

证明。 根据定义 2.4 和定义 2.5，很容易验证，对任意 $x \in \mathcal{K}_A$，$\mathcal{T}^{lin}{}_{\mathcal{K}_A}(x)^\circ \subseteq \mathcal{T}_{\mathcal{K}_A}(x)^\circ$ 均成立。因此，对于 CDP 的任意 KKT 点 x，它也是 CDP 的一阶驻点。

另一方面，对于 CDP 的任意一阶驻点 $x \in \mathcal{K}$，根据定义 2.7，有

$$0 \in \partial h(x) + \mathcal{T}_{\mathcal{K}_A}(x)^\circ = J_A(x)\partial f(x) + \mathcal{T}_{\mathcal{K}_A}(x)^\circ.$$

注意，$J_A(x)\partial f(x) \subset$是$(J_A(x))$的值域。于是有

$$\begin{aligned} 0 \in\ &\left(J_A(x)\partial f(x) + \mathcal{T}_{\mathcal{K}_A}(x)^\circ\right) \cap \text{range}(J_A(x)) \\ &= J_A(x)\partial f(x) + \left(\mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \text{range}(J_A(x))\right) \\ &= J_A(x)\partial f(x) + \left(\mathcal{T}_{\mathcal{K}_A}^{lin}(x)^\circ \cap \text{range}(J_A(x))\right) \\ &\subseteq J_A(x)\partial f(x) + \mathcal{T}_{\mathcal{K}_A}^{lin}(x)^\circ = \partial h(x) + \mathcal{T}_{\mathcal{K}_A}^{lin}(x)^\circ, \end{aligned}$$

最后一个等式由命题 3.1 得出。因此，我们可以得出结论：x 是 CDP 的 KKT 点，证明完毕。

□

A.3 小节 3.2 的证明

定理 3.7 的证明

证明。 **定理 3.7(1) 的证明**

当x 是 CDP 的一阶驻点时，$x \in \mathcal{K}$ 意味着$c(x) = 0$。于是有

$$0 \in \partial h(x) + \mathcal{T}_{\mathcal{K}_A}(x)^\circ = J_A(x)\partial f(x) + \mathcal{T}_{\mathcal{K}_A}(x)^\circ.$$

由引理 3.4 中的 (3.5) 可得

$$\begin{aligned}&0 \in \left(J_A(x)\partial f(x) + \mathcal{T}_{\mathcal{K}_A}(x)^\circ\right) \cap \text{range}(J_A(x))\\ \subseteq\ & J_A(x)\partial f(x) + \left(\mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \text{range}(J_A(x))\right)\\ \subseteq\ & J_A(x)\partial f(x) + J_A(x)\mathcal{T}_{\mathcal{K}}(x)^\circ = J_A(x)(\partial f(x) + \mathcal{T}_{\mathcal{K}}(x)^\circ).\end{aligned}$$

因此，根据引理 A.5，存在 $\hat{}\rho \in \mathbb{R}^p$ 使得 $0 \in \partial f(x) + J_c(x)\hat{\rho} + \mathcal{T}_{\mathcal{K}}(x)^\circ$。此外，引理 A.9 表明 $\mathcal{N}_x \subseteq \mathcal{T}_{\mathcal{K}}^\circ$，由此可知 $J_c(x)\hat{\rho} + \mathcal{T}_{\mathcal{K}}(x)^\circ = \mathcal{T}_{\mathcal{K}}(x)^\circ$。因此，结合定义 2.7，我们得到

$$0 \in \partial f(x) + \mathcal{T}_{\mathcal{K}}(x)^\circ,$$

因此，x 是非线性规划的一阶驻点。

定理 3.7(2) 的证明

假设$x \in \mathcal{K}$ 是非线性规划的 KKT 点，则根据定义 2.9，存在 $\rho \in \mathbb{R}^p$、$\lambda \in \mathbb{R}^{N_E}$ 和 $\mu \in \mathbb{R}_+^{N_I}$ 使得

$$0 \in \partial f(x) + \sum_{l\in[p]} \rho_l \nabla c_l(x) + \sum_{i\in[N_E]} \lambda_i \nabla u_i(x) + \sum_{j\in\mathcal{F}(x)} \mu_j \nabla v_j(x).$$

