arXiv:2304.01797v1 [math.OC] 4 Apr 2023

S基于线性规划的安全零阶优化

Baiwei Guo* Yang Wang† Yuning Jiang ‡ Maryam Kamgarpour§Giancarlo Ferrari-Trecate*

2023年4月5日

摘要

为解决带有硬约束的未建模优化问题，本文提出一种名为基于线性规划的安全零阶优化（SZO-LP）的新型零阶方法。SZO-LP 方法在每次迭代中求解一个线性规划以寻找下降方向，随后确定步长。我们证明，在温和条件下，SZO-LP 的迭代点存在一个聚点，该聚点也是一个 KKT 对的原始解。随后，我们将 SZO-LP 应用于 IEEE 30节点系统上的最优潮流（OPF）问题。结果表明，与先进方法相比，SZO-LP 所需的计算时间和样本更少。

1 引言

包括电力网络运行 [1], 机器学习 [2], 轨迹优化 [3] 和最优控制 [4], 在内的各种应用，都需要解决具有严格安全约束的复杂优化问题。然而，目标函数和约束函数的表达式并非总是可得，也可能缺乏足够的可行系统轨迹数据用于建模。在此背景下，可采用安全零阶优化方法来处理带约束的未建模优化问题。这些方法仅依靠在选定点处对未知目标函数和约束函数进行评估来采样 [5] ；其中，“安全”指样本的可行性（即满足约束）。

常见的安全零阶方法包括 SafeOPT 及其变体 [6–8]。这些方法假设已知目标函数和约束函数的 Lipschitz常数，而 [9] 则使用函数梯度的 Lipschitz常数（即光滑常数）。利用这些量，可以为约束函数构建局部代理函数。从一个可行点出发， [6, 8, 9] 利用这些代理函数搜索可能的极小点。然而，对于每次搜索，

*B. Guo 和 G. Ferrari-Trecate 隶属于瑞士 EPFL 机械工程研究所 DECODE组。电子邮箱： {baiwei.guo, giancarlo.ferraritrecate}@epfl.ch。

†Y. Wang 任职于代尔夫特理工大学代尔夫特系统与控制中心。 {y.wang-40}@tudelft.nl。

‡Y. Jiang 隶属瑞士 EPFL 机械工程研究所的 PREDICT组。电子邮箱：yuning.jiang@epfl.ch。§M. Kamgarpour 隶属瑞士 EPFL 机械工程研究所的 SYCAMORE组。电子邮箱：maryam.kamgarpour@epfl.ch。本研究由瑞士国家科学基金会通过 NCCR Automation（资助协议编号 51NF40 80545）提供支持。

都必须使用全局优化方法求解一个非凸子问题，这使得该算法在决策变量众多的问题中计算上难以处理。

为降低计算复杂度，另一种研究方向是在目标函数中引入障碍函数，以惩罚迭代点接近可行集 [10,11]边界的情况。极值搜索方法 [12] 和 LB-SGD 算法 [13] 基于估计梯度，最小化包含对数障碍惩罚项的代价函数。尽管这两种方法无需求解优化子问题，但由于对数惩罚项的存在，其性能可能并不理想。在极值搜索中，惩罚项权重的调整可能颇具挑战，因为较大的权重可能导致次优解，而较小的权重则可能导致不可行。在 LB-SGD 中，当迭代点接近可行集的边界时，对数障碍项及其导数的值可能变得很大，从而导致步长减小并减缓收敛速度。

另一种安全零阶优化方法是 [14] 中提出的 SZO-QQ。该方法避免了对数障碍惩罚，同时仍能确保样本可行性，并且比 LB-SGD [14] 具有更高的样本效率。其实现方式是利用约束函数的凸二次代理函数构建局部可行集，然后在该集合上最小化目标函数的代理函数。与 SafeOPT 不同，SZO-QQ 的子问题是凸二次约束二次规划问题（QCQPs），其求解速度远快于 SafeOPT 中的非凸子问题。然而，在处理大型问题（包含数百个约束）时，SZO-QQ 在计算效率方面不及 LB-SGD 和极值搜索（见第 5 节），因为每个 QCQP 子问题的规模几乎与原始问题相同。本文提出了一种新颖的安全零阶方法，其子问题包含的约束数量大幅减少，因此具有较高的计算效率。

最优潮流（OPF）是可从零阶优化中获益的一类大规模优化问题。其目标是在满足输电线路容量和电压水平限制等运行与安全约束的同时，分配有功功率和无功功率发电、输电线路潮流以及电压水平，以最大限度地降低成本。近年来，随着对电力系统高效、可靠运行的需求不断增长，以及可再生能源和储能系统的接入，OPF 受到了广泛关注 [15]。然而，由于难以准确推导系统模型，将 OPF 应用于电力系统运行仍是一项重大挑战。因此，我们考虑采用所提出的无模型方法来求解最优潮流问题。

本文的贡献总结如下：

- 我们提出一种名为基于线性规划的安全零阶优化（SZO-LP）的新方法。该方法通过迭代求解线性规划子问题来推导下降方向，然后通过采样确定步长；
- 我们证明，在温和的假设条件下，SZO-LP 迭代点的一个子序列收敛到 KKT 对的原始解（见定义 1）；
- 通过将 SZO-LP 应用于 IEEE 30节点基准问题，我们表明它能够高效求解包含 11 个决策变量和 158 个约束的 OPF 问题。我们将 SZO-LP 与先进方法进行比较，并展示其在计算时间和所需样本数量方面的优势。

记号：全文中，我们用 $e_i \in \mathbb{R}^d$ 定义向量空间 $\mathbb{R}^d$ 的第 i个标准基，并用 $\|\cdot\|$ 表示两个范数。给定向量 $x \in \mathbb{R}^d$ 和标量 $\epsilon > 0$，我们记作$x = [x^{(1)}, \ldots, x^{(d)}]^\top$和 $\mathcal{B}_\epsilon(x) = \{y : \|y - x\| \leq \epsilon\}$。我们用 $\mathbb{Z}_i^j = \{i, i+1, \ldots, j\}$ 定义从 i 到 j （含 $i < j$）的整数集合。对于两个向量 x、$y \in \mathbb{R}^d$，我们用 $\langle x, y\rangle := x^\top y$ 定义内积。

2 问题表述

我们考虑如下约束优化问题

$$\min_{x\in\mathbb{R}^d} \quad f_0(x) \quad \text{subject to} \quad x\in\Omega, \tag{1}$$

其中，$\Omega := \{x:\ f_i(x)\ \le 0, i\in\mathbb{Z}_1^m\}$ 是可行集。函数 $f_i:\ \mathbb{R}^d\to\mathbb{R},\ i\in\mathbb{Z}_0^m$ 未知，但可以在查询点处采样。本文始终对目标函数和约束函数的光滑性、严格可行点 x_0 的可用性，以及包含 x_0 的次水平集的有界性作如下假设。

