arXiv:2301.03026v2 [math.OC] 3 Sep 2023

Dykstra 型投影算法的收敛速率分析 *

XIAOZHOU WANG† 和 TING KEI PONG‡

摘要。给定闭凸集 $C_i, i = 1,\dots,\ell$，以及维数适当的非零线性映射 $A_i, i = 1,\dots,\ell$，多集合分裂可行性问题旨在通过计算到 C_i 的投影以及与 A_i 和 A_i^T 的乘积，找到 $\bigcap_{i=1}^{\ell} A_i^{-1}C_i$ 中的一个点。本文考虑与之相关的最佳逼近问题，即计算到 $\bigcap_{i=1}^{\ell} A_i^{-1}C_i$ 的投影；我们将该问题称为多集合分裂可行性设定下的最佳逼近问题（BA-MSF）。我们对 Dykstra 投影算法进行改造，该算法在所有 $A_i = I$ 的特殊情形下是求解 BA-MSF 的经典方法，以求解一般的 BA-MSF。我们的 Dykstra 型投影算法通过对拉格朗日对偶问题应用（近端）坐标梯度下降推导而来，每次迭代只需计算到 C_i 的投影以及与 A_i 和 A_i^T 的乘积。在标准的相对内部条件以及关于待投影点的一般性假设下，我们证明，当每个 C_i 都是某个 $\alpha \in (0,1]$: 这类集合包含 C^2-锥可约集，而后者包括所有多面体、二阶锥和半正定矩阵锥）-锥可约时，对偶目标函数在（通常无界的）对偶解集的某个邻域上满足 Kurdyka-Łojasiewicz 性质，且其指数可显式计算。由此，我们推导出 Dykstra 型投影算法生成的序列的显式收敛速率（线性或次线性）。我们构造了具体例子，以说明若干假设的必要性。

关键词。Dykstra 投影算法、Kurdyka-Łojasiewicz 性质、线性收敛 e, $C^{1,\alpha}$-锥可约性

MSC 分类号。90C25, 90C30, 90C46, 90C90

1. 引言。多集合分裂可行性（MSF）问题最初由 Censor 等人提出， [20], 是双集合分裂可行性问题的推广 [19]以及凸可行性问题的推广 [7]。这类问题自然地出现在许多当代应用领域中，例如图像重建；参见 [20] 及其中的参考文献。MSF 问题旨在寻找一个点，该点属于有限多个闭凸集的线性原像的交集，即

$$\text{Find } x \in \mathbb{R}^n \text{ such that } A_i x \in C_i \text{ for } i = 1,\dots,\ell, \tag{1.1}$$

其中 $C_i \subseteq \mathbb{R}^{m_i}$, $i = 1,\dots,\ell$ 是闭凸集，且对每个 i, $A_i \in \mathbb{R}^{m_i \times n}$；此外，假设易于计算到 C_i 的投影，而计算到集合 $A_i^{-1}C_i$ 的投影可能较为困难（参见 [20]）。关于投影的这些假设，自然要求采用利用到 C_i 的投影来求解 (1.1) 的迭代方案。其中一种方案是 [18] 提出的 CQ 算法，用于求解具有 $\ell = 2$ 的 MSF 问题；后来，该算法被推广用于求解一般的 MSF 问题，见 [20]。关于 MSF 问题求解方法的近期研究可见 [16, 17, 21, 22, 41]。本文聚焦于 MSF 问题的一种自然但相对较少研究的变体。具体而言，给定 $\bar{v} \in \mathbb{R}^n$，我们考虑寻找 $\bigcap_{i=1}^{\ell} A_i^{-1}C_i$ 中距离 $\bar{v}$ 最近的点这一问题。换言之，我们考虑

*投稿日期：2023年9月6日。**资助：** Ting Kei Pong 获得香港研究资助局 PolyU153002/21p 项目的部分资助。

†香港理工大学，中国中华人民共和国香港（xzhou.wang@connect.polyu.hk）。

‡香港理工大学，香港，中华人民共和国（tk.pong@polyu.edu.hk）。

多集合分裂可行性情形下的以下最佳逼近问题（BA-MSF ）:

$$\min_{x\in\mathbb{R}^n} f(x) := \frac{1}{2}\|x-\bar{v}\|^2 \quad \text{s.t. } A_i x \in C_i,\ i=1,\dots,\ell, \tag{1.2}$$

其中， $\bar{v}$ ∈IR^n 是给定向量，每个 $A_i \in$ IR$^{m_i\times n}\backslash\{0\}$ ，且每个 C_i 都是闭凸集；此外，我们假设

$$\bigcap_{i=1}^{\ell} A_i^{-1}C_i \neq \emptyset.$$

当 $A_i = I$ 对所有 i成立时，我们将相应的问题（1.2）称为最佳逼近（BA）问题。在这种情况下，一种经典的求解方法是 [15, 29]提出的 Dykstra 投影算法。该算法的每次迭代只需计算到各个 C_i 上的投影，而不必计算到 $\bigcap_{i=1}^{\ell} C_i$上的投影；这可能很有优势，因为一般而言，计算到 $\bigcap_{i=1}^{\ell} C_i$ 上的投影可能更加困难。Dykstra 投影算法的一个显著特点是，只要$\bigcap_{i=1}^{\ell} C_i \neq \emptyset$，所生成的序列就会收敛到 BA 问题的唯一解；参见 [15]。这与 Douglas-Rachford 分裂等分裂方法形成对比，后者通常需要额外的假设才能保证收敛；参见 [10, 推论 28。3]近年来，Dykstra 投影算法及其变体的收敛性质以及与其他算法之间的关系得到了广泛研究；参见 [6, 9, 11, 24, 27, 30, 34, 35] 及其中的参考文献。如今已知，当 $A_i = I$ 对所有 i成立时，Dykstra 投影算法可以通过将坐标梯度下降法适当地应用于（1.2）的对偶问题而导出；参见 [27, 29]。此外，对于具有多面体 C_i的 BA 问题，已知 Dykstra 投影算法生成的序列呈线性收敛 [24, 33, 39]。另一方面，对于非多面体C_i 情形，收敛速率尚未得到充分理解。

本文将改造经典 Dykstra 投影算法，以求解 (1.2)，并分析所得算法的收敛速率。按照 [27, 29]„ 我们通过对 (1.2) 的一个适当拉格朗日对偶问题应用（近端）坐标梯度下降，推导出 Dykstra 型投影算法；该算法的每次迭代只需计算到 C_i 上的投影，并与 A_i 和 A_i^T相乘。[1] 接着，在施加标准相对内部条件 $\bigcap_{i=1}^{\ell} A_i^{-1}\,\mathrm{ri}\, C_i \neq \emptyset$以及关于点 $\bar{v}$的一般性假设后，我们证明，当每个 C_i 都是 $C^{1,\alpha}$-锥可约时，对偶问题的目标函数在（通常无界的）对偶解集的某个邻域上满足指数为 $1/(\alpha+1)$ 的 Kurdyka-Lojasiewicz (KL) 性质，其中某个 $\alpha \in (0, 1]$ 。基于此，我们根据 $\alpha \in (0, 1]$的取值，证明 Dykstra 型投影算法生成的序列线性收敛或次线性收敛。我们工作的关键创新在于$both$ 收敛速率分析以及对 $C^{1,\alpha}$-锥可约条件的识别：

- 首先，我们的收敛速率结果并不能直接由标准的基于 KL 性质的收敛分析框架（如 [1–3] 中的框架）推出。事实上，Dykstra 型投影算法生成的序列通常可能无界，而标准的基于 KL 性质的收敛速率分析通常要求所生成的序列有界。我们的分析之所以可行，是因为在假设 $\bigcap_{i=1}^{\ell} A_i^{-1}\,\mathrm{ri}\, C_i \neq \emptyset$ 下，对偶问题的解集可以写成

[1]当对所有 i 均有 $A_i = I$ 时，我们的算法退化为经典 Dykstra 投影算法。

一个紧凸集与一个子空间的和；这一关键事实也已在 [4] 针对 BA 问题（即，(1.2)，其中每个 $A_i = I$）中得到证明。

- 其次，我们想指出，$C^{1,\alpha}$-锥可约（具有 $\alpha \in (0,1]$）的集合在应用中十分常见。事实上，这涵盖了 [38],[2] 中引入的 C^2-锥可约集合这一概念，其中包括多面体、二阶锥和半正定锥等集合。此外，正如我们稍后将看到的，$p \in (1,+\infty)$ 的 p- 范数球可以证明是 $C^{1,\alpha}$- 锥可约的，且 $\alpha = \min\{p-1, 1\}$；见下文例 4.1。此外，当$p \in (1, 2)$时，p-范数球是 $C^{1,\alpha}$-锥可约的，且 $\alpha = p-1$，但不是 C^2-锥可约的（见备注 4.8），这表明 $C^{1,\alpha}$-锥可约集合类严格大于 C^2-锥可约集合类。因此，我们的结果适用于分析适用于广泛集合的 Dykstra 型投影算法。此外，由于当对所有 i 均有 $A_i = I$ 时，我们的 Dykstra 型投影算法可化为经典 Dykstra 投影算法，因此可以推导出经典 Dykstra 投影算法在一大类非多面体 C_i上的收敛速率结果，而此前这些集合的收敛速率尚不明确。

本文结构安排如下。第2节介绍一些预备知识。第3节中，我们对经典 Dykstra 投影算法进行调整，以求解 BA-MSF 问题（1.2），并介绍一些基本的收敛性质。我们的主要贡献在第4 节和第5节：在适当的假设下，我们证明了对偶目标函数具有 KL 性质，并据此研究 Dykstra 型投影算法的收敛速率。我们还给出具体示例，说明在推导误差界和收敛速率结果时，某些假设不可或缺。

2. 符号与预备知识。本文中，我们用 IR^n 表示 n维欧几里得空间。对于 $x \in \mathrm{IR}^n$，记 $\|x\|_p = \sqrt[p]{|x_1|^p + \cdots + |x_n|^p}$为 p-范数，其中 $p \in [1,\infty)$；为简便起见，我们也用 $\|x\|$ 表示 2-范数。对于 x 和 $y \in \mathrm{IR}^n$，我们用 $\langle x,\ y\rangle$ 表示它们的内积。对于矩阵 $A \in \mathrm{IR}^{m\times n}$，我们用 $\|A\|$ 表示其算子范数。我们用 I表示单位矩阵，其维数由上下文确定。我们还用 $B(x,\ \eta)$表示以 x 为中心、半径为 $\eta \geq 0$的闭球，即$B(x,\ \eta) = \{u\colon \|u-x\| \leq \eta\}$。

若扩展实值函数 $h : \mathrm{IR}^n \to (-\infty,\infty]$ 的定义域 $\mathrm{dom}\, h := \{x \in \mathrm{IR}^n\colon h(x) < \infty\}$ 非空，则称其为真函数。若真函数下半连续，则称其为闭函数。对于真凸函数 h，其在 $x \in \mathrm{IR}^n$ 处的次微分 $\partial h(x)$定义为

$$\partial h(x) = \{\xi \in \mathbb{R}^n : h(y) \geq h(x) + \langle \xi, y-x\rangle \text{ for all } y \in \mathbb{R}^n\}.$$

∂h 的定义域定义为 $\mathrm{dom}\,\partial h := \{x \in \mathrm{IR}^n\colon \partial h(x) \neq \emptyset\}$，而 h 的凸共轭定义为

$$h^*(y) = \sup\{\langle x, y\rangle - h(x) : x \in \mathbb{R}^n\}.$$

对于真闭凸函数 h，关于∂h和 h^*，我们有以下等价关系（例如参见 [37, 命题 11.3]）：

$$h(x) + h^*(y) = \langle x, y\rangle \iff y \in \partial h(x) \iff x \in \partial h^*(y). \tag{2.1}$$

[2]这一概念在 [38], 中称为锥可约性，在 [14, 定义 3.135]中称为 C^2-锥可约性。这里我们采用后一种术语，以突出其可微性。

真闭凸函数 h 在 $x \in \mathbb{R}^n$ 上的近端算子定义为 a s

$$\operatorname{prox}_h(x) := \operatorname*{Arg\,min}_{u\in\mathbb{R}^n}\left\{\frac{1}{2}\|u-x\|^2 + h(u)\right\};$$

回顾一下，对于每个 $x \in \mathbb{R}^n$，上述极小化解集合都是单点集，并且 $\operatorname{prox}_h : \mathbb{R}^n \to \mathbb{R}^n$ 是非扩张的。最后，对于一个 h 极小点集合为 $h \neq \emptyset$的真闭凸函数，以下不等式成立（见 [37, 命题 10。59]）：

$$h(x) - \inf h \leq \operatorname{dist}(0, \partial h(x))\operatorname{dist}(x, \operatorname{Arg\,min} h) \quad \text{for all } x \in \mathbb{R}^n, \tag{2.2}$$

其中，$\operatorname{dist}(x, C) := \inf\{\|x-y\| : y \in C\}$ 表示 $x \in \mathbb{R}^n$ 到集合 C 的距离。

对于非空闭凸集 $C \subseteq \mathbb{R}^n$，我们用 $\operatorname{int} C$ 和 $\operatorname{ri} C$ 分别表示 C 的内部和相对内部。此类集合的示性函数和支撑函数分别定义为

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise,} \end{cases} \qquad \text{and} \qquad \sigma_C(x) = \sup\{\langle x, y\rangle : y \in C\}.$$

非空闭凸集 C 在 $x \in C$ 处的法锥定义为

$$\mathcal{N}_C(x) := \partial\delta_C(x) = \{v \in \mathbb{R}^n : \langle v, u-x\rangle \leq 0 \text{ for all } u \in C\},$$

将 $x \in \mathbb{R}^n$ 投影到 C 上记为 $\operatorname{Proj}_C(x) := \operatorname{prox}_{\delta_C}(x)$。对于闭凸锥 $K \subseteq \mathbb{R}^n$，其极锥定义为 $K^\circ = \{u \in \mathbb{R}^n : \langle u, x\rangle \leq 0 \text{ 对所有 } x \in K\}$ 均成立。

接下来，我们回顾 KL 性质的以下定义。该性质对于分析各种一阶方法的收敛速率至关重要；例如，参见 [1–3, 13, 31]。

DEFINITION 2.1（KL 性质与指数）。真闭凸函数 $h : \mathbb{R}^n \to (-\infty, \infty]$ 若在 $\hat{x} \in \operatorname{dom}\partial h$ 处满足 *KL* 性质，则存在 $c \in (0, \infty]$、$\hat{x}$ 的一个邻域 U，以及连续凹函数 $\varphi : [0, c) \to [0, \infty)$，其中 $\varphi(0) = 0$，使得

(i) φ 在 $(0, c)$ 上连续可微，且 $\varphi' > 0$ 在 $(0, c)$上连续；

(ii) $\varphi'(h(x) - h(\hat{x}))\operatorname{dist}(0, \partial h(x)) \geq 1$ 对于满足 $x \in U \cap \{x : h(\hat{x}) < h(x) < h(\hat{x}) + c\}$ 的情形均成立。

若 h 在 $\hat{x}$ 处满足 *KL* 性质，且函数 $\varphi(v) = \alpha_0 v^{1-\theta}$ 对应某个 $\alpha_0 > 0$ 和 $\theta \in [0, 1)$，则称 h 在 $\hat{x}$ 处满足指数为 θ 的 *KL* 性质。

满足 *KL* 性质且指数为 $\theta \in [0, 1)$ 的真闭凸函数 h，若该性质在 $\operatorname{dom}\partial h$ 中的每一点都成立，则称为指数为 θ 的 *KL* 函数。

指数为 $\theta \in [0, 1)$ 的 KL 性质与其他误差界概念密切相关。例如，根据 [12, 定理 5]，，真闭凸函数 h 在一个 $\bar{x} \in$ 极小点集合 h 处满足指数为 $\theta \in [0, 1)$ 的 KL 性质，当且仅当存在正常数 c_1 和 c_2，以及 $\bar{x}$ 的一个邻域 $U_{\bar{x}}$，使得

$$c_1\operatorname{dist}(x, \operatorname{Arg\,min} h) \leq (h(x) - h(\bar{x}))^{1-\theta} \text{ for all } x \in U_{\bar{x}} \cap \{x : h(\bar{x}) < h(x) < h(\bar{x}) + c_2\}.$$

在结束本节之前，我们简要回顾一下 [15, 29] 中提出的（经典）Dykstra 投影算法。该 Dykstra 投影算法旨在求解以下 BA 问题

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|x - \bar{v}\|^2 \quad \text{s.t. } x \in C_i,\ i = 1, \ldots, \ell, \tag{2.3}$$

算法 2.1用于 (2.3) 的 Dykstra 投影算法

步骤 1. 选择 $y_1^0 = \cdots = y_\ell^0 = 0 \in \mathrm{IR}^n$ 和 $x^0 = x_\ell^0 = \bar{v} \in \mathrm{IR}^n$。令 $t = 0$。**步骤 2.** 令 $x_0^{t+1} = x_\ell^t$。对于 $i = 1, \ldots, \ell$，计算

$$x_i^{t+1} = \mathrm{Proj}_{C_i}(y_i^t + x_{i-1}^{t+1}),$$
$$y_i^{t+1} = y_i^t + x_{i-1}^{t+1} - \mathrm{Proj}_{C_i}(y_i^t + x_{i-1}^{t+1}).$$

步骤 3. 令 $x^{t+1} = x_\ell^{t+1}$。更新 $t \leftarrow t+1$ 并转到步骤 2。

其中，每个 C_i 都是闭凸集，且 $\bigcap_{i=1}^{\ell} C_i \neq \emptyset$。显然，问题 (2.3) 是 (1.2) 的一个特殊实例，其中 $A_i = I$ 对所有 i均成立。下面给出了 Dykstra 投影算法（算法 2.1）。请注意，该算法在每次迭代中都会用到 Proj_{C_i}。