引理 A.4 表明，对于任意$l \in [p]$，$J_A(x)\nabla c_l(x) = 0$ 成立。因此，我们得到

$$\begin{aligned}&0 \in J_A(x)\Big(\partial f(x) + \sum_{l\in[p]} \rho_l \nabla c_l(x) + \sum_{i\in[N_E]} \lambda_i \nabla u_i(x) + \sum_{j\in\mathcal{F}(x)} \mu_j \nabla v_j(x)\Big)\\ =\ & \partial h(x) + \sum_{i\in[N_E]} \lambda_i J_A(x)\nabla u_i(x) + \sum_{j\in\mathcal{F}(x)} \mu_j J_A(x)\nabla v_j(x)\\ =\ & \partial h(x) + \sum_{i\in[N_E]} \lambda_i \nabla \tilde{u}(x) + \sum_{j\in\mathcal{F}(x)} \mu_j \nabla \tilde{v}(x),\end{aligned}$$

这表明x 是 CDP 的 KKT 点。

另一方面，当$x \in \mathcal{K}$ 是 CDP 的 KKT 点时，有

$$\begin{aligned}&0 \in \partial h(x) + \sum_{i\in[N_E]} \lambda_i \nabla \tilde{u}_i(x) + \sum_{j\in\mathcal{F}(x)} \mu_j \nabla \tilde{v}_j(x)\\ &= J_A(x)\Big(\partial f(x) + \sum_{i\in[N_E]} \lambda_i \nabla u_i(x) + \sum_{j\in\mathcal{F}(x)} \mu_j \nabla v_j(x)\Big).\end{aligned}$$

结合引理 A.5，可得

$$\text{range}(J_c(x)) \bigcap \Big(\partial f(x) + \sum_{i\in[N_E]} \lambda_i \nabla u_i(x) + \sum_{j\in\mathcal{F}(x)} \mu_j \nabla v_j(x)\Big) \neq \emptyset.$$

因此，我们可以得出存在 $\rho \in \mathbb{R}^p$ 使得

$$0 \in \partial f(x) + \sum_{l\in[p]} \rho_l c_l(x) + \sum_{i\in[N_E]} \lambda_i \nabla u_i(x) + \sum_{j\in\mathcal{F}(x)} \mu_j \nabla v_j(x),$$

这意味着x 是非线性规划的 KKT 点。证明完毕。 □

引理 A.10。对于任意给定的 $x \in \mathcal{M}$ 和任意 $y \in \overline{\Omega}_x$，有

$$\|J_A(y)(J_A(y) - I_n)\| \leq \frac{2L_{x,A}(2M_{x,A}+1)}{\sigma_{x,c}} \|c(y)\|.$$

证明。 对于任意 $y \in \overline{\Omega}_x$，有

$$\begin{aligned}
&\|(J_A(y)-I_n)J_A(y)\| \le \|J_A(y)(J_A(y)-J_A(x))\| + \|(J_A(y)-J_A(x))J_A(x)\| + \|J_A(y)-J_A(x)\| \\
&\le L_{x,A}(2M_{x,A}+1)\mathrm{dist}(y,\mathcal{M}) \\
&\le \frac{2L_{x,A}(2M_{x,A}+1)}{\sigma_{x,c}}\|c(y)\|.
\end{aligned}$$

这里第二个不等式由$J_A(y)$ 的 Lipschitz 连续性和引理 A.6 得出。最后一个不等式直接由引理 A.1 得出。至此，证明完成。 □

定理3.8的证明

证明。 对于任意 $y \in \overline{\Omega}_x$、 $\lambda \in \mathbb{R}^{N_E}$ 和 $\mu \in \mathbb{R}^{N_I}$，由引理 A.10 可知

$$\begin{aligned}
&\sup_{w\in\partial(f\circ\mathcal{A})(x)} \|(J_A(y)-I_n)w\| \le \frac{2L_{x,A}M_{x,f}(2M_{x,A}+1)}{\sigma_{x,c}}\|c(y)\|, \\
&\left\|(J_A(y)-I_n)\sum_{i\in[N_E]}\lambda_i\nabla(u_i\circ\mathcal{A})(y)\right\| \le \frac{2L_{x,A}(\|\lambda\|_1 M_{x,u})(2M_{x,A}+1)}{\sigma_{x,c}}\|c(y)\|, \\
&\left\|(J_A(y)-I_n)\sum_{j\in[N_I]}\mu_j\nabla(v_j\circ\mathcal{A})(y)\right\| \le \frac{2L_{x,A}(\|\mu\|_1 M_{x,v})(2M_{x,A}+1)}{\sigma_{x,c}}\|c(y)\|.
\end{aligned}$$

