arXiv:2301.00477v2 [math.OC] 6 Oct 2024

用于非线性等式约束随机优化的具有高概率复杂度界的序列二次规划方法*

ALBERT S. BERAHAS†、MIAOLAN XIE‡、和BAOYU ZHOU§

摘要。 本文提出一种步长搜索序列二次规划方法，用于求解非线性等式约束随机优化问题。假设约束函数值及其导数可用，但目标函数及其相关导数只能通过非精确概率零阶和一阶预言机计算得到随机近似。在合理假设下，本文推导出算法达到一阶 ε-驻定迭代点所需迭代次数的高概率界，其中 ε 由非精确概率零阶和一阶预言机的噪声水平决定，并且有一个正下界。针对标准非线性优化测试问题的数值结果展示了所提方法的优点和局限性。

关键词。非线性优化、约束随机优化、序列二次优化、步长搜索、概率预言机

AMS 学科分类。49M05, 49M10, 49M37, 65K05, 65K10, 90C15, 90C30, 90C55

1. 引言。我们提出一种步长搜索[1] 序列二次规划（SQP）算法，用于求解以下形式的非线性等式约束随机优化问题

$$\min_{x\in\mathbb{R}^n} \; f(x) \quad \text{s.t. } c(x)=0, \tag{1.1}$$

其中， $f\colon \mathbb{R}^n\to\mathbb{R}$ 和 $c\colon \mathbb{R}^n\to\mathbb{R}^m$ 均为连续可微函数。我们考虑这样一种情形：无法获得目标函数的精确函数和导数信息，只能通过不精确概率预言机获得 $\bar{f}(x;\Xi^0(x))$和 $\bar{g}(x;\Xi^1(x))$，即目标函数 $f(x)$及其一阶导数 $\nabla f(x)$的随机估计；其中， $\Xi^0(x)$和 $\Xi^1(x)$（概率空间为$(\Omega,\mathcal{F}_\Omega,P)$）分别表示目标函数和梯度估计中的底层随机性。另一方面，假设约束函数值 $c(x)$及其雅可比矩阵$\nabla c(x)^T$ 均可获得。这类具有确定性约束的随机优化问题广泛见于科学与工程应用，包括但不限于计算机视觉 [39], 、多阶段优化 [41], 、自然语言处理 [32], 、网络优化 [9], 以及 PDE 约束优化 [37]。

目前，绝大多数用于确定性求解随机等式约束优化问题的方法，都采用投影或罚函数方法。前一类方法（例如随机投影方法 [22, 24–26]）

*本材料基于美国海军研究办公室通过项目编号 N00014-21-1-2532 提供的资助。

†美国密歇根大学工业与运营工程系，安娜堡，密歇根州，美国；电子邮箱：albertberahas@gmail.com

‡美国印第安纳州西拉斐特普渡大学工业工程学院；电子邮箱：xie537@purdue.edu§美国亚利桑那州坦佩亚利桑那州立大学计算与增强智能学院；电子邮箱：baoyu.zhou@asu.edu[1]我们使用术语 *step-search* 方法，该术语由 [23] 提出，以便与线搜索方法加以区分。步长搜索方法与线搜索方法相似，但在回溯过程中，搜索（步长）方向可以改变。

要求可行域满足严格条件，以确保定义良好；但一般的非线性函数并不满足这些条件，因此这类方法不易直接应用。相比之下，随机罚函数方法 [14, 36], 不会对可行域施加此类条件。这些方法通过在目标函数中施加约束惩罚，将约束问题转化为无约束问题，再应用随机算法求解转化后的无约束问题。随机罚函数方法易于实现且经过充分研究，但其实际表现对参数选择和病态条件较为敏感，通常不如将约束作为约束处理的方法。

近期，人们开发了一类随机序列二次规划方法，用于求解（1.1）。从经验上看，这些方法优于随机罚函数方法，并且具有期望意义下的收敛保证 [7, 29]。在 [7], 中，作者针对完全随机情形提出了一种不含目标函数的随机序列二次规划方法，并采用自适应步长。相比之下，在 [29], 中，作者针对函数和导数近似误差可减小的情形，提出了一种随机步长搜索（论文中称为线搜索 [29]）序列二次规划方法。我们注意到，这两篇论文 [7, 29], 中的若干算法选择，例如优值函数和优值参数，并不相同。此后，人们提出了若干扩展方法 [3, 6, 8, 17, 30, 34], ；然而，由于约束情形所带来的困难以及随机性的影响，这些研究（以及文献中的其他研究）很少推导最坏情况迭代复杂度（或样本复杂度）。值得注意的例外包括 [16], 其中作者为 [7], 中提出的算法给出了收敛率（及复杂度保证）；以及[3, 31] ，其中作者在附加假设下，并针对误差可减小的情形，为随机序列二次规划方法的若干变体给出了复杂度界。我们注意到，除 [34], 之外，上述所有方法均假设可以获得无偏的梯度估计（以及必要时的函数值估计）。

对于上述所有方法而言，最关键的因素是目标函数及其导数的随机估计的质量和可靠性。在我们的设定中，假设目标函数及其导数均无法直接获取，算法只能通过不精确的概率零阶和一阶预言机获取它们的随机近似（其精确定义将在第 2.3 节介绍）。此类预言机已在多项研究中提出并加以应用，例如 [1, 12, 19, 23, 27, 40]。此外，这些概率预言机及其变体也已应用于直接搜索方法 [19, 38], 、信赖域方法 [1, 10, 15, 20], 和步长搜索方法 [2, 13, 29, 35]。值得注意的是，只有 [29]考虑了带（等式）约束的情形，但未给出迭代复杂度（或样本复杂度）结果。

1.1. 贡献。本文设计、分析并实现了一种步长搜索序列二次规划（SS-SQP）方法，用于求解非线性等式约束随机优化问题；在这些问题中，精确的约束函数值和导数可用，但只能计算目标函数及其相关导数的随机近似。这些随机近似通过不精确的概率零阶和一阶预言机计算得到，这些预言机与 [23], 中的预言机类似，具有控制近似精度和可靠性的参数，并允许使用有偏近似。所提出的算法受先进的线搜索序列二次规划方法 [11] 启发，并结合了近期针对无约束随机情形开发的随机自适应步长搜索框架 [23] 。在每次迭代中，该算法构造优值函数下降模型，该模型兼具两种用途：作为衡量充分进展的指标（计算工作量的一部分）和收敛性的代理。为缓解目标函数评估中的噪声所带来的挑战，我们的步长搜索方法采用了类似于 [23], 中所提出的宽松充分下降条件。

）以及收敛性的代理。为减轻目标函数评估中噪声所带来的挑战，我们的步长搜索方法采用了类似于 [4] 中所提出的宽松充分下降条件。在合理假设下，我们为所提出的算法给出了高概率迭代 ε复杂度界。具体而言，我们证明，以压倒性高的概率，我们的算法会在 ε次迭代内生成一个一阶 $\mathcal{O}(\varepsilon^{-2})$ 驻点，其中ε 有正的下界，且其下界由零阶和一阶预言机中的噪声与偏差决定。当可以计算精确的目标函数值和梯度值时，该 ε复杂度界与 [16]中的确定性算法相匹配。在 [16], 中，作者还为 [7] 中提出的随机算法提供了非渐近收敛保证（此前仅证明了其渐近收敛性）。我们的贡献与 [16] 之间存在两个关键差异：（i）我们的算法需要获得目标函数的估计值，而 [16] 中的方法无需目标函数；以及（ii）我们的一阶预言机仅以一定概率提供具有充分精度的估计值，否则可能提供任意糟糕的估计值。最后，在标准非线性等式约束测试问题上的实验 [18]展示了所提出算法的效率和有效性。

1.2. 记号。令 $\mathbb{R}$ 表示实数集，$\mathbb{R}^n$ 表示 n 维实向量集，$\mathbb{R}^{m\times n}$ 表示 m×n 维实矩阵集，$\mathbb{N}$ 表示自然数集，$\mathbb{S}^n$ 表示 n×n 维实对称矩阵集。对于任意 $a\in\mathbb{R}$，令 $\mathbb{R}_{>a}$ ($\mathbb{R}_{\geq a}$) 表示严格大于（大于或等于） a 的实数集。我们用 $\|\cdot\|$ 表示 ℓ_2-范数。我们用 $k\in\mathbb{N}$ 作为算法的迭代计数器。为简洁起见，我们用下标 k 表示第 k 个迭代点处的信息，例如 $f_k := f(x_k)$。我们分别用 $\bar{f}(x;\,\xi^0(x))$ 和 $\bar{g}(x;\,\xi^1(x))$ 表示 $\bar{f}(x;\,\Xi^0(x))$ 和 $\bar{g}(x;\,\Xi^1(x))$ 的实现；参见第 2.3 节。

1.3. 论文结构。本文其余部分安排如下。第2节介绍算法框架。第3节对算法进行分析。第4节报告数值结果。第5节给出总结性评述和未来研究方向。

2. 算法。为求解（1.1），我们设计了一种基于 SQP 范式的迭代随机算法。该算法的每次实现都会生成以下序列：（i）原始迭代序列 $\{x_k\}$，（ii）原始试探迭代序列$\{x_k^+\}$，（iii）原始搜索方向序列 $\{\bar{d}_k\}$，（iv）对偶迭代点序列 $\{\bar{y}_k\}$，（v）步长序列 $\{\alpha_k\}$，（vi）优值参数序列 $\{\bar{\tau}_k\}$，以及（vii）试探优值参数序列 $\{\bar{\tau}_k^{\text{trial}}\}$。上述序列都是某个随机过程的实现。例如，原始迭代序列 $\{x_k\}$ 是随机过程 $\{X_k\}\subset\mathbb{R}^n$的一种实现，而原始试探迭代序列 $\{x_k^+\}$ 是另一个随机过程$\{X_k^+\}\subset\mathbb{R}^n$的一种实现。下面我们将逐一讨论这些序列。在本文余下部分中，我们始终作出以下假设。

Assumption 2.1. 设 $\mathcal{X}\subseteq\mathbb{R}^n$ 为一个开凸集，包含原始迭代序列 $\{X_k\}$ 和原始试探迭代序列 $\{X_k^+\}$。目标函数$f:\ \mathbb{R}^n\to\mathbb{R}$ 在 $\mathcal{X}$ 上连续可微且有下界。目标梯度函数 $\nabla f:\ \mathbb{R}^n\to\mathbb{R}^n$ 是 *L-Lipschitz* 连续的，且在 $\mathcal{X}$ 上有界。约束函数 c: $\mathbb{R}^n\to\mathbb{R}^m$ （其中 $m\leq n$）在 $\mathcal{X}$ 上连续可微且有界，并且对所有 $i\in\{1,\ldots,m\}$，每个梯度 $\nabla c_i:\ \mathbb{R}^n\to\mathbb{R}^n$ 都是 γ_i-*Lipschitz* 连续的，且在 $\mathcal{X}$ 上有界。 $J:=\nabla c^T$ 的奇异值有界

在 $\mathcal{X}$ 上远离零。

假设 2.1 是确定性约束优化文献中的一项标准假设 [11, 33, 43]。假设 2.1 隐含地涉及算法的行为。本质上，只要算法没有恰好生成无界的迭代点序列，它就会收敛到有意义的结果。这一逻辑及其局限性同样适用于确定性情形 [43, 假设 G]和 [11, 假设 4。1]。我们的分析也需要这一假设。在假设 2.1 下，存在常数 $\{\kappa_g, \kappa_c, \kappa_J, \kappa_\sigma\} \subset \mathbb{R}_{>0}$ 和 $f_{\inf} \in \mathbb{R}$，使得对所有 $k \in \mathbb{N}$,

$$f_{\inf} \leq f_k,\ \|\nabla f_k\| \leq \kappa_g,\ \|c_k\|_1 \leq \kappa_c,\ \|J_k\| \leq \kappa_J,\ \text{and}\ \|(J_k J_k^T)^{-1}\| \leq \kappa_\sigma.$$

需要指出的是，根据假设 2.1，线性无关约束资格条件（LICQ）成立。此外，在假设 2.1 下，对所有 $x \in \mathbb{R}^n$， $d \in \mathbb{R}^n$ 以及 $\alpha \in \mathbb{R}_{\geq 0}$ 可得

$$\begin{aligned} &f(x+\alpha d) \leq f(x) + \alpha \nabla f(x)^T d + \tfrac{L}{2}\alpha^2 \|d\|^2 \\ &\text{and}\ \ \|c(x+\alpha d)\|_1 \leq \|c(x) + \alpha \nabla c(x)^T d\|_1 + \tfrac{\Gamma}{2}\alpha^2\|d\|^2,\ \ \text{where}\ \Gamma = \sum_{i=1}^{m} \gamma_i. \end{aligned} \tag{2.1}$$

在本文中，我们尤其关注寻找满足（x， y） $\in \mathbb{R}^n \times \mathbb{R}^m$ 的一阶驻点条件的原始-对偶迭代点（1.1）。为此，令 $\mathcal{L}$： $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 为（1.1）的拉格朗日函数，定义如下:

$$\mathcal{L}(x, y) = f(x) + y^T c(x), \tag{2.2}$$

其中， $y \in \mathbb{R}^m$ 为对偶变量。（1.1）的一阶驻点条件由假设2.1保证为必要条件（这是由于包含了 LICQ），具体如下:

$$0 = \begin{bmatrix} \nabla_x \mathcal{L}(x, y) \\ \nabla_y \mathcal{L}(x, y) \end{bmatrix} = \begin{bmatrix} \nabla f(x) + \nabla c(x) y \\ c(x) \end{bmatrix}. \tag{2.3}$$

本节余下部分将介绍关键算法组件：优值函数及其相关模型、搜索方向的计算与优值参数更新机制，以及非精确概率零阶和一阶预言机。主要算法为算法2.1。

2.1. 优值函数。优值函数 ϕ： $\mathbb{R}^n \times \mathbb{R}_{>0} \to \mathbb{R}$ 定义如下:

$$\phi(x, \tau) := \tau f(x) + \|c(x)\|_1, \tag{2.4}$$

其中， $\tau \in \mathbb{R}_{>0}$（即优值参数）用于平衡目标函数与约束违反程度。给定梯度（近似值） $g \in \mathbb{R}^n$ 和搜索方向$d \in \mathbb{R}^n$，优值函数模型 l： $\mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ 定义如下:

$$l(x, \tau, g, d) := \tau(f(x) + g^T d) + \|c(x) + \nabla c(x)^T d\|_1.$$

给定满足线性化可行性的搜索方向 $d \in \mathbb{R}^n$ ，即 $c(x) + \nabla c(x)^T d = 0$，优值函数模型 Δl 的缩减量： $\mathbb{R}^n \times \mathbb{R}_{>0} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ 定义为

$$\Delta l(x, \tau, g, d) := l(x, \tau, g, 0) - l(x, \tau, g, d) = -\tau g^T d + \|c(x)\|_1. \tag{2.5}$$

我们利用优值函数模型（2.5）的下降量来监测所提出算法的进展。我们将在第 2.2 节中对此进行更详细的讨论。

2.2. 算法组件。现在我们说明如何在算法 2.1 的任意实现中：（i）计算原始搜索方向序列 $\{d_k\}$；（ii）更新优值参数序列 $\{\bar{\tau}_k\}$；以及（iii）更新原始迭代序列 $\{x_k\}$。这些序列取决于目标函数梯度序列的近似值 $\{\bar{g}(x_k;\ \Xi^1(x_k))\}$。对于任意 $x_k \in \mathbb{R}^n$，令 $\bar{g}(x_k;\xi^1(x_k))$ 表示$\bar{g}(x_k;\Xi^1(x_k))$ 的一个实现。由于 $\{x_k\}$ 是随机过程 $\{X_k\} \subset \mathbb{R}^n$ 的一个实现，我们将 $G_k = \bar{g}(X_k;\ \Xi^1(X_k))$ 定义为具有实现 $\bar{g}(x_k;\xi^1(x_k))$。为简化记号，本小节中我们省略对随机性的依赖，例如 $\bar{g}_k = \bar{g}(x_k;\xi^1(x_k))$。

对于所有 $k \in \mathbb{N}$ 在算法2.1的任一实现中，原始搜索方向$d_k \in \mathbb{R}^n$ 和对偶变量 $\bar{y}_k \in \mathbb{R}^m$ 均通过求解线性方程组计算得到

$$\begin{bmatrix} H_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} \bar{d}_k \\ \bar{y}_k \end{bmatrix} = - \begin{bmatrix} \bar{g}_k \\ c_k \end{bmatrix}, \tag{2.6}$$

其中，$\{H_k\}$ 满足以下假设。

Assumption 2.2. 对于所有 $k \in \mathbb{N}$，$H_k \in \mathbb{S}^n$ 均独立于 G_k，其中 $\{G_k\}$ 是一个随机过程，其实现为 $\{\bar{g}_k\}$（梯度估计序列）。此外，存在常数 $\{\kappa_H, \zeta\} \subset \mathbb{R}_{>0}$，使得对于算法2.1的任一实现中的所有 $k \in \mathbb{N}$ $\|H_k\| \le \kappa_H$ 以及 $u^T H_k u \ge \zeta \|u\|^2$ 对于任意 $u \in$ 零空间(J_k)。

众所周知，在假设2.1 和2.2下，方程（2.6）存在唯一解（d_k，$\bar{y}_k$），因此，向量 $d_k \in \mathbb{R}^n$ 和 $\bar{y}_k \in \mathbb{R}^m$ 均有明确定义 [33]。

接下来，我们介绍优值参数更新机制。给定常数$\{\epsilon_\tau, \sigma\} \subset (0,\ 1)$，对于 $k \in \mathbb{N}$ 在算法2.1的任意一次实现中，我们通过 $\bar{\tau}_k$ 计算

$$\bar{\tau}_k \leftarrow \begin{cases} \bar{\tau}_{k-1} & \text{if } \bar{\tau}_{k-1} \le \bar{\tau}_k^{\text{trial}}; \\ \min\left\{(1-\epsilon_\tau)\bar{\tau}_{k-1}, \bar{\tau}_k^{\text{trial}}\right\} & \text{otherwise,} \end{cases} \tag{2.7}$$

其中

$$\bar{\tau}_k^{\text{trial}} \leftarrow \begin{cases} \infty & \text{if } \bar{g}_k^T \bar{d}_k + \max\left\{\bar{d}_k^T H_k \bar{d}_k, 0\right\} \le 0; \\ \frac{(1-\sigma)\|c_k\|_1}{\bar{g}_k^T \bar{d}_k + \max\left\{\bar{d}_k^T H_k \bar{d}_k, 0\right\}} & \text{otherwise.} \end{cases} \tag{2.8}$$