假设1 函数 $f_i(x),\ i\in\mathbb{Z}_0^m$ 连续可微，并且存在已知常数 $L_i, M_i>0$，使得对于任意 $x_1, x_2\in\Omega$,

$$|f_i(x_1)-f_i(x_2)|\le L_i\|x_1-x_2\|, \tag{2a}$$
$$\|\nabla f_i(x_1)-\nabla f_i(x_2)\|\le M_i\|x_1-x_2\|. \tag{2b}$$

我们假设 $L_i>\inf\{L_i:\text{(2a) 成立 }\forall x_1,x_2\in\Omega\}$，且 $M_i>\inf\{M_i:\text{(2b) 成立 }\forall x_1,x_2\in\Omega\}$。

在本文余下部分，我们还定义 $L_{\max}=\max_{i\ge1}L_i$ 和 $M_{\max}=\max_{i\ge1}M_i$。

假设2 存在一个已知的严格可行点 x_0，即对于所有 $i\in\mathbb{Z}_1^m$，$f_i(x_0)<0$。

假设3 存在 $\beta\in\mathbb{R}$，使得次水平集 $\mathcal{P}_\beta=\{x\in\Omega:\ f_0(x)\le\beta\}$ 有界，且包含初始可行点 x_0。

假设2是安全零阶方法 [6,13,14,16]中的常见假设。没有初始可行点，就无法确保所有样本都可行。假设3并不严格，因为只要可行域 Ω 有界，该假设就成立。

我们的目标是推导出一种优化算法，使迭代点的一个子序列收敛到某个 KKT 对的原始解。

定义1 若一对 (x,λ) 满足 $x\in\Omega$ 且 $\lambda\in\mathbb{R}^m_{\ge0}$，则

$$\|\nabla f_0(x)+\sum_{i=1}^m\lambda^{(i)}\nabla f_i(x)\|=0, \tag{3a}$$
$$|\lambda^{(i)}f_i(x)|=0,\quad i\in\mathbb{Z}_1^m, \tag{3b}$$

我们称 (x,λ) 为问题 (1) 的 KKT 对，称 $x\in\Omega$ 为该 KKT 对的原始解。

在下一节中，我们将设计一种安全的零阶算法。在温和的假设下，该算法的迭代点具有一个聚点，且该聚点是 (1) 的某个 KKT 对的原始解。

3 算法：SZO-LP

本节首先介绍如何估计式（1）中各函数的梯度并构造局部可行集。这些是我们零阶优化方法所使用的基本工具。

3.1 梯度估计与局部可行集构造

我们通过有限差分来估计梯度，即:

$$\nabla^{\nu} f_i(x) := \sum_{j=1}^{d} \frac{f_i(x+\nu e_j) - f_i(x)}{\nu} e_j. \tag{4}$$

下面的引理　　给出了一种控制估计误差的方法 $\Delta_i^{\nu}(x) := \nabla^{\nu} f_i(x) - \nabla f_i(x)$ 。

引理 1（ [17], 定理 3.2）在假设1下，我们有

$$\|\Delta_i^{\nu}(x)\|_2 \le \frac{\sqrt{d} M_i}{2} \nu. \tag{5}$$

令 $\nu = \nu(\epsilon) := \frac{2\epsilon}{\sqrt{d} M_{\max}}$，我们有 $\|\Delta_i^{\nu(\epsilon)}(x)\| \le \epsilon$。

利用估计梯度，我们可以在 x_0 周围构造一个局部可行集。令

$$l_0^* = \min_{i \in \{1,\dots,m\}} -f_i(x_0)/L_{\max}, \tag{6}$$

和 $\nu_0^*(\epsilon) := \min\{l_0^*/\sqrt{d}, \nu(\epsilon)\}$。由于 $\nu_0^*(\epsilon) \le \nu(\epsilon)$，对于任意 $\epsilon > 0$和$i \ge 1$，我们有 $\|\Delta_i^{\nu_0^*(\epsilon)}(x_0)\| \le \epsilon$。

那么集合

$$\begin{aligned} &\mathcal{S}^{(0)}(x_0) := \cap_{i=1}^{m} \mathcal{S}_i^{(0)}(x_0), \text{ where} \\ &\mathcal{S}_i^{(0)}(x_0) := \big\{ x : f_i(x_0) + \nabla^{\nu_0^*(\epsilon_0)} f_i(x_0)^{\top} (x - x_0) + \\ &\qquad\qquad 2M_i \|x - x_0\|^2 \le 0 \big\} \end{aligned} \tag{7}$$

如以下定理所示是可行的。

定理1（ [14], 定理1）用于构造 $\mathcal{S}^0(x_0)$ 的所有样本均可行。此外，集合 $\mathcal{S}^{(0)}(x_0)$ 是凸集，且任意 $x \in \mathcal{S}^{(0)}(x_0)$ 都是严格可行的。

在缺少显式约束函数的情况下，局部可行集是若干零阶方法 [6, 13, 14] 中常用的工具，用于确保迭代点的可行性，尽管具体形式各不相同。下文将提出我们的方法，其中使用局部可行集为所推导的下降方向选择步长。

3.2 算法：基于线性规划的安全零阶优化（SZO- LP）

SZO-LP 方法的主要思想如算法 1 所示：通过执行第 7 行中定义于式 (10) 的 LP(x_k, ϵ_k)，迭代选择一个下降方向。由于 LP(x_k, ϵ_k) 中所涉及线性规划的收紧常数为 ϵ_k ，我们得到的下降方向指向可行集的内部。沿着这一方向，我们根据局部可行集和预先定义的长度选择步长（第 9-14 行）

$$\gamma(\epsilon_k) := \frac{\epsilon_k}{4(M_{\max} + L_{\max})}.$$

主要步骤如下:

算法 1 基于线性规划的安全零阶优化（SZO-LP）

输入： ϵ_0, $\epsilon_{\min}$, K_{switch}， 初始可行点 $x_0 \in \Omega$

输出： $\tilde{x}$

$k \leftarrow 0$, $\text{TER} = 0$

当 $\epsilon_k > \epsilon_{\min}$ 时做

  $s_{\text{tmp}} \leftarrow$ 线性规划$(x_k, 2\epsilon_k)$

  如果 $\nabla^{\nu^*_k(2\epsilon_k)} f_0(x_k)^\top s_{\text{tm p}} \leq -4\epsilon_k$ 则

    $\epsilon_{k+1} \leftarrow 2\epsilon_k, \quad x_{k+1} \leftarrow x_k$

  else

    $s_k^* =$ 线性规划(x_k, ϵ_k)

    如果 $\nabla^{\nu^*_k(\epsilon_k)} f_0(x_k)^\top s_k^* \leq -2\epsilon_k$ **then**

      若 $k < K_{\text{switch}}$ 则

$$\beta_k = \arg\max_{\beta \geq 0} \beta \text{ s.t. } x_k + \beta s_k^* \in \mathcal{S}^{(k)}(x_k), \tag{8}$$

$$\alpha_k = \arg\min_{\alpha \in \{\beta_k, \gamma(\epsilon_k)\}} f_0(x_k + \alpha s_k^*) \tag{9}$$