这里最后两个不等式是由 $\|\nabla u(\mathcal{A}(y))\| \le M_{x,u}$ and $\|\nabla v(\mathcal{A}(y))\| \le M_{x,v}$ 得出的。接下来，我们在下文中估计 $\|(J_A(y)-I_n)J_c(y)c(y)\|$ 的下界，

$$\begin{aligned}
&\|(J_A(y)-I_n)J_c(y)c(y)\| \ge \|J_c(y)c(y)\| - \|J_A(y)J_c(y)c(y)\| \\
&\ge \frac{\sigma_{x,c}}{2}\|c(y)\| - \|J_A(y)J_c(y)\|\|c(y)\| \ge \frac{\sigma_{x,c}}{2}\|c(y)\| - L_{x,b}\mathrm{dist}(y,\mathcal{M})\|c(y)\| \\
&\ge \frac{\sigma_{x,c}}{2}\|c(y)\| - L_{x,b}\varepsilon_x\|c(y)\| \ge \frac{\sigma_{x,c}}{4}\|c(y)\|.
\end{aligned} \tag{A.26}$$

因此，我们可以得出结论：对于任意 $y\in\overline{\Omega}_x$，以及任意 $w \in \partial_x\mathcal{L}_{CDP}(y,\lambda,\mu)$，

$$\begin{aligned}
&\|w\| \ge \frac{1}{M_{x,A}+1}\|(J_A(y)-I_n)w\| \\
&\ge \frac{1}{M_{x,A}+1}\left(\left(\beta+\sum_{i\in[N_E]}\lambda_i\tau_i+\sum_{j\in[N_I]}\mu_j\gamma_j\right)\|(J_A(y)-I_n)J_c(y)c(y)\|\right) \\
&\quad - \frac{\|(J_A(y)-I_n)\partial(f\circ\mathcal{A})(y)\|}{M_{x,A}+1} - \frac{\left\|(J_A(y)-I_n)J_A(y)\sum_{i\in[N_E]}\lambda_i\nabla(u_i\circ\mathcal{A})(y)\right\|}{M_{x,A}+1} \\
&\quad - \frac{\left\|(J_A(y)-I_n)J_A(y)\sum_{j\in[N_I]}\mu_j\nabla(v_j\circ\mathcal{A})(y)\right\|}{M_{x,A}+1} \\
&\ge \frac{1}{M_{x,A}+1}\left(\frac{\sigma_{x,c}}{4}\left(\beta+\sum_{i\in[N_E]}\lambda_i\tau_i+\sum_{j\in[N_I]}\mu_j\gamma_j\right) - \frac{2L_{x,A}M_{x,\lambda,\mu}(2M_{x,A}+1)}{\sigma_{x,c}}\right)\|c(y)\| \\
&\ge \left(\frac{\sigma_{x,c}}{4(M_{x,A}+1)}\left(\beta+\sum_{i\in[N_E]}\lambda_i\tau_i+\sum_{j\in[N_I]}\mu_j\gamma_j\right) - \frac{4L_{x,A}M_{x,\lambda,\mu}}{\sigma_{x,c}}\right)\|c(y)\|.
\end{aligned}$$

由于w 是任意的，定理3.8中的不等式（3.7）得证。

此外，当 $\beta+\sum_{i\in[N_E]}\lambda_i\tau_i+\sum_{j\in[N_I]}\mu_j\gamma_j \geq \frac{32L_{x,A}(M_{x,A}+1)M_{x,\lambda,\mu}}{\sigma_{x,c}^2}$ 时，我们得到

$$\operatorname{dist}(0,\partial_x\mathcal{L}_{CDP}(y,\lambda,\mu)) \geq \frac{\sigma_{x,c}}{8(M_{x,A}+1)}\Big(\beta+\sum_{i\in[N_E]}\lambda_i\tau_i+\sum_{j\in[N_I]}\mu_j\gamma_j\Big)\|c(y)\|.$$