文献 [27,29] 指出，算法 2.1等价于一种用于求解 (2.3) 的（负）对偶问题的（近端）坐标梯度下降（CGD）方法，其形式如下：

$$\min_{y_1,\ldots,y_\ell} \frac{1}{2}\left\|\sum_{i=1}^{\ell} y_i - \bar{v}\right\|^2 - \frac{1}{2}\|\bar{v}\|^2 + \sum_{i=1}^{\ell} \sigma_{C_i}(y_i);$$

事实上，对于 $y_1^0 = \cdots = y_\ell^0 = 0$，当 $i = 1, \ldots, \ell$时，可以证明，算法 y_i^{t+1} 2.1 中的 可以等价地表示为

$$y_i^{t+1} := \operatorname*{Arg\,min}_{y_i \in \mathbb{R}^{m_i}} \left\{ \left\langle \sum_{j=1}^{i-1} y_j^{t+1} + \sum_{j=i}^{\ell} y_j^t - \bar{v}, y_i - y_i^t \right\rangle + \frac{1}{2}\|y_i - y_i^t\|^2 + \sigma_{C_i}(y_i) \right\}.$$

以下定理汇总了算法2.1的一些已知收敛结果。

THEOREM 2.2（算法 2.1 的收敛性质）。考虑 (2.3)。设$\{x^t\}$ 由算法 *2.1*生成。则以下结论成立。

(i) 有 $\lim_{t\to+\infty} x^t = x^*$，其中 x^* 是 (2.3)的唯一解。(ii) 如果每个 C_i 都是多面体，则存在 $a_1 > 0$， $a_0 \in (0, 1)$ 和一个正整数 t使得

$$\|x^t - x^*\| \le a_1 a_0^t, \quad \forall\, t \ge \bar{t}.$$

证明。结论 (i) 已在 [15]中证明。结论 (ii) 可由 [33,39]推出；另见下文定理 5.1。 □

3. 求解 BA-MSF 问题的 Dykstra 型投影算法。 本节将介绍如何调整算法 2.1 以求解 BA-MSF 问题（1.2），并讨论所得算法的一些基本收敛性质。我们将所得算法称为 Dykstra 型投影算法。为简化记号，下文记

$$\boldsymbol{A} \in \mathbb{R}^{m\times n} \text{ with } \boldsymbol{A}^T := \begin{bmatrix} A_1^T & \cdots & A_\ell^T \end{bmatrix} \quad \text{and} \quad D := C_1 \times \cdots \times C_\ell \subseteq \mathbb{R}^m, \tag{3.1}$$

其中 $m = \sum_{i=1}^{\ell} m_i$；此外，我们用粗体表示 IR^m 中的元素，即 $\boldsymbol{y} = (y_1, \ldots, y_\ell)$，其中 $\boldsymbol{y} \in \mathrm{IR}^m$ 和 $y_i \in \mathrm{IR}^{m_i}$， $i = 1, \ldots, \ell$。

沿着 [27], 通过考虑一种用于求解（1.2）的对偶问题的 CGD 方法，我们提出了一种用于求解 BA-MSF 问题（1.2）的 Dykstra 型投影算法。为此，我们首先定义 $\psi : \mathrm{IR}^{n+m} \to (-\infty, +\infty]$ 如下

$$\psi(x, \boldsymbol{y}) := f(x) + \delta_D(\boldsymbol{A}x + \boldsymbol{y}). \tag{3.2}$$

那么，对于每个 $\bar{x} \in \mathbb{R}^n$ 和 $\bar{\boldsymbol{y}} \in \mathbb{R}^m$，都有

$$\begin{aligned}\psi^*(\bar{x}, \bar{\boldsymbol{y}}) &= \sup_{x,\boldsymbol{y}}\{\langle(\bar{x}, \bar{\boldsymbol{y}}), (x, \boldsymbol{y})\rangle - f(x) - \delta_D(\boldsymbol{A}x + \boldsymbol{y})\} \\ &= \sup_{x,\boldsymbol{y}}\{\langle -\boldsymbol{A}^T\bar{\boldsymbol{y}} + \bar{x}, x\rangle - f(x) + \langle \bar{\boldsymbol{y}}, \mathbf{A}x + \boldsymbol{y}\rangle - \delta_D(\boldsymbol{A}x + \boldsymbol{y})\} \\ &= \sup_{x}\{\langle -\boldsymbol{A}^T\bar{\boldsymbol{y}} + \bar{x}, x\rangle - f(x)\} + \sup_{\boldsymbol{z}}\{\langle \bar{\boldsymbol{y}}, \boldsymbol{z}\rangle - \delta_D(\boldsymbol{z})\} \\ &= f^*(-\boldsymbol{A}^T\bar{\boldsymbol{y}} + \bar{x}) + \sigma_D(\bar{\boldsymbol{y}}). \qquad (3.3)\end{aligned}$$

因此，鉴于 [36, 定理 31。2]以及（3.3），（1.2）的对偶问题为

$$\max_{\boldsymbol{y} \in \mathbb{R}^m} -\psi^*(0, \boldsymbol{y}). \qquad (3.4)$$

由于 $f^*(x) = \frac{1}{2}\|x + \bar{v}\|^2 - \frac{1}{2}\|\bar{v}\|^2$，（3.4）可以进一步写为

$$\max_{\boldsymbol{y} \in \mathbb{R}^m} -\frac{1}{2}\left\|\sum_{i=1}^{\ell} A_i^T y_i - \bar{v}\right\|^2 + \frac{1}{2}\|\bar{v}\|^2 - \sum_{i=1}^{\ell}\sigma_{C_i}(y_i). \qquad (3.5)$$

注意，利用 $\boldsymbol{A}$ 和 D 在（3.1）中的定义，我们可以将上述对偶问题等价地改写为

$$d^* := \min_{\boldsymbol{y} \in \mathbb{R}^m} d(\boldsymbol{y}), \quad \text{where } d(\boldsymbol{y}) := \underbrace{\frac{1}{2}\left\|\boldsymbol{A}^T\boldsymbol{y} - \bar{v}\right\|^2 - \frac{1}{2}\|\bar{v}\|^2}_{g(\boldsymbol{y})} + \sigma_D(\boldsymbol{y}). \qquad (3.6)$$

以下命题表明，（1.2）与其对偶问题（3.5）之间的对偶间隙始终为零。这证明了通过考虑（1.2）来求解（3.6）的合理性。

PROPOSITION 3.1. 考虑（1.2）。设函数 d 由（3.6）给出。则有$\inf\{f(x) : x \in \bigcap_{i=1}^{\ell} A_i^{-1}C_i\} = -d^* < \infty$。此外，对于任意 $\boldsymbol{y}^* = (y_1^*, \ldots\ldots, y_\ell^*) \in$极小点集合 d，有

$$x^* = \bar{v} - \sum_{i=1}^{\ell} A_i^T y_i^* = \bar{v} - \boldsymbol{A}^T\boldsymbol{y}^*,$$

其中 x^* 是（1.2）的唯一解。

证明。对于 ψ 在（3.2）中定义的，注意到 $\boldsymbol{y} \mapsto \psi(\cdot, \boldsymbol{y})$是一个适当的闭凸双函数。[3] 此外，原问题（即 $\min_{x \in \mathbb{R}^n}\psi(x,\mathbf{0})$）有唯一解，这归功于 $\bigcap_{i=1}^{\ell} A_i^{-1}C_i \neq \emptyset$。由此我们可以从 [36, 定理 30.4(i)]推得对偶间隙为零，即

$$\sup_{\boldsymbol{y}} -\psi^*(0, \boldsymbol{y}) = \inf_{x}\psi(x, \mathbf{0}) = \inf\left\{f(x) : x \in \bigcap_{i=1}^{\ell} A_i^{-1}C_i\right\} = -d^* < \infty.$$

接下来，若 $\boldsymbol{y}^* = (y_1^*, \ldots\ldots, y_\ell^*) \in$极小点集合 d，则由《》中的定理 10.1 和练习 8.8 可得 [37] 该结论

$$\mathbf{0} \in \partial d(\boldsymbol{y}^*) = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{y}^* - \bar{v}) + \partial\sigma_D(\boldsymbol{y}^*).$$

[3]关于双函数的定义，请参见 [36, 第 291-292] 页。若其图函数 $(x, \boldsymbol{y}) \mapsto \psi(x, \boldsymbol{y})$ 是适当闭凸的，则称双函数 $\boldsymbol{y} \mapsto \psi(\cdot, \boldsymbol{y})$ 为适当闭凸；参见 [36, 第 293]

结合上述推导式与 [37,例 11。4] 可得

$$\boldsymbol{y}^* \in \mathcal{N}_D(\boldsymbol{A}(\bar{v} - \boldsymbol{A}^T\boldsymbol{y}^*)).$$

为简化记号，记 $\hat{x} := \bar{v} - \boldsymbol{A}^T\boldsymbol{y}^*$。根据 [37, 命题 10.5], ，我们有 $y_i^* \in \mathcal{N}_{C_i}(A_i\hat{x})$, 对所有 i均成立。因此，对所有 i，我们有 $A_i^T y_i^* \in A_i^T\mathcal{N}_{C_i}(A_i\hat{x}) \subseteq \mathcal{N}_{A_i^{-1}C_i}(\hat{x})$，其中集合包含关系直接由法锥的定义可得。因此，

$$\bar{v} - \hat{x} = \boldsymbol{A}^T\boldsymbol{y}^* = \sum_{i=1}^{\ell} A_i^T y_i^* \subseteq \sum_{i=1}^{\ell} \mathcal{N}_{A_i^{-1}C_i}(\hat{x}) \subseteq \mathcal{N}_{\bigcap_{i=1}^{\ell} A_i^{-1}C_i}(\hat{x}),$$

其中，最后一个集合包含关系直接由定义可得。上式表明，$\hat{x}$ 是 $\bar{v}$ 在 $\bigcap_{i=1}^{\ell} A_i^{-1}C_i$上的投影，因此我们得到了所需的 $\hat{x} = x^*$。 □

现在，我们通过将（近端）CGD 方法应用于求解（3.6），推导 Dykstra 型投影算法。令起始点为

$$\boldsymbol{y}^0 = (y_1^0, \ldots, y_\ell^0) = (0, \ldots, 0). \tag{3.7}$$

对于任意 $t \geq 0$ 和任意 $i \in \{1, \ldots, \ell\}$，为简化记号，记作

$$\tilde{\boldsymbol{y}}_{i-1}^{t+1} := (y_1^{t+1}, \ldots, y_{i-1}^{t+1}, y_i^t, \ldots, y_\ell^t). \tag{3.8}$$

按照循环顺序将（近端）CGD 应用于 (3.6)，并加入由 $\gamma_i I - A_i A_i^T$诱导的近端项，其中 $\gamma_i = \lambda_{\max}(A_i A_i^T)$，可得 $y_i, i = 1, \ldots, \ell$ 的如下更新公式：

$$y_i^{t+1} := \operatorname*{Arg\,min}_{y_i \in \mathbb{R}^{m_i}} \left\{ \langle \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}), y_i - y_i^t \rangle + \frac{\gamma_i}{2}\|y_i - y_i^t\|^2 + \sigma_{C_i}(y_i) \right\}. \tag{3.9}$$

遵循 [27, 第 33-34]页中的类似推导，并利用 (3.7)-(3.9)，可推导出用于求解 (1.2) 的 Dykstra 型投影算法，如算法 3.1 所示。特别地，可以通过归纳法证明

$$x^t := x_\ell^t = \bar{v} - \boldsymbol{A}^T\boldsymbol{y}^t \quad \text{for all } t \geq 0. \tag{3.10}$$

详细推导见附录A。

算法 3.1求解 BA-MSF 问题 (1.2) 的 Dykstra 型投影算法

步骤 1. 选取 $y_i^0 = 0 \in \mathbb{R}^{m_i}$ 用于 $i = 1, \ldots, \ell$, $x^0 = \bar{v} \in \mathbb{R}^n$，并令 $\gamma_i = \lambda_{\max}(A_i^T A_i)$ 用于 $i = 1, \ldots, \ell$。令 $\boldsymbol{y}^0 = (y_1^0, \ldots, y_\ell^0)$, $x_\ell^0 = x^0$ 和 $t = 0$。

步骤2. 令 $x_0^{t+1} = x_\ell^t$。对于 $i = 1, \ldots, \ell$，计算

$$\begin{aligned} x_i^{t+1} &= (I - \gamma_i^{-1} A_i^T A_i) x_{i-1}^{t+1} + \gamma_i^{-1} A_i^T \mathrm{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}), \\ y_i^{t+1} &= y_i^t + \gamma_i^{-1} A_i x_{i-1}^{t+1} - \gamma_i^{-1} \mathrm{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}). \end{aligned} \tag{3.11}$$

步骤3. 令 $x^{t+1} = x_\ell^{t+1}$ 和 $\boldsymbol{y}^{t+1} = (y_1^{t+1}, \ldots, y_\ell^{t+1})$。更新 $t \leftarrow t+1$ 并转到步骤 2。

可以看出，算法 3.1 在每个 $A_i = I$ 时都会退化为算法 2.1 。如上所述，算法 3.1 正是用于求解（3.6）的 CGD 方法。以下命题汇总了算法 $\{\boldsymbol{y}^t\}$ 3.1 中对偶迭代 $\{\boldsymbol{y}^t\}$ 的一些性质，这些性质是 CGD 相关已知结果的直接推论；例如，参见 [39]。为便于后续的收敛速率分析，我们在此列出这些性质及其简单证明。

PROPOSITION 3.2. 考虑 (1.2)。设函数 d 由 (3.6) 给出，且 $\{\boldsymbol{y}^t\}$为算法 *3.1* 生成的序列。则以下结论成立：(i) 对每个 $i=1,\ldots,\ell$，均有

$$d(\tilde{\boldsymbol{y}}_i^{t+1})-d(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\le\frac{1}{2}\Delta_i^t\le-\frac{\gamma_i}{2}\|y_i^{t+1}-y_i^t\|^2,$$

其中 $\Delta_i^t:=\langle\nabla_{y_i}g(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\ y_i^{t+1}-y_i^t\rangle+\sigma_D(\tilde{\boldsymbol{y}}_i^{t+1})-\sigma_D(\tilde{\boldsymbol{y}}_{i-1}^{t+1})$，$\tilde{\boldsymbol{y}}_{i-1}^{t+1}$与 (3.8)中 $i=1,\ldots,\ell$和 $\tilde{\boldsymbol{y}}_\ell^{t+1}:=\boldsymbol{y}^{t+1}$。
(ii) 存在 $c>0$ 使得 $d(\boldsymbol{y}^{t+1})-d(\boldsymbol{y}^t)\le-c\|\boldsymbol{y}^{t+1}-\boldsymbol{y}^t\|^2$。(iii) $\lim_{t\to\infty}\|\boldsymbol{y}^{t+1}-\boldsymbol{y}^t\|=0$。(iv) $\{\boldsymbol{y}^t\}$ 是驻点序列，即 $\lim_{t\to\infty}\operatorname{dist}(\mathbf{0},\partial d(\boldsymbol{y}^t))=0$。(v) 的每个聚点 $\{\boldsymbol{y}^t\}$ 都是 (3.6)的解。

证明。 我们首先注意到，根据（ $\tilde{\boldsymbol{y}}_{i-1}^{t+1}$ 3.8）中的定义、（ y_i^{t+1} 3.9）中的定义，以及（3.9）中目标函数的强凸性，对于 $i=1,\ldots,\ell$

$$\langle\nabla_{y_i}g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}),y_i^{t+1}-y_i^t\rangle+\frac{\gamma_i}{2}\|y_i^{t+1}-y_i^t\|^2+\sigma_D(\tilde{\boldsymbol{y}}_i^{t+1})\le\sigma_D(\tilde{\boldsymbol{y}}_{i-1}^{t+1})-\frac{\gamma_i}{2}\|y_i^{t+1}-y_i^t\|^2.\tag{3.12}$$

此外，若令 $z_i^t:=\sum_{j=1}^{i-1}A_j^Ty_j^{t+1}+\sum_{j=i+1}^{\ell}A_j^Ty_j^t-\bar{v}$，则

$$\begin{aligned}
d(\tilde{\boldsymbol{y}}_i^{t+1})&=g(\tilde{\boldsymbol{y}}_i^{t+1})+\sigma_D(\tilde{\boldsymbol{y}}_i^{t+1})=\frac{1}{2}\left\|A_i^Ty_i^{t+1}+z_i^t\right\|^2-\frac{1}{2}\|\bar{v}\|^2+\sigma_D(\tilde{\boldsymbol{y}}_i^{t+1})\\
&\overset{(a)}{=}\frac{1}{2}\left\|A_i^Ty_i^t+z_i^t\right\|^2-\frac{1}{2}\|\bar{v}\|^2+\langle\nabla_{y_i}g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}),y_i^{t+1}-y_i^t\rangle+\frac{1}{2}\|A_i^T(y_i^{t+1}-y_i^t)\|^2+\sigma_D(\tilde{\boldsymbol{y}}_i^{t+1})\\
&\le\frac{1}{2}\left\|A_i^Ty_i^t+z_i^t\right\|^2-\frac{1}{2}\|\bar{v}\|^2+\langle\nabla_{y_i}g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}),y_i^{t+1}-y_i^t\rangle+\frac{\gamma_i}{2}\|y_i^{t+1}-y_i^t\|^2+\sigma_D(\tilde{\boldsymbol{y}}_i^{t+1})\\
&\overset{(b)}{=}d(\tilde{\boldsymbol{y}}_{i-1}^{t+1})+\Delta_i^t+\frac{\gamma_i}{2}\|y_i^{t+1}-y_i^t\|^2\overset{(c)}{\le}d(\tilde{\boldsymbol{y}}_{i-1}^{t+1})-\frac{\gamma_i}{2}\|y_i^{t+1}-y_i^t\|^2,
\end{aligned}$$