优值参数更新机制可确保优值参数值序列单调不增。此外，该更新机制旨在确保优值函数模型的下降量充分为正。我们将在以下引理中具体说明这些结论。

Lemma 2.3. 假设条件2.1 和2.2 成立。根据(2.7)–(2.8)中描述的优值参数更新机制可知，对于 $k \in \mathbb{N}$ 在算法2.1的任意一次实现中

$$\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \ge \bar{\tau}_k \max\left\{\bar{d}_k^T H_k \bar{d}_k, 0\right\} + \sigma \|c_k\|_1. \tag{2.9}$$

此外，如果 $\bar{\tau}_k \ne \bar{\tau}_{k-1}$，那么 $0 < \bar{\tau}_k \le (1-\epsilon_\tau)\bar{\tau}_{k-1}$。

证明。由 (2.7)，我们有 $\bar{\tau}_k \le \bar{\tau}_k^{\text{trial}}$。此外，由 (2.5)、(2.7) 和 (2.8)，可知对于所有 $k \in \mathbb{N}$，(2.9) 均成立。由 (2.7)，若 $\bar{\tau}_k \ne \bar{\tau}_{k-1}$，则 $\bar{\tau}_k = \min\left\{(1-\epsilon_\tau)\bar{\tau}_{k-1}, \bar{\tau}_k^{\text{trial}}\right\} \le (1-\epsilon_\tau)\bar{\tau}_{k-1}$。此外，当 $c_k = 0$ 时，由假设 2.2、(2.6) 和 (2.8) 可得 $\bar{d}_k \in$ 零空间(J_k) 且 $\bar{g}_k^T \bar{d}_k + \max\{\bar{d}_k^T H_k \bar{d}_k, 0\} = \bar{g}_k^T \bar{d}_k + \bar{d}_k^T H_k \bar{d}_k = c_k^T \bar{y}_k = 0$，这意味着 $\bar{\tau}_k^{\text{trial}} = \infty$。因此，对于所有 $k \in \mathbb{N}$，我们有 $\bar{\tau}_k^{\text{trial}} > 0$。最后，由 $\bar{\tau}_{-1} \in \mathbb{R}_{>0}$ 和 (2.7)，对于所有 $k \in \mathbb{N}$，我们有 $\bar{\tau}_k > 0$。

□

我们注意到，在确定性情形下，优值函数模型的减少量仅在满足 (2.3) 的迭代点处为零。

在每次迭代中 $k \in \mathbb{N}$，对于算法2.1的任意一次实现，在更新优值参数 $\bar{\tau}_k$后，我们计算 $\Delta l(x_k,\ \bar{\tau}_k,\ \bar{g}_k,\ d_k)$，即优值函数的随机模型约化，并用它来检查是否取得了充分进展。具体而言，给定步长 α_k，我们计算候选迭代点$x_k^+ := x_k + \alpha_k \bar{d}_k$ 并通过修正的充分下降条件检查能否取得充分进展

$$\bar{\phi}(x_k^+, \bar{\tau}_k; \xi^0(x_k^+)) \leq \bar{\phi}(x_k, \bar{\tau}_k; \xi^0(x_k)) - \alpha_k \theta \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + 2\bar{\tau}_k \epsilon_f, \tag{2.10}$$

其中，$\bar{\phi}(x_k^+, \bar{\tau}_k; \xi^0(x_k^+))$ 和 $\bar{\phi}(x_k, \bar{\tau}_k; \xi^0(x_k))$ 是优值函数估计，$\theta \in (0, 1)$ 是用户定义参数，ϵ_f 是目标函数近似中期望噪声的上界。注意，$\bar{\phi}(x_k^+, \bar{\tau}_k; \xi^0(x_k^+))$ 和 $\bar{\phi}(x_k, \bar{\tau}_k; \xi^0(x_k))$ 是第 2.3 节详细描述的零阶预言机的实现。右侧的正项允许充分下降条件有所放宽，即优值函数在采取一步后可能增大，并用于校正优值函数近似中的噪声。若满足 (2.10)，我们接受候选点 x_k^+，令 $x_{k+1} \leftarrow x_k^+$，并可能增大下一次迭代的步长，即 $\alpha_{k+1} \geq \alpha_k$。若不满足 (2.10)，算法则不接受该候选迭代点，而是令 $x_{k+1} \leftarrow x_k$ 并缩小下一次迭代的步长，即 $\alpha_{k+1} < \alpha_k$。这一 步长更新规则是我们步长搜索方法的核心，与传统的线搜索策略有本质区别；参见 [5, 13, 23]。与先计算搜索方向、再沿该方向寻找步长的线搜索方法不同，在我们的方法中，搜索方向在每次迭代中都会改变（即使没有采取步长）。

本节最后，我们将无约束情形与本文方法作几点比较。首先，在无约束情形下（其中 $H_k = I$），量 $\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, d_k)$ 化简为 $\|\bar{g}_k\|^2$，它既提供了一个充分下降度量，也是一个近似的一阶驻点性度量。在约束情形下，优值函数模型的降低将发挥类似作用。其次，在无约束情形下，(2.10) 恢复了一类带噪声的线搜索/步长搜索无约束优化算法所采用的充分下降条件；例如，参见 [4, 5, 23]。

2.3. 概率预言机。在许多现实应用中，无法轻易精确计算目标函数和导数信息。因此，这些量可通过非精确概率零阶和一阶预言机来近似获得。这些预言机能够生成精度和可靠性各异的近似值，下面将对其作形式化介绍。

Oracle 0 （概率零阶预言机）。给定 $x \in \mathbb{R}^n$，预言机计算 $\bar{f}(x; \xi^0(x))$，这是 $\bar{f}(x; \Xi^0(x))$ 的一次实现；后者是目标函数值 $f(x)$ 的一个（随机）估计，其中 $\Xi^0(x)$ 表示底层随机性（可能依赖于 x），其对应的概率空间为 $(\Omega, \mathcal{F}_\Omega, P)$。令$E(x; \Xi^0(x)) := |\bar{f}(x; \Xi^0(x)) - f(x)|$。对于任意 $x \in \mathbb{R}^n$，$E(x; \Xi^0(x))$ 是一个参数为 $\{\nu, b\} \subset \mathbb{R}_{\geq 0}$、均值有界于某个常数 $\epsilon_f \in \mathbb{R}_{\geq 0}$ 的“单侧”次指数随机变量。具体而言，对于所有 $x \in \mathbb{R}^n$ 和 $\lambda \in [0, 1/b)$,

$$\begin{aligned} &\mathbb{E}_{\Xi^0(x)}\left[E(x; \Xi^0(x))\right] \leq \epsilon_f \\ &\text{and } \mathbb{E}_{\Xi^0(x)}\left[\exp(\lambda(E(x; \Xi^0(x)) - \mathbb{E}_{\Xi^0(x)}\left[E(x; \Xi^0(x))\right]))\right] \leq \exp\left(\frac{\lambda^2\nu^2}{2}\right). \end{aligned} \tag{2.11}$$

优值函数值的随机近似定义为 $\bar{\phi}(x, \tau;\ \xi^0(x)) = \tau \bar{f}(x;\ \xi^0(x)) + \|c(x)\|_1$。

ORACLE 1 (**概率一阶预言机**)。给定 $x \in \mathbb{R}^n$ 和 $\alpha \in \mathbb{R}_{>0}$，预言机计算 $\bar{g}(x;\ \xi^1(x))$，它是 $\bar{g}(x;\ \Xi^1(x))$ 的一个实现，也是目标函数 $\nabla f(x)$ 梯度的一个（随机）估计，使得(2.12)

$$\mathbb{P}_{\Xi^1(x)}\left[\|\bar{g}(x;\Xi^1(x)) - \nabla f(x)\| \leq \max\left\{\epsilon_g, \kappa_{\mathrm{FO}}\alpha\sqrt{\Delta l(x,\bar{\tau}(x;\Xi^1(x)),\bar{g}(x;\Xi^1(x)),\bar{d}(x;\Xi^1(x)))}\right\}\right] \geq 1-\delta,$$

其中 $\Xi^1(x)$ 表示底层随机性（可能依赖于 x），其对应的概率空间为$(\Omega\ \mathcal{F}_\Omega, P)$；$(1-\delta) \in (\frac{1}{2}, 1]$表示预言机生成“足够准确”的梯度估计的概率（即预言机可靠性），而$\{\epsilon_g, \kappa_{\mathrm{FO}}\} \subset \mathbb{R}_{\geq 0}$是预言机固有的常数（即预言机精度）。

在本文其余部分，为简化记号，我们省略 $\Xi^0(x)$（分别为 $\xi^0(x)$）和 $\Xi^1(x)$（分别为 $\xi^1(x)$）中对 x 的依赖。此外，我们用（Ξ_k^0，Ξ_k^+，Ξ_k^1）和（ξ_k^0，ξ_k^+，ξ_k^1）分别表示（$\Xi^0(x_k)$ $\Xi^0(x_k^+)$ $\Xi^1(x_k)$）和（$\xi^0(x_k)$ $\xi^0(x_k^+)$ $\xi^1(x_k)$）。

REMARK 2.4. 关于预言机 *0* 和 *1*，我们作几点说明：

- 预言机*0*和*1* 与 *[12, 23]*中定义的预言机类似。有关这些预言机的完整讨论和示例，请感兴趣的读者参阅 *[23,* 第 *5]*节。
- 预言机 *1* 将 *[12, 23]*中定义的预言机推广到了等式约束情形。事实上，在无约束情形下，预言机 *1*右侧可化为 $\max\{\epsilon_g, \kappa_{\mathrm{FO}}\alpha\|\bar{g}(x;\ \Xi^1)\|\}$，这正是 *[12, 23]*中采用的形式。
- 在预言机 $\epsilon_g \in \mathbb{R}_{\geq 0}$ 的最大项中包含*1*，使得梯度近似可以带有偏差；偏差的大小与ϵ_g成正比。

2.4. 算法框架。现在我们可以介绍随机步长搜索序列二次规划方法（SS-SQP），见算法 2.1。

算法 2.1自适应步长搜索序列二次规划 (SS-SQP)

要求： 初始迭代点 $x_0 \in \mathbb{R}^n$；初始优值参数 $\bar{\tau}_{-1} \in \mathbb{R}_{>0}$；最大步长大小 $\alpha_{\max} \in (0, 1]$；初始步长 $\alpha_0 \in (0,\ \alpha_{\max}]$；参数 $\epsilon_f \in \mathbb{R}_{\geq 0}$ 的零阶预言机（预言机 0）；以及其他参数 $\{\gamma, \theta, \sigma, \epsilon_\tau\} \subset (0,\ 1)$

1: **对所有** $k \in \mathbb{N}$ **执行**
2: 通过 Oracle 1 和 $\alpha = \alpha_k$，生成 $\bar{g}_k = \bar{g}(x_k; \xi_k^1)$ $\bar{d}_k = \bar{d}(x_k; \xi_k^1)$，如（2.6）所示，并且 $\bar{\tau}_k = \bar{\tau}(x_k;\ \xi_k^1)$，如（2.7）–（2.8）所示
3: 令 $x_k^+ = x_k + \alpha_k\bar{d}_k$，并通过 Oracle 0生成 $\bar{\phi}(x_k,\ \bar{\tau}_k;\ \xi_k^0)$ 和 $\bar{\phi}(x_k^+,\ \bar{\tau}_k;\ \xi_k^+)$
4: **如果** (2.10) 成立，**则**
5: 集合 $x_{k+1} \leftarrow x_k^+$ 和 $\alpha_{k+1} \leftarrow \min\{\alpha_{\max}, \gamma^{-1}\alpha_k\}$
6: **else**
7: 集合 $x_{k+1} \leftarrow x_k$ 和 $\alpha_{k+1} \leftarrow \gamma\alpha_k$

REMARK 2.5. 关于 *SS-SQP*，我们作出如下说明：

- （步长搜索）算法 *2.1*是一种步长搜索算法，它与传统线搜索方法的主要区别在于，每次迭代只测试一个试探迭代点。也就是说，如果 （2.10）不满足，则会缩短步长，并在下一次迭代中计算新的搜索方向和候选迭代点。其他论文也采用了这种策略，例如参见 *[5, 13, 23, 29]*。需要注意的是，在每次

迭代中，即使迭代点没有变化，我们的算法也要求在下一次迭代中重新计算目标函数和梯度估计。

- （修正的充分下降条件 (2.10)） $2\bar{\tau}_k\epsilon_f$ 式(2.10) 右侧的项是一个校正项，用于补偿概率零阶预言机（预言机 *0*）的不精确性。该校正项放宽了充分下降要求。与传统的充分下降条件不同，修正条件 (2.10) 允许按预言机 *0*的噪声水平成比例地放宽要求。

- （目标函数评估；第 *3* 行）在候选迭代点 x_k^+ （第 *3* 行）处评估目标函数值所涉及的随机性，与在当前点 x_k处评估时的随机性不同。也就是说，在候选迭代点 x_k^+ 处对目标函数的评估独立于在当前迭代点处对目标函数的评估 x_k。我们还注意到，实际上只需函数估计值之差中的噪声（即$\left|\bar{f}(x_k,\ \Xi_k^0)\ -\bar{f}(x_k^+,\ \Xi_k^+)\ -(f(x_k)-f(x_k^+))\ \right|$）满足次指数性质，因为这是分析中需要控制的基本量。此外，我们注意到，即使在未成功的迭代中（迭代点没有变化），也会重新评估目标函数值。

- （目标梯度评估；第 *2* 行）为了生成满足预言机 *1* 条件的目标函数梯度估计，可以采用类似于 *[40,* 算法 *2]* 的过程（一个循环）。其思路是逐步改进估计，以生成满足该条件的估计。在许多实际问题中，包括机器学习中的经验风险最小化，都可以通过逐步增加采样数量来改进梯度近似。

- （最大步长 $\alpha_{\max}$）我们选择 $\alpha_{\max}\in(0,1]$ 主要是为了简化分析。尽管如此，对 $\alpha_{\max}$ 设置为 *1* 的上界，依据的是确定性约束情形。在确定性情形（没有噪声）下，优值函数的下降由一个非光滑函数从上方界定，该函数唯一的非光滑点位于 $\alpha=1$，这使分析变得复杂；参见 *[7,* 引理 *2.13]*。

- （参数 ϵ_f 和 ϵ_g 见预言机 *0* 和 *1*）要实现并运行算法*2.1*，需要（对）ϵ_f （预言机 *0* 中的偏差）进行估计，而不需要ϵ_g （预言机 *1* 中的偏差）。目标梯度估计中的噪声是预言机*1* 的固有量；虽然实现算法 *2.1* 并不需要它，但它在分析中起着核心作用，因为它决定了收敛邻域。关于 ϵ_f （目标函数估计中的噪声），算法需要对该量进行估计（或给出其上界）。在实践中，可以通过采样或其他技术来估计，例如*[21]*。最后需要指出的是，在无约束情形下，许多随机自适应线搜索/步长搜索（及信赖域）方法也需要类似的常数（预言机条件）；例如参见 *[4, 5, 12, 34, 42]*。可以合理预期，这些要求也适用于约束随机情形。

在继续之前，我们先定义与该算法相关的随机过程。令 M_k 表示具有实现 $\{\xi_k^0,\xi_k^+,\xi_k^1\}$的 $\{\Xi_k^0,\Xi_k^+,\Xi_k^1\}$。该算法生成一个随机过程： $\{(G_k,D_k,\mathcal{T}_k,\bar{\phi}(X_k,\mathcal{T}_k;\ \Xi_k^0)\ \bar{\phi}(X_k^+,\mathcal{T}_k;\ \Xi_k^+)\ X_k,A_k)\}$，其实现为 $\{(\bar{g}_k,\bar{d}_k,\bar{\tau}_k,\bar{\phi}(x_k,\bar{\tau}_k;\xi_k^0)\ \bar{\phi}(x_k^+,\bar{\tau}_k;\ \xi_k^+)\ x_k,\alpha_k)\}$，并适应于滤过$\{\mathcal{F}_k\colon\ k\geq 0\}$，其中 $\mathcal{F}_k=\boldsymbol{\sigma}(M_0,M_1,\ldots,M_k)$，而 $\boldsymbol{\sigma}$ 表示 $\boldsymbol{\sigma}$-代数。在第 k次迭代中， G_k 是随机梯度， D_k 是随机原始搜索方向， $\mathcal{T}_k$ 是随机优值参数， $\bar{\phi}(X_k,\mathcal{T}_k;\ \Xi_k^0)$和 $\bar{\phi}(X_k^+,\mathcal{T}_k;\ \Xi_k^+)$分别是当前点和候选点处的随机含噪优值函数评估，$X_k$ 是第 k 次迭代的随机迭代点， A_k 是随机步长。注意，

G_k， D_k， $\mathcal{T}_k$ 由以下内容决定： Ξ_k^1（Oracle 1），含噪优值函数评估则由以下内容决定： Ξ_k^0 和Ξ_k^+ （Oracle0）。

3. 理论分析。 本节分析算法 2.1 的行为。为简洁起见，在本节中，我们假设假设条件2.1 和2.2均成立，并且不在每个引理和定理中重复说明这一事实。我们首先给出一些预备知识、定义和假设，然后给出算法 2.1 的最坏情况下的迭代复杂度界。

3.1. 预备知识、定义与假设。 我们首先定义一些 确定性 量，这些量用于算法 2.1的分析，但在算法的实现中从不显式计算。对于算法 $k \in \mathbb{N}$ 的任意一次实现，令 2.1 中的 $(d_k, y_k) \in \mathbb{R}^n \times \mathbb{R}^m$ 为 (2.6) 的确定性对应形式的解，即：

$$\begin{bmatrix} H_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ y_k \end{bmatrix} = - \begin{bmatrix} \nabla f_k \\ c_k \end{bmatrix}. \tag{3.1}$$

拉格朗日函数的梯度范数（定义于 (2.2) 中）作为一阶驻点性度量，用于1.1)；在每个原始-对偶迭代点 (x_k, y_k) 处均有界：

$$\left\| \begin{bmatrix} \nabla f_k + J_k^T y_k \\ c_k \end{bmatrix} \right\| = \left\| \begin{bmatrix} -H_k d_k \\ -J_k d_k \end{bmatrix} \right\| \leq (\kappa_H + \kappa_J)\|d_k\|, \tag{3.2}$$