        $x_{k+1} \leftarrow x_k + \alpha_k s_k^*, \quad \epsilon_{k+1} \leftarrow \epsilon_k$

      else

        $x_{k+1} \leftarrow x_k + \gamma\ (\epsilon_k)\ s_k^*, \quad \epsilon_{k+1} \leftarrow \epsilon_k$

      结束条件判断

  else

    $\epsilon_{k+1} \leftarrow \epsilon_k / 2, \quad x_{k+1} \leftarrow x_k$

  结束条件判断

结束条件判断

$k \leftarrow k + 1$

结束 while 循环

3.2.1 提供输入数据

输入包括一个初始严格可行点 x_0（见假设2）和一个收紧常数ϵ_0。算法的每次迭代都会生成一个新的收紧常数 ϵ_k，它可以等于ϵ_{k-1}、 $2\epsilon_{k-1}$ 或 $\epsilon_{k-1}/2$。由于 ϵ_k 收敛到 0（见定理2），用户可以通过为 ϵ_k提供一个下界 $\epsilon_{\min}$ 来控制终止。参数 K_{switch} 标示了两种步长选择方法的分界，见最后一个项目。

3.2.2 构建局部可行集

对于严格可行的 x_k，我们使用 (6) 定义 l_k^* 和

$$\nu_k^*(\epsilon_k) := \min\{l_k^*/\sqrt{d}, \nu(\epsilon_k)\}.$$

然后，我们使用 $\nu_k^*(\epsilon_k)$ 和 (7) 定义 $\mathcal{S}^{(k)}(x_k)$，即围绕 x_k的一个局部可行集。根据定理1，我们知道，如果 $x_{k+1} \in \mathcal{S}^{(k)}(x_k)$，那么 x_{k+1} 也严格可行。

3.2.3 求解下降方向的子问题

在每次迭代中,我们在第7行执行 LP(x_k, ϵ_k) 以求得一个搜索方向，其返回

$$\begin{aligned} \arg\min_{\|s\|_1 \le 1} \ & (\nabla^{\nu_k^*(\epsilon_k)} f_0(x_k))^\top s \\ \text{s.t.} \ & (\nabla^{\nu_k^*(\epsilon_k)} f_i(x_k))^\top s + 2\epsilon_k \le 0, \\ & \forall i \in \mathcal{A}(x_k, \epsilon_k), \end{aligned} \tag{10}$$

如果（10）不可行，则返回 NaN。这里， $\mathcal{A}(x,\epsilon) := \{i : f_i(x) \ge -2\epsilon\}$ 是近活跃约束指标集。（10）的解是一个方向，它不仅能快速下降，还指向可行域 Ω 的内部（远离边界）。在（10）中，由于收紧常数 ϵ_k，沿着第7行的方向 s_k^*，约束函数值会减小。因此，沿方向 s_k^* 移动确实能使我们远离 Ω 的边界。这个方向有助于避免 $-f_i(x_k)$ 取值过小，因为这会导致局部可行集 $\mathcal{S}^{(k)}(x_k)$ 过于保守。此外，只纳入近活跃约束使得（10）规模较小且易于求解。稍后我们将在定理2中看到，ϵ_k 收敛到0。因此，迭代点的某个子序列仍可能收敛到可行集边界上的一点。

我们还令 $s_{\text{tmp}} = \text{LP}(x_k, 2\epsilon_k)$，并在第 4 行检查 $\nabla^{\nu_k^*(2\epsilon_k)} f_0(x_k)^\top s_{\text{tmp}} \le -4\epsilon_k$，从而可以得到命题 1，其证明见附录 A。稍后，我们将使用该命题来证明定理 3 中 $\{x_k\}_{k\ge1}$ 在 k 趋于无穷时的性质。

命题 1 任何 ϵ_k 进入第 7 行时都满足

$$\begin{aligned} \epsilon_k \ge \frac{1}{8}\sup\{\epsilon : & s = \text{LP}(x_k, \epsilon) \ \textit{verifies} \\ & \nabla^{\nu_k^*(\epsilon)} f_0(x_k)^\top s \le -2\epsilon\}. \end{aligned} \tag{11}$$

3.2.4 确定步长

当第 7 行得到的方向 s_k^* 能够实现充分下降（即 $\nabla^{\nu_k^*(\epsilon_k)} f_0(x_k)^\top s_k^* \le -2\epsilon_k$）时，我们沿试探方向 s_k^*移动。为确定步长，我们考虑局部可行集以及预先设定的步长 $\gamma(\epsilon_k)$，该步长可保证实现非平凡下降（见引理 2）。在式 (8) 中，我们通过二分法计算局部可行集内的最大步长，以

在式 (9) 中求得 α_k。当 x_k 不接近可行集边界时，第 10 行使用局部可行集使我们能够获得比 $\gamma(\epsilon)$更大的步长。这是因为，根据式 (7) 的表述， $f_i(x_k)$的值越小， $\mathcal{S}_i^{(k)}(x_k)$的范围就越大，而 $\gamma(\epsilon_k)$与迭代点距离可行集边界的远近无关。当 $k > K_{\text{switch}}$时，我们按照第 12 行的设定，将步长设为 $\gamma(\epsilon_k)$；这有助于证明当 k 趋于无穷大时迭代点的性质（见定理 3）。 K_{switch} 的选择并不重要，因为我们仅在算法早期迭代中使用第 10 行的步长，而非第 12 行定义的步长，以加快下降。

另一方面，如果方向 s_k^* 无法实现充分下降，我们会在第 15 行令 $\epsilon_{k+1} = \epsilon_k/2$ 放宽 (10) 中收紧的约束。这样放宽后，s_{k+1}^* 更容易实现充分下降，即满足 $\nabla^{\nu_{k(\epsilon_{k+1})}^*} f_0(x_{k+1})^\top s_{k+1}^* \le -2\epsilon_{k+1}$。只有当 s_{k+1}^* 实现充分下降时，我们才会沿着 s_{k+1}^* 移动到一个新点。

关于 SZO-LP 与一些当前最先进方法的比较，请参阅备注。

4 方法的收敛性质

本节旨在说明，在温和条件下并令 $\epsilon_{\min} = 0$时，算法 1 生成的序列 $\{x_k\}_{k\ge1}$ 具有一个聚点 x_c，该聚点同时也是问题 (1) 的某个 KKT 对的原始解。首先，我们将在引理 2 中证明，只要 $x_{k+1} \ne x_k$，新的迭代点 x_{k+1} 就是严格可行的，且目标函数值会有非平凡的下降。

引理 2 假设在算法 1 第 7 行得到的 s_k^* 满足

$$\nabla^{\nu_k^*(\epsilon_k)} f_0(x_k)^\top s_k^* \le -2\epsilon_k.$$

我们有， $x_k + \gamma(\epsilon_k)s_k^*$ 严格可行。此外， $x_k + \gamma(\epsilon_k)s_k^*$ 满足

$$f_0(x_k + \gamma(\epsilon_k)s_k^*) - f_0(x_k) < -\epsilon_k^2/(8(M_{\max} + L_{\max})). \tag{12}$$