其中，(a) 成立是因为 $A_i(z_i^t+A_i^Ty_i^t)=\nabla_{y_i}g(\tilde{\boldsymbol{y}}_{i-1}^{t+1})$；(b) 根据$\Delta_i^t$的定义成立；而 (c) 根据（3.12）以及 Δ_i^t 的定义成立。 这就证明了第 (i) 项。第 (ii) 项是 (i) 的直接推论，而第 (iii) 项由第 (ii) 项推出（因为 $\inf d>-\infty$，根据命题 3.1）。

对于第 (iv) 项，由（3.9）的一阶最优性条件可知

$$\begin{aligned}
0&\in\nabla_{y_i}g(\tilde{\boldsymbol{y}}_{i-1}^{t+1})+\gamma_i(y_i^{t+1}-y_i^t)+\partial\sigma_{C_i}(y_i^{t+1})\\
&=[\gamma_i(y_i^{t+1}-y_i^t)+(\nabla_{y_i}g(\tilde{\boldsymbol{y}}_{i-1}^{t+1})-\nabla_{y_i}g(\boldsymbol{y}^{t+1}))]+\nabla_{y_i}g(\boldsymbol{y}^{t+1})+\partial\sigma_{C_i}(y_i^{t+1}).
\end{aligned}$$

利用 $\boldsymbol{y}^{t+1}-\boldsymbol{y}^t\to0$ （见第 (iii) 项）这一事实、 在 IR ∇g 上的一致连续性[m] 以及定义 $\boldsymbol{y}^{t+1}=(y_1^{t+1},\ldots,y_\ell^{t+1})$ 和 $\tilde{\boldsymbol{y}}_{i-1}^{t+1}=(y_1^{t+1},\ldots,y_{i-1}^{t+1},y_i^t,\ldots,y_\ell^t)$（见（3.8））， 上述推导表明 $\{\boldsymbol{y}^t\}$ 是一个驻点序列。最后，第 (v) 项由第 (iv) 项以及次微分作为集值映射的闭性立即得出 [36,定理 24。4]。 □

命题3.2(v) 只有在聚点存在时才有用。对于每个 $A_i=I$ 属于 (1.2) 的情形， [29, 引理 46] 表明， d 是水平有界的，条件是$\bigcap_{i=1}^{\ell}\operatorname{int}C_i\neq\emptyset$，这将进一步推出 $\{\boldsymbol{y}^t\}$ 的有界性，因而聚点存在。然而，一般而言，可能出现 $\bigcap_{i=1}^{\ell}\operatorname{int}C_i=\emptyset$ 且序列 $\{\boldsymbol{y}^t\}$ 无界的情况。幸运的是，命题 3.2(iv) 指出， $\{\boldsymbol{y}^t\}$ 始终是 驻点序列。下文与 [4] 研究该情形时一样

当 $A_i = I$ 对所有 i均成立时，我们将证明，若交集集合满足温和条件，则函数d 在 (3.6) 中的每个驻点序列都是极小化序列。根据 [5, 定义 4.22] ，所有驻点序列都是极小化序列的真闭凸函数称为渐近良态（AWB）。

PROPOSITION 3.3（d的 *AWB* 性质）。考虑 *(1.2)*，并假设条件$\bigcap_{i=1}^{\ell} A_i^{-1}\, ri\, C_i \neq \emptyset$成立。设 d 由 *(3.6)* 给出，且 $q(y) := \inf_x \psi(x, y)$，其中 ψ 由 *(3.2)* 定义。则 $\mathbf{0} \in ri\, dom\, q$。此外，对所有 y，有 $d(y) = q^*(y)$，并且

$$\operatorname{Arg\,min} d = \partial q_E(\mathbf{0}) + E^{\perp} \neq \emptyset, \tag{3.13}$$

其中 $q_E(y) := q(\mathrm{Proj}_E(y))$ 且

$$E := \operatorname{span}(\operatorname{dom} q) = \operatorname{span}(D - \operatorname{Range}(\boldsymbol{A})) \tag{3.14}$$

其中 D 且 $\boldsymbol{A}$ 由以下给出(3.1)。此外，每个驻点序列 $\{z^t\}$ 的 d 都满足$d(z^t) \to d^*$ 且 $\operatorname{dist}(z^t,$极小点集合 $d) \to 0$，其中 d^* 由以下给出(3.6)。

证明。根据 的定义，我们有 $q(y) < +\infty$ 当且仅当存在$x \in \mathrm{IR}^n$ 使得 $A_i x + y_i \in C_i$ 对所有 i （因为 $\operatorname{dom} f = \mathrm{IR}^n$），这意味着 $\operatorname{dom} q = D - \operatorname{Range}(\boldsymbol{A})$。根据 [37, 练习 2. 45], 上面的等式推出

$$\operatorname{ri} \operatorname{dom} q = \operatorname{ri} D - \operatorname{Range}(\boldsymbol{A}).$$

因此，假设 $\bigcap_{i=1}^{\ell} A_i^{-1} \operatorname{ri} C_i \neq \emptyset$蕴含 $\mathbf{0} \in \operatorname{ri} \operatorname{dom} q$。由（3.3）、（3.6）以及 q 的定义可得 d（y） $= \psi^*(0, y)$ $= q^*$（y）对所有 y均成立。所需结论现在可由 [4, 定理 3。1]得出。 □

现在我们可以给出一个全局收敛结果，即使 $\{y^t\}$可能无界，该结果仍然具有参考价值。证明可直接由命题 3.2(iv) 和 3.3 得出。

THEOREM 3.4（无有界性条件下的全局收敛）。考虑(1.2) 并假设 $\bigcap_{i=1}^{\ell} A_i^{-1} \operatorname{ri} C_i \neq \emptyset$。令 $\{y^t\}$ 为由算法 *3.1*生成的序列。则有 d（y^t） $\to d^*$ 且 $\operatorname{dist}(y^t,$极小点集合$d) \to 0$，其中 d和 d^* 由 （(3.6)中定义。

上述定理指出，序列 $\{d(y^t) - d^*\}$ 和 $\{\operatorname{dist}(y^t,$ 极小点集合 $d)\}$ 收敛于零。本文关注它们的收敛速率。为此，我们回顾一下，KL 性质广泛用于分析一阶方法的收敛速率，其速率通常显式依赖于某个合适势函数的 KL 指数；例如参见 [1–3, 31, 42]。下一节中，我们将确定一大类集合 C_i，并证明相应函数 d 满足 KL 性质，且其指数显式已知。此外，我们还将证明，在可能无界的极小点集合d的某个邻域内，KL 指数可以统一选取。第 5 节将进一步发展这些结果，以推导 $\{d(y^t) - d^*\}$ 和 $\{\operatorname{dist}(y^t,$ 极小点集合 $d)\}$ 的显式收敛速率。

4. KL 性质与 $C^{1,\alpha}$-锥可约集合。

4.1. $C^{1,\alpha}$-锥可约集。 本小节将介绍一类集合，以供后续收敛速率分析使用。具体而言，我们对闭集 $\mathfrak{D} \subseteq \mathbb{X}$定义以下 $C^{1,\alpha}$-锥可约性的概念，它可以视为 C^2-锥可约性概念 [38]的推广；在此处及全文中，我们用字体 $\mathbb{X}, \mathbb{Y}, \mathbb{Z}$等表示有限维希尔伯特空间，其相应的内积和范数分别记为 $\langle \cdot, \cdot \rangle$ 和 $\| \cdot \|$,

为简化记号。 我们还用 $\mathcal{T}^*$ 表示线性算子 $\mathcal{T}$: $\mathbb{X} \to \mathbb{Z}$的伴随算子，其定义为 $\langle \mathcal{T}x, z\rangle = \langle x,$对所有 $x \in \mathbb{X}$和 $z \in \mathbb{Z}$均成立；并且 $\|\mathcal{T}\|$ 是 $\mathcal{T}$的算子范数。

Definition 4.1（C^1, $^\alpha$-锥可约集）。设 $\alpha \in (0, 1]$。若存在正常数 ρ、有限维希尔伯特空间 $\mathbb{Y}$ 中的闭凸尖锥 K 以及映射 Ξ: $\mathbb{X} \to \mathbb{Y}$，使得该映射将 x 映射到 0，并且在 $B(x, \rho)$ 中连续可微，且 $D\Xi(x)$ 是满射，[4]映射 $x \mapsto D\Xi(x)$ 在 $B(x, \rho)$ 中是 α-*H¨older* 连续的，则称闭集 $\mathfrak{D} \subseteq \mathbb{X}$ 在 $\hat{x} \in \mathfrak{D}$ 处是 C^1, $^\alpha$-锥可约的，并且

$$\mathfrak{D} \cap B(\hat{x}, \rho) = \{x : \Xi(x) \in K\} \cap B(\hat{x}, \rho).$$

若闭集 $\mathfrak{D}$ 在每个$x \in \mathfrak{D}$处都是 C^1, $^\alpha$-锥可约的，则称其为 C^1, $^\alpha$-锥可约集。

显然，每个 C^2-锥可约集（见 [38,定义 3.1]）对于任意 $\alpha \in (0, 1]$ 都是 C^1, $^\alpha$-锥可约的。根据 [38], , 所有多面体、二阶锥和半正定矩阵锥都是 C^2-锥可约集；因此，对于任意 $\alpha \in (0, 1]$，它们也都是 C^1, $^\alpha$-锥可约的。下一个引理为一个 C^1, $^\alpha$-锥可约闭凸集的法锥映射建立一种“镇定性类型”的性质。这可以看作是对 [42, 定理 4.4]，的扩展，后者考虑了 C^2-锥可约集。我们的证明也类似于 [42, 定理 4.4] 的证明。

Lemma 4.2. 设 $\alpha \in (0, 1]$ 和 $\mathfrak{D} \subseteq \mathbb{X}$ 是闭凸集，且在 $\hat{x} \in \mathfrak{D}$处为 $C^{1,\alpha}$-锥可约。令 $\hat{y} \in \mathcal{N}_{\mathfrak{D}}(\hat{x})$。那么存在 $\hat{\rho} > 0$、$\hat{\delta} > 0$ 和 $\hat{\kappa} > 0$，使得

$$\mathcal{N}_{\mathfrak{D}}(x) \cap B(\hat{y}, \hat{\delta}) \subseteq \mathcal{N}_{\mathfrak{D}}(\hat{x}) + \hat{\kappa}\|x - \hat{x}\|^\alpha B(0, 1) \quad \textit{for all } x \in B(\hat{x}, \hat{\rho}). \tag{4.1}$$

证明。 由于 $\mathfrak{D}$ 在 $C^{1,\alpha}$-锥可约于 $\hat{x}$，存在一个正数 ρ、闭凸尖锥 $K \subseteq \mathbb{Y}$ 以及连续可微映射 $\Xi : \mathbb{X} \to \mathbb{Y}$，它们满足定义 4.1 中的条件，使得

$$\mathfrak{D} \cap B(\hat{x}, \rho) = \{x : \Xi(x) \in K\} \cap B(\hat{x}, \rho). \tag{4.2}$$

由于 $D\Xi(\hat{x})$ 是满射，可以选取一个正数 $\tilde{\rho} \in (0, \rho)$，使得对于任意$x \in B(\hat{x}, \tilde{\rho})$，线性映射 $D\Xi(x)$ 都是满射（即 $D\Xi(x)D\Xi(x)^*$ 可逆），并且

$$\sup_{x \in B(\hat{x}, \tilde{\rho})} \|[D\Xi(x)D\Xi(x)^*]^{-1}\| =: \tau < \infty. \tag{4.3}$$

关系（4.2）表示

$$\delta_{\mathfrak{D}}(x) = \delta_K(\Xi(x)) \quad \text{for all } x \in B(\hat{x}, \tilde{\rho}).$$

利用以下事实：只要 $x \in B(\hat{x}, \tilde{\rho})$，$D\Xi(x)$ 就是满射；再运用 [37, Exercise 10.7], , 我们便可由上面的等式推出

$$\mathcal{N}_{\mathfrak{D}}(x) = D\Xi(x)^*\mathcal{N}_K(\Xi(x)) \quad \text{for all } x \in \mathfrak{D} \cap B(\hat{x}, \tilde{\rho}). \tag{4.4}$$

˜任取 $\delta > 0$ 并取

$$x \in \mathfrak{D} \cap B(\hat{x}, \tilde{\rho}), \quad y \in \mathcal{N}_{\mathfrak{D}}(x) \cap B(\hat{y}, \tilde{\delta}). \tag{4.5}$$

[4]回忆一下， $D\Xi(x)$ 是 Ξ 在某个 $x \in \mathbb{X}$处的导数映射，它是从$\mathbb{X}$ 到 $\mathbb{Y}$ 的线性映射，满足 $[D\Xi(x)](p) = \lim_{t \downarrow 0} \frac{\Xi(x+tp) - \Xi(x)}{t}$ 对所有 $p \in \mathbb{X}$ 都成立。

关系（4.4）意味着存在一个 $v_x \in \mathcal{N}_K(\Xi(x))$，使得

$$y = D\Xi(x)^* v_x. \tag{4.6}$$

回顾一下，对于任意 $z_0 \in K$，有

$$\mathcal{N}_K(z_0) \overset{\text{(a)}}{=} \{u : \langle u, z - z_0\rangle \le 0, \forall\, z \in K\} \overset{\text{(b)}}{\subseteq} \{u : \langle u, z\rangle \le 0, \forall\, z \in K\} \overset{\text{(c)}}{=} K^\circ, \tag{4.7}$$

其中，(a) 由法锥的定义得出，而 (b) 和 (c) 成立是因为 K是一个闭凸锥。特别地，当 4.7 中的 $z_0 = 0.$ 时，等式 4.7 两边处处成立。

现在，对于（ v_x 4.6）中的 v_x ，有

$$D\Xi(\hat{x})^* v_x \overset{\text{(a)}}{\in} D\Xi(\hat{x})^* \mathcal{N}_K(\Xi(x)) \overset{\text{(b)}}{\subseteq} D\Xi(\hat{x})^* K^\circ \overset{\text{(c)}}{=} D\Xi(\hat{x})^* \mathcal{N}_K(\Xi(\hat{x})) \overset{\text{(d)}}{=} \mathcal{N}_{\mathfrak{D}}(\hat{x}), \tag{4.8}$$

其中，(a) 成立是因为 $v_x \in \mathcal{N}_K(\Xi(x))$；(b) 由 $\Xi(x) \in K$ 和（4.7）得出；(c) 得益于 $\Xi(\hat{x})$ $= 0$；(d) 则由（4.4）得出。

接下来，对于（4.5）中的任意 x 和 y，由（4.6）以及 $D\Xi(x)$ 的满射性，可得

$$v_x = [D\Xi(x)D\Xi(x)^*]^{-1} D\Xi(x) y =: H(x) D\Xi(x) y, \tag{4.9}$$

其中 $H(x) := [D\Xi(x)D\Xi(x)^*]^{-1}$。

设 L 为 α-H¨连续性模数 $D\Xi(\cdot)$在集合 $x \in \mathfrak{D} \cap B(\hat{x}, \tilde{\rho})$上的值。于是，对于任意按照(4.5)选取的 x 和 y，有

$$\begin{aligned}
\|v_x - v_{\hat{x}}\| &= \|[H(x)D\Xi(x)y - H(\hat{x})D\Xi(\hat{x})\hat{y}\| \\
&\le \|H(x)D\Xi(x)y - H(\hat{x})D\Xi(\hat{x})y\| + \|H(\hat{x})D\Xi(\hat{x})(y - \hat{y})\| \\
&\le \|H(x)D\Xi(x) - H(x)D\Xi(\hat{x}) + H(x)D\Xi(\hat{x}) - H(\hat{x})D\Xi(\hat{x})\|\|y\| \\
&\quad + \|H(\hat{x})D\Xi(\hat{x})(y - \hat{y})\| \\
&\le (\|H(x)\|\|D\Xi(x) - D\Xi(\hat{x})\| + \|H(x) - H(\hat{x})\|\|D\Xi(\hat{x})\|)\|(y - \hat{y}) + \hat{y}\| \\
&\quad + \|H(\hat{x})D\Xi(\hat{x})\|\|y - \hat{y}\| \\
&\le (\tau L\|x - \hat{x}\|^\alpha + 2\tau\|D\Xi(\hat{x})\|)(\|\hat{y}\| + \|y - \hat{y}\|) + \tau\|D\Xi(\hat{x})\|\|y - \hat{y}\| \\
&\le (\tau L\tilde{\rho}^\alpha + 2\tau\|D\Xi(\hat{x})\|)(\|\hat{y}\| + \tilde{\delta}) + \tau\|D\Xi(\hat{x})\|\tilde{\delta},
\end{aligned} \tag{4.10}$$

其中，第一个等式由(4.9)得到，第四个不等式由(4.3)以及 α-H¨连续映射 $D\Xi(\cdot)$的 α-H¨连续性得到，最后一个不等式由(4.5)得到。不等式(4.10)意味着，当 v_x 以及 x 和 y 按照(4.5)选取时， $\tilde{\kappa} := \sup\|v_x\| < +\infty$有界。