因此，对于满足$y\in\overline{\Omega}_x$的 CDP 的任意 KKT 点，以上不等式的左端为 0，因此 $\|c(y)\|=0$。由此$y\in\mathcal{K}$。结合定理 3.7，我们得出y 是非线性规划的 KKT 点。证明完毕。

□

定理3.10的证明

证明。首先，由中值定理、引理 A.2 和引理 A.3 可直接得出 t

$$\begin{aligned}&\Big|\Big(f(\mathcal{A}^2(y))+\lambda^\top u(\mathcal{A}^2(y))+\mu^\top v(\mathcal{A}^2(y))\Big)-\Big(f(\mathcal{A}(y))+\lambda^\top u(\mathcal{A}(y))+\mu^\top v(\mathcal{A}(y))\Big)\Big|\\ \leq\ & M_{x,\lambda,\mu}\left\|\mathcal{A}^2(y)-\mathcal{A}(y)\right\| \leq \frac{2M_{x,\lambda,\mu}(M_{x,A}+1)}{\sigma_{x,c}}\|c(\mathcal{A}(y))\| \leq \frac{8M_{x,\lambda,\mu}(M_{x,A}+1)L_{x,b}}{\sigma_{x,c}^3}\|c(y)\|^2.\end{aligned}$$

回顾引理 A.3，我们知道，对于任意$y\in\Omega_x$，不等式 $\|c(\mathcal{A}(y))\|\leq\frac{1}{2}\|c(y)\|$ 成立。因此，

$$\begin{aligned}&\mathcal{L}_{CDP}(\mathcal{A}(y),\lambda,\mu)-\mathcal{L}_{CDP}(y,\lambda,\mu)\\ \leq\ &\Big|\Big(f(\mathcal{A}^2(y))+\lambda^\top u(\mathcal{A}^2(y))+\mu^\top v(\mathcal{A}^2(y))\Big)-\Big(f(\mathcal{A}(y))+\lambda^\top u(\mathcal{A}(y))+\mu^\top v(\mathcal{A}(y))\Big)\Big|\\ &+\frac{1}{2}\Big(\beta+\sum_{i\in[N_E]}\lambda_i\tau_i+\sum_{j\in[N_I]}\mu_j\gamma_j\Big)\Big(\|c(\mathcal{A}(y))\|^2-\|c(y)\|^2\Big)\\ \leq\ &-\left(\frac{1}{4}\Big(\beta+\sum_{i\in[N_E]}\lambda_i\tau_i+\sum_{j\in[N_I]}\mu_j\gamma_j\Big)-\frac{8M_{x,\lambda,\mu}(M_{x,A}+1)L_{x,b}}{\sigma_{x,c}^3}\right)\|c(y)\|^2\leq 0.\end{aligned}$$

这里最后一个不等式利用了以下事实：

$$\beta+\sum_{i\in[N_E]}\lambda_i\tau_i+\sum_{j\in[N_I]}\mu_j\gamma_j\geq\frac{8M_{x,\lambda,\mu}(M_{x,A}+1)L_{x,b}}{\sigma_{x,c}^3}.$$

至此，证明完成。 □

A.4 第 3.3 节的证明

引理 A.11. 对于任意 $x\in\mathcal{M}$，有

$$J_{\mathcal{A}^\infty}(x)=J_A(x).$$

证明. 引理 A.6 表明，对于任意 $x\in\mathcal{M}$，$J_A(x)^2=J_A(x)$ 成立。因此，对于任意 $k\geq 1$，有

$$J_{A^k}(x)=J_{A^{k-1}}(x)J_A(A^{k-1}(x))=J_{A^{k-1}}(x)J_A(x)=J_{A^{k-2}}(x)J_A(x)^2=J_{A^{k-2}}(x)J_A(x)=J_{A^{k-1}}(x).$$

因此，我们可以得出，对于任意 $k\geq 1$，$J_{A^k}(x)=J_A(x)$ 成立。由此可得

$$J_{\mathcal{A}^\infty}(x)=J_A(x),$$

至此，证明完成。 □

以下 上述命题可由 [35], 直接得到推论，因此为简便起见，我们省略其证明 .