引理 2 的证明见附录 B。主要思路是利用假设1中的光滑常数，对 $f_i(x_k + \gamma(\epsilon_k))$ 关于 $i \in \mathbb{Z}_0^m$ 给出上界。基于此引理，当 k 趋于无穷时，我们得到关于序列 $\{x_k\}_{k\ge1}$ 和 $\{\epsilon_k\}_{k\ge1}$ 的以下定理。

定理2 以下结论成立：

1. 序列 $\{f_0(x_k)\}_{k\ge1}$ 单调不增；
2. 序列 $\{x_k\}_{k\ge1}$ 至少存在一个聚点。对于任意聚点 x_c，

$$\lim_{k\to\infty} f_0(x_k) = f_0(x_c) > -\infty.$$

3. 序列 $\{\epsilon_k\}_{k\ge1}$ 收敛到 0。

证明。第一点是引理 2 的直接推论，它意味着每当迭代点移动到一个新点时，目标函数值都会下降。

命题 2 的证明。由于 $\{f_0(x_k)\}_{k\ge1}$ 单调不增，因此对任意 $k \ge 1$， $f_0(x_k) \le f_0(x_0)$，从而 $x_k \in \mathcal{P}_\beta$。由于 $\mathcal{P}_\beta$ 有界，根据博尔查诺–魏尔斯特拉斯定理，我们知道存在

至少一个 $\{x_k\}_{k\geq 1}$的聚点。对于任意聚点 x_c，存在一个子序列 $\{x_{k_p}\}_{p\geq 1}$ 收敛到 x_c。由于 $f_0(x)$ 连续，

$$\lim_{p\to\infty} f_0(x_{k_p}) = f_0(x_c) > -\infty.$$

通过利用一个 获得 $\{f_0(x_k)\}_{k\geq 1}$ 的单调性，我们有 $\lim_{k\to\infty} f_0(x_k) = f_0(x_c) > -\infty$.

命题3的证明。我们通过反证法证明这一结果，假设 $\{\epsilon_k\}_{k\geq 1}$ 不会随着 k 趋于无穷而减小。基于这一假设，我们可以证明 $\{\epsilon_k\}_{k\geq 1}$ 不会收敛到任何非零值。如果 $\{\epsilon_k\}_{k\geq 1}$ 收敛到一个非零值，注意到 $\epsilon_k \in \{\epsilon_0 * 2^i : i \in \mathbb{Z}\}$，我们可知存在 $K > 0$，使得任意 $k > K$ 都满足 $\epsilon_k = \epsilon_{k-1}$。那么，对于任意 $k > K$，新的迭代点 x_{k+1} 由算法 1 的第 10 行或第 12 行得到。根据引理 2，$f_0(x_{k+1}) \leq f_0(x_k) - \epsilon_k^2/(8(M_{\max} + L_{\max}))$，因此 $f_0(x_k)$ 在 k 趋于$+\infty$时趋于 $-\infty$，这与第 2 点矛盾。因此，ϵ_k 不收敛。

由于算法 1 $\{\epsilon_k\}_{k\geq 1}$ 有界，我们可以得出结论：$\{\epsilon_k\}_{k\geq 1}$ 有多个聚点。于是存在 $\epsilon > 0$ 以及无限多个 k，使得 $\epsilon_k = \epsilon$ 和 $\epsilon_{k-1} = \epsilon/2$。对于任意这类 k，都存在 $k' \geq k$ 满足 $f_0(x_{k'+1}) \leq f_0(x_{k'}) - \epsilon^2/(8(M_{\max} + L_{\max}))$。因此，存在无限多个 k' 满足 $f_0(x_{k'+1}) \leq f_0(x_{k'}) - \epsilon^2/(8(M_{\max} + L_{\max}))$，这再次与要点 2 矛盾。■

定理2为我们提供了必要的工具，以便在假设 4 下证明 $\{x_k\}_{k\geq 1}$ 的聚点所具有的性质，这将在定理 3 中给出。

假设 4：$\{x_k\}_{k\geq 1}$ 至少有一个聚点 x_c 满足线性无关约束资格条件（LICQ），也就是说，梯度 $\nabla f_i(x_c)$（其中 $i \in \mathcal{A}(x_c, 0)$）线性无关。

假设 4 在优化中被广泛使用 [18]。例如，它被用于证明内点法 [19] 的极限点性质。

定理 3 关于假设 4 中的聚点 x_c，存在 $\lambda_c \in \mathbb{R}^m_{\geq 0}$，使得 (x_c, λ_c) 是 (1) 的一个 KKT 对。

附录 C 中的证明基于矛盾。若 x_c 不是某个 KKT 对的原始解，我们可以找到 $r > 0$、$\epsilon > 0$ 和 $s_\epsilon \in \mathbb{R}^d$，使得对于任意 $x_k \in \mathcal{B}_r(x_c)$，线性规划 $s = \text{LP}(x_k, \epsilon)$ 的解满足$_{\nabla\nu_k^*(\epsilon)} f_0(x_k)^\top s \leq -2\epsilon$。存在无穷多个 k，使得 $x_k \in \mathcal{B}_r(x_c)$，且 s_k^* 是通过算法 1 的第 7 行推导得出的。对于这些 k 中的任意一个，根据 (11)，$\epsilon_k \geq \epsilon/8$，这与定理 2 的第 3 点矛盾。

备注 与 SZO-QQ *[14]* 和 LB-SGD *[13],* 类似，SZO-LP 中的样本均为可行点；在温和假设下，迭代点存在一个聚点，且该聚点也是某个 KKT 对的原始解。相比之下，SZO-LP 的收紧常数 ϵ_k 使迭代点远离可行集的边界，并且其局部可行集比 SZO-QQ 和 LB-SGD 中使用的局部可行集保守性更低。此外，由于使用了近活跃集 $\mathcal{A}(x_k, \epsilon_k)$，子问题 (10) 的规模更小，也比 SZO-QQ 中的二次约束二次规划问题和安全贝叶斯优化方法 *[6, 20]*中的非凸子问题更容易求解。然而，要严格证明这些优势，我们需要在给定精度要求的情况下，对 SZO-LP 所需的迭代次数给出上界；这留待未来工作。

5 最优潮流问题实验

为了展示 SZO-LP 的性能，我们考虑将其应用于 IEEE 30节点系统上的一个 OPF 问题。

5.1 最优潮流问题的建模

为建立最优潮流问题的模型，我们引入以下符号和假设：

- 设 $B = \{b_1, b_2, \ldots, b_n\}$ 为母线集合，并令 $T = \{(b_i, b_j)$：当 b_i 与 $b_j\}$ 之间存在输电线路时，该集合包含一条表示该输电线路的无向边；
- 我们将 P_{G_i}、P_{L_i}、Q_{L_i}、U_i 和 θ_i 分别表示 b_i处的有功功率发电、有功功率消耗、无功功率消耗、电压和电压相角；
- 从 b_i 到 b_j传输的有功功率和无功功率分别记为 $P_{ij}(U_i, U_j, \theta_i, \theta_j)$ 和 $Q_{ij}(U_i, U_j, \theta_i, \theta_j)$，电流记为 $I_{ij}(U_i, U_j, \theta_i, \theta_j)$。这些函数的具体表达式请参阅 [21]；
- 我们还假设母线 b_i 和 $i \in \mathbb{Z}_1^{n_G}$ 上有 n_G 台发电机，而 b_1 是平衡母线，通过提供有功功率来维持电网内的功率平衡，且其电压相角为 0。