对于任意根据(4.5)选取的 x 和 y，有

$$\operatorname{dist}(y, \mathcal{N}_{\mathfrak{D}}(\hat{x})) \le \|y - D\Xi(\hat{x})^* v_x\| = \|D\Xi(x)^* v_x - D\Xi(\hat{x})^* v_x\| \le L\tilde{\kappa}\|x - \hat{x}\|^\alpha,$$

其中，第一个不等式由(4.8)得到，等式由(4.6)得到，最后一个不等式由 α-H¨连续性以及 $D\Xi$ 和 $\tilde{\kappa}$的定义得到。 $\hat{}$ ˜令$\hat{\rho} := \tilde{\rho}$， $\delta := \delta$ 以及$\hat{\kappa} := L\tilde{\kappa}$，则上述推导式意味着(4.1)。 □

借助引理4.2，现在我们可以证明本小节的主要定理，该定理将用于推导下一小节中式（ d 3.6）给出的函数的 KL 指数。它可视为 [42, 定理 4。4], 的推广，该定理研究了 C^2-锥可约集。在陈述定理之前，我们先回顾一下，若一对闭凸集 $\{\mathfrak{D}_1, \mathfrak{D}_2\}$ 满足有界线性正则性（见 [7, 定义

5.6]) 在 $\hat{x} \in \mathfrak{D}_1 \cap \mathfrak{D}_2$ 处，若对于 $\hat{x}$ 的任意有界邻域 $\mathfrak{U}$，都存在 $c > 0$ 使得

$$\operatorname{dist}(x, \mathfrak{D}_1 \cap \mathfrak{D}_2) \leq c(\operatorname{dist}(x, \mathfrak{D}_1) + \operatorname{dist}(x, \mathfrak{D}_2)) \quad \text{for all } x \in \mathfrak{U}.$$

已知，当$x \in \mathfrak{D}_1 \cap \mathfrak{D}_2$ 满足 $\mathfrak{D}_1$ 和 $\mathfrak{D}_2$为多面体集，或者当 $\mathfrak{D}_1$ 为多面体集且 $\mathfrak{D}_1 \cap \operatorname{ri} \mathfrak{D}_2 \neq \emptyset$时，有界线性正则性在任意 $\hat{x}$ 处成立；参见 [8, 推论 3]。

THEOREM 4.3. 设 $\mathfrak{D} \subseteq \mathbb{X}$ 是一个非空的 C^1、α-锥可约闭凸集，且满足 $\alpha \in (0, 1]$，$\mathcal{T}: \mathbb{X} \to \mathbb{Y}$ 为线性映射，$l: \mathbb{Y} \to \mathrm{IR}$ 在任意紧凸集上均为强凸函数且具有局部利普希茨梯度，并且 $v \in \mathbb{X}$。考虑

$$h(x) := l(\mathcal{T}x) + \langle v, x \rangle + \sigma_{\mathfrak{D}}(x).$$

那么，对任意 $\bar{x} \in$ 极小点集合 h，均有

$$\bar{x} \in \mathcal{N}_{\mathfrak{D}}(\bar{w}) \text{ with } \bar{w} := -\mathcal{T}^* \nabla l(\mathcal{T}\bar{x}) - v. \tag{4.11}$$

此外，存在正常数 ρ、δ 和 κ，使得

$$\mathcal{N}_{\mathfrak{D}}(w) \cap B(\bar{x}, \delta) \subseteq \mathcal{N}_{\mathfrak{D}}(\bar{w}) + \kappa \|w - \bar{w}\|^{\alpha} B(0, 1) \quad \text{for all } w \in B(\bar{w}, \rho). \tag{4.12}$$

此外，若 $\{\mathcal{T}^{-1}\{\mathcal{T}\bar{x}\}, \mathcal{N}_{\mathfrak{D}}(-\mathcal{T}^* \nabla l(\mathcal{T}\bar{x}) - v)\}$ 在 $\bar{x}$处有界线性正则，则 h 在 $\bar{x}$ 处满足 *KL* 性质，指数为 $\frac{1}{\alpha+1}$。

证明。关系（4.11）可由与 [42, 定理 4。4] 中相同的论证得到。关系（4.12）可将引理 4.2 应用于（$\bar{x}$，$\bar{w}$），其中 $\bar{x} \in \mathcal{N}_{\mathfrak{D}}$（$\bar{w}$）。

现在证明在 $\bar{x}$ 处所声称的 KL 性质。首先证明（4.12）蕴含

$$\operatorname{dist}(x, (\partial \sigma_{\mathfrak{D}})^{-1}(\bar{w})) \leq \kappa \operatorname{dist}(\bar{w}, \partial \sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho))^{\alpha} \quad \text{for all } x \in B(\bar{x}, \delta). \tag{4.13}$$

我们的论证与 [25, 定理 3H.3] 的证明类似。具体而言，请注意，当 $\partial \sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho) = \emptyset$ 时，上述公式显然成立。现在考虑情形$\partial \sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho) \neq \emptyset$。利用 $(\partial \sigma_{\mathfrak{D}})^{-1} = \mathcal{N}_{\mathfrak{D}}$（见 [37, 例 11.4]），可得

$$w \in \partial \sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho), x \in B(\bar{x}, \delta) \Longleftrightarrow w \in B(\bar{w}, \rho), x \in \mathcal{N}_{\mathfrak{D}}(w) \cap B(\bar{x}, \delta).$$

因此，对于任意满足 x，w 的 $w \in \partial \sigma_{\mathfrak{D}}$（$x$）$\cap B$（$\bar{w}$，$\rho$）以及 $x \in B$（$\bar{x}$，δ），由上述公式和（4.12）可得

$$\operatorname{dist}(x, (\partial \sigma_{\mathfrak{D}})^{-1}(\bar{w})) \leq \kappa \|w - \bar{w}\|^{\alpha}.$$

对 $w \in \partial \sigma_{\mathfrak{D}}(x) \cap B(\bar{w}, \rho)$ 取下确界，得到 (4.13)。

由 [25, 练习 3H，4]，不等式 (4.13) 可知存在 s $\delta' \in (0, \delta]$ 使得

$$\operatorname{dist}(x, (\partial \sigma_{\mathfrak{D}})^{-1}(\bar{w})) \leq \kappa \operatorname{dist}(\bar{w}, \partial \sigma_{\mathfrak{D}}(x))^{\alpha} \quad \text{for all } x \in B(\bar{x}, \delta'). \tag{4.14}$$

现在，定义一个真闭凸函数

$$F(x) := \sigma_{\mathfrak{D}}(x) - \langle \bar{w}, x \rangle.$$

利用（2.1），可得以下等价关系

$$z \in (\partial \sigma_{\mathfrak{D}})^{-1}(\bar{w}) \Leftrightarrow \bar{w} \in \partial \sigma_{\mathfrak{D}}(z) \Leftrightarrow 0 \in \partial F(z) \Leftrightarrow z \in (\partial F)^{-1}(0). \tag{4.15}$$

利用（4.15），可将（4.14）改写为

$$\text{dist}(x,(\partial F)^{-1}(0)) \le \kappa \text{dist}(0,\partial F(x))^{\alpha} \quad \text{for all } x \in B(\bar{x},\delta'). \tag{4.16}$$

那么，对于任意 $x \in B(\bar{x},\ \delta')$，有

$$F(x)-F(\bar{x}) \le \text{dist}(0,\partial F(x))\text{dist}(x,(\partial F)^{-1}(0)) \le \kappa \text{dist}(0,\partial F(x))^{1+\alpha},$$

其中，第一个不等式由 (2.2) 得出，[5] 第二个不等式由 (4.16) 得出。上述推导表明，F 在 $\bar{x}$ 处具有 KL 性质，$\frac{1}{1+\alpha}$指数为 [12,；根据定理 5]，还可得存在 $\bar{\kappa}>0$ 使得

$$\text{dist}(x,(\partial F)^{-1}(0)) \le \bar{\kappa}(F(x)-F(\bar{x}))^{1-\frac{1}{1+\alpha}} \quad \text{for all } x \in B(\bar{x},\delta').$$

根据 F 的定义以及 (4.15)，上述推导等价地可写为

$$\begin{aligned} \text{dist}(x,\mathcal{N}_{\mathfrak{D}}(\bar{w})) &= \text{dist}(x,(\partial\sigma_{\mathfrak{D}})^{-1}(\bar{w})) = \text{dist}(x,(\partial F)^{-1}(0)) \\ &\le \bar{\kappa}(\sigma_{\mathfrak{D}}(x)-\sigma_{\mathfrak{D}}(\bar{x})-\langle \bar{w},x-\bar{x}\rangle)^{1-\frac{1}{1+\alpha}} \quad \text{for all } x \in B(\bar{x},\delta'), \end{aligned} \tag{4.17}$$

其中，第一个等式由 [37, 例 11.4] 得出。接下来，注意到对于所有 $x \in$ 极小点集合 h，由于 l 的严格凸性，我们有$\mathcal{T}x=\mathcal{T}\bar{x}$ [6] 此外

$$\begin{aligned} \text{Arg min}\, h &= \{x : 0 \in \partial h(x)\} = \{x : \mathcal{T}x=\mathcal{T}\bar{x}, x \in \mathcal{N}_{\mathfrak{D}}(-\mathcal{T}^*\nabla l(\mathcal{T}\bar{x})-v)\} \\ &= \{x : \mathcal{T}x=\mathcal{T}\bar{x}, x \in \mathcal{N}_{\mathfrak{D}}(\bar{w})\}. \end{aligned}$$

此外，对于 $\bar{x}$ 的任意有界凸邻域 $U \subseteq B(\bar{x},\ \delta')$，对于任意 $x \in U$，有

$$\begin{aligned} \|\mathcal{T}x-\mathcal{T}\bar{x}\|^{\frac{1+\alpha}{\alpha}} &= \|\mathcal{T}x-\mathcal{T}\bar{x}\|^{\frac{1}{\alpha}-1}\|\mathcal{T}x-\mathcal{T}\bar{x}\|^2 \le \left(\sup_{u\in U}\|\mathcal{T}u-\mathcal{T}\bar{x}\|^{\frac{1}{\alpha}-1}\right)\|\mathcal{T}x-\mathcal{T}\bar{x}\|^2 \\ &\le \bar{M}(l(\mathcal{T}x)-l(\mathcal{T}\bar{x})-\langle \mathcal{T}^*\nabla l(\mathcal{T}\bar{x}),x-\bar{x}\rangle), \end{aligned} \tag{4.18}$$

其中，最后一个不等式对某个 $M>0$ 成立，这归因于 l 在 U 上的强凸性，以及 $\sup_{u\in U}\|\mathcal{T}u-\mathcal{T}\bar{x}\|^{\frac{1}{\alpha}-1}<\infty$ 这一事实（由 $\alpha \in (0,1]$ 可知）。现在，利用有界线性正则条件和 Hoffman 误差界，存在$\beta_1>0$ 和 $\beta_2>0$，使得只要 $x \in U$，就有

$$\text{dist}(x,\mathcal{T}^{-1}\{\mathcal{T}\bar{x}\}\cap\mathcal{N}_{\mathfrak{D}}(\bar{w})) \le \beta_1[\text{dist}(x,\mathcal{T}^{-1}\{\mathcal{T}\bar{x}\})+\text{dist}(x,\mathcal{N}_{\mathfrak{D}}(\bar{w}))], \tag{4.19}$$

$$\text{dist}(x,\mathcal{T}^{-1}\{\mathcal{T}\bar{x}\}) \le \beta_2\|\mathcal{T}x-\mathcal{T}\bar{x}\|. \tag{4.20}$$

于是，对任意 $x \in U$

$$\begin{aligned} \text{dist}(x,\text{Arg min}\, h) = \text{dist}(x,\mathcal{T}^{-1}\{\mathcal{T}\bar{x}\}\cap\mathcal{N}_{\mathfrak{D}}(\bar{w})) &\le \beta_1[\beta_2\|\mathcal{T}x-\mathcal{T}\bar{x}\|+\text{dist}(x,\mathcal{N}_{\mathfrak{D}}(\bar{w}))] \\ &\le \beta_1[\beta_2\bar{M}^{\frac{\alpha}{\alpha+1}}(l(\mathcal{T}x)-l(\mathcal{T}\bar{x})-\langle \mathcal{T}^*\nabla l(\mathcal{T}\bar{x}),x-\bar{x}\rangle)^{1-\frac{1}{\alpha+1}} \end{aligned}$$

---

[5] $\partial F^{-1}0 \neq \emptyset$ $\bar{x}$ 4 11 4 15 $\partial\sigma_{\mathfrak{D}}^{-1}=$注意到 ( ) ( ), $\mathcal{N}_{\mathfrak{D}}$ [37, 11 4]

因为根据 (.)、(.) 及 ( ) 这一事实，它包含（见例 . ）。[6]事实上，假设对于某个 $x \in$，有 $\mathcal{T}x \neq \mathcal{T}\bar{x}$ Argmin h。那么我们得到以下矛盾：inf $h=(h(x)+h(\bar{x}))/2>l((\mathcal{T}x+\mathcal{T}\bar{x})/2)+\langle v,\ (x+\bar{x})/2\rangle+\sigma_{\mathfrak{D}}((x+\bar{x})/2) \ge \inf\ h,$

其中，第一个不等式由 l 的严格凸性、$\langle v,\cdot\rangle+\sigma_{\mathfrak{D}}(\cdot)$ 的凸性以及 $\mathcal{T}x \neq \mathcal{T}\bar{x}$ 这一假设得出；第二个不等式则由 $(x+\bar{x})/2 \in$ dom h as h 是凸集这一事实得出。

$$+\bar{\kappa}(\sigma_{\mathfrak{D}}(x)-\sigma_{\mathfrak{D}}(\bar{x})-\langle\bar{w},x-\bar{x}\rangle)^{1-\frac{1}{\alpha+1}}], \tag{4.21}$$

其中，第一个不等式由（4.19）-（4.20）推出，最后一个不等式由（4.17）和（4.18）推出。最后，注意到

$$h(x)-h(\bar{x})=l(\mathcal{T}x)-l(\mathcal{T}\bar{x})-\langle\mathcal{T}^*\nabla l(\mathcal{T}\bar{x}),x-\bar{x}\rangle+\sigma_{\mathfrak{D}}(x)-\sigma_{\mathfrak{D}}(\bar{x})-\langle\bar{w},x-\bar{x}\rangle. \tag{4.22}$$

结合 4.21）-（4.22）以及不等式 $a^p+b^p\le 2^{1-p}(a+b)^p$ 对任意 a，$b\ge 0$以及 $p\in(0,1]$, 我们可以推得存在正常数 c 使得

$$\operatorname{dist}(x,\operatorname{Arg\,min}h)\le c(h(x)-h(\bar{x}))^{1-\frac{1}{\alpha+1}}\quad\text{for all } x\in U.$$

调用 [12,定理 5]即可得到所需结果。 □

直接根据定义验证 $C^{1,\alpha}$-锥可约性可能并不容易。以下命题有助于判断一个集合是否为 $C^{1,\alpha}$-锥可约的。这是 [38, 命题 3.2], 的类比，该命题研究了 C^2-锥可约集。

PROPOSITION 4.4. 设 $V:=\{x:\ G(x)\ \in K\}$，其中 $G:\ \mathbb{X}\to\mathbb{Y}$ 连续可微，且其导数局部 α-*H¨older* 连续，其中某个 $\alpha\in(0,1]$，，并且$K\subseteq\mathbb{Y}$ 是一个闭凸集。若 K 在 $y_0=G(x_0)$ 处是 $C^{1,\alpha}$-锥可约的，并且

$$DG(x_0)\mathbb{X}+\operatorname{lin}T_K(y_0)=\mathbb{Y},$$

其中 $T_K(y_0)$ 是 K 在 y_0 处的切锥，而 $\operatorname{lin}T_K(y_0):=T_K(y_0)\cap -T_K(y_0)$。那么 V 在 x_0处是 $C^{1,\alpha}$-锥可约的。

证明。 证明沿用 [38, 命题 3.2]中的相同论证。 □

当 $\mathbb{X}=\mathrm{IR}^n$、 $\mathbb{Y}=\mathrm{IR}^m$、 $K=-\mathrm{IR}^m_+$，且 $G(x)=(g_1(x)\ldots,g_m(x))$ 中每个 g_i均连续可微，且其导数局部 α-H¨older 连续时，由于$K=-\mathrm{IR}^m_+$ 是 C^2-锥可约的，我们根据命题 4.4 推得集合

$$V:=\{x\in\mathrm{IR}^n:g_i(x)\le 0,i=1,\ldots,m\}$$

若 $\{\nabla g_i(x_0):\ i\in\mathcal{I}\}$ 线性无关，则该集合在 x_0 处是 $C^{1,\alpha}$-锥可约的，其中$\mathcal{I}:=\{i:\ g_i(x_0)=0\}$。作为运用命题 4.4的更多具体例，下面证明，对于某个 $\alpha\in(0,1]$， p-锥以及带有 $p\in(1,\infty)$ 的 p-范数球均是 $C^{1,\alpha}$-锥可约的。回顾一下， p-锥 $\mathcal{K}_p^{n+1}\subseteq\mathrm{IR}^{n+1}$ 定义为

$$\mathcal{K}_p^{n+1}:=\{(x,r)\in\mathrm{IR}^n\times\mathrm{IR}:r\ge\|x\|_p\}.$$

EXAMPLE 4.1（$\mathcal{K}_p^{n+1}$, $p\in(1,\infty)$是 $C^{1,\alpha}$-锥可约的）。注意， $\mathcal{K}_p^{n+1}=\{(x,r)\in\mathrm{IR}^n\times\mathrm{IR}:\ G(x,r)\le 0\}$ with $G(x,r):=\|x\|_p-r$。正如 [38],所指出的，任意闭凸锥在原点及其相对内部处都是 C^2-锥可约的，相应的 Ξ 可以选为线性的。因此，我们只需证明 $\mathcal{K}_p^{n+1}$在每个非零边界点处都是$C^{1,\alpha}$-锥可约的。注意，当 $x\ne 0$时，对于任意满足 $r\ge\|x\|_p$ 的(x,r)，有

$$\nabla G(x,r)=\|x\|_p^{1-p}\cdot\begin{bmatrix}\operatorname{sgn}(x_1)|x_1|^{p-1} & \cdots & \operatorname{sgn}(x_n)|x_n|^{p-1} & -\|x\|_p^{p-1}\end{bmatrix}^T.$$