命题 A.12。 对于任意给定的 $x \in \mathcal{M}$，假设 β 和 $\{\gamma_j\}$ 满足条件*3.12*，则对于任意 $y \in \overline{\Omega}_x$，以下不等式成立

$$h(\mathcal{A}^\infty(y)) - h(y) \leq -\frac{\beta}{4}\|c(y)\|^2 \tag{A.27}$$

$$v_i(\mathcal{A}^\infty(y)) \leq v_i(y) + \frac{\gamma_i}{4}\|c(y)\|^2. \tag{A.28}$$

引理 3.14 的证明

证明。 根据非线性规划和 CDP 的表达式，可得 $\mathcal{K} \subseteq \mathcal{K}_A$，因此

$$\mathcal{T}_{\mathcal{K}}(x) \subseteq \mathcal{T}_{\mathcal{K}_A}(x), \quad \text{and} \quad \mathcal{T}^{lin}_{\mathcal{K}}(x) \subseteq \mathcal{T}^{lin}_{\mathcal{K}_A}(x).$$

然后结合引理 3.4，我们得到

$$\mathcal{T}_{\mathcal{K}}(x) \subseteq J_A(x)^\top \mathcal{T}_{\mathcal{K}_A}(x), \quad \text{and} \quad \mathcal{T}^{lin}_{\mathcal{K}}(x) = J_A(x)^\top \mathcal{T}^{lin}_{\mathcal{K}_A}(x).$$

此外，对于任意 $d \in \mathcal{T}_{\mathcal{K}_A}(x)$，存在一个序列 $\{x_k\} \subset \mathcal{K}_A \cap \overline{\Omega}_x$ 收敛到 x，以及一个序列 $\{t_k\} \subset \mathbb{R}_+$ 收敛到 0，使得 $\lim_{k\to+\infty} \frac{x_k - x}{t_k} = d$。于是，根据命题 A.12，对于任意$x_k$ 以及任意 $j \in [N_I]$，有如下结论：

$$\begin{aligned} v_j(\mathcal{A}^\infty(x_k)) = v_j(\mathcal{A}^\infty(\mathcal{A}(x_k))) &\leq v_j(\mathcal{A}(x_k)) + \frac{\gamma_i}{4}\|c(\mathcal{A}(x_k))\|^2 \\ &\leq v_j(\mathcal{A}(x_k)) + \frac{\gamma_i}{16}\|c(x_k)\|^2 \leq \tilde{v}_j(x_k) \leq 0, \end{aligned}$$

这意味着 $\{\mathcal{A}^\infty(x_k)\} \subset \mathcal{K}$。注意，我们在第二个不等式中使用了引理 A.3 中的事实：$\|c(\mathcal{A}(x_k))\| \leq \frac{1}{2}\|c(x_k)\|$

因此，由引理 A.11 可知，

$$\mathcal{T}_{\mathcal{K}}(x) \ni \lim_{k\to+\infty} \frac{\mathcal{A}^\infty(x_k) - \mathcal{A}^\infty(x)}{t_k} = J_{\mathcal{A}^\infty}(x)^\top d = J_A(x)^\top d,$$

由$d \in \mathcal{T}_{\mathcal{K}_A}(x)$的任意性可知，$J_A(x)^\top \mathcal{T}_{\mathcal{K}_A}(x) \subseteq \mathcal{T}_{\mathcal{K}}(x)$。再结合引理 3.4，可得

$$\mathcal{T}_{\mathcal{K}}(x) = J_A(x)^\top \mathcal{T}_{\mathcal{K}_A}(x).$$

此外，直接由引理 A.8 可得

$$\mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \text{range}(J_A(x)) = J_A(x)\mathcal{T}_{\mathcal{K}}(x)^\circ,$$

从而完成证明。 □

定理3.15的证明

证明。 由定理 3.7 可直接得出，CDP 的任一一阶驻点都是非线性规划的一阶驻点。

另一方面，假设x 是非线性规划的一个一阶驻点，则有

$$0 \in \partial f(x) + \mathcal{T}_{\mathcal{K}}(x)^\circ.$$

因此，由引理 3.14 可知

$$\begin{aligned} & 0 \in J_A(x)\left(\partial f(x) + \mathcal{T}_{\mathcal{K}}(x)^\circ\right) \\ &= J_A(x)\partial f(x) + \mathcal{T}_{\mathcal{K}_A}(x)^\circ \cap \text{range}(J_A(x)) \\ &= (\partial h(x) + \mathcal{T}_{\mathcal{K}_A}(x)^\circ) \cap \text{range}(J_A(x)) \\ &\subseteq \partial h(x) + \mathcal{T}_{\mathcal{K}_A}(x)^\circ. \end{aligned}$$