因此，最优潮流问题的形式为 [21]

$$\min_{P_{G_i}, U_i, \theta_i} \sum_{i=1}^{n_G} C_i(P_{G_i}) \tag{13a}$$

满足以下约束

$$P_{G_i} = P_{L_i} + \sum_{(i,j)\in T} P_{ij}(U_i, U_j, \theta_i, \theta_j), \quad \forall i \tag{13b}$$

$$-Q_{L_i} = \sum_{(i,j)\in T} Q_{ij}(U_i, U_j, \theta_i, \theta_j), \quad i > n_G \tag{13c}$$

$$P_{G_i} = 0, \quad \text{for } i > n_G, \quad \theta_1 = 0, \tag{13d}$$

$$P_{G,\min} \le P_{G_i} \le P_{G,\max}, \text{for } \quad i \le n_G, \tag{13e}$$

$$I_{ij,\min} \le I_{ij}(U_i, U_j, \theta_i, \theta_j) \le I_{ij,\max}, \quad \forall (i,j) \in T, \tag{13f}$$

$$U_{\min} \le U_i \le U_{\max}, \forall i. \tag{13g}$$

其中，$C_i(\cdot)$ 是一个用于计算发电成本的二次函数，方程 (13e)–(13g) 给出了相应变量的安全区间。

OPF 应用的主要挑战在于系统建模以及推导式 (13) 的精确表达式。困难包括设备动态特性的非线性、缓慢变化的物理参数以及扰动 [22]。不准确的模型可能导致 OPF 求解结果次优（从而增加发电成本），或违反实际的硬性约束（导致设备受损） [23]。因此，我们考虑黑箱设定并采用 SZO-LP。

为此，我们将（13）重新表述为仅含不等式约束的优化问题，以适配 SZO-LP 所采用的（1）。令 $\{P_{G_i}\}_{i=2}^{n_G}$ 和 $\{U_i\}_{i=1}^{n_G}$ 为主要决策变量。随后，通过为$\{P_{G_i}\}_{i=2}^{n_G}$ 和 $\{U_i\}_{i=1}^{n_G}$赋值，可以求解潮流方程（13b）-（13d），从而推导出

所有t （13）中的其他决策变量。因此，（13b）至（13d）为我们提供了函 ns

$$
\begin{aligned}
U_i &= U_i(\{P_{G_j}\}_{j=2}^{n_G}, \{U_j\}_{j=1}^{n_G}), \quad i = n_G + 1, \ldots, n, \\
\theta_i &= \theta_i(\{P_{G_j}\}_{j=2}^{n_G}, \{U_j\}_{j=1}^{n_G}), \quad i = 1, \ldots, n.
\end{aligned} \tag{14}
$$

将（14）代入（13），我们可以得到一种重新表述，其中 $\{\{P_{G_j}\}_{j=2}^{n_G}, \{U_j\}_{j=1}^{n_G}\}$ 是唯一的决策变量，并且不存在等式约束。

5.2 实验结果

我们使用 SZO-LP 求解 IEEE 30节点系统上的一个特定最优潮流问题，其中 $n_G = 6$ 。总共有 11 个决策变量和 158 个约束。在此优化任务中，我们不假设已知系统模型。不过，给定全部 11 个决策变量的取值后，我们可以使用 Matpower [24] 中的黑盒仿真模型，对全部 30 个母线的电压以及电网中所有输电线路的电流进行采样。此外，我们假设可以获得所有决策变量的初始值，以便从一个可行点开始运行 SZO-LP 算法。

我们采用 SZO-LP 来降低由初始决策值引起的二次成本。数值实验在配备 Intel Core i9 处理器的 PC 上进行。我们在 Matlab 中采用的子问题 (10) 求解器为 LINPROG 。我们令 $M_i = M_{\max} = 0.13$ 和 $L_i = L_{\max} = 0.5$。这两个参数的参数调节方法见 [14]。此外，我们设定 $\epsilon_0 = 0.05$、 $\epsilon_{\min} = 10^{-6}$ 和 $K_{\text{switch}} = 200$。

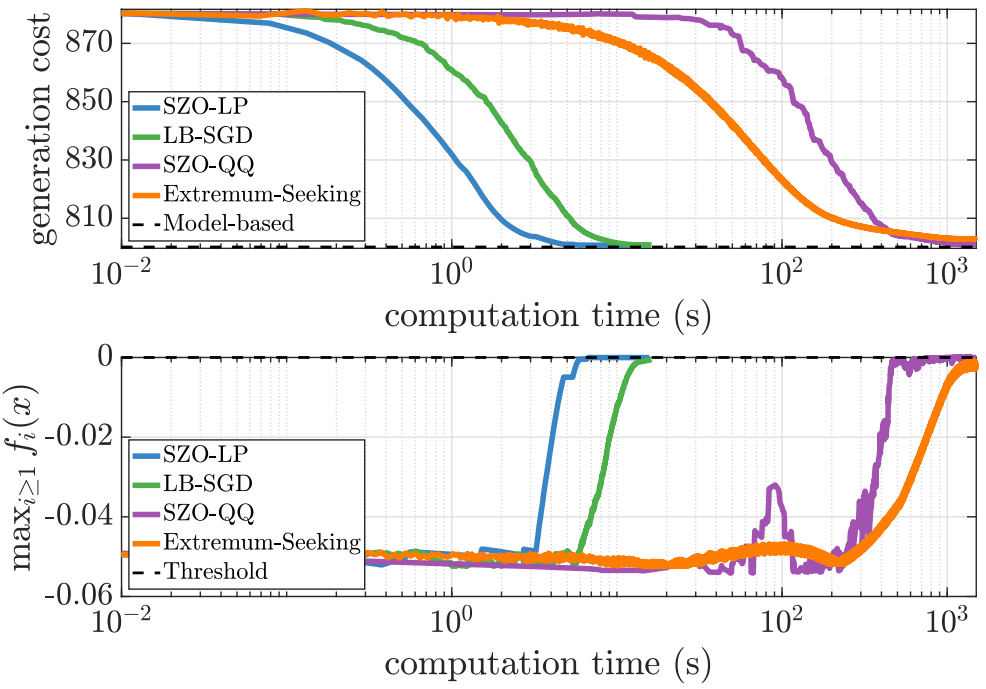


图 1：成本随计算时间减少，最大约束函数值随计算时间增长

在图 1 和图 2 中，我们展示了数值实验结果，并比较了 SZO-LP、SZO-QQ [14], LB-SGD [13] 和极值搜索 [25]的性能。二次约束二次规划