其中等式由（3.1）得出，不等式则由假设 2.1 和 2.2推出。因此，（3.2）意味着 d_k （原始搜索方向）可作为一阶驻点性度量的代理量。下面的引理表明，对于算法 2.1的任意一次实现，元组（d_k， y_k）均有界。

LEMMA 3.1. 存在常数 $\{\kappa_d, \kappa_y\} \subset \mathbb{R}_{>0}$ ，使得 $\|d_k\| \leq \kappa_d$ 且$\|y_k\| \leq \kappa_y$ 对于算法 *2.1*的任意一次实现中的所有 $k \in \mathbb{N}$ 均成立。

证明。 根据柯西–施瓦茨不等式和（3.1），可得

$$\left\| \begin{bmatrix} d_k \\ y_k \end{bmatrix} \right\| = \left\| \begin{bmatrix} H_k & J_k^T \\ J_k & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla f_k \\ c_k \end{bmatrix} \right\| \leq \left\| \begin{bmatrix} H_k & J_k^T \\ J_k & 0 \end{bmatrix}^{-1} \right\| \left\| \begin{bmatrix} \nabla f_k \\ c_k \end{bmatrix} \right\|,$$

不等式右侧的两项均由假设 2.1 和2.2有界，由此证明得证。 □

此外，对于算法 2.1的任意一次实现中的所有 $k \in \mathbb{N}$ ，我们定义 $\tau_k \in \mathbb{R}_{>0}$和 $\tau_k^{\text{trial}} \in \mathbb{R}_{>0}$，分别为（2.7）和（2.8）的确定性对应形式，

$$\tau_k \leftarrow \begin{cases} \bar{\tau}_k & \text{if } \bar{\tau}_k \leq \tau_k^{\text{trial}}; \\ \min\left\{(1-\epsilon_\tau)\bar{\tau}_k, \tau_k^{\text{trial}}\right\} & \text{otherwise,} \end{cases} \tag{3.3}$$

其中

$$\tau_k^{\text{trial}} \leftarrow \begin{cases} \infty & \text{if } \nabla f_k^T d_k + \max\left\{d_k^T H_k d_k, 0\right\} \leq 0; \\ \frac{(1-\sigma)\|c_k\|_1}{\nabla f_k^T d_k + \max\left\{d_k^T H_k d_k, 0\right\}} & \text{otherwise.} \end{cases} \tag{3.4}$$

我们再次强调， $\{(\tau_k, \tau_k^{\text{trial}})\}_{k\in\mathbb{N}}$ 仅为分析之目的而引入，在算法2.1 中从不计算它们（即使在使用真实梯度的情形下也不计算，见

其中使用真实梯度，即 $\bar{g}_k = \nabla f(x_k)$ ）。我们还要指出，此定义与 [7, 16]中的定义不同。差异在于，计算 τ_k时，比较对象是 $\bar{\tau}_k$，而不是 $\bar{\tau}_{k-1}$。这对分析很重要，因为这保证了 $\tau_k \le \bar{\tau}_k$，对于算法 $k \in \mathbb{N}$ 任意一次实现中的所有 2.1。

我们在一个与优值参数相关的特定事件发生这一背景下分析算法 2.1。具体而言，我们考虑由算法 2.1 生成的、与优值参数序列 $\{\mathcal{T}_k\}$ （其实现为 $\{\bar{\tau}_k\}$）相关的随机过程远离零的事件（假设 3.2）。先前文献 [6–8, 16, 17]也采用了此类事件；关于详细讨论，请参阅[7, 第 3 节2] 和 [16, 第 4 节2] 。我们的数值实验也支持这一事件发生的假设；参见第 4.5 节的讨论。我们强调，算法的实现不需要了解 $\bar{\tau}_{\min}$ （定义见假设 3.2）。此外，如果随机梯度恰好一致有界，则可以证明，在假设 2.1 和 2.2 下，优值参数序列远离零；例如参见 [7, 命题 3.18] 和 [8, 引理 4.6]。

ASSUMPTION 3.2. 事件 $\mathcal{E} := \mathcal{E}(k_{\max}, \bar{\tau}_{\min})$ 发生，给定常数 $k_{\max} \in \mathbb{N}$ 和 $\bar{\tau}_{\min} \in \mathbb{R}_{>0}$，其含义是存在 $\mathcal{T}' \in \mathbb{R}_{>0}$，使得

$$\mathcal{T}_k = \mathcal{T}' \ge \bar{\tau}_{\min} > 0, \quad \textit{for all} \quad k \ge \bar{k}_{\max},$$

其中，$\{\mathcal{T}_k\}$ 是一个随机过程，其实现为 $\{\bar{\tau}_k\}$，这是由算法*2.1*生成的优值参数值序列。此外，在事件 $\mathcal{E}$发生的条件下，(2.11)–(2.12) 对算法 $\{X_k\}$ （其实现为 $\{x_k\}$）生成的所有原始迭代点以及 $\{X_k^+\}$ （其实现为 $\{x_k^+\}$）生成的所有原始试探迭代点*2.1*均以相同的常数成立。

在本文余下部分中，我们假设事件 $\mathcal{E}$ 发生，即假设 3.2成立。接下来，我们陈述并证明关于确定性优值参数序列 $\{\tau_k\}$ （定义于（3.3）中）的一个有用性质。

LEMMA 3.3. 假设 *3.2* 成立，则存在一个正常数$\tau_{\min} \in \mathbb{R}_{>0}$，使得对于算法 *2.1* 的任意一种实现，只要事件 $\mathcal{E}$ 发生，$\tau_k \ge \tau_{\min}$ 对所有 $k \in \mathbb{N}$均成立。

证明。 根据 [7, 引理 2。16]， $\{\tau_k^{\text{trial}}\} \subset \mathbb{R}_{>0} \cup \{+\infty\}$ 始终远离零。我们定义 $\tau_{\min}^{\text{trial}} \in \mathbb{R}_{>0}$ 使得 $\tau_{\min}^{\text{trial}} \le \tau_{\text{试探}k}$对于算法 $k \in \mathbb{N}$的每一种实现中的所有 2.1。当事件 $\mathcal{E}$ 发生时（见假设 3.2），由（3.3）–（3.4），可取 $\tau_{\min} = \text{最小值}\{(1-\epsilon_\tau)\tau_{\min}^{\text{trial}}, \ \bar{\tau}_{\min}\}$从而得证。 □

设 E_k和 E_k^+ 分别为预言机 0 对目标函数评估所得的误差，即 $E_k := \left|\bar{f}(X_k; \Xi_k^0) - f(X_k)\right|$ 和 $E_k^+ := \left|\bar{f}(X_k^+; \Xi_k^+) - f(X_k^+)\right|$，其实现分别为 e_k 和 e_k^+。接下来，我们引入若干定义，以便分析算法 2.1。具体而言，我们定义真/假迭代（定义 3.4）、成功/不成功迭代（定义 3.5）和大/小步长（定义 3.6），并分别引入三个指示变量。

DEFINITION 3.4. 对于算法 *2.1*的任意实现，迭代 $k \in \mathbb{N}$ 当且仅当 **.....** 时为真

$$\|\bar{g}_k - \nabla f_k\| \le \max\left\{\epsilon_g, \kappa_{\text{FO}}\alpha_k\sqrt{\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)}\right\} \ \textit{and} \ e_k + e_k^+ \le 2\epsilon_f, \tag{3.5}$$

其中 $\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, d_k)$ 定义见 (2.5)，常数 ϵ_f， ϵ_g和 κ_{FO} 是第 e

与预言机*0* 和 *1*相同。如果 (3.5) 不成立，则称该迭代为错误迭代。我们用随机指示变量 I_k 表示一次迭代是否为真。

DEFINITION 3.5. 给定 $\theta \in (0,1)$，对于算法 *2.1* 的任意实现中的所有 $k \in \mathbb{N}$，令 $\bar{\phi}(x_k, \bar{\tau}_k; \xi_k)$ 和 $\bar{\phi}(x_k^+, \bar{\tau}_k; \xi_k^+)$ 为通过预言机 *0* 获得的结果。若 *(2.10)* 成立，则迭代 k 为成功迭代；否则，为不成功迭代。我们使用随机指示变量 Θ_k 表示一次迭代是否成功。

DEFINITION 3.6. 对于 $k \in \mathbb{N}$ 算法 *2.1*的任意实现，若$\min\{\alpha_k, \alpha_{k+1}\} \geq \tilde{\alpha}$ 其中 $\tilde{\alpha}$ 是某个依赖于问题的正实数（定义于引理 *3.17*中），则称该步为大步，并将指示变量设为$U_k = 1$。否则，称该步为 k 小步，并将 $U_k = 0$。

我们将证明，在适当条件下，如果步长是一个小步且迭代是真实的，那么该迭代必定是成功的（见引理3.17）。最后一个定义针对停止时刻（$T_{\varepsilon_{\Delta l}}$）和进展度量（$\{Z_k\}$）。

DEFINITION 3.7. 对于算法 *2.1* 的任意实现，定义 $T_{\varepsilon_{\Delta l}} = \min\{k : \sqrt{\Delta l(x_k, \tau_k, \nabla f_k, d_k)} \leq \varepsilon_{\Delta l}\}$，即达到一阶$\varepsilon$-驻点迭代所需的迭代次数，其中 $\varepsilon = \Omega(\varepsilon_{\Delta l})$。我们在注释 *3.8* 中讨论 ε 与 $\varepsilon_{\Delta l}$ 之间的明确关系。此外，对于算法 *2.1* 的任意实现中的所有 $k \in \mathbb{N}$，若事件 $\mathcal{E}$ 发生（见假设 *3.2*），则令 $Z_k := \phi(x_k, \bar{\tau}_k) - \phi_{\min} - (\bar{\tau}_k f_{\inf} - \bar{\tau}_{\min} f_{\inf})$，其中 $\phi_{\min}$ 是 $\phi(\cdot, \bar{\tau}_{\min})$ 在 $\mathcal{X}$ 上的下界，而 $\bar{\tau}_{\min}$ 是假设 *3.2* 中定义的常数。

REMARK 3.8. 我们算法的一个关键要素是停止时刻 $T_{\varepsilon_{\Delta l}}$，它与 $\Delta l(x_k, \tau_k, \nabla f_k, d_k)$ 有关。事实上，根据 (3.2)、假设 *2.1*、假设 *2.2*、假设 *3.2*、引理 *3.3* 和引理 *3.11* （见下文），在定义 $T_{\varepsilon_{\Delta l}}$*3.7* 中定义的停止时刻是达到一阶$\varepsilon$-驻点迭代点所需的迭代次数，即：

$$\max\{\|\nabla f_k + J_k^T y_k\|, \sqrt{\|c_k\|}\} \leq \varepsilon, \quad \textit{where} \quad \varepsilon = \frac{\max\{\kappa_H, 1\}}{\sqrt{\kappa_l \tau_{\min}}} \cdot \varepsilon_{\Delta l} \tag{3.6}$$

并且 $\{\kappa_H, \tau_{\min}, \kappa_l\} \subset \mathbb{R}_{>0}$ 在假设 *2.2* 以及引理 *3.3* 和 *3.11* 中定义。由于零阶和一阶预言机（预言机 *0* 和 *1*）存在噪声和偏差，ε 必须与零保持一定距离。我们注意到，*(3.6)* 与 *[16,* 式 *(5)],* 中使用的驻点性度量相同；与 $\left\| \begin{bmatrix} \nabla f_k + J_k^T y_k \\ c_k \end{bmatrix} \right\|$ 相比，它是一种非标准的一阶驻点性度量。尽管如此，可以证明 $\left\| \begin{bmatrix} \nabla f_k + J_k^T y_k \\ c_k \end{bmatrix} \right\| \leq 2 \max\{\|\nabla f_k + J_k^T y_k\|, \|c_k\|\} \leq {}^2\max_{\sqrt{\kappa_l \tau_{\min}}}^{\{\kappa_H, \kappa_J\}} \varepsilon_{\Delta l} = \Omega(\varepsilon)$。本文始终关注 *(3.6)*，并给出其 ε-复杂度界，因为当 $\varepsilon < 1$ 时，它对可行性（$\|c_k\|$）给出了更强的结果。

3.2. 主要技术结果。 我们通过一系列技术引理逐步推导本文的主要结果（定理 3.20）。我们的第一个引理表明，Z_k （定义见定义 3.7）始终非负。

LEMMA 3.9. 对于所有 $k \in \mathbb{N}$ 在算法 *2.1* 的任意一次实现中，若事件 $\mathcal{E}$发生（见假设 *3.2*），$Z_k \geq 0$。

证明。 由式（2.4）和定义3.7 可知

$$\begin{aligned} Z_k &= \phi(x_k, \bar{\tau}_k) - \phi_{\min} - (\bar{\tau}_k f_{\inf} - \bar{\tau}_{\min} f_{\inf}) \\ &= (\bar{\tau}_k(f_k - f_{\inf}) + \|c_k\|_1) - \phi_{\min} + \bar{\tau}_{\min} f_{\inf} \\ &\geq (\bar{\tau}_{\min}(f_k - f_{\inf}) + \|c_k\|_1) - \phi_{\min} + \bar{\tau}_{\min} f_{\inf} = \phi(x_k, \bar{\tau}_{\min}) - \phi_{\min} \geq 0, \end{aligned}$$

由此完成证明。 □

根据引理 3.9，如果事件 $\mathcal{E}$ 发生，则 $\{Z_k\}$ 始终非负，并且 $Z_k = 0$ 当且仅当 $\bar{\tau}_k(f_k - f_{\inf}) = \bar{\tau}_{\min}(f_k - f_{\inf})$ 以及 $\phi(x_k, \bar{\tau}_{\min}) = \phi_{\min}$时成立。事实上，$\{Z_k\}$ 只需证明其一致地以下界常数为界，即足以得到我们的主要理论结果（不一定需要 $\{Z_k\}$ 收敛到零）。在无约束情形下也存在类似性质；例如参见 $\{Z_k\}$。 [23]

下一个引理给出了优值函数模型下的下降量的一个有用下界，$\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, d_k)$，该下界与原始搜索方向（$\|\bar{d}_k\|^2$）和不可行性度量（$\|c_k\|$）有关。

LEMMA 3.10. 存在某个常数 $\kappa_l \in \mathbb{R}_{>0}$ 使得对于所有 $k \in \mathbb{N}$ 在算法 *2.1* 的任意实现中，$\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \geq \kappa_l \bar{\tau}_k(\|\bar{d}_k\|^2 + \|c_k\|_1)$。

证明。对于任意迭代 $k \in \mathbb{N}$，根据 [7, 引理 3.4], ，存在某个常数$\kappa_l \in \mathbb{R}_{>0}$使得 $-\bar{\tau}_k(\bar{g}_k^T \bar{d}_k + \frac{1}{2}\max\{\bar{d}_k^T H_k \bar{d}_k, 0\}) + \|c_k\|_1 \geq \kappa_l \bar{\tau}_k(\|\bar{d}_k\|^2 + \|c_k\|_1)$。根据 $\bar{\tau}_k \in \mathbb{R}_{>0}$（由引理 2.3 得出），这意味着 $\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) = -\bar{\tau}_k \bar{g}_k^T \bar{d}_k + \|c_k\|_1 \geq -\bar{\tau}_k(\bar{g}_k^T \bar{d}_k + \frac{1}{2}\max\{\bar{d}_k^T H_k \bar{d}_k, 0\}) + \|c_k\|_1$，从而证明得证。 □

LEMMA 3.11. 存在某个常数 $\kappa_l \in \mathbb{R}_{>0}$ 使得对于所有 $k \in \mathbb{N}$ 在算法 *2.1*的任意一次实现中，$\Delta l(x_k, \tau_k, g_k, d_k) \geq \kappa_l \tau_k(\|d_k\|^2 + \|c_k\|_1)$。

证明。 证明过程与引理 3.10 相同，只需将随机量替换为其确定性对应形式。根据 [7, 引理 3.4], ，所需不等式对引理 κ_l 3.10 中定义的同一常数 成立。 □

下一个引理分别给出了随机搜索方向和对偶变量的误差相对于梯度近似误差的界。

LEMMA 3.12. 对于所有 $k \in \mathbb{N}$ 在算法 *2.1*的任意一次实现中，存在常数 $\{\zeta, \zeta_y\} \subset \mathbb{R}_{>0}$ 使得 $\|\bar{d}_k - d_k\| \leq \zeta^{-1}\|\bar{g}_k - \nabla f_k\|$ 且 $\|\bar{y}_k - y_k\| \leq \zeta_y \|\bar{g}_k - \nabla f_k\|$，其中 ζ 在假设 *2.2*中定义。

证明。 由柯西–施瓦茨不等式、假设 2.2、(3.1) 以及以下事实即 $(d_k - d_k) \in \mathrm{Null}(J_k)$，可得

$$\begin{aligned} \|\bar{d}_k - d_k\|\|\bar{g}_k - \nabla f_k\| &\geq (\bar{d}_k - d_k)^T(\nabla f_k - \bar{g}_k) \\ &= (\bar{d}_k - d_k)^T(H_k(\bar{d}_k - d_k) + J_k^T(\bar{y}_k - y_k)) \\ &= (\bar{d}_k - d_k)^T H_k(\bar{d}_k - d_k) \geq \zeta\|\bar{d}_k - d_k\|^2, \end{aligned} \tag{3.7}$$

以及 $\|\bar{d}_k - d_k\| \leq \zeta^{-1}\|\bar{g}_k - \nabla f_k\|$。由三角不等式和柯西–施瓦茨不等式、假设 2.1 和 2.2，以及 (3.1) 和 (3.7)，可得

$$\begin{aligned} \|\bar{y}_k - y_k\| &= \|(J_k J_k^T)^{-1} J_k \left((\bar{g}_k - \nabla f_k) + H_k(\bar{d}_k - d_k)\right)\| \\ &\leq \|(J_k J_k^T)^{-1}\|\|J_k\|(\|\bar{g}_k - \nabla f_k\| + \|H_k\|\|\bar{d}_k - d_k\|) \\ &\leq \kappa_\sigma \kappa_J(1 + \kappa_H \zeta^{-1})\|\bar{g}_k - \nabla f_k\|. \end{aligned}$$