这里第二个等式用到了$c(x) = 0$这一事实。因此，x 是 CDP 的一阶驻点。由此可得，$x \in \mathcal{K}$ 是非线性规划的一阶驻点，当且仅当x 是 CDP 的一阶驻点。证明完毕。

□

定理3.16的证明

证明。我们用反证法证明该定理。假设存在$y \in \overline{\Omega}_x$，它是 CDP 的一阶驻点，且$c(y) \neq 0$。根据定义 2.7，我们首先得到$y \in \mathcal{K}_A$。接下来，我们要找到一个方向 $\tilde{d} \in \mathcal{T}_{\mathcal{K}_A}(y)$，使$h$充分下降。

令 $\tilde{d} := (J_A(y)^\top - I_n)J_c(y)c(y)$，我们首先证明 $\tilde{d} \in \mathcal{T}_{\mathcal{K}_A}(y)$。对于任意 $j \in [N_I]$，由引理 A.10 可知

$$\begin{aligned} &\left\langle \tilde{d}, \nabla(v_j \circ \mathcal{A})(y) \right\rangle \\ =& \left\langle \tilde{d}, J_A(y)\nabla v_j(\mathcal{A}(y)) \right\rangle = \left\langle J_c(y)c(y), (J_A(y) - I_n)J_A(y)\nabla v_j(\mathcal{A}(y)) \right\rangle \\ \leq& \frac{2L_{x,A}(2M_{x,A}+1)M_{x,v}}{\sigma_{x,c}} \|c(y)\| \|J_c(y)c(y)\| \leq \frac{4L_{x,A}(2M_{x,A}+1)M_{x,v}}{\sigma_{x,c}^2} \|J_c(y)c(y)\|^2 . \end{aligned} \tag{A.29}$$

另一方面，由 (A.26) 可得

$$\begin{aligned} &\left\langle (J_A(y)^\top - I_n)J_c(y)c(y), J_c(y)c(y) \right\rangle = -\|J_c(y)c(y)\|^2 + \langle J_c(y)c(y), J_A(y)J_c(y)c(y) \rangle \\ \leq& -\|J_c(y)c(y)\|^2 + \|J_A(y)J_c(y)\| \|c(y)\| \|J_c(y)c(y)\| \\ \leq& -\|J_c(y)c(y)\|^2 + \frac{\varepsilon_x L_{x,b}}{\sigma_{x,c}} \|J_c(y)c(y)\|^2 \leq -\frac{1}{2}\|J_c(y)c(y)\|^2 . \end{aligned} \tag{A.30}$$

于是可得

$$\begin{aligned} \left\langle \tilde{d}, \nabla\tilde{v}_j(y) \right\rangle &= \left\langle \tilde{d}, \nabla(v_j \circ \mathcal{A})(y) \right\rangle + \gamma_i \left\langle (J_A(y)^\top - I_n)J_c(y)c(y), J_c(y)c(y) \right\rangle \\ &\leq \left(\frac{4L_{x,A}(2M_{x,A}+1)M_{x,v}}{\sigma_{x,c}^2} - \frac{\gamma_i}{2} \right) \|J_c(y)c(y)\|^2 < 0. \end{aligned} \tag{A.31}$$

这里最后一个不等式由条件3.12得出。因此，我们可以得出结论：对于任意 $1 \leq j \leq N_I$，$\langle \tilde{d}, \nabla\tilde{v}_j(y) \rangle < 0$ 成立。由 $\tilde{v}_j$ 的 Lipschitz光滑性，存在常数 $\hat{\varepsilon}_0 > 0$，使得对于任意 $t \in [0, \hat{\varepsilon}_0]$，$\tilde{v}(y + t\tilde{d}) < \tilde{v}(y)$ 成立。结合 $y \in \mathcal{K}_A$ 这一事实，我们可以得出，对于任意 $t \in [0, \hat{\varepsilon}_0]$，$y + t\tilde{d} \in \mathcal{K}_A$ 成立，因此 $\tilde{d} \in \mathcal{T}_{\mathcal{K}_A}(y)$。