因此，对于边界上的任意（x,r）$\ne 0$，有 $r=\|x\|_p$且 ∇G（$x,\|x\|_p$）$\ne 0$；因此，｛∇G（x, $\|x\|_p$）｝线性无关。

根据本例前面的讨论，现在只需证明，对于 $\mathcal{K}_p^{n+1}$ 的任意非零边界点， ∇G 在该点局部 $H¨older$ 连续。利用上面的公式，不难验证，当 $p\in(1,2)$ 时， ∇G 局部 $(p-1)$-$H¨older$ 连续；当 $p\in[2,\infty)$ 时， ∇G 局部利普希茨连续（因此也是 1-$H¨older$ 连续）。因此，对于任意满足 $\alpha=\min\{1,p-1\}$ 的 $p\in(1,+\infty)$，$\mathcal{K}_p^{n+1}$ 都是 $C^{1,\alpha}$-锥可约的。

EXAMPLE 4.2（ p-范数球 $p\in(1,\infty)$ 是 $C^{1,\alpha}$-锥可约的）。与例4.1中的讨论类似，可以证明， p-范数球 $\{x\in\mathrm{IR}^n:\ \|x-x_0\|_p\le\beta\}$ （中心为 x_0，半径为 $\beta>0$）是 $C^{1,\alpha}$-锥可约的，其中 $\alpha=\min\{1,p-1\}$。

4.2. 函数的 KL 性质 d 在 (3.6)中。在本小节中，我们证明，在对 (1.2) 作出适当假设的条件下，函数 d 在 (3.6) 中满足 KL 性质，且具有显式的 KL 指数。具体而言，我们考虑以下假设。

ASSUMPTION 4.1. 考虑 (1.2)。假设以下条件成立。

(i) 每个 C_i 都是 $C^{1,\alpha}$-锥可约闭凸集，且满足 $\alpha\in(0,1]$; [7](ii) $\bigcap_{i=1}^{\ell}A_i^{-1}\,\mathrm{ri}\,C_i\neq\emptyset$; (iii) $0\in x^*-\bar v+\mathrm{ri}\,\partial(\sum_{i=1}^{\ell}\delta_{A_i^{-1}C_i})(x^*)$，其中 x^* 是 (1.2)的唯一解。

以下命题给出了函数 d 在 (3.6) 中的 KL 指数。

PROPOSITION 4.5（ d的 KL 指数）。考虑(1.2)。假设4.1 成立。则 d 中(3.6) 的函数是 *KL* 函数，其 *KL* 指数为 $\frac{1}{\alpha+1}$。

证明。任取 $\boldsymbol{y}^*=(y_1^*,\dots,y_\ell^*)\in$极小点集合 d。于是命题3.1 推出

$$x^*=\bar v-\boldsymbol{A}^T\boldsymbol{y}^*, \tag{4.23}$$

其中 x^* 是方程（1.2）的唯一解。

由式（1.2）可知 $f(x)=\frac12\|x-\bar v\|^2$。我们首先证明

$$\{(\boldsymbol{A}^T)^{-1}\{\boldsymbol{A}^T\boldsymbol{y}^*\},\mathcal{N}_D(-\boldsymbol{A}\nabla f(\boldsymbol{A}^T\boldsymbol{y}^*))\} \tag{4.24}$$

在 $\boldsymbol{y}^*$ 处有界线性正则。为此，我们首先注意到

$$(\boldsymbol{A}^T)^{-1}\{\boldsymbol{A}^T\boldsymbol{y}^*\}=(y_1^*,\dots,y_\ell^*)+\ker\boldsymbol{A}^T. \tag{4.25}$$

此外，利用式（4.23）并注意到 $\nabla f(\boldsymbol{A}^T\boldsymbol{y}^*)=\boldsymbol{A}^T\boldsymbol{y}^*-\bar v$，可得

$$\mathcal{N}_D(-\boldsymbol{A}\nabla f(\boldsymbol{A}^T\boldsymbol{y}^*))=\mathcal{N}_D(\boldsymbol{A}x^*)=\mathcal{N}_{C_1}(A_1x^*)\times\cdots\times\mathcal{N}_{C_\ell}(A_\ell x^*). \tag{4.26}$$

此外，根据假设 4.1(iii)，我们推导出

$$\bar v-x^*\in\mathrm{ri}\,\partial\left(\sum_{i=1}^{\ell}\delta_{A_i^{-1}C_i}\right)(x^*)=\mathrm{ri}\,\left(\sum_{i=1}^{\ell}A_i^T\mathcal{N}_{C_i}(A_ix^*)\right)=\sum_{i=1}^{\ell}A_i^T\,\mathrm{ri}\,\mathcal{N}_{C_i}(A_ix^*),$$

其中，第一个等式由假设4.1(ii) 以及 [36], 中的定理 23.8 和 23.9 得出，而第二个等式由 [36, 定理 6 得出。6]结合上述推导式与 (4.26)，我们得出存在

$$\boldsymbol{u}^*=(u_1^*,\dots,u_\ell^*)\in\mathrm{ri}\,\mathcal{N}_{C_1}(A_1x^*)\times\cdots\times\mathrm{ri}\,\mathcal{N}_{C_\ell}(A_\ell x^*)=\mathrm{ri}\,\mathcal{N}_D(-\boldsymbol{A}\nabla f(\boldsymbol{A}^T\boldsymbol{y}^*)) \tag{4.27}$$

使得 $x^*-\bar v+\boldsymbol{A}^T\boldsymbol{u}^*=0$。此最后一个等式与 (4.23) 一起表明$\boldsymbol{u}^*-\boldsymbol{y}^*\in$核 $\boldsymbol{A}^T$。因此，

$$\boldsymbol{u}^*=\boldsymbol{y}^*+(\boldsymbol{u}^*-\boldsymbol{y}^*)\in\boldsymbol{y}^*+\ker\boldsymbol{A}^T. \tag{4.28}$$

利用 (4.25)、(4.27) 和 (4.28)，我们看到 $\boldsymbol{u}^*\in(\boldsymbol{A}^T)^{-1}\{\boldsymbol{A}^T\boldsymbol{y}^*\}\cap\mathrm{ri}\,\mathcal{N}_D(-\boldsymbol{A}\nabla f(\boldsymbol{A}^T\boldsymbol{y}^*))$，这意味着

$$(\boldsymbol{A}^T)^{-1}\{\boldsymbol{A}^T\boldsymbol{y}^*\}\cap\mathrm{ri}\,\mathcal{N}_D(-\boldsymbol{A}\nabla f(\boldsymbol{A}^T\boldsymbol{y}^*))\neq\emptyset.$$

[7]由此，可以直接根据定义验证，对于相同的 α， D 是 $C^{1,\alpha}$-锥可约的。

这与 $(\boldsymbol{A}^T)^{-1}\{\boldsymbol{A}^T\boldsymbol{y}^*\}$ 的多面性共同保证了 (4.24) 在 $\boldsymbol{y}^*$ (见 [8, 推论 3]) 处具有有界线性正则性。进而，由假设 4.1(i) 和定理 4.3 可推得 d 在 $\boldsymbol{y}^*$处满足指数为 $\frac{1}{\alpha+1}$ 的 KL 性质。由于 $\boldsymbol{y}^* \in \operatorname{Arg\,min} d$ 的任意性以及 [31, 引理 21]，所需结论即得证。 □

下一个定理证明了函数 d 满足一致化增长条件和 KL 性质，这有助于在下一节推导 Luo-Tseng 型误差界。我们指出，[13, 引理 6]中关于一致化KL性质的已有结果，讨论的是在紧集上的 KL 性质。由于集合 极小点集合 d 一般可能无界，因此无法直接应用这些已有结果。相反，我们通过注意到沿着 d 移动时保持不变，建立了一致性结果 $E^\perp$。这一点正式表述为以下辅助引理，它直接推论自 [5] 中的推论2.5.5 和定理2.5.3，因为 $d = q^*$ （见命题3.3）。

LEMMA 4.6. 考虑(1.2)。设函数 d 由 (3.6)给出，且 E 由 (3.14)给出。假设条件 *4.1(ii)* 成立。那么对于任意 $\boldsymbol{y} \in \operatorname{dom} d$，当 $d(\boldsymbol{y}+\boldsymbol{u}) = d(\boldsymbol{y})$ 时， $\boldsymbol{u} \in E^\perp$。

THEOREM 4.7（一致化KL性质与增长条件）。考虑(1.2)。设函数 d 由(3.6)给出。假设*4.1*成立。那么存在正常数 ϵ 和 c ，使得

$$\operatorname{dist}(\boldsymbol{y}, \operatorname{Arg\,min} d) \le c\,(d(\boldsymbol{y}) - d^*)^{1-\frac{1}{1+\alpha}} \tag{4.29}$$

$$\text{and} \qquad (d(\boldsymbol{y}) - d^*)^{\frac{1}{1+\alpha}} \le c \operatorname{dist}(\boldsymbol{0}, \partial d(\boldsymbol{y})) \tag{4.30}$$

只要 $\boldsymbol{y} \in Y_\epsilon := \{\boldsymbol{y} : \operatorname{dist}(\boldsymbol{y},$ 极小点集合 $d) \le \epsilon,\ d^* \le d(\boldsymbol{y}) \le d^* + \epsilon\}$。

证明。在这些假设下，由命题4.5可知， d 是一个指数为$\frac{1}{1+\alpha}$的KL函数。在 [4, 引理2.1(a)], 中已证明

$$\operatorname{dom} q_E = \operatorname{dom} q + E^\perp \text{ with } \operatorname{int}(\operatorname{dom} q_E) = \operatorname{ri} \operatorname{dom} q + E^\perp,$$

其中 q_E和 q 在命题3.3中定义。现在，注意到 $\partial q_E(\boldsymbol{0})$(⊆ 极小点集合 d)非空且紧，因为根据上面的等式以及 $\boldsymbol{0} \in \operatorname{int}(\operatorname{dom} q_E$这一事实，我们有 $\boldsymbol{0} \in \operatorname{ri} \operatorname{dom} q$ (参见命题3.3)。于是，利用 [13, 引理 6]和 [12, 定理 5], ，存在正常数 ϵ 和 c ，使得

$$\operatorname{dist}(\boldsymbol{z}, \operatorname{Arg\,min} d) \le c(d(\boldsymbol{z}) - d^*)^{1-\frac{1}{1+\alpha}}, \tag{4.31}$$

¯当 $\boldsymbol{z} \in Y_\epsilon := \{\boldsymbol{z} : \operatorname{dist}(\boldsymbol{z}, \partial q_E(\boldsymbol{0})) \le \epsilon, d^* \le d(\boldsymbol{z}) \le d^* + \epsilon\}$时。接下来，我们证明

$$Y_\epsilon = \bar{Y}_\epsilon + E^\perp. \tag{4.32}$$

首先，回顾（3.13）。对于任意 $\boldsymbol{y} \in Y_\epsilon$，令 $\operatorname{Proj}_{\operatorname{Arg\,min} d}(\boldsymbol{y}) = \hat{\boldsymbol{z}} + \hat{\boldsymbol{u}}$，其中 $\hat{\boldsymbol{z}} \in \partial q_E(\boldsymbol{0})$ 且 $\hat{\boldsymbol{u}} \in E^\perp$。于是

$$\operatorname{dist}(\boldsymbol{y} - \hat{\boldsymbol{u}}, \partial q_E(\boldsymbol{0})) \le \|(\boldsymbol{y} - \hat{\boldsymbol{u}}) - \hat{\boldsymbol{z}}\| = \|\boldsymbol{y} - \operatorname{Proj}_{\operatorname{Arg\,min} d}(\boldsymbol{y})\| \le \epsilon;$$

此外，得益于引理 4.6，我们还有 $d(\boldsymbol{y} - \hat{\boldsymbol{u}}) = d(\boldsymbol{y})$。由此可得 $\boldsymbol{y} - \hat{\boldsymbol{u}} \in \bar{Y}_\epsilon$。于是 $\boldsymbol{y} = (\boldsymbol{y} - \hat{\boldsymbol{u}}) + \hat{\boldsymbol{u}} \in \bar{Y}_\epsilon + E^\perp$，这意味着 $Y_\epsilon \subseteq \bar{Y}_\epsilon + E^\perp$。为证明反向包含，任取满足 $\boldsymbol{z} \in \bar{Y}_\epsilon$ 和 $\boldsymbol{u} \in E^\perp$ 的 $\boldsymbol{z} + \boldsymbol{u} \in \bar{Y}_\epsilon + E^\perp$。根据引理 4.6，有 $d(\boldsymbol{z} + \boldsymbol{u}) = d(\boldsymbol{z})$，并且

$$\operatorname{dist}(\boldsymbol{z} + \boldsymbol{u}, \operatorname{Arg\,min} d) \le \operatorname{dist}(\boldsymbol{z} + \boldsymbol{u}, \partial q_E(\boldsymbol{0}) + \boldsymbol{u}) = \operatorname{dist}(\boldsymbol{z}, \partial q_E(\boldsymbol{0})) \le \epsilon,$$

其中，第一个不等式由 $\partial q_E(\mathbf{0}) + \boldsymbol{u} \subseteq$ 极小点集合 d (参见 (3.13))。于是，$\bar{Y}_\epsilon + E^\perp \subseteq Y_\epsilon$ 且关系 (4.32) 成立。

现在，对于任意 $\boldsymbol{y} \in Y_\epsilon$，根据（4.32），存在 $\boldsymbol{z} \in \bar{Y}_\epsilon$ 和 $\boldsymbol{u} \in E^\perp$，使得 $\boldsymbol{y} = \boldsymbol{z} + \boldsymbol{u}$。因此，对于任意这样的 $\boldsymbol{y}$,

$$\begin{aligned}\operatorname{dist}(\boldsymbol{y}, \operatorname{Arg\,min} d) &\overset{\text{(a)}}{=} \operatorname{dist}(\boldsymbol{z}+\boldsymbol{u}, (\operatorname{Arg\,min} d) + \boldsymbol{u}) = \operatorname{dist}(\boldsymbol{z}, \operatorname{Arg\,min} d) \\ &\overset{\text{(b)}}{\leq} c(d(\boldsymbol{z}) - d^*)^{1-\frac{1}{1+\alpha}} \overset{\text{(c)}}{=} c(d(\boldsymbol{y}) - d^*)^{1-\frac{1}{1+\alpha}},\end{aligned} \tag{4.33}$$

其中，(a) 是由极小点集合 $d =$（极小点集合 d）$+\boldsymbol{u}$ 得出的（这是因为极小点集合 $d = \partial q_E(\mathbf{0}) + E^\perp$ 且 $\boldsymbol{u} \in E^\perp$），(b) 是由（4.31）得出的，(c) 则由引理 4.6得出。由此证明了（4.29）。最后，（4.29）与（2.2）共同推出（4.30）。 □

在结束本节之前，我们通过给出 BA-MSF（1.2）的一个实例来说明，在假设 4.1下，（4.29）中的指数是“最优的”：存在一条路径，使得（4.29）两边以相同的数量级趋于零。证明该指数紧性的下例，与近期关于紧误差界的研究一致：参见文献 [40, 定义 6] 中广义凹 KL 性质的精确模数，以及文献 [32, 定义 31]。

EXAMPLE 4.3（式（4.29）中指数的紧性）。考虑

$$\min_{x \in \mathbb{R}^2} \frac{1}{2}\|x - \bar{v}\|^2 \quad \text{s.t.} \quad A_1 x \in C_1,$$

其中 $\bar{v} = (2, 0)$，$p \in (1, 2)$，$C_1 = \{x \in \mathbb{R}^2 : \|x\|_p \leq 1\}$，且 $A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$。由于 $A_1^{-1} C_1 = \{x \in \mathbb{R}^2 : A_1 x \in C_1\} = [-1, 1] \times \mathbb{R}$，因此 $x^* = (1, 0)$ 是唯一解。此外，根据例 4.2，C_1 是 $C^{1,\alpha}$-锥可约的，且 $\alpha = p - 1$；并且有 $A_1^{-1}\operatorname{ri} C_1 \neq \emptyset$，因为 $(0, 0) \in (\operatorname{ri} C_1) \cap \operatorname{Range}(A_1)$。此外，

$$\begin{bmatrix} 1 & 0 \end{bmatrix}^T = \bar{v} - x^* \in \left\{ \begin{bmatrix} t & 0 \end{bmatrix}^T : t > 0 \right\} = \operatorname{ri} \mathcal{N}_{[-1,1]\times\mathbb{R}}(x^*) = \operatorname{ri} \mathcal{N}_{A_1^{-1}C_1}(x^*).$$

因此，假设 4.1 得以满足。于是，定理 4.7 表明，（4.29）以指数 $1 - \frac{1}{1+\alpha} = 1 - \frac{1}{p} = \frac{p-1}{p}$ 成立，其中式 d （4.29）中的现在取如下形式

$$d(y_1) = (1/2)\|A_1 y_1 - \bar{v}\|^2 - (1/2)\|\bar{v}\|^2 + \|y_1\|_{\frac{p}{p-1}}.$$