令 $\zeta_y = \kappa_\sigma \kappa_J(1 + \kappa_H \zeta^{-1})$，证明即完成。 □

下面的引理将随机梯度与随机搜索方向的内积，与优值函数模型中的随机减少量联系起来。我们考虑两种情况，它们分别对应于 Oracle 1 中最大项的两种情况。

LEMMA 3.13. 对于任意 $k \in \mathbb{N}$ 在算法*2.1:*

- 若 $\|\bar{g}_k - \nabla f_k\| \le \kappa_{\mathrm{FO}}\alpha_k\sqrt{\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)}$，则

$$\bar{\tau}_k|\bar{g}_k^T\bar{d}_k| \le \left(\frac{\max\{\kappa_H,\kappa_y\}}{\kappa_l} + \frac{\sqrt{\bar{\tau}_k}\left(1+\kappa_H\zeta^{-1}\right)\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l}}\right)\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k).$$

- 若 $\|\bar{g}_k - \nabla f_k\| \le \epsilon_g$，

$$\bar{\tau}_k|\bar{g}_k^T\bar{d}_k| \le \frac{\max\{\kappa_H,\kappa_y\}+1}{\kappa_l}\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \frac{\bar{\tau}_k\left(1+\kappa_H\zeta^{-1}\right)^2}{4}\epsilon_g^2.$$

证明。若 $\|\bar{g}_k - \nabla f_k\| \le \kappa_{\mathrm{FO}}\alpha_k\sqrt{\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)}$，根据三角不等式，(2.6)、假设 2.2和引理 3.1、3.10 以及3.12，可得

$$\begin{aligned}
\bar{\tau}_k|\bar{g}_k^T\bar{d}_k| &= \bar{\tau}_k|(H_k\bar{d}_k + J_k^T y_k + J_k^T(\bar{y}_k - y_k))^T\bar{d}_k| \\
&\le \bar{\tau}_k(|\bar{d}_k^T H_k\bar{d}_k| + |y_k^T J_k\bar{d}_k| + |(\bar{y}_k - y_k)^T J_k\bar{d}_k|) \\
&\le \bar{\tau}_k(\kappa_H\|\bar{d}_k\|^2 + \|y_k\|\|c_k\| + \|(\bar{g}_k - \nabla f_k) + H_k(\bar{d}_k - d_k)\|\|\bar{d}_k\|) \\
&\le \max\{\kappa_H,\kappa_y\}\bar{\tau}_k(\|\bar{d}_k\|^2 + \|c_k\|) + \bar{\tau}_k(\|\bar{g}_k - \nabla f_k\| + \kappa_H\|\bar{d}_k - d_k\|)\|\bar{d}_k\| \\
&\le \frac{\max\{\kappa_H,\kappa_y\}}{\kappa_l}\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \bar{\tau}_k\left(1+\kappa_H\zeta^{-1}\right)\|\bar{g}_k - \nabla f_k\|\|\bar{d}_k\| \\
&\le \frac{\max\{\kappa_H,\kappa_y\}}{\kappa_l}\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \frac{\sqrt{\bar{\tau}_k}\left(1+\kappa_H\zeta^{-1}\right)\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l}}\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k),
\end{aligned}$$

这就完成了证明的第一部分。

采用类似的推理，若 $\|\bar{g}_k - \nabla f_k\| \le \epsilon_g$，根据三角不等式，(2.6)、假设 2.2、引理 3.1、3.10、3.12以及 $ab \le a^2 + \frac{b^2}{4}$ 对任意$\{a, b\} \subset \mathbb{R}$成立这一事实，可得

$$\begin{aligned}
\bar{\tau}_k|\bar{g}_k^T\bar{d}_k| &\le \frac{\max\{\kappa_H,\kappa_y\}}{\kappa_l}\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \bar{\tau}_k\left(1+\kappa_H\zeta^{-1}\right)\|\bar{g}_k - \nabla f_k\|\|\bar{d}_k\| \\
&\le \frac{\max\{\kappa_H,\kappa_y\}}{\kappa_l}\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \frac{\sqrt{\bar{\tau}_k}\left(1+\kappa_H\zeta^{-1}\right)}{\sqrt{\kappa_l}}\epsilon_g\sqrt{\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)} \\
&\le \frac{\max\{\kappa_H,\kappa_y\}+1}{\kappa_l}\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \frac{\bar{\tau}_k\left(1+\kappa_H\zeta^{-1}\right)^2}{4}\epsilon_g^2,
\end{aligned}$$

由此完成证明。 □

下一个引理针对与随机搜索方向（及梯度）相关的误差，给出了与引理 3.13 中相同两种情形下的一个有用上界。

LEMMA 3.14. 对于 $k \in \mathbb{N}$ 在算法 *2.1*的任意一次实现中：

- 若 $\|\bar{g}_k - \nabla f_k\| \le \kappa_{\mathrm{FO}}\alpha_k\sqrt{\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)}$，则

$$|\nabla f_k^T d_k - \bar{g}_k^T\bar{d}_k| \le \left(\frac{(1+\kappa_H\zeta^{-1})\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l\bar{\tau}_k}} + \frac{\kappa_{\mathrm{FO}}^2\alpha_k^2}{\zeta}\right)\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)$$

$$\textit{and}\quad |d_k^T H_k d_k - \bar{d}_k^T H_k\bar{d}_k| \le \left(\frac{2\kappa_H\zeta^{-1}\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l\bar{\tau}_k}} + \frac{\kappa_H\kappa_{\mathrm{FO}}^2\alpha_k^2}{\zeta^2}\right)\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k).$$

- 若 $\|\bar{g}_k - \nabla f_k\| \le \epsilon_g$，则

$$|\nabla f_k^T d_k - \bar{g}_k^T\bar{d}_k| \le \frac{(1+\kappa_H\zeta^{-1})\epsilon_g}{\sqrt{\kappa_l\tau_k}}\sqrt{\Delta l(x_k, \tau_k, \nabla f_k, d_k)} + \zeta^{-1}\epsilon_g^2$$

$$\textit{and}\quad |d_k^T H_k d_k - \bar{d}_k^T H_k\bar{d}_k| \le \frac{2\kappa_H\zeta^{-1}\epsilon_g}{\sqrt{\kappa_l\tau_k}}\sqrt{\Delta l(x_k, \tau_k, \nabla f_k, d_k)} + \kappa_H\zeta^{-2}\epsilon_g^2.$$

证明。 我们从 $\|\bar{g}_k - \nabla f_k\| \le \kappa_{\mathrm{FO}}\alpha_k\sqrt{\Delta l(x_k,\bar{\tau}_k,\bar{g}_k,\bar{d}_k)}$。根据三角不等式和柯西–施瓦茨不等式、假设2.1以及引理 3.1、3.10 和 3.12，

$$\begin{aligned}
&|\nabla f_k^T d_k - \bar{g}_k^T\bar{d}_k| \\
&= |(\bar{g}_k - \nabla f_k)^T\bar{d}_k + (\nabla f_k - \bar{g}_k)^T(\bar{d}_k - d_k) + \bar{g}_k^T(\bar{d}_k - d_k)| \\
&= |(\bar{g}_k - \nabla f_k)^T\bar{d}_k + (\nabla f_k - \bar{g}_k)^T(\bar{d}_k - d_k) - (H_k\bar{d}_k + J_k^T\bar{y}_k)^T(\bar{d}_k - d_k)| \\
&\le \|\bar{g}_k - \nabla f_k\|\|\bar{d}_k\| + \|\nabla f_k - \bar{g}_k\|\|\bar{d}_k - d_k\| + \kappa_H\|\bar{d}_k\|\|\bar{d}_k - d_k\| \\
&\le \left(\tfrac{(1+\kappa_H\zeta^{-1})\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l\bar{\tau}_k}} + \zeta^{-1}\kappa_{\mathrm{FO}}^2\alpha_k^2\right)\Delta l(x_k,\bar{\tau}_k,\bar{g}_k,\bar{d}_k).
\end{aligned}$$

此外，根据假设2.2 可得

$$\begin{aligned}
|d_k^T H_k d_k - \bar{d}_k^T H_k\bar{d}_k| &= |2\bar{d}_k^T H_k(\bar{d}_k - d_k) - (\bar{d}_k - d_k)^T H_k(\bar{d}_k - d_k)| \\
&\le 2|\bar{d}_k^T H_k(\bar{d}_k - d_k)| + |(\bar{d}_k - d_k)^T H_k(\bar{d}_k - d_k)| \\
&\le 2\kappa_H\|\bar{d}_k\|\|\bar{d}_k - d_k\| + \kappa_H\|\bar{d}_k - d_k\|^2 \\
&\le \left(\tfrac{2\kappa_H\zeta^{-1}\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l\bar{\tau}_k}} + \kappa_H\zeta^{-2}\kappa_{\mathrm{FO}}^2\alpha_k^2\right)\Delta l(x_k,\bar{\tau}_k,\bar{g}_k,\bar{d}_k),
\end{aligned}$$

这就完成了证明的第一部分。

若 $\|\bar{g}_k - \nabla f_k\| \le \epsilon_g$，按照证明第一部分的类似推理，通过 t he 三角不等式和柯西–施瓦茨不等式，（3.1），以及引理 3.1、3.11 和3.12，

$$\begin{aligned}
&|\nabla f_k^T d_k - \bar{g}_k^T\bar{d}_k| \\
&= |(\bar{g}_k - \nabla f_k)^T(\bar{d}_k - d_k) + (\bar{g}_k - \nabla f_k)^T d_k + \nabla f_k^T(\bar{d}_k - d_k)| \\
&= |(\bar{g}_k - \nabla f_k)^T(\bar{d}_k - d_k) + (\bar{g}_k - \nabla f_k)^T d_k - (H_k d_k + J_k^T y_k)^T(\bar{d}_k - d_k)| \\
&\le |(\bar{g}_k - \nabla f_k)^T(\bar{d}_k - d_k)| + |(\bar{g}_k - \nabla f_k)^T d_k| + |d_k^T H_k(\bar{d}_k - d_k)| + |y_k^T J_k(\bar{d}_k - d_k)| \\
&\le \zeta^{-1}\|\bar{g}_k - \nabla f_k\|^2 + (1+\kappa_H\zeta^{-1})\|d_k\|\|\bar{g}_k - \nabla f_k\| \\
&\le \zeta^{-1}\epsilon_g^2 + \tfrac{(1+\kappa_H\zeta^{-1})\epsilon_g}{\sqrt{\kappa_l\bar{\tau}_k}}\sqrt{\Delta l(x_k,\tau_k,\nabla f_k,d_k)}.
\end{aligned}$$

此外，根据假设2.2 可得

$$\begin{aligned}
|d_k^T H_k d_k - \bar{d}_k^T H_k\bar{d}_k| &= |(d_k - \bar{d}_k)^T H_k(d_k - \bar{d}_k) + 2d_k^T H_k(\bar{d}_k - d_k)| \\
&\le \kappa_H\|d_k - \bar{d}_k\|^2 + 2\kappa_H\|d_k\|\|d_k - \bar{d}_k\| \\
&\le \kappa_H\zeta^{-2}\epsilon_g^2 + \tfrac{2\kappa_H\zeta^{-1}\epsilon_g}{\sqrt{\kappa_l\bar{\tau}_k}}\sqrt{\Delta l(x_k,\tau_k,\nabla f_k,d_k)},
\end{aligned}$$

这就完成了证明。 □

下一个引理给出了优值函数在一次迭代过程中的界。

LEMMA 3.15. 对于所有 $k \in \mathbb{N}$ 在算法 *2.1* 的任意实现中，

$$\begin{aligned}
&\phi(x_k + \alpha_k\bar{d}_k,\bar{\tau}_k) - \phi(x_k,\bar{\tau}_k) \\
&\le -\alpha_k\Delta l(x_k,\bar{\tau}_k,\bar{g}_k,\bar{d}_k) + \alpha_k\bar{\tau}_k(\nabla f_k - \bar{g}_k)^T\bar{d}_k + \tfrac{\bar{\tau}_k L+\Gamma}{2}\alpha_k^2\|\bar{d}_k\|^2.
\end{aligned}$$

证明。 根据算法 2.1，对于任意 $k \in \mathbb{N}$， $0 < \alpha_k \le \alpha_{\max} \le 1$。此外，由三角不等式以及式（2.1）、（2.4）和（2.6），可得

$$\phi(x_k + \alpha_k\bar{d}_k,\bar{\tau}_k) - \phi(x_k,\bar{\tau}_k)$$

$$
\begin{aligned}
&= \bar{\tau}_k(f(x_k+\alpha_k\bar{d}_k)-f_k)+(\|c(x_k+\alpha_k\bar{d}_k)\|_1-\|c_k\|_1)\\
&\le \bar{\tau}_k(\alpha_k\nabla f_k^T\bar{d}_k+\tfrac{L}{2}\alpha_k^2\|\bar{d}_k\|^2)+(\|c_k+\alpha_k J_k\bar{d}_k\|_1-\|c_k\|_1+\tfrac{\Gamma}{2}\alpha_k^2\|\bar{d}_k\|^2)\\
&\le \alpha_k\bar{\tau}_k\nabla f_k^T\bar{d}_k+|1-\alpha_k|\|c_k\|_1+\alpha_k\|c_k+J_k\bar{d}_k\|_1-\|c_k\|_1+\tfrac{\bar{\tau}_kL+\Gamma}{2}\alpha_k^2\|\bar{d}_k\|^2\\
&= -\alpha_k\Delta l(x_k,\bar{\tau}_k,\bar{g}_k,\bar{d}_k)+\alpha_k\bar{\tau}_k(\nabla f_k-\bar{g}_k)^T\bar{d}_k+\tfrac{\bar{\tau}_kL+\Gamma}{2}\alpha_k^2\|\bar{d}_k\|^2,
\end{aligned}
$$

至此，证明完成。 □

由于零阶和一阶预言机（预言机0和1）的质量和可靠性，只能保证收敛到解的一个邻域。假设3.16根据 ε （以及 $\varepsilon_{\Delta l}$）给出了收敛邻域大小的下界。我们再次指出，在采用不精确概率预言机的无约束含噪情形下，也需要类似的限制；例如参见 [5, 23]。

ASSUMPTION 3.16. 当事件 $\mathcal{E}$ 发生时（见假设*3.2*），令

$$
\varepsilon > \max\left\{\tfrac{\epsilon_g}{\eta}, \omega_{10}\sqrt{\epsilon_f}\right\}\tfrac{\max\{\kappa_H,1\}}{\sqrt{\kappa_l\tau_{\min}}},
$$

根据注释 *3.8*，这等价于 $\varepsilon_{\Delta l} > \max\left\{\tfrac{\epsilon_g}{\eta}, \omega_{10}\sqrt{\epsilon_f}\right\}$，其中 $\{\kappa_H,\tau_{\min},\kappa_l\}\subset\mathbb{R}_{>0}$ 在假设 *2.2* 和引理 *3.3* 和 *3.11* 中定义，(ϵ_f,ϵ_g) 是预言机 *0* 和 *1* 在条件事件 $\mathcal{E}$ 下的内在参数，$0<\eta<2(1-\theta)\min\left\{\tfrac{1}{\eta_1+\eta_2},\tfrac{1}{\eta_3+\eta_4}\right\}$，而 $\{\eta_1,\eta_2,\eta_3,\eta_4,\omega_{10}\}\subset\mathbb{R}_{>0}$ 在表 *A.1*（附录 *A*）以及 $p\in\left(\tfrac{1}{2},1\right]$ 中定义。

假设 3.16涉及许多常数，确实难以理解。为展示其对收敛邻域的精确依赖关系，我们将所有常数明确写出。尽管如此，重要的是，下界 ε 与梯度近似中的偏差成正比，并与函数近似中噪声水平的平方根成正比。

现在我们可以给出本节的关键引理。在引理 3.17 中，我们首先定义 $(p,\tilde{\alpha},h(\cdot))$，其中 $p\in\left(\tfrac{1}{2},1\right]$ 是给定过去信息（停止时刻之前）时真实迭代概率的下界，$\tilde{\alpha}\in\mathbb{R}_{>0}$ 是大步长阈值，而 $h\colon\mathbb{R}_{>0}\to\mathbb{R}_{>0}$ 是一个关于 α 的单调递增函数，用于界定任意一次迭代所取得的势函数进展。此外，我们证明了五个结果，可概括如下：*(i)* 步长为 $\tilde{\alpha}$ 时，势函数进展的下界与 ϵ_f 成正比；*(ii)* 给定过去信息和事件 $\mathcal{E}$ （见假设 3.2），下一次迭代为真实迭代的概率至少为 p；*(iii)* 界定事件$\mathcal{E}$下真实且成功的迭代所取得的势函数进展；*(iv)* 在事件 $\mathcal{E}$ 下，步长较小的真实迭代是成功迭代；以及 *(v)* 界定任意迭代 k 中 Z_k 的势函数增量，其上界与 ϵ_f 成正比。

LEMMA 3.17. 假设 *3.2* 和 *3.16* 成立。对于所有 $k<T_{\varepsilon_{\Delta l}}$，令

- $p=1-\delta$ 当噪声由 ϵ_f 界定时，否则为 $p=1-\delta-$ 指数$\left(-\text{最小值}\{\tfrac{u^2}{2\nu^2},\tfrac{u}{2b}\}\right)$
（其中 $u=$ 下确界$_{x\in\mathcal{X}}\{\epsilon_f-\mathbb{E}_{\Xi^0}[E(x,\Xi^0)\,|\mathcal{E}]\}$，
- $h(\alpha)=\alpha\theta\varepsilon_{\Delta l}^2$ 最小值$\left\{\tfrac{1-\eta\omega_4}{1+\epsilon_\tau\omega_1},\ 1-\eta\omega_5,\ \tfrac{1}{1+\omega_2+\omega_3}\right\}$，
- 且，$\tilde{\alpha}=$ 最小化$\left\{\tfrac{1-\theta}{\omega_7},\ \bar{\tau}^{\ \omega_8}_{\text{最小化}L+\Gamma}\right\}$。常数

$\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_7,\omega_8\}\subset\mathbb{R}_{>0}$ 定义见表 *A.1* (附录 *A*）。那么，以下结果成立：

(i) $h(\tilde{\alpha})>\tfrac{4\bar{\tau}^{f\ g}_{-1}}{p-1/2}\epsilon_f$.