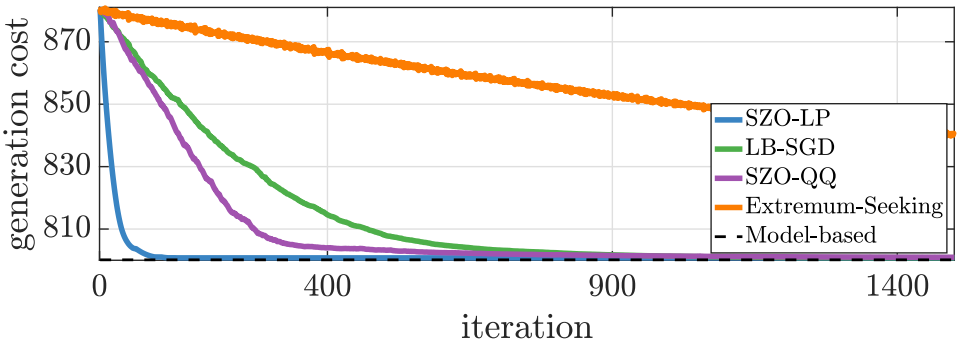


图 2：成本随迭代次数增加而下降

SZO-QQ 中的子问题使用 MOSEK 求解。基于真实模型进行优化并使用 Gurobi [26] 作为求解器，得到最优潮流问题的参考解。图 1 中的计算时间包括通过 Matpower 进行电网仿真的耗时，即查询目标函数和约束函数时所消耗的时间。我们观察到，所有四种方法都能保持迭代点可行，并最终达到非常接近基于真实模型得出的发电成本（800.14）。然而，SZO-LP 的发电成本下降速度快于其他方法。

SZO-LP 在计算时间方面优于 SZO-QQ（如图 1 所示），一个主要原因是线性规划子问题求解得更快。我们注意到，完成前 60 个子问题时，SZO-LP 用时 5.63 秒，而 SZO-QQ 用时 72.06 秒。首先，SZO-LP 的子问题仅考虑近活跃约束，而 SZO-QQ 的子问题则涉及所有约束。在 SZO-LP 的各次迭代中，约束数量最多为 2。其次，尽管图 1 所示的效率差距可能是由我们所选的特定求解器造成的，但总体而言，线性规划可供选择的求解器更多，因此能够实现更高效的求解。

与 SZO-LP 和 SZO-QQ 不同，LB-SGD 和极值搜索无需求解任何子问题，因此能在一定时间内进行更多迭代。这也是 LB-SGD 能在短时间内实现较低发电成本的原因。然而，考虑到这四种方法每次迭代使用的样本数相同，LB-SGD 和极值搜索的样本效率低于 SZO-LP 和 SZO-QQ，因为如图 2 所示，它们需要更多迭代。此外，由于 LB-SGD 和极值搜索基于对数障碍，这两种方法都需要调节障碍函数系数。调节不当可能导致 LB-SGD 的结果次优，甚至使极值搜索不可行。

与 SZO-QQ 相比，SZO-LP 还有另一项优势：在接近 KKT 对的原始解之前，SZO-LP 能使迭代点远离可行集边界。迭代点过于接近可行集边界，可能会阻碍成本下降。从图 1 可以看出，在 SZO-QQ 中，当最大的约束函数值大于 -0.005 时，发电成本的下降会放缓。原因是，当最大的约束函数值接近 0 时，SZO-QQ 构造的局部可行集会变得保守，因此步长会变小。当最大的约束函数值首次大于 -0.005 时，SZO-QQ 的发电成本为 805.27，而 SZO-LP 中相应的成本为 801.77，后者更接近基于真实模型进行优化得出的 800.14。因此，我们可以看到，SZO-QQ 中目标函数值的下降可能会在更早阶段就放缓。

总之，从实验结果可以看出，在四种方法中，SZO-LP 是计算效率最高且样本效率最高的方法。

6 结论

本文提出了一种安全零阶方法 SZO-LP，该方法通过迭代求解线性规划来获得下降方向并确定步长。我们证明，在温和条件下，SZO-LP 的迭代点存在一个聚点，该聚点也是某个 KKT 对的原始解。通过使用 SZO-LP 求解 IEEE 30节点系统上的一个 OPF 问题，并与另外三种方法进行比较的实验，我们发现 SZO-LP 兼具计算高效性和样本高效性。未来研究方向包括推导 SZO-LP 的计算复杂度，以检验其一般情况下是否高效，以及扩展 SZO-LP 以考虑测量噪声。

参考文献

[1] Z. Chu、N. Zhang 和 F. Teng，《微电网的频率约束韧性调度：一种分布鲁棒方法》，IEEE Transactions on Smart Grid，第 12 卷，第 6 期，第 4914–4925 页，2021 年。[2] Y. Chen、A. Orvieto 和 A. Lucchi，《有限和凸函数的加速无导数优化算法》，载于第 37 届国际机器学习会议论文集，Proceedings of Machine Learning Research 第 119 卷，第 1681–1690 页，2020 年 7 月 13–18 日。[3] Z. Manchester 和 S. Kuindersma，《基于无迹动态规划的无导数轨迹优化》，载于 2016 年 IEEE 第 55 届决策与控制会议（CDC），第 3642–3647 页，IEEE，2016 年。[4] C. V. Rao、S. J. Wright 和 J. B. Rawlings，《内点法在模型预测控制中的应用》，Journal of Optimization Theory and Applications，第 99 卷，第 3 期，第 723–757 页，1998 年。[5] I. Bajaj、A. Arora 和 M. Hasan，《黑箱优化：方法与应用》，第 35–65 页。Springer，2021 年。[6] Y. Sui、A. Gotovos、J. Burdick 和 A. Krause，《利用高斯过程进行安全优化探索》，载于国际机器学习会议，第 997–1005 页，PMLR，2015 年。[7] M. Turchetta、F. Berkenkamp 和 A. Krause，《交互式机器学习的安全探索》，Advances in Neural Information Processing Systems，第 32 卷，第 2887–2897 页，2019 年。[8] L. Sabug Jr、F. Ruiz 和 L. Fagiano，《Smgo-δ：在带黑箱约束的全局优化中平衡谨慎与收益》，Information Sciences，第 605 卷，第 15–42 页，2022 年。[9] A. P. Vinod、A. Israel 和 U. Topcu，《利用利普希茨连续梯度对未知函数进行约束全局优化》，SIAM Journal on Optimization，第 32 卷，第 2 期，第 1239–1264 页，2022 年。[10] R. M. Lewis 和 V. Torczon，《一种用于带一般约束和简单界优化问题的全局收敛增广拉格朗日模式搜索算法》，SIAM Journal on Optimization，第 12 卷，第 4 期，第 1075–1089 页，2002 年。