接下来，根据命题 3.1，我们有 $\begin{bmatrix} 1 & 0 \end{bmatrix}^T = \bar{v} - x^* = A_1 \hat{y}_1$ 当 $\hat{y}_1 \in$极小点集合 d时。这意味着 $\hat{y}_1$ 的第一个坐标为 1。此外，根据 d 的定义可知，$\hat{y}_1$ 的第二个坐标为 0。因此，极小点集合 $d = \{\begin{bmatrix} 1 & 0 \end{bmatrix}^T\}$。

现在，令 $y_1^\epsilon = (1, \epsilon)$ 对于 $\epsilon \downarrow 0$。通过直接计算，我们得到 $\epsilon \downarrow 0$，即

$$d(y_1^\epsilon) - d(\hat{y}_1) = (1 + \epsilon^{\frac{p}{p-1}})^{\frac{p-1}{p}} - 1 = \Theta(\epsilon^{\frac{p}{p-1}}) \quad \textit{and} \quad \operatorname{dist}(y_1^\epsilon, \operatorname{Arg\,min} d) = \Theta(\epsilon),$$

这表明，沿着 (4.29) 两边以相同的数量级趋于零 y_1^ϵ。

备注 4.8. 当假设4.1 中的 $\alpha = 1$ 满足条件 (i) 时（若每个 C_i 都是 C^2-锥可约的，则后者成立），定理4.7 断言，(3.6) 中给出的函数 d 满足二次增长条件；也就是说，(4.29) 以指数$1 - \frac{1}{1+\alpha} = \frac{1}{2}$成立。然而，对于例4.2 中满足 $p \in (1, 2)$ 的 p-范数球，根据例4.3，d 以紧确指数 $1 - \frac{1}{\alpha+1} = \frac{p-1}{p} < \frac{1}{2}$ 满足 (4.29)。因此，例4.2 中满足 $p \in (1, 2)$ 的p-范数球是 $C^{1,\alpha}$-锥可约的，其中 $\alpha = p - 1 \in (0, 1)$，但并非 C^2-锥可约。

5. 收敛速率分析。 本节研究算法 3.1 的收敛速率。 根据定理 2.2 可知，当每个 C_i 都是多面体时，（经典）Dykstra 投影算法已知具有线性收敛性。我们将论证，算法 3.1 也有相同的结论。证明技巧将揭示如何进一步发展第 4 节中的工具，以分析算法 3.1 在更一般的集合类上的收敛速率。

下面的分析依赖于以下映射 $\mathcal{G}:\mathrm{IR}^m \to \mathrm{IR}^m$,

$$\mathcal{G}(\boldsymbol{y}) := \boldsymbol{y} - \mathrm{Prox}_{\sigma_D}(\boldsymbol{y} - \nabla g(\boldsymbol{y})), \tag{5.1}$$

其中 g 和 D 分别由式（3.6）和（3.1）给出。请注意， $\mathcal{G}$ 式（5.1）中的条件对于 [33, 39] 中一阶方法的收敛速率分析框架至关重要。

THEOREM 5.1（多面体 C_i下的线性收敛）。考虑（1.2），并令函数 d 由式（3.6）给出。假设每个 C_i 都是多面体的。那么 $\mathit{Arg\ min}\, d \neq \emptyset$。此外，若 $\{x^t\}, \{\boldsymbol{y}^t\}$ 是由算法 3.1 生成的序列，则存在$\boldsymbol{y}^* \in \mathit{Arg\ min}\, d$, $r \in (0,1)$, $a_0 \in (0,1)$, $a_1 > 0$ 以及正整数 $\bar{t}$，使得

$$\|x^t - x^*\| \le a_1 a_0^t, \quad d(\boldsymbol{y}^{t+1}) - d^* \le r(d(\boldsymbol{y}^t) - d^*), \quad \|\boldsymbol{y}^t - \boldsymbol{y}^*\| \le a_1 a_0^t, \quad \forall\, t \ge \bar{t},$$

其中 x^* 是（1.2）的唯一解，且 d^* 由 （3.6）给出。

证明。 首先，由 [37, 命题 8.29] 可知 σ_{C_i} 对每个 i 都是分段线性函数，因此函数 d 是分段线性二次函数。由于 d 根据命题 3.1有下界，利用 Frank-Wolfe 定理（见 [23, 定理 2.81]），可知极小点集合 $\mathrm{Arg\ min}\, d \neq \emptyset$。结合 [39,定理 4]，，我们断言满足以下一阶误差界条件[8]： 存在正常数 ϵ_0 和 c_0，使得

$$\mathrm{dist}(\boldsymbol{y}, \mathrm{Arg\,min}\, d) \le c_0 \|\mathcal{G}(\boldsymbol{y})\| \quad \text{whenever } \|\mathcal{G}(\boldsymbol{y})\| \le \epsilon_0 \text{ and } d(\boldsymbol{y}) \le d^* + \epsilon_0, \tag{5.2}$$

其中 $\mathcal{G}$ 由（5.1）给出。于是， $\{d(\boldsymbol{y}^{t+1})\}$ 和 $\{\boldsymbol{y}^t\}$的线性收敛结果可由 [39, 定理 2(b)]得出。[9] 此外，由（3.10）和命题 3.1，可得

$$\|x^t - x^*\| = \left\|\bar{v} - \textstyle\sum_{j=1}^{\ell} A_j^T y_j^t - \left(\bar{v} - \sum_{j=1}^{\ell} A_j^T y_j^*\right)\right\| \le \|\boldsymbol{A}^T\| \|\boldsymbol{y}^t - \boldsymbol{y}^*\|. \tag{5.3}$$

$\{x^t\}$ 于是立即得到所声称的收敛结果。 □

请注意，在上述证明中，关键要素是一阶误差界条件（5.2）；众所周知， [26, 第 3] 节指出，该条件本质上与二阶增长条件相关。下面，我们将借助第 4 节对 $C^{1,\alpha}$-锥可约集的增长条件的研究，构造（5.2）的一个类似条件（在 $\|\mathcal{G}(\boldsymbol{y})\|$ 上采用一般指数），以分析当每个 C_i 对某个 $\alpha \in (0,1]$）是 $C^{1,\alpha}$-锥可约时，算法 3.1 的收敛速率。

[8]在文献中，这一条件也称为 Luo-Tseng 误差界；参见 [31, 33, 39, 43]。[9]需要说明的是，近端 CGD（3.7）和（3.9）是 [39] 中所研究算法的一个特例，因此 [39, 定理 2(b)] 适用。具体而言，由命题 3.2(i) 可知，若在 [39] 中使用的 Armijo 规则中选取 $\sigma = 0.5$、$\gamma = 0$ 和 $\alpha_{\mathrm{init}}^k \equiv 1$，并考虑搜索方向 $y_i^{t+1} - y_i^t$，则步长取 1 时 Armijo 规则成立，因此 y_i^{t+1} 会出现在下一个迭代点中。由此，近端 CGD（3.7）和（3.9）与 [39] 中考虑的算法一致，其中采用上述对 σ、γ 和 $\{\alpha_{\mathrm{init}}^k\}$ 的选取。

5.1. $C^{1,\alpha}$**-锥可约** C_i**的收敛速率分析。我们首先基于第** 4节对增长条件的研究，推导出一个 Luo-Tseng 型误差界。

LEMMA 5.2（Luo-Tseng 型误差界）。考虑(1.2)。设 d^* 和函数 d 分别由(3.6)以及 $\mathcal{G}$ 由(5.1)定义。假设*4.1*成立。那么存在 $c>0$ 和 $\epsilon>0$，使得

$$\operatorname{dist}(\boldsymbol{y}, \operatorname{Arg\,min} d) \le c\|\mathcal{G}(\boldsymbol{y})\|^{\alpha} \quad \textit{whenever } \|\mathcal{G}(\boldsymbol{y})\| \le \epsilon,\ d^* \le d(\boldsymbol{y}) \le d^* + \epsilon.$$

证明。由定理4.7，存在 $\epsilon \in (0, 1/2)$ 和 $c>0$，使得对任意 $\boldsymbol{y} \in Y_\epsilon$

$$\text{dist}(\boldsymbol{y},\ \text{极小点集合 } d) \le c(d(\boldsymbol{y}) - d^*)^{1-\frac{1}{1+\alpha}} \le c\,\text{dist}(\boldsymbol{0},\ \partial d(\boldsymbol{y}))^{\frac{\alpha}{1+\alpha}} \le \text{dist}(\boldsymbol{y},\ \text{极小点集合 } d)^{\frac{\alpha}{1+\alpha}},$$

其中第二个不等式由（2.2）得出。这进一步意味着

$$\operatorname{dist}(\boldsymbol{y}, \operatorname{Arg\,min} d) \le c_1 \operatorname{dist}(\boldsymbol{0}, \partial d(\boldsymbol{y}))^{\alpha} \quad \text{for all } \boldsymbol{y} \in Y_\epsilon \tag{5.4}$$

其中 $c_1 := c^{1+\alpha}$。接下来，对于任意 $\boldsymbol{y} \in Y_\epsilon$，有

$$\begin{cases} d^* \le d(\operatorname{prox}_d(\boldsymbol{y})) \le d(\operatorname{prox}_d(\boldsymbol{y})) + \dfrac{1}{2}\|\boldsymbol{y} - \operatorname{prox}_d(\boldsymbol{y})\|^2 \\ \qquad\qquad = \inf\limits_{\boldsymbol{u}\in\mathbb{R}^m}\{d(\boldsymbol{u}) + \dfrac{1}{2}\|\boldsymbol{y}-\boldsymbol{u}\|^2\} \le d(\boldsymbol{y}) \le d^* + \epsilon, \\ \operatorname{dist}(\operatorname{prox}_d(\boldsymbol{y}), \operatorname{Arg\,min} d) \le \|\operatorname{prox}_d(\boldsymbol{y}) - \bar{\boldsymbol{y}}\| = \|\operatorname{prox}_d(\boldsymbol{y}) - \operatorname{prox}_d(\bar{\boldsymbol{y}})\| \\ \qquad\qquad \overset{\text{(a)}}{\le} \|\boldsymbol{y} - \bar{\boldsymbol{y}}\| = \operatorname{dist}(\boldsymbol{y}, \operatorname{Arg\,min} d) \le \epsilon, \end{cases}$$

其中 $\bar{\boldsymbol{y}} := \text{Proj}_{\text{Arg 极小值 } d}(\boldsymbol{y})$，(a) 成立是因为近端算子是非扩张的。上述公式意味着 $\text{prox}_d(\boldsymbol{y}) \in Y_\epsilon$。类似地，对于任意 $\boldsymbol{y} \in Y_\epsilon$，有

$$\|\boldsymbol{y} - \operatorname{prox}_d(\boldsymbol{y})\| \le \|\boldsymbol{y} - \bar{\boldsymbol{y}}\| + \|\operatorname{prox}_d(\boldsymbol{y}) - \bar{\boldsymbol{y}}\| \le 2\|\boldsymbol{y} - \bar{\boldsymbol{y}}\| \le 2\epsilon < 1. \tag{5.5}$$

于是我们推得，对于任意 $\boldsymbol{y} \in Y_\epsilon$，有

$$\begin{aligned} &\text{dist}(\boldsymbol{y},\ \text{极小点集合 } d) \le \|\boldsymbol{y} - \text{prox}_d(\boldsymbol{y})\| + \text{dist}(\text{prox}_d(\boldsymbol{y}),\ \text{极小点集合 } d) \\ &\overset{\text{(a)}}{\le} \|\boldsymbol{y} - \operatorname{prox}_d(\boldsymbol{y})\| + c_1 \operatorname{dist}(\boldsymbol{0}, \partial d(\operatorname{prox}_d(\boldsymbol{y})))^{\alpha} \overset{\text{(b)}}{\le} \|\boldsymbol{y} - \operatorname{prox}_d(\boldsymbol{y})\| + c_1\|\boldsymbol{y} - \operatorname{prox}_d(\boldsymbol{y})\|^{\alpha} \\ &\overset{\text{(c)}}{\le} (1+c_1)\|\boldsymbol{y} - \operatorname{prox}_d(\boldsymbol{y})\|^{\alpha} \overset{\text{(d)}}{\le} (1+c_1)c_2\|\mathcal{G}(\boldsymbol{y})\|^{\alpha}, \end{aligned}$$

其中，(a) 由 (5.4) 以及 $\text{prox}_d(\boldsymbol{y}) \in Y_\epsilon$ 这一事实推出；(b) 由 $\boldsymbol{y} - \text{prox}_d(\boldsymbol{y}) \in \partial d(\text{prox}_d(\boldsymbol{y}))$ 这一事实推出；(c) 成立是因为 $\|\boldsymbol{y} - \text{prox}_d(\boldsymbol{y})\| < 1$（由 (5.5) 得出）且 $\alpha \in (0, 1]$，，而 (d) 根据定理 3.5], 对某个 $c_2 > 0$ 成立。

现在可以得出结论：存在 $c_3 > 0$，使得只要 $\boldsymbol{y} \in Y_\epsilon$，就有 $\text{dist}(\boldsymbol{y},\ \text{极小点集合 } d) \le c_3\|\mathcal{G}(\boldsymbol{y})\|^{\alpha}$。为完成证明，只需证明存在 $\nu \in (0,\ \epsilon)$，使得

$$\{\boldsymbol{y} : \|\mathcal{G}(\boldsymbol{y})\| \le \nu\} \subseteq \{\boldsymbol{y} : \operatorname{dist}(\boldsymbol{y}, \operatorname{Arg\,min} d) \le \epsilon\}.$$

反设结论不成立。于是存在 $\{\boldsymbol{z}^t\}$，使得对所有 t 和 $\mathcal{G}(\boldsymbol{z}^t) \to \boldsymbol{0}$，都有 $\text{dist}(\boldsymbol{z}^t,$ 极小点集合 $d) > \epsilon$。根据后一个极限以及[26, 定理 3.5]，，可知 $\|\boldsymbol{z}^t - \text{prox}_d(\boldsymbol{z}^t)\| \to 0$。由于 $\boldsymbol{z}^t - \text{prox}_d(\boldsymbol{z}^t) \in \partial d(\text{prox}_d(\boldsymbol{z}^t))$，我们进一步推出 $\text{dist}(0, \partial d(\text{prox}_d(\boldsymbol{z}^t))) \to 0$，因此根据命题 3.3，有 $\text{dist}(\text{prox}_d(\boldsymbol{z}^t),$ 极小点集合 $d) \to 0$。由此得到如下矛盾：

$$\epsilon < \operatorname{dist}(\boldsymbol{z}^t, \operatorname{Arg\,min} d) \le \|\boldsymbol{z}^t - \operatorname{prox}_d(\boldsymbol{z}^t)\| + \operatorname{dist}(\operatorname{prox}_d(\boldsymbol{z}^t), \operatorname{Arg\,min} d) \to 0.$$

至此，证明完成。 □

下面的定理给出了算法3.1的收敛速率。推导 $\|\mathcal{G}(\boldsymbol{y}^t)\|$ （参见（5.12））和 $d(\boldsymbol{y}^{t+1}) - d^*$ （参见（5.13））上界的步骤与 [39, 定理 2]的证明类似。

THEOREM 5.3（C^1、$^\alpha$-锥可约 C_i的收敛速率）。考虑（*1.2*），并假设假设 *4.1* 成立。设 d 为（*3.6*）中给出的函数，$\{x^t\}$和 $\{\boldsymbol{y}^t\}$ 为由算法 *3.1* 生成的序列。则有 $x^t \to x^*$且 $\mathit{dist}(\boldsymbol{y}^t,$ 极小点集合 $d) \to 0$，其中 x^* 是（*1.2*）的唯一解。此外，以下结论成立：

(i) 若每个 C_i 是 C^1, 1-锥可约，则存在 $\boldsymbol{y}^* \in$ 极小点集合 d、$a_1 > 0$、$a_0 \in (0, 1)$ 以及一个正整数 $\bar{t}$，使得对任意 $t \geq \bar{t}$，

$$\|x^t - x^*\| \leq a_1 a_0^t, \quad \|\boldsymbol{y}^t - \boldsymbol{y}^*\| \leq a_1 a_0^t.$$

(ii) 若每个 C_i 是 C^1, $^\alpha$-锥可约且 $\alpha \in (0, 1)$，则存在 $a_1 > 0$ 和一个正整数 $\bar{t}$，使得对任意 $t \geq \bar{t}$，

$$\begin{aligned}
&\|x^t - x^*\| \leq a_1 t^{-\frac{\alpha}{2}\cdot\frac{1-\theta}{2\theta-1}},\\
&d(\boldsymbol{y}^t) - d^* \leq a_1 t^{-\frac{1-\theta}{2\theta-1}},\\
&\operatorname{dist}(\boldsymbol{y}^t, \operatorname{Arg\,min} d) \leq a_1 t^{-\frac{\alpha}{2}\cdot\frac{1-\theta}{2\theta-1}},
\end{aligned}$$

其中 d^* 由 (3.6) 给出，且 $\theta := \frac{1}{1+\alpha} \in (1/2, 1)$ 是 d 的 *KL* 指数。

证明。由定理 $\boldsymbol{y}^t$，极小点集合 $d) \to 0$ 立即可得 3.4。由 (3.1) 可知 $\boldsymbol{A}^T = [A_1^T \ \cdots \ A_\ell^T]$。然后，注意到$\{x^t\}$ 收敛到 x^* 即可推出

$$\|x^t - x^*\| \overset{(a)}{=} \left\|\bar{v} - \textstyle\sum_{j=1}^{\ell} A_j^T y_j^t - \left(\bar{v} - \sum_{j=1}^{\ell} A_j^T \bar{y}_j^t\right)\right\| \leq \|\boldsymbol{A}^T\| \operatorname{dist}(\boldsymbol{y}^t, \operatorname{Arg\,min} d), \tag{5.6}$$