(ii) $\mathbb{P}[I_k = 1|\mathcal{F}_{k-1} \cap \mathcal{E}] \geq p$ 满足某个 $p \in \left(\frac{1}{2} + \frac{4\bar{\tau}_{-1}\epsilon_f}{h\ \tilde{\alpha}}(\ ) \quad 1\right]$。*(iii)* 对于任意一个算法 *2.1* 的实现，若事件 $\mathcal{E}$ 发生（见假设 *3.2*），如果某次迭代 k 为真且成功，则 $Z_{k+1} \leq Z_k - h(\alpha_k) + 4\bar{\tau}_{-1}\epsilon_f$。*(iv)* 对于任意一个算法 *2.1* 的实现，若事件 $\mathcal{E}$ 发生（见假设 *3.2*），如果某次迭代 k 为真且 $\alpha_k \leq \tilde{\alpha}$，则迭代 k 也成功。*(v)* 对于算法 *2.1* 的任意实现中的所有 $k \in \mathbb{N}$， $Z_{k+1} \leq Z_k + 2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_{-1}(e_k + e_k^+)$。

证明。 为简洁起见，证明推迟至附录B。 □

接下来的两个引理将用于后续的 ε-复杂度分析。

LEMMA 3.18. 假设*3.2* 成立。对于所有 $t \geq 1$ 以及任意 $\hat{p} \in [0, p)$，有

$$\mathbb{P}\left[\sum_{k=0}^{t-1} I_k < \hat{p}t \middle| \mathcal{E}\right] \leq e^{-\frac{(p-\hat{p})^2}{2p^2}t},$$

其中，事件 $\mathcal{E}$ 定义见假设*3.2*。

证明。 证明与 [23, 引理 3.1]相同。 □

LEMMA 3.19. 假设*3.2*成立。对于任意正整数 t 以及任意 $\hat{p} \in \left(\frac{1}{2},\ 1\right]$，有

$$\mathbb{P}\left[T_{\varepsilon_{\Delta l}} > t,\ \sum_{k=0}^{t-1} I_k \geq \hat{p}t, \sum_{k=0}^{t-1} \Theta_k I_k U_k < \left(\hat{p} - \tfrac{1}{2}\right)t - \tfrac{l}{2} \middle| \mathcal{E}\right] = 0,$$

其中 $l =$ 最大 $\left\{-\frac{\ln \alpha_0 - \ln \tilde{\alpha} \ln \gamma}{}, 0\right\}$ 且事件 $\mathcal{E}$ 在假设 *3.2*中定义。

证明。 证明与 [23,引理 3.5]相同。 □

现在给出主定理；算法 ε2.1 的迭代复杂度。

THEOREM 3.20. 假设条件 *2.1*、*2.2*、*3.2* 和*3.16* 成立，并且预言机*0*和*1*的条件在事件 $\mathcal{E}$ 发生时满足（见假设*3.2*）。那么，对于任意 $s \geq 0$， $\hat{p} \in \left(\frac{1}{2} + \frac{4\bar{\tau}_{-1}\epsilon_f + s}{h\ \tilde{\alpha}}(\ )\right.$ $\left.p\right)$，以及 $t \geq$ $\frac{R}{\hat{p} - \frac{1}{2} - \frac{4\bar{\tau}_{-1}\epsilon_f + s}{h\ \tilde{\alpha}}(\ )}$，

$$\mathbb{P}[T_{\varepsilon_{\Delta l}} \leq t|\mathcal{E}] \geq 1 - e^{-\frac{(p-\hat{p})^2}{2p^2}t} - e^{-\min\left\{\frac{s^2 t}{2(2\bar{\tau}_{-1}\nu)^2}, \frac{st}{2(2\bar{\tau}_{-1}b)}\right\}},$$

其中 $R = \frac{Z_0}{h\ \tilde{\alpha}(\ ) +}$ 最大值 $\left\{\frac{}{\text{自然对数 } \tilde{\alpha} - \text{自然对数 } 2}, \frac{}{\alpha_0 \text{ 自然对数 } \gamma},\ 0\right\}$ 和 $(p,\ \tilde{\alpha},\ h(\cdot))$ 定义同引理 *3.17*.

证明。根据全概率公式，

$$\mathbb{P}[T_{\varepsilon_{\Delta l}} > t|\mathcal{E}] = \underbrace{\mathbb{P}\left[T_{\varepsilon_{\Delta l}} > t,\ \tfrac{1}{t}\sum_{k=0}^{t-1}(2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_{-1}(E_k + E_k^+)) > 4\bar{\tau}_{-1}\epsilon_f + s \middle| \mathcal{E}\right]}_{\boldsymbol{A}}$$
$$+ \underbrace{\mathbb{P}\left[T_{\varepsilon_{\Delta l}} > t,\ \tfrac{1}{t}\sum_{k=0}^{t-1}(2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_{-1}(E_k + E_k^+)) \leq 4\bar{\tau}_{-1}\epsilon_f + s \middle| \mathcal{E}\right]}_{\boldsymbol{B}}.$$

首先，我们对 $\mathbb{P}[\boldsymbol{A}]$进行界定。在事件 $\mathcal{E}$ （见假设3.2）条件下，对于每次迭代 k，由于 E_k 和E_k^+ 满足参数为（ν， b的单侧次指数界（2.11），可以证明 $\bar{\tau}_{-1}$（$E_k+E_k^+$）满足参数为（$2\bar{\tau}_{-1}\nu$， $2\bar{\tau}_{-1}b$的（2.11）。此外，由于 $\mathbb{E}[\bar{\tau}_{-1}$（$E_k+E_k^+$）$|\mathcal{E}]$ 有界于 $2\bar{\tau}_{-1}\epsilon_f$，应用单侧伯恩斯坦不等式，对任意 $s\geq 0$

$$\mathbb{P}[\boldsymbol{A}] \leq \mathbb{P}\left[\frac{1}{t}\sum_{k=0}^{t-1}\bar{\tau}_{-1}(E_k+E_k^+) > 2\bar{\tau}_{-1}\epsilon_f + s \middle| \mathcal{E}\right] \leq e^{-\min\left\{\frac{s^2 t}{2(2\bar{\tau}_{-1}\nu)^2}, \frac{st}{2(2\bar{\tau}_{-1}b)}\right\}}.$$

令 $l=\max\left\{-\frac{\ln\alpha_0}{\ln\tilde{\alpha}\ln\gamma}, 0\right\}$。为界定 $\mathbb{P}[B]$，我们应用全概率公式，

$$\mathbb{P}[\boldsymbol{B}] = \underbrace{\mathbb{P}\left[\sum_{k=0}^{t-1}\Theta_k I_k U_k < \left(\hat{p}-\tfrac{1}{2}\right)t - \tfrac{l}{2},\ \boldsymbol{B}\right]}_{\boldsymbol{B}_1} + \underbrace{\mathbb{P}\left[\sum_{k=0}^{t-1}\Theta_k I_k U_k \geq \left(\hat{p}-\tfrac{1}{2}\right)t - \tfrac{l}{2},\ \boldsymbol{B}\right]}_{\boldsymbol{B}_2}.$$

我们首先证明 $\mathbb{P}[\mathrm{B}_2]=0$。根据引理 3.17，对于算法 2.1 的任何一种实现，只要事件 $\mathcal{E}$ 发生（见假设 3.2），对于任意迭代 $k<T_{\varepsilon_{\Delta l}}$，便有 $Z_{k+1}\leq Z_k - h(\tilde{\alpha}) + 2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_{-1}(E_k+E_k^+) \leq Z_k - h(\tilde{\alpha}) + 4\bar{\tau}_{-1}\epsilon_f$，当 $U_k I_k \Theta_k = 1$时成立；当 $U_k I_k \Theta_k = 0$时，则有 $Z_{k+1}\leq Z_k + 2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_{-1}(E_k+E_k^+)$。根据假设 3.2，
$\mathbb{E}[E_k|\mathcal{E}]$ $\mathbb{E}[E_k^+|\mathcal{E}]$ ϵ_f k $\mathcal{E}$ and are bounded above by for all 。在以事件 （见假设 3.2）为条件时，事件 $T_{\varepsilon_{\Delta l}} > t$意味着 $Z_t > 0$（因为根据引理 3.9 的证明，只有当 $T_{\varepsilon_{\Delta l}} \leq t$ 时才会发生 $Z_t = 0$）。结合 $\frac{1}{t}\sum_{k=0}^{t-1}(2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_{-1}(E_k+E_k^+)) \leq 4\bar{\tau}_{-1}\epsilon_f + s$，便可推出事件 $\sum_{k=0}^{t-1}\Theta_k I_k U_k < (\hat{p}-\frac{1}{2})t - \frac{l}{2}$。为说明这一点，假设 $-\sum_{k=0}^{t-1}\Theta_k I_k U_k \geq (\hat{p}-\frac{1}{2})t - \frac{l}{2}$，则 $-$ $-$

$$\begin{aligned} Z_t &\leq Z_0 - \left[\left(\left(\hat{p}-\tfrac{1}{2}\right)t - \tfrac{l}{2}\right)h(\tilde{\alpha}) - \sum_{k=0}^{t-1}(2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_{-1}(E_k+E_k^+))\right] \\ &\leq Z_0 - \left(\left(\hat{p}-\tfrac{1}{2}\right)t - \tfrac{l}{2}\right)h(\tilde{\alpha}) + t(4\bar{\tau}_{-1}\epsilon_f + s) \\ &= Z_0 - \left(\left(\hat{p}-\tfrac{1}{2}\right)h(\tilde{\alpha}) - (4\bar{\tau}_{-1}\epsilon_f + s)\right)t + \tfrac{l}{2}h(\tilde{\alpha}) \leq 0. \end{aligned}$$

上面的最后一个不等式源于以下假设：$\hat{p} > \frac{1}{2} + \frac{4\bar{\tau}_{-1}\epsilon_f + s}{h\ \tilde{\alpha}}$ () 且 $t \geq \frac{l}{2\hat{p}-1} - \frac{4\bar{\tau}_{-1}\epsilon_f + s}{h\ \tilde{\alpha}}$ ()。因此，$\mathbb{P}[\boldsymbol{B}_2]=0$。现在我们对 $\mathbb{P}[\boldsymbol{B}_1]$进行界定；根据引理 3.18和 3.19，

$$\begin{aligned} \mathbb{P}[\boldsymbol{B}_1] &\leq \mathbb{P}\left[T_{\varepsilon_{\Delta l}} > t,\ \sum_{k=0}^{t-1}\Theta_k I_k U_k < \left(\hat{p}-\tfrac{1}{2}\right)t - \tfrac{l}{2}\middle|\mathcal{E}\right] \\ &= \mathbb{P}\left[T_{\varepsilon_{\Delta l}} > t,\ \sum_{k=0}^{t-1}\Theta_k I_k U_k < \left(\hat{p}-\tfrac{1}{2}\right)t - \tfrac{l}{2}, \sum_{k=0}^{t-1} I_k < \hat{p}t\middle|\mathcal{E}\right] \\ &\quad + \mathbb{P}\left[T_{\varepsilon_{\Delta l}} > t,\ \sum_{k=0}^{t-1}\Theta_k I_k U_k < \left(\hat{p}-\tfrac{1}{2}\right)t - \tfrac{l}{2}, \sum_{k=0}^{t-1} I_k \geq \hat{p}t\middle|\mathcal{E}\right] \\ &\leq \mathbb{P}\left[\sum_{k=0}^{t-1} I_k < \hat{p}t\middle|\mathcal{E}\right] + \mathbb{P}\left[T_{\varepsilon_{\Delta l}} > t,\ \sum_{k=0}^{t-1}\Theta_k I_k U_k < \left(\hat{p}-\tfrac{1}{2}\right)t - \tfrac{l}{2}, \sum_{k=0}^{t-1} I_k \geq \hat{p}t\middle|\mathcal{E}\right] \\ &\leq e^{-\frac{(p-\hat{p})^2}{2p^2}t} + 0 = e^{-\frac{(p-\hat{p})^2}{2p^2}t}. \end{aligned}$$

结合 $\mathbb{P}[\boldsymbol{A}]$和 $\mathbb{P}[\boldsymbol{B}]$即可完成证明。 □

COROLLARY 3 21 3 20, $s \geq 0$ $p \in$ ..在 – $\left(\frac{1}{2} + \frac{4\bar{\tau}_{-1}\epsilon_f + s}{\tilde{\alpha}\theta\omega_p\varepsilon^2_{\Delta l}}, p\right)$ $t \geq \frac{R}{\quad}$ 定理 .对任意， ˆ和 $\hat{p} - \frac{1}{2} - \frac{4\bar{\tau}_{-1}\epsilon_f + s}{\tilde{\alpha}\theta\omega_p\varepsilon^2_{\Delta l}}$， $\frac{\quad}{\quad}$

$$\mathbb{P}\left[T_{\varepsilon_{\Delta l}} \leq t | \mathcal{E}\right] \geq 1 - e^{-\frac{(p-\hat{p})^2}{2p^2}t} - e^{-\min\left\{\frac{s^2 t}{2(2\bar{\tau}_{-1}\nu)^2}, \frac{st}{2(2\bar{\tau}_{-1}b)}\right\}}, \tag{3.8}$$

其中 ˆ$R = \frac{\phi(x_0,\bar{\tau}_{-1}) - \phi_{\min} - (\bar{\tau}_{-1} - \bar{\tau}_{\min})}{\tilde{\alpha}\theta\omega_p\varepsilon^2_{\Delta l}}$ $\frac{f_{\inf}}{\quad} + max\left\{ln\ \tilde{\alpha} - ln\ 2^{\alpha_0} ln\ \gamma,\right.$ $\left.0\right\}$；等价地，根据注释 *3.8*，ˆ$R = max_{\{\kappa_H^2,1\}} \frac{\phi(x_0,\bar{\tau}_{-1}) - \phi_{\min} - (\bar{\tau}_{-1} - \bar{\tau}_{\min})}{\tilde{\alpha}\theta\omega_p\varepsilon^2} + max\left\{ln\ \tilde{\alpha} - ln\ 2^{\alpha_0^{\kappa_l \tau_{\min}}} ln\ \gamma, 0\right\}$；事件 $\mathcal{E}$定义于假设 *3.2* 中，$\omega_p = min\left\{\frac{1-\eta\omega_4}{1+\epsilon_\tau\omega_1}, 1 - \eta\omega_5, \frac{1}{1+\omega_2+\omega_3}\right\}$；其余常数定义见表 *A.1*（附录 *A*）。

REMARK 3.22. 我们对论文的主要理论结果（定理 *3.20* 和推论 *3.21*）作几点说明。

- （迭代 ε 复杂度）根据定义 *3.7*（以及注释 *3.8*）和推论 *3.21*，我们得出结论：在事件 $\mathcal{E}$ （见假设 *3.2*）成立的条件下，以极高的概率，算法 *2.1* 生成满足 $max\{\|\nabla f_k + J_k^T y_k\|, \sqrt{\|c_k\|}\} \leq \varepsilon$ 的原始-对偶迭代点 $(x_k, y_k) \in \mathbb{R}^n \times \mathbb{R}^m$ 所需的迭代 ε 复杂度为 $\mathcal{O}(\varepsilon^{-2})$。就对 ε 的依赖而言，这一迭代 ε 复杂度与确定性对应形式（例如， *[16]*）的迭代复杂度同阶，但由于预言机（预言机 *0* 和 *1*）中的噪声和偏差，还需满足 ε 远离零（假设 *3.16*）的限制。

- （几乎必然收敛）我们注意到，在事件 $\mathcal{E}$ 发生时（见假设 *3.2*），算法 *2.1* 会在有限次迭代内找到一个 ε-驻点迭代点，且其概率为 *1*，即 $\mathbb{P}[\cap_{k=1}^{\infty} \cup_{t=k}^{\infty} (T_{\varepsilon_{\Delta l}} > t) | \mathcal{E}] = 0$。这是博雷尔–坎泰利引理的直接推论，因为由 (3.8) 可知，失败事件的概率之和有限，即 $\sum_{t=1}^{\infty} \mathbb{P}[T_{\varepsilon_{\Delta l}} > t | \mathcal{E}] = \sum_{t=1}^{\infty} (1 - \mathbb{P}[T_{\varepsilon_{\Delta l}} \leq t | \mathcal{E}]) < \infty$。

- （无约束情形）本文中的高概率 ε-复杂度界是无约束版本（例如， *[23]*）的泛化。在无约束情形下，各参数化简为 $\sigma = 0$, $\zeta_y = 0$, $\omega_1 = 1$, $\Gamma = 0$, $\zeta = 1$, $\kappa_H = 1$,$\kappa_l = 1$, $\epsilon_\tau = 0$，以及 $\bar{\tau}_k = 1$ 对所有 $k \in \mathbb{N}$均成立。虽然将这些取值代入推论 *3.21* 的结果并不能完全还原无约束情形下的结论（例如，*[23]*），但就对 ε的依赖而言，结果的阶相同。差距源于自适应优越性参数带来的额外复杂度。我们强调，尽管与 *[23]*, 相比结果存在常数差异，我们的算法仍可恢复确定性变体 *[16]*的复杂度界。

4. 数值结果。本节展示所提出算法在标准等式约束非线性优化问题上的数值结果。数值实验旨在考察 SS-SQP 算法在一组多样化测试问题上的效率和稳健性，这些问题的目标函数和梯度评估受到不同程度的噪声影响。所有实验均在 MATLAB 中进行。在展示数值结果之前，我们先介绍测试问题、实现细节和评估指标。

4.1. 测试问题。我们在 CUTEst 集合 [18]中的一部分等式约束优化问题上进行了数值实验。我们选取了满足以下标准的问题：(*i*) 目标函数不是常数函数，(*ii*) 变量和约束的总数不超过 10^3, ，并且 (*iii*) 在所有

所有运行中的迭代点处，约束的雅可比矩阵的奇异值均大于 10^{-8}。由此得到 35 个不同维度的测试问题。

我们考虑了 35 个 CUTEst 测试问题的含噪版本（目标函数和梯度评估带有噪声）。具体而言，每当需要进行目标函数或目标函数梯度评估时，分别采用了近似值 $\bar{f}(x;\xi) = \mathcal{N}\left(f(x), \epsilon_{f,N}^2\right)$ 和 $\bar{g}(x;\xi') = \mathcal{N}\left(\nabla f(x), \frac{\epsilon_{g,N}^2}{n} I\right)$。在目标函数和梯度评估中，我们考虑了 4 种不同的噪声水平，分别由常数 $\epsilon_{f,N} \in \{0, 10^{-4}, 10^{-2}, 10^{-1}\}$ 和 $\epsilon_{g,N} \in \{0, 10^{-4}, 10^{-2}, 10^{-1}\}$ 决定。每个 CUTEst 测试问题都有一个唯一的初始起点，该起点被用作所有算法所有运行的起点。此外，对于每个选定的噪声水平元组 $(\epsilon_{f,N}, \epsilon_{g,N}) \in \{0, 10^{-4}, 10^{-2}, 10^{-1}\} \times \{10^{-4}, 10^{-2}, 10^{-1}\} \cup \{0\} \times \{0\}$，在适用情况下，我们使用五个不同的随机种子运行每个问题。