[11] C. Audet 和 J. E. Dennis Jr，《无导数非线性规划的渐进屏障》，SIAM Journal on Optimization，第 20 卷，第 1 期，第 445–472 页，2009 年。[12] L. Hazeleger、D. Ne˘si´c 和 N. van de Wouw，《用于非线性动力系统约束优化的采样数据极值寻优框架》，《自动化》，第 142 卷，第 110415 页，2022 年。[13] I. Usmanova、Y. As、M. Kamgarpour 和 A. Krause，《用于安全黑箱优化的对数障碍及其在安全强化学习中的应用》，arXiv 预印本 arXiv:2207.10415，2022 年。[14] B. Guo、Y. Jiang、G. Ferrari-Trecate 和 M. Kamgarpour，《基于二次局部近似的安全零阶优化》，arXiv 预印本 arXiv:2303.16659，2023 年。[15] H. Abdi、S. D. Beigvand 和 M. La Scala，《智能电网和微电网最优潮流研究综述》，《可再生与可持续能源评论》，第 71 卷，第 742–766 页，2017 年。[16] I. Usmanova、A. Krause 和 M. Kamgarpour，《不确定约束下的安全凸学习》，载于第 22 届国际人工智能与统计学会议，第 2106–2114 页，PMLR，2019 年。[17] A. S. Berahas、L. Cao、K. Choromanski 和 K. Scheinberg，《无导数优化中梯度近似的理论与实证比较》，Foundations of Computational Mathematics，第 22 卷，第 2 期，第 507–560 页，2022 年。[18] G. Wachsmuth，《关于线性无关约束资格条件与拉格朗日乘子的唯一性》，《运筹学快报》，第 41 卷，第 1 期，第 78–80 页，2013 年。[19] J. Nocedal 和 S. J. Wright，《数值优化》。Spinger，2006 年。[20] F. Berkenkamp、A. Krause 和 A. P. Schoellig，《带安全约束的贝叶斯优化：机器人学中的安全自动参数调优》，《机器学习》，第 1–35 页，2021 年。[21] J. Das，《潮流优化与最优潮流》。CRC出版社，2017 年。[22] Z. Chu、S. Lakshminarayana、B. Chaudhuri 和 F. Teng，《利用网络弹性经济调度缓解针对电网的负荷改变攻击》，IEEE Transactions on Smart Grid，2022 年，提前出版。[23] D. Lee、K. Turitsyn、D. K. Molzahn 和 L. A. Roald，《采用鲁棒凸限制的鲁棒交流最优潮流》，《IEEE电力系统汇刊》，第 36 卷，第 6 期，第 4953–4966 页，2021 年。[24] R. D. Zimmerman、C. E. Murillo-S´anchez 和 R. J. Thomas，《Matpower：用于电力系统研究与教育的稳态运行、规划和分析工具》，《IEEE电力系统汇刊》，第 26 卷，第 1 期，第 12–19 页，2010 年。[25] D. B. Arnold、M. Negrete-Pincetic、M. D. Sankur、D. M. Auslander 和 D. S. Callaway，《配电系统中无功资源的无模型最优控制：一种极值搜索方法》，《IEEE电力系统汇刊》，第 31 卷，第 5 期，第 3583–3593 页，2015 年。[26] Gurobi Optimization, LLC，《Gurobi优化器参考手册》，2022 年。

附录

A 命题 1 的证明

我们首先证明，若存在某个 $k>0$ 和 $\epsilon_\alpha>0$

$$s_1=\mathtt{LP}(x_k,\epsilon_\alpha)\text{ verifies }\nabla^{\nu_k^*(\epsilon_\alpha)}f_0(x_k)^\top s_2\le -2\epsilon_\alpha \tag{15}$$

则对于任意 $\epsilon_\beta\le\epsilon_\alpha/4$

$$s_2=\mathtt{LP}(x_k,\epsilon_\beta)\text{ verifies }\nabla^{\nu_k^*(\epsilon_\beta)}f_0(x_k)^\top s_2\le -2\epsilon_\beta. \tag{16}$$

由 (15) 可知， s_1 与 $\|s_1\|_1\le 1$ 是 LP(x_k, ϵ_β) 中所涉及的线性规划的一个可行解。这是因为对于任意 $i\in\mathcal{A}(x_k,\epsilon_\beta)\subset\mathcal{A}(x_k,\epsilon_\alpha)$，我们有

$$\begin{aligned}
&\langle\nabla^{\nu_k^*(\epsilon_\beta)}f_i(x_k),s_1\rangle\\
\le&\langle\nabla f_i(x_k),s_1\rangle+|\langle\nabla f_i(x_k)-\nabla^{\nu_k^*(\epsilon_\beta)}f_i(x_k),s_1\rangle|\\
\le&\langle\nabla^{\nu_k^*(\epsilon_\alpha)}f_i(x_k),s_1\rangle+|\Delta_i^{\nu_k^*(\epsilon_\alpha)}(x_k)|+|\Delta_i^{\nu_k^*(\epsilon_\beta)}(x_k)|\\
\le&-2\epsilon_\alpha+\epsilon_\alpha+\epsilon_\alpha/4<-\epsilon_\alpha/2\le-2\epsilon_\beta.
\end{aligned}$$

类似地，我们可以证明

$$\langle\nabla^{\nu_k^*(\epsilon_\beta)}f_0(x_k),s_1\rangle\le-2\epsilon_\beta.$$

考虑到 s_2 是 LP(x_k,ϵ_β) 中涉及的线性规划的最优解，(16) 成立。

那么，如果 ϵ_k 进入算法 1 的第 7 行，第 4 行的条件“$s_{\text{tmp}}=\mathtt{LP}(x_k,2\epsilon_k)$ verifying$_{\nabla^{\nu_k^*(2\epsilon_k)}f_0(x_k)^\top s_{\text{tmp}}\le-4\epsilon_k}$”不成立。令

$$\epsilon_\alpha=\sup\{\epsilon:s=\mathtt{LP}(x_k,\epsilon)\text{ verifies }\nabla^{\nu_k^*(\epsilon)}f_0(x_k)^\top s\le-2\epsilon\},$$

对于任意 $\epsilon_\beta\le\epsilon_\alpha/4$，(16) 成立。因此， $2\epsilon_k\ge\epsilon_\alpha/4$。 ■

B 引理 2 的证明

首先，我们证明 $x_k+\gamma(\epsilon_k)s_k^*$ 确实是严格可行的。利用中值定理并注意到 $\|s_k^*\|\le\|s_k^*\|_1\le 1$，可得对于任意 $\gamma>0$

$$\begin{aligned}
&f_i(x_k+\gamma s_k^*)\\
<&f_i(x_k)+\gamma\nabla f_i(x_k)^\top s_k^*+2\gamma^2M_{\max}\|s_k^*\|^2\\
<&\gamma\nabla^{\nu_k^*(\epsilon_k)}f_i(x_k)^\top s_k^*+\gamma\|\Delta_i^{\nu_k^*(\epsilon_k)}(x)\|\cdot\|s_k^*\|+2M_{\max}\gamma^2\\
<&-2\epsilon_k\gamma+\epsilon_k\gamma+2M_{\max}\gamma^2\\
<&2(M_{\max}+L_{\max})\gamma^2-\epsilon_k\gamma,\ \forall i\in\mathcal{A}(x_k,\epsilon_k),
\end{aligned}\tag{17}$$

$$\begin{aligned}
&f_i(x_k+\gamma s_k^*)\\
<&f_i(x_k)+L_{\max}\gamma,\ \forall i\in\mathbb{Z}_1^m\setminus\mathcal{A}(x_k,\epsilon_k).
\end{aligned}\tag{18}$$