其中 $\mathbf{Proj}_{\operatorname{Arg\,min} d}(\boldsymbol{y}^t) =: (\bar{y}_1^t, \ldots, \bar{y}_\ell^t)$，且 (a) 由命题 3.1 和 (3.10) 得出。由 (3.1) 可知 $D = C_1 \times \cdots \times C_\ell$。定义

$$\mathcal{G}_i^\gamma(\boldsymbol{y}) := y_i - \operatorname{Prox}_{\gamma^{-1}\sigma_{C_i}}(y_i - \gamma^{-1}\nabla_{y_i} g(\boldsymbol{y})), \quad i = 1, \ldots, \ell,$$

其中 g 如 (3.6) 中所述。因此我们可以写成 $\mathcal{G}(\boldsymbol{y}) = (\mathcal{G}_1^1(\boldsymbol{y}) \ldots, \mathcal{G}_\ell^1(\boldsymbol{y}))$，且 (3.9) 变为

$$y_i^{t+1} = \operatorname{Prox}_{\gamma_i^{-1}\sigma_{C_i}}(y_i^t - \gamma_i^{-1}\nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1})) = y_i^t - \mathcal{G}_i^{\gamma_i}(\tilde{\boldsymbol{y}}_{i-1}^{t+1}), \tag{5.7}$$

其中 $\tilde{\boldsymbol{y}}_{i-1}^{t+1}$ 与（3.8）中相同。令 $\hat{}y_i^{t+1} := y_i^t - \mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) =$ 近端算子$^{\sigma_{C_i}}(y_i^t - \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}))$。根据近端算子的定义，我们可知 $\hat{}y_i^{t+1}$ 和 y_i^{t+1} 分别满足以下一阶最优性条件：

$$\begin{aligned}
&0 \in \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) + (\hat{y}_i^{t+1} - y_i^t) + \partial\sigma_{C_i}(\hat{y}_i^{t+1}),\\
&0 \in \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) + \gamma_i(y_i^{t+1} - y_i^t) + \partial\sigma_{C_i}(y_i^{t+1}),
\end{aligned}$$

这意味着

$$\begin{aligned}
&\hat{y}_i^{t+1} \in \operatorname*{Arg\,min}_{z\in\mathbb{R}^{m_i}} \langle \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) + (\hat{y}_i^{t+1} - y_i^t), z\rangle + \sigma_{C_i}(z),\\
&y_i^{t+1} \in \operatorname*{Arg\,min}_{z\in\mathbb{R}^{m_i}} \langle \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) + \gamma_i(y_i^{t+1} - y_i^t), z\rangle + \sigma_{C_i}(z).
\end{aligned}$$

那么，对于任意 $z \in \mathbb{R}^{m_i}$，均有

$$\langle \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) + (\hat{y}_i^{t+1} - y_i^t), \hat{y}_i^{t+1} - z\rangle + \sigma_{C_i}(\hat{y}_i^{t+1}) - \sigma_{C_i}(z) \le 0, \tag{5.8}$$

$$\langle \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) + \gamma_i(y_i^{t+1} - y_i^t), y_i^{t+1} - z\rangle + \sigma_{C_i}(y_i^{t+1}) - \sigma_{C_i}(z) \le 0. \tag{5.9}$$

将 $z = y_i^{t+1}$ 和 $z = \hat{y}_i^{t+1}$ 分别代入（5.8）和（5.9），并将两式相加。回顾（5.7）以及 $\hat{y}_i^{t+1} - y_i^t = -\mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1})$ 这一事实，可得

$$\langle \mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1}), \mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) - \mathcal{G}_i^{\gamma_i}(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\rangle + \gamma_i\langle \mathcal{G}_i^{\gamma_i}(\tilde{\boldsymbol{y}}_{i-1}^{t+1}), \mathcal{G}_i^{\gamma_i}(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) - \mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\rangle \le 0,$$

这给出 $\|\mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\|^2 + \gamma_i\|\mathcal{G}_i^{\gamma_i}(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\|^2 \le (1+\gamma_i)\langle \mathcal{G}_i^{\gamma_i}(\tilde{\boldsymbol{y}}_{i-1}^{t+1}), \mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\rangle$，因此

$$\|\mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\| \le (1+\gamma_i)\|\mathcal{G}_i^{\gamma_i}(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\| = (1+\gamma_i)\|y_i^{t+1} - y_i^t\|, \tag{5.10}$$

其中，最后一个等式由（5.7）可得。回顾 $\boldsymbol{y}^t = (y_1^t, \ldots, y_\ell^t)$。于是

$$\begin{aligned}
&\|\mathcal{G}_i^1(\boldsymbol{y}^t) - \mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\| \\
=&\|y_i^t - \mathrm{prox}_{\sigma_{C_i}}(y_i^t - \nabla_{y_i} g(\boldsymbol{y}^t)) - (\tilde{\boldsymbol{y}}_{i-1}^{t+1})_i + \mathrm{prox}_{\sigma_{C_i}}((\tilde{\boldsymbol{y}}_{i-1}^{t+1})_i - \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}))\| \\
\le& L_g\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|,
\end{aligned} \tag{5.11}$$

其中，不等式由 $(\tilde{\boldsymbol{y}}_{i-1}^{t+1})_i = y_i^t$、近端算子的非扩张性，以及 ∇g 的 Lipschitz 连续性（Lipschitz 模数为 $L_g := \|\boldsymbol{A}^T\boldsymbol{A}\|$）得出。进一步可推得

$$\begin{aligned}
\|\mathcal{G}(\boldsymbol{y}^t)\| &\le \textstyle\sum_{i=1}^{\ell} \|\mathcal{G}_i^1(\boldsymbol{y}^t)\| \le \sum_{i=1}^{\ell} \left(\|\mathcal{G}_i^1(\boldsymbol{y}^t) - \mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\| + \|\mathcal{G}_i^1(\tilde{\boldsymbol{y}}_{i-1}^{t+1})\|\right) \\
&\overset{(a)}{\le} \textstyle\sum_{i=1}^{\ell} \left(L_g\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\| + (1+\gamma_i)\|y_i^{t+1} - y_i^t\|\right) \le \hat{M}\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|,
\end{aligned} \tag{5.12}$$

其中 (a) 由（5.10）和（5.11）得出，且 $\hat{M} := \ell L_g + (\ell + \sum_{i=1}^{\ell}\gamma_i)\sqrt{\ell}$。

接下来，令 $\bar{\boldsymbol{y}}^t = (\bar{y}_1^t, \ldots, \bar{y}_\ell^t) := \mathrm{Proj}_{\mathrm{Arg\,min}\, d}(\boldsymbol{y}^t)$。于是可得

$$\begin{aligned}
d(\boldsymbol{y}^{t+1}) - d^* &\overset{(a)}{=} g(\boldsymbol{y}^{t+1}) + \sigma_D(\boldsymbol{y}^{t+1}) - g(\bar{\boldsymbol{y}}^t) - \sigma_D(\bar{\boldsymbol{y}}^t) \\
&\overset{(b)}{=} \langle \nabla g(\boldsymbol{u}^t), \boldsymbol{y}^{t+1} - \bar{\boldsymbol{y}}^t\rangle + \textstyle\sum_{i=1}^{\ell}[\sigma_{C_i}(y_i^{t+1}) - \sigma_{C_i}(\bar{y}_i^t)] \\
&= \langle \nabla g(\boldsymbol{u}^t) - \nabla g(\boldsymbol{y}^t), \boldsymbol{y}^{t+1} - \bar{\boldsymbol{y}}^t\rangle + \textstyle\sum_{i=1}^{\ell}\langle \nabla_{y_i} g(\boldsymbol{y}^t) - \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}), y_i^{t+1} - \bar{y}_i^t\rangle \\
&\quad + \textstyle\sum_{i=1}^{\ell}[\langle \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) + \gamma_i(y_i^{t+1} - y_i^t), y_i^{t+1} - \bar{y}_i^t\rangle + \sigma_{C_i}(y_i^{t+1}) - \sigma_{C_i}(\bar{y}_i^t)] \\
&\quad - \textstyle\sum_{i=1}^{\ell}\gamma_i\langle y_i^{t+1} - y_i^t, y_i^{t+1} - \bar{y}_i^t\rangle \\
&\overset{(c)}{\le} L_g\|\boldsymbol{u}^t - \boldsymbol{y}^t\|\|\boldsymbol{y}^{t+1} - \bar{\boldsymbol{y}}^t\| + \ell L_g\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|\|\boldsymbol{y}^{t+1} - \bar{\boldsymbol{y}}^t\| \\
&\quad + \gamma_{\mathrm{sum}}\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|\|\boldsymbol{y}^{t+1} - \bar{\boldsymbol{y}}^t\| \\
&\le [(L_g + \ell L_g + \gamma_{\mathrm{sum}})\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\| + L_g\|\boldsymbol{y}^t - \bar{\boldsymbol{y}}^t\|](\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\| + \|\boldsymbol{y}^t - \bar{\boldsymbol{y}}^t\|),
\end{aligned} \tag{5.13}$$

其中 (a) 由（d 在（3.6）中的定义得出；(b) 由中值定理得出，其中 $\boldsymbol{u}^t = \tau_t\boldsymbol{y}^{t+1} + (1-\tau_t)\bar{\boldsymbol{y}}^t$ 对某个 $\tau_t \in (0,1)$ 成立；(c) 由（5.9）以及 ∇g 的 Lipschitz 连续性得出，其中 $\gamma_{\mathrm{sum}} := \sum_{i=1}^{\ell}\gamma_i$。

另一方面，我们有 $\|\mathcal{G}(\boldsymbol{y}^t)\| \to 0$ (得益于 (5.12) 和命题3.2(iii))，以及 $d(\boldsymbol{y}^t) \to d^*$ (得益于定理 3.4)。因此，存在一个足够大的正

整数 $\bar{t}$ 使得 $\|\mathcal{G}(\boldsymbol{y}^t)\| \le \epsilon$ 并且 $d(\boldsymbol{y}^t) \le d^* + \epsilon$ 对任意 $t \ge \bar{t}$，其中 ϵ 由引理 5.2指定。于是，对任意 $t \ge \bar{t}$，

$$\|\boldsymbol{y}^t - \bar{\boldsymbol{y}}^t\| \overset{\text{(a)}}{=} \operatorname{dist}(\boldsymbol{y}^t, \operatorname{Arg\,min} d) \overset{\text{(b)}}{\le} c\|\mathcal{G}(\boldsymbol{y}^t)\|^\alpha \overset{\text{(c)}}{\le} \tilde{c}\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|^\alpha, \tag{5.14}$$

其中 (a) 成立，因为 $\bar{\boldsymbol{y}}^t = 投影_{\operatorname{Arg\,min} d}(\boldsymbol{y}^t)$，(b) 由引理 5.2 得出，而 (c) 对$\tilde{c} := c\hat{}M^\alpha$ 成立，这是因为 (5.12)。现在，由于 $\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\| \to 0$ (见命题 3.2(iii))，且$\alpha \in (0, 1]$, 由 (5.13) 和 (5.14) 可知，存在 $c_1 > 0$ 使得

$$d(\boldsymbol{y}^{t+1}) - d^* \le c_1\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|^{2\alpha} \quad \text{for all } t \ge \bar{t}. \tag{5.15}$$

此外，命题3.2(ii) 表明，对所有 t,

$$d(\boldsymbol{y}^{t+1}) - d(\boldsymbol{y}^t) \le -c_2\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|^2, \tag{5.16}$$

对于某个 $c_2 > 0$。然后，结合以上两个式子可得，对所有 $t \ge \bar{t}$,

$$(d(\boldsymbol{y}^{t+1}) - d^*) - (d(\boldsymbol{y}^t) - d^*) \le -c_2\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|^2 \le -c_1^{-1/\alpha}c_2(d(\boldsymbol{y}^{t+1}) - d^*)^{1/\alpha},$$

这意味着

$$(d(\boldsymbol{y}^{t+1}) - d^*) + c_1^{-1/\alpha}c_2(d(\boldsymbol{y}^{t+1}) - d^*)^{1/\alpha} \le d(\boldsymbol{y}^t) - d^* \quad \text{for all } t \ge \bar{t}. \tag{5.17}$$

接下来，我们分别考虑 $\alpha = 1$ 和 $\alpha \in (0, 1)$ 的情形。

情形 1: $\alpha = 1$。在此情形下，由（5.17）可得 $t \ge \bar{t}$

$$d(\boldsymbol{y}^{t+1}) - d^* \le \tfrac{1}{1+c_1^{-1}c_2}(d(\boldsymbol{y}^t) - d^*),$$

这意味着 $\{d(\boldsymbol{y}^{t+1}) - d^*\}$ Q-线性收敛于零。结合（5.16），最后一个不等式表明

$$\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|^2 \le c_2^{-1}[(d(\boldsymbol{y}^t) - d^*) - (d(\boldsymbol{y}^{t+1}) - d^*)] \le c_2^{-1}(d(\boldsymbol{y}^t) - d^*),$$

结合 $\{d(\boldsymbol{y}^t) - d^*\}$ 的 Q-线性收敛（于零），这意味着存在 $c_3 > 0$, $\alpha_0 \in (0, 1)$ 使得 $\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\| \le c_3\alpha_0^t$。因此，我们有

$$\|\boldsymbol{y}^{t_1} - \boldsymbol{y}^{t_2}\| \le \textstyle\sum_{j=t_1}^{t_2-1}\|\boldsymbol{y}^{j+1} - \boldsymbol{y}^j\| \le \tfrac{c_3}{1-\alpha_0}\alpha_0^{t_1} \quad \text{for all } t_2 > t_1 \ge \bar{t},$$

这意味着 $\{\boldsymbol{y}^t\}$ 收敛。令 $\boldsymbol{y}^*$ 表示其极限。则 $\boldsymbol{y}^* \in$ 极小点集合 d由定理 3.4可知。令 $t_2 \to \infty$ 趋于上述显示式中的极限，可得

$$\|\boldsymbol{y}^{t_1} - \boldsymbol{y}^*\| \le (c_3\alpha_0^{t_1})/(1 - \alpha_0).$$

利用这一点、（3.10）和命题 3.1，我们进一步推出，对 $t \ge \bar{t}$

$$\|x^t - x^*\| = \left\|\bar{v} - \textstyle\sum_{j=1}^{\ell} A_j^T y_j^t - \left(\bar{v} - \sum_{j=1}^{\ell} A_j^T y_j^*\right)\right\| \le \|\boldsymbol{A}^T\|\tfrac{c_3}{1-\alpha_0}\alpha_0^t, \tag{5.18}$$

这意味着 $\{x^t\}$ 线性收敛到 x^*。

情形 2: $\alpha \in (0,1)$. 我们首先证明序列 $\{d(\boldsymbol{y}^t) - d^*\}$ 次线性收敛至零。 为此，回顾一下 $\theta = \frac{1}{\alpha+1}$, (5.17) 可以写成

$$(d(\boldsymbol{y}^{t+1}) - d^*)^{\frac{\theta}{1-\theta}} \le c_2^{-1}c_1^{\frac{\theta}{1-\theta}}[(d(\boldsymbol{y}^t) - d^*) - (d(\boldsymbol{y}^{t+1}) - d^*)] \quad \text{for all } t \ge \bar{t}. \tag{5.19}$$

然后，按照从 [1, 方程 (13)],开始的 [1, 定理 2] 中的相同论证，我们可以推导出存在 $c_4 > 0$ 使得

$$d(\boldsymbol{y}^t) - d^* \le c_4 t^{-\frac{1-\theta}{2\theta-1}} \quad \text{for all } t \ge \bar{t}; \tag{5.20}$$

其中，最后一个不等式对应于 [1, 第 15] 页上的第一个不等式。

对于 $\{\mathrm{dist}(\boldsymbol{y}^t,$ 极小点集合 $d)\}$ 的次线性收敛，我们看到，对于任意 $t \ge \bar{t}$

$$\begin{aligned} c_2\tilde{c}^{-\frac{2}{\alpha}}\mathrm{dist}(\boldsymbol{y}^t, \operatorname{Arg\,min} d)^{\frac{2}{\alpha}} &\overset{\text{(a)}}{\le} c_2\tilde{c}^{-\frac{2}{\alpha}}(\tilde{c}\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|^\alpha)^{2/\alpha} = c_2\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\|^2 \\ &\overset{\text{(b)}}{\le} d(\boldsymbol{y}^t) - d(\boldsymbol{y}^{t+1}) \le d(\boldsymbol{y}^t) - d^*, \end{aligned} \tag{5.21}$$

其中，(a) 由（5.14）得出，(b) 由（5.16）得出。因此，$\{\mathrm{dist}(\boldsymbol{y}^t,$ 极小点集合 $d)\}$ 的次线性收敛速率由（5.20）得出。最后，（5.6）表明，$\{x^t\}$ 的收敛速率可由 $\{\mathrm{dist}(\boldsymbol{y}^t,$ 极小点集合 $d)\}$ 的收敛速率得到。 □

备注 5.4. 我们再次指出，基于 KL 性质的收敛框架（参见例如 [1, 定理 1]）在此并不直接适用，因为 $\{\boldsymbol{y}^t\}$ 可能没有聚点。另一方面，若 $\{\boldsymbol{y}^t\}$ 有界，则定理 $(\boldsymbol{y}^t) - d^*\}$ 5.3(ii) 中的收敛速率与基于 KL 的分析所导出的标准速率相差一个 $1-\theta$ 因子；例如，参见 [28, 定理 4.1(iv)]。