4.2. 实现细节。我们在前文介绍的含噪 CUTEst 问题上，将 SS-SQP（算法 2.1）与 [7]提出的自适应随机序列二次规划算法（我们称之为 AS-SQP）进行了比较。SS-SQP 的用户定义参数设置如下：$\epsilon_f = \epsilon_{f,N}, \epsilon_g = \epsilon_{g,N}, \epsilon_\tau = 10^{-2}, \bar{\tau}_{-1} = \sigma = 0.1, \gamma = 0.5, \theta = 10^{-4}, \alpha_0 = \alpha_{\max} = 1$，并且对所有 $k \in \mathbb{N}$均如此。对于 AS-SQP，[7] 我们将参数设置如下（参数选择参考了 [7]中的参数设置）：$\bar{\tau}_{-1} = \sigma = 0.1, \bar{\xi}_{-1} = 1, \epsilon = 10^{-2}, \theta = 10^4, H_k = I$，并且对所有 $k \in \mathbb{N}$均如此。AS-SQP 的步长规则需要已知（或估计）Lipschitz 常数 L 和 Γ。为此，我们利用初始点附近的梯度差估计这些常数，并对所有 $k \in \mathbb{N}$设置 $L_k = L$ 和 $\Gamma_k = \Gamma$。需要指出的是，尽管 SS-SQP 算法的分析要求满足预言机 1 的条件，但实验中既未强制满足也未检查这些条件；相反，在每次实验中，算法都使用了精度相同且预先指定的随机梯度估计（如上文所述）。尽管如此，SS-SQP 与 AS-SQP 的区别在于，前者需要目标函数的估计，而后者不需要（AS-SQP 是一种无目标函数的方法）。具体而言，SS-SQP 每次迭代需要 2 次函数评估和 1 次梯度评估，而 AS-SQP 每次迭代只需要 1 次梯度评估。我们将在介绍数值结果时进一步讨论这一点。

4.3. 终止条件和评估指标。在所有实验中，我们都根据不同的评估指标（迭代次数和计算工作量）报告不可行性（$\|c(x_k)\|_\infty$）和驻点性（KKT）（$\max\{\|c(x_k)\|_\infty, \min_{y\in\mathbb{R}^m} \|\nabla f(x_k) + \nabla c(x_k)y\|_\infty\}$）。所有算法的迭代预算均为 10^3 次，只有在找到近似驻点时才会提前终止运行；近似驻点定义为满足 $x_* \in \mathbb{R}^n$ 使得 $\|c(x_*)\|_\infty \le 10^{-6}$ 且 $\min_{y\in\mathbb{R}^m} \|\nabla f(x_*) + \nabla c(x_*)y\|_\infty \le 10^{-4}$ 的点。

我们以性能剖面的形式展示结果，分别考察迭代次数和计算工作量（定义为函数和梯度评估的总次数，或等价地，概率预言机调用的总次数）。每次迭代中，SS-SQP 都需要预言机 0 和 1，而 AS-SQP 只需要预言机 1，这意味着 SS-SQP 每次迭代的成本更高。此外，我们采用 [28], 所述的收敛度量，即 $m(x_0) - m(x) \ge (1 - \epsilon_{pp})(m(x_0) - m_b)$，其中 $m(x)$ 或为$\|c(x)\|_\infty$ (不可行性)，或为 $\max\{\|c(x)\|_\infty, \min_{y\in\mathbb{R}^m} \|\nabla f(x) + \nabla c(x)y\|_\infty\}$ (驻点性（KKT）)；x_0 为初始迭代点，m_b 为给定问题实例在预算范围内由任一算法得到的该度量的最佳值，$\epsilon_{pp} \in (0, 1)$为容差。对于所展示的所有实验，我们均取 $\epsilon_{pp} = 10^{-3}$。

4.4. 含噪梯度与精确函数 ($\epsilon_f = 0$)。在第一组实验中，我们考虑目标函数评估精确、目标梯度评估含噪的问题，并比较 SS-SQP与 AS-SQP。本实验旨在展示梯度噪声的影响，以及使用（精确）函数值的优势。图4.1 中的每一行展示了不同梯度噪声水平下的性能剖面（最下面一行的噪声水平最高），每一列则展示了不同的评估指标。从无噪声基准情形（$\epsilon_f = 0$ 和 $\epsilon_g = 0$，图 4.1第一行）开始，可以看出，两种方法在不可行性误差和 KKT误差方面的表现相近；就 KKT误差而言，SS-SQP 在有效性（可求解的问题总数）方面略占优势。随着梯度噪声增大，两种方法在所有指标上的表现差距也随之扩大，且有利于 SS-SQP。这并不意外，因为SS-SQP 还利用了额外信息（精确函数值）。这些结果凸显了可靠的函数信息对方法性能的影响。

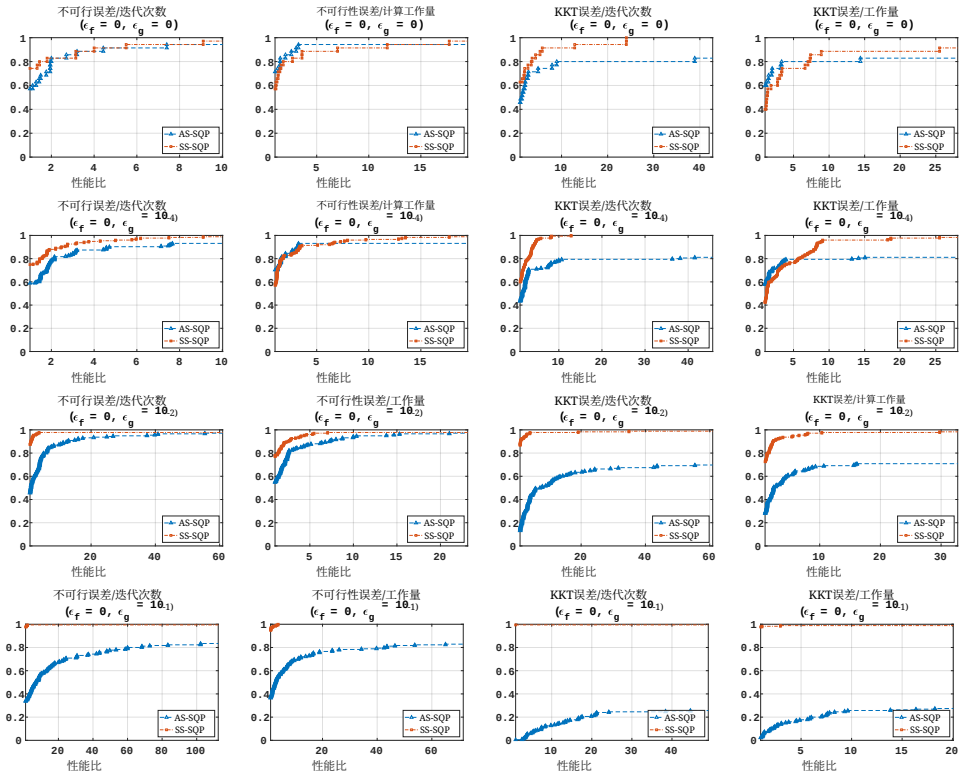


FIG. 4.1。 *Performance profiles for AS-SQP and SS-SQP on CUTEst collection [18] with deterministic objective function evaluations ($\epsilon_f = 0$) and noisy objective gradient evaluations. Each column corresponds to a different evaluation metric (infeasibility and KKT errors vs. iterations and work). The noise in the objective gradient evaluations ϵ_g increases from top to bottom (First row: $\epsilon_g = 0$; Second row: $\epsilon_g = 10^{-4}$; Third row: $\epsilon_g = 10^{-2}$; Fourth row: $\epsilon_g = 10^{-1}$).*

4.5. 含噪函数与梯度。这里给出了目标函数和梯度评估均含有噪声时的结果。如图4.1所示，在图4.2中，不同的行展示了梯度中不同噪声水平下的结果（底行的噪声最高），不同的列展示了不同评估指标下的结果。

每个性能剖面图包含 4 条线：AS-SQP（即无目标函数，且不受函数评估中噪声的影响）以及采用不同目标函数评估噪声水平的三种 SS-SQP 方法变体。我们可以观察到以下几点。首先，不出所料，随着目标函数评估中的噪声增加，SS-SQP 方法的性能会下降。其次，在不可行性误差方面，AS-SQP 和 SS-SQP 具有竞争力，并达到相近的稳健性水平。第三，也是最有意思的一点是，方法的性能取决于函数和梯度评估的相对误差。具体而言，当目标函数噪声水平相对于目标梯度偏差足够小时，SS-SQP 表现更好。另一方面，当函数估计相对于梯度评估中的噪声水平过于嘈杂时，AS-SQP 的表现略好。这些结果凸显了在存在噪声时无目标函数优化方法的优势（尤其是在目标函数评估噪声较高时），也体现了在需要零阶信息的方法中，高质量（或至少相对高质量）函数评估的价值。

最后，我们对优值参数的表现作几点说明。在无噪声情形（$\epsilon_f = \epsilon_g = 0$）下，在所有问题中，优值参数的最小值约为 10^{-4}。在有噪声情形下，针对全部 2100 个问题实例（35 个问题、12 种噪声水平、5 次重复试验），优值参数的最小值约为 10^{-6}（该值对应于噪声水平最高的问题），且最终优值参数值中不到 5% 小于约 10^{-4}。此外，对于具有相同噪声水平的同一问题的不同实现，最终优值参数值的方差较小（且取决于噪声水平）。

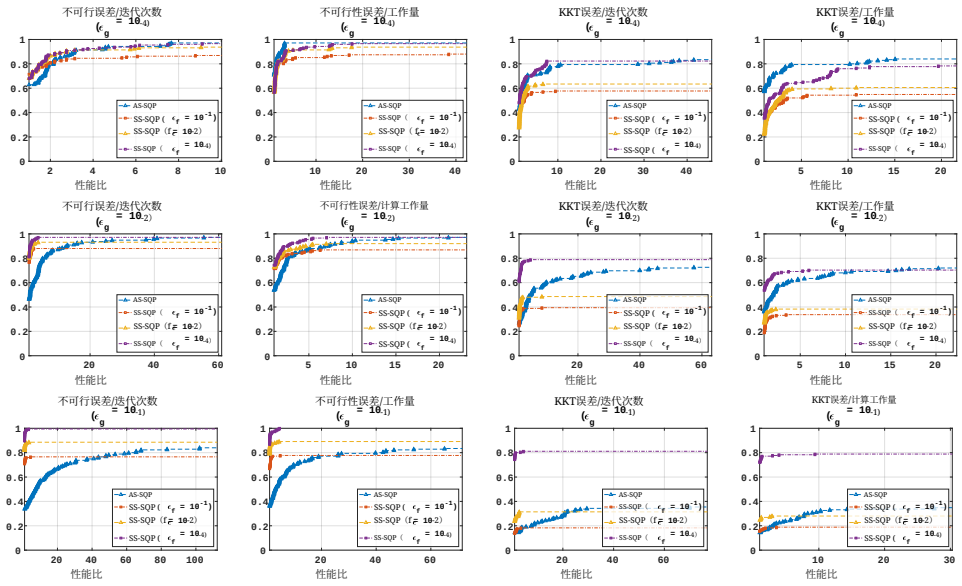


FIG. 4.2。

Performance profiles for `AS-SQP` *and* `SS-SQP` *on CUTEst collection [18] with noise in both the objective function and gradient evaluations. Each column corresponds to a different evaluation metric (infeasibility, KKT, work, and work+). The second row corresponds to the objective gradient evaluations, ranging from top to bottom (first row: $\epsilon_g = 10^{-1}$, second row: $\epsilon_g = 10^{-2}$, third row: $\epsilon_g = 10^{-1}$). The different variants of* `SS-SQP` *correspond to different levels of noise in the objective function evaluations.*

5. 结论。我们提出了一种步长搜索 SQP 算法（SS-SQP），用于求解具有确定性等式约束的随机优化问题。我们证明，在对非精确概率零阶和一阶预言机作出合理假设的条件下，对于任意 ε，只要其大于由预言机中的噪声和偏差决定的某个正数，我们的算法就能以极高的概率，在 $O(\varepsilon^{-2})$ 次迭代内产生一个满足一阶 ε-驻点性的迭代点，其迭代复杂度与 SQP 算法的确定性对应形式相当 [16]。数值结果有力地证明了所提方法的效率和有效性。未来的研究方向包括但不限于：（1）在算法设计和分析中纳入随机约束评估；（2）将该框架扩展到具有不等式约束的情形。上述两个方向都是未来的研究课题，因为它们要求对算法的设计、分析和实现作出重大调整。

致谢。本文基于美国海军研究办公室通过编号为 N00014-21-1-2532 的奖项提供的支持。我们谨向 Frank E. Curtis 教授和 Katya Scheinberg 教授致以谢意，感谢他们给予的宝贵支持和反馈。

参考文献。

[1] Bandeira, K. Scheinberg, and L. N. Vicente, *Convergence of trust-region methods based on probabilistic models*, SIAM《优化杂志》, 24 (2014), 第 1238–1264 页。[2] S. Bellavia, E. Fabrizi, and B. Morini, *Linesearch Newton-CG methods for convex optimization with noise*,《费拉拉大学学报》, 68 (2022), 第 483–504 页。[3] A. S. Berahas, R. Bollapragada, and B. Zhou, *An adaptive sampling sequential quadratic programming method for equality constrained stochastic optimization*, arXiv 预印本 arXiv:2206.00712, (2022)。[4] A. S. Berahas, R. H. Byrd, and J. Nocedal, *Derivative-free optimization of noisyfunctions via quasi-Newton methods*, SIAM《优化杂志》, 29 (2019), 第 965–993 页。[5] A. S. Berahas, L. Cao, and K. Scheinberg, *Global convergence rate analysis of ageneric line search algorithm with noise*, SIAM《优化杂志》, 31 (2021), 第 1489–1518 页。[6] A. S. Berahas, F. E. Curtis, M. J. O'Neill, and D. P. Robinson, *A stochastic sequential quadratic optimization algorithm for nonlinear-equality-constrained optimization with rank-deficient jacobians*,《运筹学研究》, (2023)。[7] A. S. Berahas, F. E. Curtis, D. Robinson, and B. Zhou, *Sequential quadratic optimization for nonlinear equality constrained stochastic optimization*, SIAM《优化杂志》, 31 (2021), 第 1352–1379 页。[8] A. S. Berahas, J. Shi, Z. Yi, and B. Zhou, *Accelerating stochastic sequential quadratic programming for equality constrained optimization using predictive variance$_{reduction}$*,《计算优化与应用》, (2023), 第 1–38 页。[9] D. Bertsekas, *Network optimization: continuous and discrete models*,第 8 卷，Athena Scientific，1998。[10] J. Blanchet, C. Cartis, M. Menickelly, and K. Scheinberg, *Convergence rate analysis of a stochastic trust-region method via supermartingales*,《INFORMS 优化期刊》, 1 (2019), 第 92–119 页。[11] R. H. Byrd, F. E. Curtis, and J. Nocedal, *An inexact SQP method for equalityconstrained optimization*, SIAM《优化杂志》, 19 (2008), 第 351–369 页。[12] L. Cao, A. S. Berahas, and K. Scheinberg, *First-and second-order high probabilitycomplexity bounds for trust-region methods with noisy oracles*,《数学规划》, (2023), 第 1–52 页。[13] C. Cartis and K. Scheinberg, *Global convergence rate analysis of unconstrained op-timization methods based on probabilistic models*,《数学规划》, 169 (2018), 第 337–375 页。

[14] C. Chen, F. Tung, N. Vedula, and G. Mori, *Constraint-aware deep neural network compression* 2018, 400 415 , 收录于欧洲计算机视觉会议论文集，第 – 页。[15] R. Chen, M. Menickelly, and K. Scheinberg, *Stochastic optimization using a trust-region method and random models* 169 2018 447 487 ,《数学规划》, ( ), 第 – 页。[16] F. E. Curtis, M. J. O'Neill, and D. P. Robinson, *Worst-case complexity of an SQP method for nonlinear equality constrained stochastic optimization* ,《数学规划》, (2023), 第 1–53 页。[17] F. E. Curtis, D. P. Robinson, and B. Zhou, *A stochastic inexact sequential quadratic optimization algorithm for nonlinear equality-constrained optimization*,《INFORMS 优化期刊》, (2024)。[18] N. I. Gould, D. Orban, and P. L. Toint, *CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization* ,《计算优化与应用》, 60 (2015), 第 545–557 页。[19] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang, *Direct search based on probabilistic descent* 25 2015 1515 1541 , SIAM《优化杂志》, ( ), 第 – 页。[20] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang, *Complexity and global rates of trust-region methods based on probabilistic models* 38 , IMA数值分析期刊, (2018), 第 1579–1597 页。[21] R. W. Hamming, *Introduction to applied numerical analysis*, Courier Corporation, 2012。[22] E. Hazan and H. Luo, *Variance-reduced and projection-free stochastic optimization*, 收录于国际机器学习会议，机器学习研究论文集，2016，第 1263–1271 页。[23] B. Jin, K. Scheinberg, and M. Xie, *High probability complexity bounds for adap-tive step search based on stochastic oracles*, SIAM优化期刊, 34 (2024), 第 2411–2439 页。[24] H. J. Kushner and D. S. Clark, *Stochastic approximation methods for constrained and unconstrained systems* 26, 2012 ,第 Springer Science & Business Media 卷。[25] G. Lan, *First-order and stochastic optimization methods for machine learning*,施普林格，2020。[26] H. Lu and R. M. Freund *Generalized stochastic Frank–Wolfe algorithm with stochas-* ,, 《数学规划》, *tic "substitute" gradient for structured convex optimization* 187 2021 ( ), 第 317–349 页。[27] M. Menickelly, S. M. Wild, and M. Xie, *A stochastic quasi-newton method in the absence of common random numbers* 2302 09128, 2023 , arXiv 预印本 arXiv: . ( )。[28] J. J. Moré and S. M. Wild, *Benchmarking derivative-free optimization algorithms*, SIAM《优化杂志》, 20 (2009), 第 172–191 页。[29] S. Na, M. Anitescu, and M. Kolar, *An adaptive stochastic sequential quadratic programming with differentiable exact augmented lagrangians* 2022 ,《数学规划》, ( ), 第 1–71 页。[30] S. Na, M. Anitescu, and M. Kolar, *Inequality constrained stochastic nonlinear optimization via active-set sequential quadratic programming* 2023 ,《数学规划》, ( ), 第 1–75 页。[31] S. Na and M. W. Mahoney, *Asymptotic convergence rate and statistical inference for stochastic sequential quadratic programming* 2205 13687, 2022 , arXiv 预印本 arXiv: . ( )。[32] Y. Nandwani, A. Pathak, and P. Singla, *A primal dual formulation for deep learning with constraints* 32 2019 ,神经信息处理系统进展, ( )。[33] J. Nocedal and S. Wright, *Numerical optimization*,施普林格运筹学与金融工程丛书，施普林格纽约出版社，2006。[34] F. Oztoprak, R. Byrd, and J. Nocedal, *Constrained optimization in the presence of noise* 33 2023 2118 2136 , SIAM《优化杂志》, ( ), 第 – 页。[35] C. Paquette and K. Scheinberg, *A stochastic line search method with expected complexity analysis* 30 2020 349 376 , SIAM《优化杂志》, ( ), 第 – 页。[36] S. N. Ravi, T. Dinh, V. S. Lokhande, and V. Singh, *Explicitly imposing constraints in deep networks via conditional gradients gives improved generalization and faster convergence* 33, 2019, , 收录于AAAI人工智能会议论文集，第