因此，我们有

$$f_i(x_k+\gamma(\epsilon_k)s_k^*)<-\epsilon_k^2/(8(M_{\max}+L_{\max}))<0$$

对于任意 $i \in \mathcal{A}(x_k, \epsilon_k)$ 以及

$$f_i(x_k + \gamma(\epsilon_k)s_k^*) < -\epsilon_k/2$$

对于任意 $i \in \mathbb{Z}_1^m \setminus \mathcal{A}(x_k, \epsilon_k)$。因此， $x_k + \gamma(\epsilon_k)s_k^*$ 是严格可行的。

类似地，对于 $\gamma = \gamma(\epsilon_k)$，目标函数满足

$$f_0(x_k + \gamma s_k^*) < f_0(x_k) + 2(M_{\max} + L_{\max})\gamma^2 - \epsilon_k\gamma.$$

因此， $f_0(x_k + \gamma s_k^*) < f_0(x_k) - \epsilon_k^2/(8(M_{\max} + L_{\max}))$。

C 定理 3 的证明

我们仅考虑 $\mathcal{A}(x_c, 0)$ 非空的情形。证明可轻易推广至$\mathcal{A}(x_c,0) = \emptyset$。

我们通过反证法证明该结果，假设不存在 $\lambda_c \in \mathbb{R}^m_{\geq 0}$ 使得 (x_c, λ_c) 是一个 KKT 对。于是，以下论述中恰有一个成立:

1) $\nabla f_0(x_c)$ 不是 $\nabla f_i(x_c)$、 $i \in \mathcal{A}(x_c, 0)$ 的线性组合,
2) $\nabla f_0(x_c) = \sum_{i \in \mathcal{A}(x_c,\ 0)} \lambda_i \nabla f_i(x_c)$，并且存在 $i^* \in \mathcal{A}(x_c, 0)$，使得 $\lambda_{i^*} > 0$.

下面我们证明，无论哪种情况成立，我们总能找到 $s \in \mathbb{R}^d$ ，使得

$$\langle \nabla f_0(x_c), s\rangle < 0,\ \langle \nabla f_i(x_c), s\rangle \leq 0,\ \forall i \in \mathcal{A}(x_c, 0). \tag{19}$$

若 1) 成立，则令 $g_{\|}$ 为 $\nabla f_0(x_c)$ 在张成空间 $\{\nabla f_i(x_c)\ i \in \mathcal{A}(x_c,0)\}$ 上的投影，且 $g_\perp := \nabla f_0(x_c) - g_{\|}$。于是 $g_\perp \neq 0$, $\langle \nabla f_0(x_c)\ g_\perp\rangle > 0$ 且 $\langle \nabla f_i(x_c)\ g_\perp\rangle = 0$, $\forall i \in \mathcal{A}(x_c,0)$。因此， $s = -g_\perp$ 满足 (19)。

若条件 2) 成立，不妨设 $\mathcal{A}(x_c,0) \neq \{i^*\}$。然后令 $h_{\|}$ 为 $\nabla f_{i^*}(x_c)$ 在 $\mathrm{span}\{\nabla f_i(x_c)\ i \in \mathcal{A}(x_c,0)$ 和 $i \neq i^*\}$ 上的投影，并令 $h_\perp := \nabla f_{i^*}(x_c) - h_{\|}$。由于满足线性无关约束资格条件， $h_\perp \neq 0$。可以验证， $s = -h_\perp$ 也满足 (19)。

然后我们注意到，由于集合 $\{s : (19)\ \mathrm{holds}\}$ 非空，存在 $\epsilon > 0$ 和 s_ϵ ，使得 $\|s_\epsilon\|_1 = 1$

$$\langle \nabla f_i(x_c), s_\epsilon\rangle \leq -4\epsilon,\ \forall i \in \mathcal{A}(x_c, 0) \cup \{0\}. \tag{20}$$

为说明这一结果，我们假设 $\bar{s} \in \mathbb{R}^d$ 满足 $(\nabla f_0(x_c))^\top \bar{s} < 0$ 和 $(\nabla f_i(x_c))^\top \bar{s} \leq 0$ ，对任意 $i \in \mathcal{A}(x_c,0)$ 均成立。令 $\mathcal{A}(x_c, 0) = \{i_1, \ldots, i_l\}$。存在 $y \in \mathbb{R}^d$ 使得

$$Jy = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix}, \text{ where } J = \begin{bmatrix} \nabla f_{i_1}(x_c)^\top \\ \vdots \\ \nabla f_{i_l}(x_c)^\top \end{bmatrix}, \tag{21}$$

因为根据线性无关约束资格条件， J 是满行秩的。因此，存在 $\sigma > 0$ 使得 t

$$\delta_i := -(\nabla f_i(x_c))^\top(\bar{s} + \sigma y) > 0,\ \forall i \in \mathcal{A}(x_c, 0) \cup \{0\}.$$

Then $s_\epsilon = s_\epsilon^* := (\bar{s} + \sigma y)/\|\bar{s} + \sigma y\|_1$ 和 $\epsilon = \epsilon^* := \frac{1}{4}\min_i \delta_i/\|\bar{s} + \sigma y\|_1$ 满足（20） .

由于 $\nabla f_i(x)$ 关于 $i \in \mathbb{Z}_0^m$ 连续，存在 $r > 0$ ，使得任意 $x \in \mathcal{B}_r(x_c)$ 都满足

$$\langle \nabla f_i(x), s_\epsilon^*\rangle \leq -3\epsilon^*,\ \forall i \in \mathcal{A}(x_c, 0) \cup \{0\}. \tag{22}$$

由于 x_c 是一个聚点，且 $\{\epsilon_k\}_{k\geq 1}$ 收敛于 0，因此存在无穷多个 k，使得

$$\begin{aligned} &k > K_{\text{switch}},\ x_k \neq x_{k+1}, \\ &\mathcal{A}(x_c, \epsilon_k) \subset \mathcal{A}(x_c, 0),\ x_k \in \mathcal{B}_r(x_c). \end{aligned} \tag{23}$$

对于任意这样的 ks，考虑式 (22) 以及任意的 i

$$\langle \Delta_i^{\nu_k^*(\epsilon^*)}(x), s_\epsilon^* \rangle \leq |\Delta_i^{\nu_k^*(\epsilon^*)}(x)| \cdot \|s_\epsilon^*\| \leq \epsilon^*,$$

我们有

$$\langle \nabla^{\nu_k^*(\epsilon^*)} f_i(x), s_\epsilon^* \rangle \leq -2\epsilon^*,\ \forall i \in \mathcal{A}(x_c, 0) \cup \{0\}. \tag{24}$$

根据算法 1 可知，对于任意满足式 (23) 的 k s，x_{k+1} 由第 12 行得出，而s_k^* 由第 7 行得出。因此，我们可以利用命题 1 和式 (24) 得出结论：对于无穷多个 k，$\epsilon_k \geq \epsilon^*/8$ 。然而，这一结论与定理 2 的第 3 点矛盾。

■