此外，由于 $0 \in \partial h(y) + \mathcal{T}_{\mathcal{K}_A}(y)^\circ$，存在 $w \in \partial f(\mathcal{A}(y))$，使得 $-(J_A(y)w + \beta J_c(y)c(y)) \in \mathcal{T}_{\mathcal{K}_A}(x)^\circ$。由于 $\tilde{d} \in \mathcal{T}_{\mathcal{K}_A}(y)$，因此有

$$\begin{aligned} 0 \leq& \left\langle \tilde{d}, J_A(y)w + \beta J_c(y)c(y) \right\rangle \\ \overset{(i)}{\leq}& \frac{2L_{x,A}M_{x,f}(2M_{x,A}+1)}{\sigma_{x,c}} \|c(y)\| \|J_c(y)c(y)\| - \frac{\beta}{2}\|J_c(y)c(y)\|^2 \\ \leq& \left(\frac{2L_{x,A}M_{x,f}(2M_{x,A}+1)}{\sigma_{x,c}^2} - \frac{\beta}{2} \right) \|J_c(y)c(y)\|^2 \\ \overset{(ii)}{\leq}& -\frac{\beta}{4}\|J_c(y)c(y)\|^2 < 0, \end{aligned} \tag{A.32}$$

这里 (i) 由引理 A.10 和 (A.30) 得出。此外，(ii) 由条件3.12得出，这说明

$$\beta \geq \frac{64M_{x,f}(M_{x,A}+1)(L_{x,b}+\sigma_{x,c}L_{x,A})}{\sigma_{x,c}^3} \geq \frac{64M_{x,f}(M_{x,A}+1)L_{x,A}}{\sigma_{x,c}^2} \geq \frac{8L_{x,A}M_{x,f}(2M_{x,A}+1)}{\sigma_{x,c}^2}. \tag{A.33}$$

因此，由（A.32）可得矛盾，并得到$c(y) = 0$。因此$y \in \mathcal{K}$。因此，由定理3.15可知，y 是非线性规划的一阶驻点。 □

推论3.17的证明

证明。对于任意一个是 CDP 一阶驻点的 y $\in \overline{\Omega_x}$，令 $\tilde{d} := (J_A(y)^\top - I_n)J_c(y)c(y)$，则由（A.31）可知，对任意 j $\in [N_I]$，$\langle \tilde{d}, \nabla \tilde{v}_j(y) \rangle \leq 0$ 成立。因此可得 $\tilde{d} \in \mathcal{T}\mathrm{lin}\mathcal{K}_A(y)$。

注意到 y 是 CDP 的 KKT 点，因此存在 $\tilde{\mu} \in \mathbb{R}_+^{N_I}$ 和 $w \in \partial f(\mathcal{A}(y))$，使得 $0 = J_A(y)w + \beta J_c(y)c(y) + \sum_{j\in[N_I]} \tilde{\mu}_j \nabla \tilde{v}_j(y)$。因此，类似于 (A.32)，我们有

$$
\begin{aligned}
0 &= \left\langle \tilde{d}, J_A(y)w + \beta J_c(y)c(y) + \sum_{j\in[N_I]} \tilde{\mu}_j \nabla \tilde{v}_j(y) \right\rangle \\
&\leq \left\langle \tilde{d}, J_A(y)w + \beta J_c(y)c(y) \right\rangle \\
&\leq \left(\frac{2L_{x,A}M_{x,f}(2M_{x,A}+1)}{\sigma_{x,c}^2} - \frac{\beta}{2} \right) \|J_c(y)c(y)\|^2 \\
&\leq -\frac{\beta}{4} \|J_c(y)c(y)\|^2 \leq 0.
\end{aligned}
\tag{A.34}
$$

这说明 $c(y) = 0$，因此 $y \in \mathcal{K}$。所以，根据定理 3.7，我们可以得出结论：y 是非线性规划的 KKT 点。另一方面，根据定理 3.7，任何非线性规划的 KKT 点都直接是 CDP 的 KKT 点。证明完毕。

□