第 4772–4779 页。[37] T. Rees, H. S. Dollar, and A. J. Wathen, *Optimal solvers for PDE-constrained optimization* 32 2010 271 298 ，《SIAM科学计算杂志》，（），第 – 页。[38] L. Roberts and C. W. Royer, *Direct search based on probabilistic descent in reduced spaces* 33 2023 3057 3082 ，SIAM《优化杂志》，（），第 – 页。[39] S. K. Roy, Z. Mhammedi, and M. Harandi, *Geometry aware constrained optimization techniques for deep learning* 2018, 4460 4469 ，载于《计算机视觉与模式识别会议》论文集，第 – 页。[40] K. Scheinberg and M. Xie, *Stochastic adaptive regularization method with cubics: Ahigh probability complexity bound*，载于 2023 年冬季仿真会议（WSC），IEEE，2023 年，第 3520–3531 页。[41] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on stochastic programming: modeling and theory* 2021 ，SIAM，。[42] S. Sun and J. Nocedal, *A trust region method for noisy unconstrained optimization*，《数学规划》，（2023），第 1–28 页。[43] A. Wächter and L. T. Biegler, *Line search filter methods for nonlinear programming: Motivation and global convergence* 16 2005 1 31 ，SIAM《优化杂志》，（），第 – 页。

附录 A。假设3.16 和引理 3.17。

在本附录中，我们给出假设 3.16 和引理3.17中出现的所有常数的定义。

Table A.1
Table of Constants from Assumption 3.16 and Lemma 3.17.

常数	定义
η_0	$1+\frac{\kappa_H}{\zeta}$
η_1	$\sqrt{(1-\theta)^2\bar{\tau}_{-1}\eta_0^2\left(\frac{(1+\epsilon_\tau)\bar{\tau}_{-1}\eta_0}{\sqrt{\kappa_l\tau_{\min}}}+\epsilon_\tau\right)+4\bar{\tau}_{-1}\left(\frac{1+\epsilon_\tau\omega_1}{\kappa_l}+\frac{(1-\theta)^2(1+\epsilon_\tau}{\zeta}\right)}$
η_2	$\frac{(1-\theta)\bar{\tau}_{-1}(\bar{\tau}_{-1}(3\eta_0-2)+(1-\sigma)\tau_{\min}\eta_0)}{(1-\sigma)\tau_{\min}\sqrt{\kappa_l\tau_{\min}}}\left(\frac{\bar{\tau}_{-1}\eta_0}{1-\sigma-(\cdot}\right.$
η_3	$\sqrt{\frac{(1-\theta)^2\bar{\tau}_{-1}^2(\bar{\tau}_{-1}(3\eta_0-2)+(1-\sigma)\tau_{\min}\eta_0)^2}{(1-\sigma)^2\tau_{\min}^3\kappa_l}+\frac{4\bar{\tau}_{-1}}{\kappa_l}+\frac{4(1-\theta)^2\bar{\tau}_{-1}}{\zeta}\left.\frac{}{\tau_{\min}}+1\right)}$
η_4	$\frac{\text{最大值}\{\kappa_H,\kappa_y\}+1}{\kappa_l}$
ω_1	$\frac{\eta_0\kappa_{\rm FO}\sqrt{\bar{\tau}_{-1}}\alpha_{\max}}{\sqrt{\kappa_l}}+\frac{\bar{\tau}_{-1}\kappa^2{}_{\rm FO}\alpha^2{}_{\text{最大值}}}{\zeta}$
ω_2	$\text{最大值}\left\{\epsilon_\tau\left(\frac{\max\{\kappa_H,\kappa_y\}}{\frac{\kappa_l}{\bar{\tau}_{-1}}}+\frac{\sqrt{\bar{\tau}_{-1}}\eta_0\kappa_{\rm FO}\alpha_{\max}}{\sqrt{\kappa_l}}+\omega_2\right),\right.$
ω_3	$\left.\frac{\kappa_l}{(1-\sigma)\tau_{\min}}\left(\frac{(3\eta_0-2)\kappa_{\rm FO}\sqrt{\bar{\tau}_{-1}\alpha_{\max}}}{\sqrt{\kappa_l}}+\frac{\eta_0\bar{\tau}_{-1}\kappa^2{}_{\rm FO}\alpha^2}{\zeta}\right)\right\}^{\text{最大值}}$
ω_4	$(1+\epsilon_\tau)\bar{\tau}_{-1}\left(\frac{\eta}{\zeta}+\frac{\eta_0}{\sqrt{\kappa_l\tau_{\min}}}\right)+\frac{\epsilon_\tau\bar{\tau}_{-1}\eta_0^2\eta}{4}$
ω_5	$\frac{\bar{\tau}_{-1}^2\left(\frac{\eta_0\eta}{\zeta}+\frac{3\eta_0-2}{\sqrt{\kappa_l\tau_{\min}}}\right)}{\frac{4\bar{\tau}_{-1}}{(1-\sigma)\tau_{\min}}}+\bar{\tau}_{-1}\left(\frac{\eta}{\zeta}+\frac{\eta_0}{\sqrt{\kappa_l\tau_{\min}}}\right)$
ω_6	$\sqrt{\frac{1}{p-2}()\theta}\text{最大值}\left\{\frac{1+\epsilon_\tau\omega_1}{1-\eta\omega_4},\frac{1}{1-\eta\omega_5},1+\omega_2+\omega_3\right\}$
ω_7	$\sqrt{\frac{\bar{\tau}_{-1}}{\kappa_l}}\kappa_{\rm FO}+\frac{\bar{\tau}_{\min}L+\Gamma}{2\bar{\tau}_{\min}\kappa_l}$
ω_8	$2\bar{\tau}_{\min}\kappa_l\left(1-\theta-\eta\sqrt{\frac{\bar{\tau}_{-1}}{\kappa_l}}\text{最大值}\left\{\sqrt{\frac{1+\epsilon_\tau\omega_1}{1-\eta\omega_4}},\sqrt{\frac{1}{1-\eta\omega_5}}\right\}\right)$
ω_9	$\sqrt{\text{最大值}\left\{\frac{\omega_7}{1-\theta},\frac{\bar{\tau}_{\min}L+\Gamma}{\omega_8}\right\}}$
ω_{10}	$\omega_6\omega_9$

附录 B. 引理 3.17的证明

在本附录中，我们给出引理 3.17的证明。根据引理陈述，我们注意到：(1) 由于常数及 p 的形式， p是一个有效的概率，即 $p \in (\frac{1}{2}, 1]$,(2) $\tilde{\alpha} > 0$由假设 η中的限制保证3.16，以及 (3) h： $\mathbb{R}_{>0} \to \mathbb{R}_{>0}$是一个正函数，用于衡量迭代真实且成功时所取得的势函数进展。接下来，我们将陈述 (i)–(v) 的证明分别展开如下。

Proof of (i). 该结果可直接由 $h(\tilde{\alpha})$ 的定义以及 $\varepsilon_{\Delta l}$的下界推出；参见假设 3.16。

□

Proof of (ii). 此证明与 [23, 命题 3(ii)]的证明基本相同。令

$$J_k := \mathbb{1}\left\{\|G_k - \nabla f(X_k)\| \le \max\left\{\epsilon_g, \kappa_{\mathrm{FO}} A_k \sqrt{\Delta l(X_k, \mathcal{T}_k, G_k, D_k)}\right\}\right\}.$$

显然，根据定义3.4,

$$\begin{aligned}\mathbb{P}\left[I_k = 0 \mid \mathcal{F}_{k-1} \cap \mathcal{E}\right] &= \mathbb{P}\left[J_k = 0 \ \text{ or } \ E_k + E_k^+ > 2\epsilon_f \mid \mathcal{F}_{k-1} \cap \mathcal{E}\right] \\ &\le \mathbb{P}\left[J_k = 0 \mid \mathcal{F}_{k-1} \cap \mathcal{E}\right] + \mathbb{P}\left[E_k + E_k^+ > 2\epsilon_f \mid \mathcal{F}_{k-1} \cap \mathcal{E}\right].\end{aligned}$$

根据假设 3.2 和一阶概率预言机（预言机 1），不等式右侧第一项的上界为 δ。若 ϵ_f是噪声的确定性界，则第二项为零。否则，由于在事件 $\mathcal{E}$条件下（见假设 3.2）， E_k和 E_k^+分别满足参数为 ϵ_f和 (ν, b) 的 (2.11) 中的单侧次指数界，可以证明在事件 $\mathcal{E}$条件下， $E_k + E_k^+$ 满足参数为 $2\epsilon_f$和 $(2\nu, 2b)$ 的 (2.11)。因此，根据单侧伯恩斯坦不等式，第二项的上界为 $e^{-\min\{u^2/2\nu^2, u/2b\}}$，其中 $u = \inf_{x \in \mathcal{X}}\{\epsilon_f - \mathbb{E}[E(x)]\}$。因此，对所有 k，当 p如陈述中所定义时， $\mathbb{P}\left[I_k = 1 \mid \mathcal{F}_{k-1} \cap \mathcal{E}\right] \ge p$。 $p \in \left(\frac{1}{2} + \frac{4\bar{\tau}^{-1}\epsilon_f}{h(\tilde{\alpha})}, 1\right]$的取值范围可由陈述中 $h(\cdot)$ 和 $\tilde{\alpha}$ 的定义，以及假设 3.16 中关于 $\varepsilon_{\Delta l}$的不等式推出。

□

Proof of (iii). 假设迭代 k 成立且成功。根据定义 3.4，我们分别考虑两种情况，即 $\|\bar{g}_k - \nabla f_k\| \le \kappa_{\mathrm{FO}} \alpha_k \sqrt{\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)}$ 和 $\|\bar{g}_k - \nabla f_k\| \le \epsilon_g$。我们进一步将分析细分为 $\nabla f_k^T d_k \le 0$ 和 $\nabla f_k^T d_k > 0$ 的情形。情形 A 当 $\|\bar{g}_k - \nabla f(x_k)\| \le \kappa_{\mathrm{FO}} \alpha_k \sqrt{\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)}$ 时，根据引理 3.12,

$$\|\bar{d}_k - d_k\| \le \zeta^{-1}\|\bar{g}_k - \nabla f(x_k)\| \le \zeta^{-1}\kappa_{\mathrm{FO}}\alpha_k\sqrt{\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)}.$$

情况 A.1 若 $\nabla f_k^T d_k \le 0$，根据 $\bar{\tau}_k \ge \tau_k$、三角不等式、（2.5）和引理 3.14，可得

$$\begin{aligned}&\Delta l(x_k, \tau_k, \nabla f_k, d_k) - \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \\ &= \bar{\tau}_k \bar{g}_k^T \bar{d}_k - \tau_k \nabla f_k^T d_k \le \bar{\tau}_k(\bar{g}_k^T \bar{d}_k - \nabla f_k^T d_k) \le \bar{\tau}_k |\bar{g}_k^T \bar{d}_k - \nabla f_k^T d_k| \\ &\le \bar{\tau}_k \left(\tfrac{(1+\kappa_H \zeta^{-1})\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l \bar{\tau}_k}} + \tfrac{\kappa_{\mathrm{FO}}^2 \alpha_k^2}{\zeta}\right) \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k).\end{aligned} \tag{B.1}$$

情形 A.2 若 $\nabla f_k^T d_k > 0$，由三角不等式可得（2.5）及引理 3.14，（B.2）

$$\begin{aligned}&\Delta l(x_k, \tau_k, \nabla f_k, d_k) - \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \\ &= \bar{\tau}_k \bar{g}_k^T \bar{d}_k - \tau_k \nabla f_k^T d_k \le |\bar{\tau}_k \bar{g}_k^T \bar{d}_k - \tau_k \nabla f_k^T d_k| \le |(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| + \bar{\tau}_k |\bar{g}_k^T \bar{d}_k - \nabla f_k^T d_k| \\ &\le |(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| + \bar{\tau}_k \left(\tfrac{(1+\kappa_H \zeta^{-1})\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l \bar{\tau}_k}} + \tfrac{\kappa_{\mathrm{FO}}^2 \alpha_k^2}{\zeta}\right) \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k).\end{aligned}$$

现在我们对该项 $|(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k|$进行界定；根据优值参数的更新公式 ((2.7)–(2.8) 和 (3.3)–(3.4))，我们考虑三种情形。

情况 A.2.1 若 $\tau_k = \bar{\tau}_k$，则 $|(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| = 0$。

情形 A.2.2 若 $\tau_k = (1-\epsilon_\tau)\bar{\tau}_k$，由三角不等式和引理 3.13 与 3.14，

$$\begin{aligned}|(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| &= \epsilon_\tau \bar{\tau}_k |\nabla f_k^T d_k| \le \epsilon_\tau \bar{\tau}_k (|\bar{g}_k^T \bar{d}_k| + |\nabla f_k^T d_k - \bar{g}_k^T \bar{d}_k|) \\ &\le \epsilon_\tau \left(\frac{\max\{\kappa_H, \kappa_y\}}{\kappa_l} + \frac{\sqrt{\bar{\tau}_k}\left(1+\frac{\kappa_H}{\zeta}\right)\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l}} \right) \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \\ &\quad + \epsilon_\tau \bar{\tau}_k \left(\frac{(1+\frac{\kappa_H}{\zeta})\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l \bar{\tau}_k}} + \frac{\kappa_{\mathrm{FO}}^2 \alpha_k^2}{\zeta} \right) \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k).\end{aligned}$$

情况 A.2.3 若 $\bar{\tau}_k > \tau_k = \left(\nabla f_k^T d_k + \frac{1-\sigma}{\|c_k\|_1}\max\left\{d_k^T H_k d_k, 0\right\}\right)$，由 (2.7)–(2.8)，

$$\nabla f_k^T d_k + \max\left\{d_k^T H_k d_k, 0\right\} > \frac{(1-\sigma)\|c_k\|_1}{\bar{\tau}_k} \ge \bar{g}_k^T \bar{d}_k + \max\left\{\bar{d}_k^T H_k \bar{d}_k, 0\right\}. \tag{B.3}$$

在事件 $\mathcal{E}$(见假设 3.2) 条件下，由引理 3.3，我们有 $\tau_k \ge \tau_{\min}$ 对所有 $k \in \mathbb{N}$。此外，由 (2.5) 和引理 3.11 可得 $0 \le \Delta l(x_k, \tau_k, \nabla f_k, d_k)$，这意味着 $\tau_k \nabla f_k^T d_k \le \|c_k\|_1$。利用 $\tau_k \in \mathbb{R}_{>0}$ 和 $\nabla f_k^T d_k > 0$，

$$\frac{|\nabla f_k^T d_k|}{\|c_k\|_1} = \frac{\nabla f_k^T d_k}{\|c_k\|_1} \le \frac{1}{\tau_k}. \tag{B.4}$$

由引理 3.14、(B.3) 和 (B.4)，可得

$$\begin{aligned}&|(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| \\ =\ & \left(\bar{\tau}_k - \frac{(1-\sigma)\|c_k\|_1}{\nabla f_k^T d_k + \max\left\{d_k^T H_k d_k, 0\right\}}\right)|\nabla f_k^T d_k| \\ \le\ & \frac{(\nabla f_k^T d_k + \max\left\{d_k^T H_k d_k, 0\right\}) - (\bar{g}_k^T \bar{d}_k + \max\{\bar{d}_k^T H_k \bar{d}_k, 0\})}{\nabla f_k^T d_k + \max\left\{d_k^T H_k d_k, 0\right\}} \bar{\tau}_k |\nabla f_k^T d_k| \\ \le\ & \frac{|\nabla f_k^T d_k - \bar{g}_k^T \bar{d}_k| + |\max\left\{d_k^T H_k d_k, 0\right\} - \max\{\bar{d}_k^T H_k \bar{d}_k, 0\}|}{(1-\sigma)\|c_k\|_1} \bar{\tau}_k^2 |\nabla f_k^T d_k| \\ \le\ & \frac{\bar{\tau}_k^2}{(1-\sigma)\tau_k}\left(|\nabla f_k^T d_k - \bar{g}_k^T \bar{d}_k| + |\max\left\{d_k^T H_k d_k, 0\right\} - \max\{\bar{d}_k^T H_k \bar{d}_k, 0\}|\right) \\ \le\ & \frac{\bar{\tau}_k^2}{(1-\sigma)\tau_k}\left(|\nabla f_k^T d_k - \bar{g}_k^T \bar{d}_k| + |d_k^T H_k d_k - \bar{d}_k^T H_k \bar{d}_k|\right) \\ \le\ & \frac{\bar{\tau}_k^2}{(1-\sigma)\tau_{\min}}\left(\frac{(1+3\kappa_H\zeta^{-1})\kappa_{\mathrm{FO}}\alpha_k}{\sqrt{\kappa_l \bar{\tau}_k}} + \frac{(1+\kappa_H\zeta^{-1})}{\zeta}\kappa_{\mathrm{FO}}^2\alpha_k^2\right)\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k).\end{aligned}$$