定理 5.3(i) 表明，在对 $\bar{v}$作一般性假设，即 $0 \in x^* - \bar{v} + \mathrm{ri}\,\partial(\sum_{i=1}^{\ell}\delta_{A_i^{-1}C_i})(x^*)$ 时，算法 3.1具有线性收敛速率。为说明该条件不可或缺，我们给出一个满足定理 5.3(i) 中除假设 4.1(iii) 之外所有假设的例子，但线性收敛并不成立。

Example 5.1（线性收敛不成立）。考虑以下问题

$$\min_{y_1,y_2\in\mathbb{R}^3} d(y_1,y_2) := \tfrac{1}{2}\|y_1 + y_2 - \bar{v}\|^2 + \sigma_{C_1}(y_1) + \sigma_{C_2}(y_2), \tag{5.22}$$

其中 $\bar{v} = (1, -1, 1)$ 以及

$$C_1 := \{(x_1,x_2,x_3) \in \mathbb{R}^3 : x_3 \le -\|(x_1,x_2)\|\}, \quad C_2 := \{(0,x_2,x_3) \in \mathbb{R}^3 : x_2, x_3 \in \mathbb{R}\}.$$

注意，(5.22) 是 *BA* 问题 $\min_{x\in C_1\cap C_2}\frac{1}{2}\|x-\bar{v}\|^2$. 的拉格朗日对偶的相反数（相差一个加性常数）。根据 C_1和 C_2的定义，可以看出 C_1和 C_2是 $C^{1,1}$-锥可约的（另见例 *4.1*），且 $\mathrm{ri}\,C_1 \cap \mathrm{ri}\,C_2 \ne \emptyset$。

注意，$C_1 = K_1^\circ$和 $C_2 = K_2^\circ$，其中

$$K_1 := \{(x_1,x_2,x_3) \in \mathbb{R}^3 : x_3 \ge \|(x_1,x_2)\|\}, \quad K_2 := \{(x_1,0,0) \in \mathbb{R}^3 : x_1 \in \mathbb{R}\}.$$

那么 $\sigma_{C_1} = \delta_{K_1}$ 并且 $\sigma_{C_2} = \delta_{K_2}$ 根据 *[37,* 例 ***11.4(b)****]*, 以及(5.22) 可以等价地写成

$$\min_{y_1,y_2\in\mathbb{R}^3} d(y_1,y_2) = \tfrac{1}{2}\|y_1 + y_2 - \bar{v}\|^2 + \delta_{K_1}(y_1) + \delta_{K_2}(y_2). \tag{5.23}$$

可以验证，上述问题的最优值为零，并且

$$y_1^* = (0,-1,1) \in K_1 \quad \text{and} \quad y_2^* = (1,0,0) \in K_2$$

是 (5.22) 的唯一解，即 极小点集合 $d = \{(y_1^*, y_2^*)\}$，且 $x^* = (0,0,0)$ 根据命题 *3.1*，是原问题的唯一解。此外，我们有

$$0 \notin x^* - \bar{v} + \mathrm{ri}\,\partial(\delta_{C_1} + \delta_{C_2})(x^*),$$

由于 $\mathrm{ri}\,\partial(\delta_{C_1}+\delta_{C_2})(x^*) = \mathrm{ri}\,\mathcal{N}_{C_1\cap C_2}(x^*) \subseteq \mathcal{N}_{C_1\cap C_2}(x^*) = \{(x_1,x_2,x_3)\colon x_3 \geq |x_2|\}$。

回顾一下， y算法 3.1 中的迭代点可以通过对(5.23)应用 *CGD* 得到，初始点为 $y_1^0 = y_2^0 = (0,0,0)$。由归纳法可知，算法 *y3.1* 生成的迭代点可以写成[10]

$$
\text{(5.24)}\quad \begin{cases} y_1^{t+1} = \mathrm{Proj}_{K_1}(\bar{v} - y_2^t) = \left(a_{t+1}, -\frac{1}{2}\left(1 + \frac{1}{\sqrt{a_t^2+1}}\right), \frac{1}{2}\left(1+\sqrt{a_t^2+1}\right)\right), \\ y_2^{t+1} = \mathrm{Proj}_{K_2}(\bar{v} - y_1^{t+1}) = (1 - a_{t+1}, 0, 0), \\ \textit{with } a_{t+1} = \frac{1}{2}\left(1 + \frac{1}{\sqrt{a_t^2+1}}\right) a_t \textit{ and } a_0 = 1. \end{cases}
$$

由上文可知， $\{(y_1^t, y_2^t)\}$ 是收敛的，因为 $\{a_t\}$ 有界且单调不增，因此我们知道 $(y_1^t, y_2^t) \to (y_1^*, y_2^*)$ 由命题 *3.2(v)* 得出。这意味着极限$_{t\to\infty} a_t = 0$。根据 a_t 中 (5.24) 的定义，并注意到 $a_t > 0$，可得

$$
\frac{1}{a_{t+1}^2} - \frac{1}{a_t^2} = \frac{3a_t^2 + 2 - 2\sqrt{a_t^2+1}}{a_t^2(1+\sqrt{a_t^2+1})^2} = \frac{3a_t^2 + 2 - 2(1 + 0.5a_t^2 + O(a_t^4))}{a_t^2(1+\sqrt{a_t^2+1})^2} = \frac{2 + O(a_t^2)}{(1+\sqrt{a_t^2+1})^2}.
$$

将上述等式两边从 $t=0$到 $t=N$ 求和，得到 $\frac{1}{a_N^2} - \frac{1}{a_0^2} = \Theta(N)$，这意味着 $a_t = \Theta(1/\sqrt{t})$。然后，利用这一关系、*(5.24)* 以及事实$\sqrt{h^2+1} - 1 = \Theta(h^2)$（当 $h \to 0$ 时），我们推得 $\mathrm{dist}((y_1^{t+1}, y_2^{t+1}),$ 极小点集合 $d)^2 = \|(y_1^{t+1}, y_2^{t+1}) - (y_1^*, y_2^*)\|^2$ 等于

$$
2a_{t+1}^2 + \left(\frac{\sqrt{a_t^2+1}-1}{2\sqrt{a_t^2+1}}\right)^2 + \left(\frac{\sqrt{a_t^2+1}-1}{2}\right)^2 = 2a_{t+1}^2 + \Theta(a_t^4) = \Theta\left(\frac{1}{t+1}\right),
$$

这意味着 $\{(y_1^t, y_2^t)\}$ 到 (y_1^*, y_2^*) 的收敛速率不是线性的。

附录 A. 算法 3.1 与近端 CGD 方案的等价性。 我们首先从（3.7）和（3.9）推导出算法 3.1。为此，注意到（3.9）等价于

$$
y_i^{t+1} = \mathrm{prox}_{\gamma_i^{-1}\sigma_{C_i}}(y_i^t - \gamma_i^{-1}\nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1})). \tag{A.1}
$$

定义 $x_\ell^0 := \bar{v}$并且，对于 $i = 0,\ 1,\ \ldots,\ \ell$,

$$
x_i^{t+1} := \bar{v} - \textstyle\sum_{j=1}^{i} A_j^T y_j^{t+1} - \sum_{j=i+1}^{\ell} A_j^T y_j^t. \tag{A.2}
$$

那么，根据（A.2）以及 $\tilde{\boldsymbol{y}}_{i-1}^{t+1}$在（3.8）中的定义，对于 $i = 1,\ \ldots,\ \ell$，成立:

$$
A_i x_{i-1}^{t+1} = -\nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}), \tag{A.3}
$$

$$
x_i^{t+1} = x_{i-1}^{t+1} - A_i^T(y_i^{t+1} - y_i^t). \tag{A.4}
$$

因此，关系（A.1）可以进一步改写为

$$
y_i^{t+1} \overset{\text{(a)}}{=} y_i^t - \gamma_i^{-1}\nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1}) - \gamma_i^{-1}\mathrm{Proj}_{C_i}(\gamma_i y_i^t - \nabla_{y_i} g(\tilde{\boldsymbol{y}}_{i-1}^{t+1})) \tag{A.5}
$$

$$
\overset{\text{(A.3)}}{=} y_i^t + \gamma_i^{-1} A_i x_{i-1}^{t+1} - \gamma_i^{-1}\mathrm{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}), \tag{A.6}
$$

[10] $(u,t) \in \mathrm{IR}^{n+1}$ 上的投影 K_1 可通过 [10, 练习 2911]求得。

其中 (a) 成立，是因为对于任意闭凸集 $K \subseteq \mathrm{IR}^n$， $r > 0$ 和 $u \in \mathrm{IR}^n$，都有 $\mathrm{prox}_{r\sigma_K}(u) = u - \mathrm{prox}_{(\sigma_{rK})^*}(u) = u - \mathrm{Proj}_{rK}(u) = u - r\mathrm{Proj}_K(r^{-1}u)$。然后，利用 (A.4) 以及由 (A.6) 得出的 $y_i^{t+1} - y_i^t$ 表达式，对于 $i = 1, \ldots..., \ell$,

$$x_i^{t+1} = (I - \gamma_i^{-1} A_i^T A_i) x_{i-1}^{t+1} + \gamma_i^{-1} A_i^T \mathrm{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}). \tag{A.7}$$

因此，我们证明了 (3.9) 给出了 (A.6) 和 (A.7)，即算法 3.11 中的式子3.1。此外，由 (3.7)、(A.2) 以及定义 $x_\ell^0 := \bar{v}$，可知

$$x_0^{t+1} = \bar{v} - \textstyle\sum_{j=1}^{\ell} A_j^T y_j^t = x_\ell^t \quad \text{for all } t \geq 0. \tag{A.8}$$

结合 (A.6)、(A.7) 和 (A.8)，我们推导出算法 3.1。

反过来，我们证明 (3.7) 和 (3.9) 可由算法 3.1 推导得到。根据算法 3.1 的步骤 1，只需证明 (A.1)。为此，设 $\{x^t\}$,$\{x_i^t\}$ 和 $\{y_i^t\}$, $i = 1, \ldots, \ell$, 是由算法 3.1 生成的。由 (3.11) 可得 (A.4)。我们通过归纳法证明 (3.10) 成立。首先，根据算法 3.1 的步骤 1，显然 $t = 0$ 满足该式。假设对于某个 $t \geq 0$， $x_\ell^t = \bar{v} - \boldsymbol{A}^T \boldsymbol{y}^t$ 成立。于是，算法 3.1 的步骤 2 开始部分表明 $x_0^{t+1} = \bar{v} - \boldsymbol{A}^T \boldsymbol{y}^t$，结合 (A.4) 可知 $x_\ell^{t+1} = \bar{v} - \boldsymbol{A}^T \boldsymbol{y}^{t+1}$，从而通过归纳法证明 (3.10) 成立。

现在，(3.10)、(A.4) 和 $x_0^{t+1} = x_\ell^t$ (算法 3.1的步骤 2) 给出 (A.2)，因此 (A.3)。(y在 (3.11) 中的更新随后表明 (A.5) 成立，而这正是 (A.1)。

参考文献

[1] H. Attouch and J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features* 116 2009 5 16 ,《数学规划》， （），第–页。

[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality* 35 2010 438 457 ,《数学运筹研究》， （），第–页。

[3] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms forward–backward splitting and regularized Gauss–Seidel methods* 137 2013 91 129 ,, ，《数学规划》， （），第–页。

[4] A. Auslender, R. Cominetti, and J. Crouzeix, *Convex functions with unbounded level sets and applications to duality theory* 3 1993 669 687 ,SIAM优化杂志， （），第–页。

[5] A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities* 2006 ,施普林格科学与商业媒体，。

[6] H. H. Bauschke and J. M. Borwein, *Dykstra's alternating projection algorithm for two sets*,《逼近理论杂志》，79（1994年），第418–443页。[7] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems* 38 1996 367 426 ,《SIAM 评论》， （），第–页。

[8] H. H. Bauschke, J. M. Borwein, and W. Li, *Strong conical hull intersection property bounded linear regularity Jameson's property (G) and error bounds in convex optimization* ,, , ,《数学规划》，86（1999年），第135–160页。[9] H. H. Bauschke, R. S. Burachik, D. B. Hermann, and C. Y. Kaya, *On Dykstra's algorithm: finite convergence stalling and the method of alternating projections* 14 ,, ,《优化快报》，（2020年），第1975–1987页。[10] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* 2017 ,施普林格，。

[11] H. H. Bauschke and A. S. Lewis, *Dykstra's algorithm with Bregman projections: A convergence proof* 48 2000 409 427 ,《优化》， （），第–页。

[12] J. Bolte, T. P. Nguyen, J. Peypouquet, and B. W. Suter, *From error bounds to the complexity of first-order descent methods for convex functions* 165 2017 ,《数学规划》， （），第471–507页。[13] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems* 146 2014 459 494 ,《数学规划》， （），第–页。

[14] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*,施普林格，2000年。

[15] J. P. Boyle and R. L. Dykstra, *A method for finding projections onto the intersection of convex sets in Hilbert spaces*,载于《序约束统计推断进展》，施普林格，1986年，第28–47页。[16] M. Brooke, Y. Censor, and A. Gibali, *Dynamic string-averaging CQ-methods for the split feasibility problem with percentage violation constraints arising in radiation therapy treatment planning* 30 2020 181 205 ,国际运筹学交易，（），第–页。

[17] J. V. Burke, F. E. Curtis, H. Wang, and J. Wang, *Iterative reweighted linear least squares for exact penalty subproblems on product sets* 25 2015 261 294 ,SIAM优化杂志，（），第–页。

[18] C. Byrne, *Iterative oblique projection onto convex sets and the split feasibility problem*,逆问题，18（2002年），第441–453页。[19] Y. Censor and T. Elfving, *A multiprojection algorithm using Bregman projections in a product space* 8 1994 221 239 ,数值算法，（），第–页。

[20] Y. Censor, T. Elfving, N. Kopf, and T. Bortfeld, *The multiple-sets split feasibility problem and its applications for inverse problems* 21 2005 2071 2084 ,逆问题，（），第–页。

[21] Y. Censor, D. Reem, and M. Zaknoon, *A generalized block-iterative projection method for the common fixed point problem induced by cutters* 84 2022 967 987 ,全局优化杂志，（），第–页。

[22] A. Chambolle and T. Pock, *On the ergodic convergence rates of a first-order primal–dual algorithm* 159 2016 253 287 ,《数学规划》，（），第–页。

[23] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The Linear Complementarity Problem*,学术出版社，美国马萨诸塞州波士顿，1992年。[24] F. Deutsch and H. Hundal, *The rate of convergence of Dykstra's cyclic projections algorithm: The polyhedral case* 15 1994 537 565 ,数值泛函分析与优化，（），第–页。

[25] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*,第543卷，施普林格，2009年。

[26] D. Drusvyatskiy and A. S. Lewis, *Error bounds quadratic growth and linear convergence of proximal methods* 43 2018 919 948 ,,，《数学运筹研究》，（），第–页。[27] N. Gaffke and R. Mathar, *A cyclic projection algorithm via duality*,Metrika，36（1989年），第29–54页。

[28] G. Garrigos, L. Rosasco, and S. Villa, *Convergence of the forward-backward algorithm: beyond the worst-case with the help of geometry* 198 2023 937 996 ,《数学规划》，（），第–页。

[29] S.-P. Han, *A successive projection method*,《数学规划》，40（1988年），第1–14页。[30] H. Hundal and F. Deutsch, *Two generalizations of Dykstra's cyclic projections algorithm*,《数学规划》，77（1997年），第335–355页。

[31] G. Li and T. K. Pong, *Calculus of the exponent of Kurdyka–Łojasiewicz inequality and its applications to linear convergence of first-order methods* 18 2018 ,计算数学基础，（），第1199–1232页。[32] T. Liu and B. F. Lourenço, *Convergence analysis under consistent error bounds*, 计算数学基础，（2022年），https://doi.org/https://doi.org/10.1007/s10208-022-09586-4。[33] Z.-Q. Luo and P. Tseng, *Error bounds and convergence analysis of feasible descent methods: a general approach* 46 1993 157 178 ,运筹学年刊，（），第–页。

[34] C. J. Pang, *Set intersection problems: supporting hyperplanes and quadratic programming*,《数学规划》，149（2015年），第329–359页。[35] C. J. Pang, *Distributed deterministic asynchronous algorithms in time-varying graphs through Dykstra splitting* 29 2019 484 510 ,SIAM优化杂志，（），第–页。

[36] R. T. Rockafellar, *Convex Analysis*,普林斯顿大学出版社，1970年。[37] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*,施普林格，2009年。[38] A. Shapiro, *Sensitivity analysis of generalized equations*,数学科学杂志，115（2003年），第2554–2565页。[39] P. Tseng and S. Yun, *A coordinate gradient descent method for nonsmooth separable minimization* 117 2009 387 423 ,《数学规划》，（），第–页。

[40] X. Wang and Z. Wang,*property* 47 2022 2765 2783 ,《数学运筹研究》，（），第–页。[41] Y. Yang, Q. Yang, and S. Zhang, *Modified alternating direction methods for the modified multiple-sets split feasibility problems* 163 2014 130 147 ,优化理论与应用杂志，（），第–页。

[42] P. Yu, G. Li, and T. K. Pong, *Kurdyka–Łojasiewicz exponent via inf-projection*,计算数学基础，22（2022年），第1171–1217页。[43] M.-G. Yue, Z. Zhou, and A. M.-C. So, *A family of inexact SQA methods for non-smooth convex minimization with provable convergence guarantees based on the Luo–Tseng error bound property* 174 2019 327 358 ,《数学规划》，（），第–页。