结合 **情形 A.2.1–A.2.3**，并根据 (B.1)、(B.2) 以及 $\{\omega_2, \omega_3\} \subset \mathbb{R}_{>0}$ (假设 3.16 和表 A.1) 的定义，可得

$$\Delta l(x_k, \tau_k, \nabla f_k, d_k) - \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \le (\omega_2 + \omega_3)\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k).$$

根据 $\{\omega_2, \omega_3\} \subset \mathbb{R}_{>0}$，$\frac{\Delta l(x_k, \tau_k, 1+\omega_2+\omega_3}{}$ $\le \Delta l_{, \nabla f_k, d_k)}(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)$。由于迭代 k 是成功的，且根据定义 3.5，可得

$$\begin{aligned}\bar{\phi}(x_k^+, \bar{\tau}_k; \xi_k^+) - \bar{\phi}(x_k, \bar{\tau}_k; \xi_k) &\le -\alpha_k \theta \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + 2\bar{\tau}_k \epsilon_f \\ &\le -\alpha_k \theta \tfrac{\Delta l(x_k, \tau_k, \nabla f_k, d_k)}{1+\omega_2+\omega_3} + 2\bar{\tau}_{-1}\epsilon_f.\end{aligned}$$

因此可得

$$\begin{aligned}&Z_{k+1} - Z_k \\ =\ & \phi(x_{k+1}, \bar{\tau}_{k+1}) - \phi(x_k, \bar{\tau}_k) - \bar{\tau}_{k+1} f_{\inf} + \bar{\tau}_k f_{\inf} \\ \le\ & \phi(x_{k+1}, \bar{\tau}_{k+1}) - \bar{\phi}(x_k, \bar{\tau}_k; \xi_k) - \bar{\tau}_{k+1} f_{\inf} + \bar{\tau}_k f_{\inf} + \bar{\tau}_k e_k \\ =\ & \phi(x_{k+1}, \bar{\tau}_{k+1}) - \bar{\phi}(x_{k+1}, \bar{\tau}_k; \xi_k^+) + \bar{\phi}(x_{k+1}, \bar{\tau}_k; \xi_k^+) - \bar{\phi}(x_k, \bar{\tau}_k; \xi_k) - \bar{\tau}_{k+1} f_{\inf} + \bar{\tau}_k f_{\inf} + \bar{\tau}_k e_k \\ \le\ & -\alpha_k \theta \tfrac{\Delta l(x_k, \tau_k, \nabla f_k, d_k)}{1+\omega_2+\omega_3} + 2\bar{\tau}_{-1}\epsilon_f + (\bar{\tau}_{k+1} - \bar{\tau}_k)(f(x_{k+1}) - f_{\inf}) + \bar{\tau}_k(e_k + e_k^+) \\ \le\ & -\alpha_k \theta \tfrac{\Delta l(x_k, \tau_k, \nabla f_k, d_k)}{1+\omega_2+\omega_3} + 2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_k(e_k + e_k^+).\end{aligned} \tag{B.5}$$

情况 B 当 $\|\bar{g}_k - \nabla f(x_k)\| \leq \epsilon_g$时，根据 $k < T_{\varepsilon_{\Delta l}}$和定义 3.7，可得 $\sqrt{\Delta l(x_k, \tau_k, \nabla f_k, d_k)} > \varepsilon_{\Delta l} > \frac{\epsilon_g}{\eta}$。根据引理 3.12，

$$\|\bar{d}_k - d_k\| \leq \zeta^{-1}\|\bar{g}_k - \nabla f_k\| \leq \zeta^{-1}\epsilon_g < \zeta^{-1}\eta\sqrt{\Delta l(x_k, \tau_k, \nabla f_k, d_k)}.$$

情况 B.1 与 **情况 A.1**类似，若 $\nabla f_k^T d_k \leq 0$，根据 $\bar{\tau}_k \geq \tau_k$、三角不等式、（2.5）和引理 3.14，可得

$$\begin{aligned} &\Delta l(x_k, \tau_k, \nabla f_k, d_k) - \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \\ &\leq \bar{\tau}_k|\bar{g}_k^T\bar{d}_k - \nabla f_k^T d_k| \leq \bar{\tau}_k\left(\zeta^{-1}\epsilon_g^2 + \tfrac{(1+\kappa_H\zeta^{-1})\epsilon_g}{\sqrt{\kappa_l\tau_k}}\sqrt{\Delta l(x_k, \tau_k, \nabla f_k, d_k)}\right) \\ &\leq \bar{\tau}_k\left(\tfrac{\eta}{\zeta} + \tfrac{1+\kappa_H\zeta^{-1}}{\sqrt{\kappa_l\tau_k}}\right)\eta\Delta l(x_k, \tau_k, \nabla f_k, d_k). \end{aligned} \tag{B.6}$$

情况 B.2 若 $\nabla f_k^T d_k > 0$，利用前述论证、（2.5）和引理 3.14，

$$\begin{aligned} &\Delta l(x_k, \tau_k, \nabla f_k, d_k) - \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \\ &\leq |(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| + \bar{\tau}_k|\bar{g}_k^T\bar{d}_k - \nabla f_k^T d_k| \\ &\leq |(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| + \bar{\tau}_k\left(\tfrac{\epsilon_g^2}{\zeta} + \tfrac{(1+\kappa_H\zeta^{-1})\epsilon_g}{\sqrt{\kappa_l\tau_k}}\sqrt{\Delta l(x_k, \tau_k, \nabla f_k, d_k)}\right) \\ &\leq |(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| + \bar{\tau}_k\left(\tfrac{\eta}{\zeta} + \tfrac{1+\kappa_H\zeta^{-1}}{\sqrt{\kappa_l\tau_k}}\right)\eta\Delta l(x_k, \tau_k, \nabla f_k, d_k). \end{aligned} \tag{B.7}$$

我们接下来对项 $|(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k|$进行估计。
情况 B.2.1 若 $\tau_k = \bar{\tau}_k$，则 $|(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| = 0$。
情况 B.2.2 若 $\tau_k = (1-\epsilon_\tau)\bar{\tau}_k$，则根据引理 3.13 和 3.14 以及假设 3.16，

$$\begin{aligned} &|(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| \\ &= \epsilon_\tau\bar{\tau}_k|\nabla f_k^T d_k| \leq \epsilon_\tau\bar{\tau}_k\left(|\bar{g}_k^T\bar{d}_k| + |\nabla f_k^T d_k - \bar{g}_k^T\bar{d}_k|\right) \\ &\leq \epsilon_\tau\omega_2\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \tfrac{\epsilon_\tau\bar{\tau}_k(1+\kappa_H\zeta^{-1})^2}{4}\epsilon_g^2 \\ &\quad + \epsilon_\tau\bar{\tau}_k\left(\tfrac{\epsilon_g^2}{\zeta} + \tfrac{(1+\kappa_H\zeta^{-1})\epsilon_g}{\sqrt{\kappa_l\tau_k}}\sqrt{\Delta l(x_k, \tau_k, \nabla f_k, d_k)}\right) \\ &\leq \epsilon_\tau\omega_2\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \epsilon_\tau\bar{\tau}_k\eta\left(\tfrac{(1+\kappa_H\zeta^{-1})^2\eta}{4} + \tfrac{\eta}{\zeta} + \tfrac{1+\kappa_H\zeta^{-1}}{\sqrt{\kappa_l\tau_k}}\right)\Delta l(x_k, \tau_k, \nabla f_k, d_k). \end{aligned}$$

情况 B.2.3 若 $\bar{\tau}_k > \tau_k = (\frac{1-\sigma-\|c_k\|_1)\text{乘}\{d_k^T H_k d_k, 0\}}{\nabla f_k^T d_k +}$，沿用 **情况 A.2.3**中的相同推理，根据引理 3.14、（B.3）和（B.4），

$$\begin{aligned} &|(\bar{\tau}_k - \tau_k)\nabla f_k^T d_k| \\ &\leq \tfrac{\bar{\tau}_k^2}{(1-\sigma)\tau_k}\left(|\nabla f_k^T d_k - \bar{g}_k^T\bar{d}_k| + |d_k^T H_k d_k - \bar{d}_k^T H_k\bar{d}_k|\right) \\ &\leq \tfrac{\bar{\tau}_k^2}{(1-\sigma)\tau_{\min}}\left(\tfrac{(1+\kappa_H\zeta^{-1})\eta}{\zeta} + \tfrac{1+3\kappa_H\zeta^{-1}}{\sqrt{\kappa_l\tau_k}}\right)\eta\Delta l(x_k, \tau_k, \nabla f_k, d_k). \end{aligned}$$

结合 **情形 B**。 **2 1** **2 3**, 6 7 $\{\omega_1, \omega_4, \omega_5\} \subset$ **.–B.** 。并根据 $\mathbb{R}_{>0}$ （B.）、（B.）以及（假设 3 16 。和表 A.）的定义，可得 1

$$\begin{aligned} &\Delta l(x_k, \tau_k, \nabla f_k, d_k) - \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \\ &\leq \max\left\{\epsilon_\tau\omega_1\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \eta\omega_4\Delta l(x_k, \tau_k, \nabla f_k, d_k), \eta\omega_5\Delta l(x_k, \tau_k, \nabla f_k, d_k)\right\}, \end{aligned} \tag{B.8}$$

其中 $\{\omega_1, \omega_4, \omega_5\} \subset \mathbb{R}_{>0}$ 在假设 3.16中定义。因此可得:

$$\Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \geq \min\left\{\tfrac{1-\eta\omega_4}{1+\epsilon_\tau\omega_1}, 1-\eta\omega_5\right\}\Delta l(x_k, \tau_k, \nabla f_k, d_k). \tag{B.9}$$

根据假设 η 进行选取 3.16，利用迭代 k 是 成功的 以及定义3.5，

$$\begin{aligned}
&\bar{\phi}(x_k^+, \bar{\tau}_k; \xi_k^+) - \bar{\phi}(x_k, \bar{\tau}_k; \xi_k) \\
\leq\ & -\alpha_k \theta \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + 2\bar{\tau}_k \epsilon_f \\
\leq\ & -\alpha_k \theta \min\left\{\tfrac{1-\eta\omega_4}{1+\epsilon_\tau \omega_1}, 1-\eta\omega_5\right\} \Delta l(x_k, \tau_k, \nabla f_k, d_k) + 2\bar{\tau}_{-1}\epsilon_f.
\end{aligned}$$

因此，按照与（B.5）中类似的逻辑，可得

$$\begin{aligned}
&Z_{k+1} - Z_k \\
\leq\ & \phi(x_{k+1}, \bar{\tau}_{k+1}) - \bar{\phi}(x_{k+1}, \bar{\tau}_k; \xi_k^+) + \bar{\phi}(x_{k+1}, \bar{\tau}_k; \xi_k^+) - \bar{\phi}(x_k, \bar{\tau}_k; \xi_k) \\
& - \bar{\tau}_{k+1} f_{\inf} + \bar{\tau}_k f_{\inf} + \bar{\tau}_k e_k \\
\leq\ & -\alpha_k \theta \min\left\{\tfrac{1-\eta\omega_4}{1+\epsilon_\tau \omega_1}, 1-\eta\omega_5\right\} \Delta l(x_k, \tau_k, \nabla f_k, d_k) + 2\bar{\tau}_{-1}\epsilon_f \\
& + (\bar{\tau}_{k+1} - \bar{\tau}_k)(f(x_{k+1}) - f_{\inf}) + \bar{\tau}_k(e_k + e_k^+) \\
\leq\ & -\alpha_k \theta \min\left\{\tfrac{1-\eta\omega_4}{1+\epsilon_\tau \omega_1}, 1-\eta\omega_5\right\} \Delta l(x_k, \tau_k, \nabla f_k, d_k) + 2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_k(e_k + e_k^+).
\end{aligned}$$

结合**情形 A** 和**情况 B**的结果，并结合迭代为真的假设，可得

$$\begin{aligned}
Z_{k+1} - Z_k \leq\ & -\alpha_k \theta \min\left\{\tfrac{1-\eta\omega_4}{1+\epsilon_\tau \omega_1}, 1-\eta\omega_5, \tfrac{1}{1+\omega_2+\omega_3}\right\} \Delta l(x_k, \tau_k, \nabla f_k, d_k) \\
& + 2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_{-1}(e_k + e_k^+) \\
\leq\ & -h(\alpha_k) + 4\bar{\tau}_{-1}\epsilon_f,
\end{aligned}$$

wh最后一个不等式来自条件 $\Delta l(x_k,\ \tau_k,\ \nabla f_k,\ d_k) > \varepsilon_{\Delta l}^2$ 以及 $e_k + e_k^+ \leq 2\epsilon_f$.*Proof of (iv).* □

我们首先证明，对于任意 $k \in \mathbb{N}$，如果 $\alpha_k \leq \tilde{\alpha}$ 且迭代 k 为真，则

$$\phi(x_k + \alpha \bar{d}_k, \bar{\tau}_k) \leq \phi(x_k, \bar{\tau}_k) - \alpha_k \theta \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k).$$

由于迭代 k 为真，根据定义 3.4，我们再次分别考虑两种情况： $\|\bar{g}_k - \nabla f_k\| \leq \kappa_{\mathrm{FO}} \alpha_k \sqrt{\Delta l}$ $(x_k,\ \bar{\tau}_k,\ \bar{g}_k,\ \bar{d}_k)$ 以及 $\|\bar{g}_k - \nabla f_k\| \leq \epsilon_{g^\circ}$

情形 A 当 $\|\bar{g}_k - \nabla f_k\| \leq \kappa_{\mathrm{FO}} \alpha_k \sqrt{\Delta l}\ (x_k,\ \bar{\tau}_k,\ \bar{g}_k,\ \bar{d}_k)$ 时，根据 $\alpha_k \leq \tilde{\alpha}$，由 $\tilde{\alpha}$的定义、假设 3.2 （事件 $\mathcal{E}$发生）以及引理3.10 和3.15，

$$\begin{aligned}
&\phi(x_k + \alpha_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) \\
\leq\ & -\alpha_k \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \alpha_k \bar{\tau}_k (\nabla f_k - \bar{g}_k)^T \bar{d}_k + \tfrac{\bar{\tau}_k L + \Gamma}{2} \alpha_k^2 \|\bar{d}_k\|^2 \\
\leq\ & -\alpha_k \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \alpha_k \bar{\tau}_k \|\nabla f_k - \bar{g}_k\| \|\bar{d}_k\| + \tfrac{\bar{\tau}_k L + \Gamma}{2} \alpha_k^2 \|\bar{d}_k\|^2 \\
\leq\ & -\left(1 - \left(\sqrt{\tfrac{\bar{\tau}_{-1}}{\kappa_l}} \kappa_{\mathrm{FO}} + \tfrac{L}{2\kappa_l} + \tfrac{\Gamma}{2\bar{\tau}_{\min}\kappa_l}\right) \tilde{\alpha}\right) \alpha_k \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) \\
\leq\ & -\alpha_k \theta \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k).
\end{aligned}$$

情况 B 当 $\|\bar{g}_k - \nabla f_k\| \leq \epsilon_g$ 且迭代 k 成立时，式（B.9）成立。此外，由于条件 $k < T_{\varepsilon_{\Delta l}}$ 和定义 3.7，可得

$$\|\bar{g}_k - \nabla f_k\| \leq \epsilon_g < \eta \varepsilon_{\Delta l} < \eta \sqrt{\Delta l(x_k, \tau_k, \nabla f_k, d_k)}.$$

因此，根据 $\alpha_k \leq \tilde{\alpha}$、假设 3.2 （事件 $\mathcal{E}$ 的发生）、（B.9）以及引理 3.10 和3.15，

$$\begin{aligned}
&\phi(x_k + \alpha_k \bar{d}_k, \bar{\tau}_k) - \phi(x_k, \bar{\tau}_k) \\
\leq\ & -\alpha_k \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \alpha_k \bar{\tau}_k \|\nabla f_k - \bar{g}_k\| \|\bar{d}_k\| + \tfrac{\bar{\tau}_k L + \Gamma}{2} \alpha_k^2 \|\bar{d}_k\|^2
\end{aligned}$$

$$\leq -\alpha_k \left(\left(1 - \eta \sqrt{\frac{\bar{\tau}_{-1}}{\kappa_l}} \max \left\{ \sqrt{\frac{1+\epsilon_\tau \omega_1}{1-\eta\omega_4}}, \frac{1}{\sqrt{1-\eta\omega_5}} \right\} \right) - \frac{\bar{\tau}_{\min} L + \Gamma}{2\bar{\tau}_{\min}\kappa_l} \tilde{\alpha} \right) \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k)$$
$$\leq -\alpha_k \theta \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k).$$

结合**情况 A** 和**B**，并且该迭代为真，我们得出了 (iv) 的证明

$$\begin{aligned} \bar{\phi}(x_k + \alpha_k \bar{d}_k, \bar{\tau}_k; \xi_k^+) - \bar{\phi}(x_k, \bar{\tau}_k; \xi_k) &\leq -\alpha_k \theta \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + \bar{\tau}_k e_k + \bar{\tau}_k e_k^+ \\ &\leq -\alpha_k \theta \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + 2\bar{\tau}_k \epsilon_f. \end{aligned}$$

□

Proof of (v). 如果迭代 k 未 成功，则根据定义 $Z_{k+1} = Z_k$，因此该不等式显然成立。否则，从（B.5）中的第二个方程开始

$$\begin{aligned} Z_{k+1} - Z_k &\leq \phi(x_{k+1}, \bar{\tau}_{k+1}) - \bar{\phi}(x_{k+1}, \bar{\tau}_k; \xi_k^+) + \bar{\phi}(x_{k+1}, \bar{\tau}_k; \xi_k^+) - \bar{\phi}(x_k, \bar{\tau}_k; \xi_k) \\ &- \bar{\tau}_{k+1} f_{\inf} + \bar{\tau}_k f_{\inf} + \bar{\tau}_k e_k \leq -\alpha_k \theta \Delta l(x_k, \bar{\tau}_k, \bar{g}_k, \bar{d}_k) + (\bar{\tau}_{k+1} - \bar{\tau}_k)(f(x_{k+1}) - f_{\inf}) \\ &+ 2\bar{\tau}_k \epsilon_f + \bar{\tau}_k (e_k + e_k^+) \leq 2\bar{\tau}_{-1}\epsilon_f + \bar{\tau}_{-1}(e_k + e_k^+)。 \end{aligned}$$

□