arXiv:2603.03625v1 [math.OC] 4 Mar 2026

随机非凸优化中具有高概率复杂度保证的负曲率方法

Albert S. Berahas[†*] Raghu Bollapragada[‡] Wanping Dong[†]

2026年3月5日

摘要

本文针对随机环境下的连续非线性无约束优化问题，提出了负曲率方法。在这类问题中，函数、梯度和海森信息只能通过概率预言机获得，即通过能够返回具有一定精度和可靠性的近似值的预言机获取。我们针对这些预言机提出了条件，并设计了一个系统地结合梯度步和负曲率步的两步框架。该框架采用提前停止机制来保证取得充分进展，并使用基于 Armijo 型准则的自适应机制来选择这两种步的步长。我们为求得二阶驻点建立了高概率迭代复杂度保证，并推导出明确的尾界，以量化收敛邻域及其对预言机噪声的依赖。重要的是，这些界与确定性收敛速率相匹配，仅相差依赖于噪声的项；此外，该框架将确定性结果作为特例涵盖在内。最后，数值实验表明，即使存在噪声，利用负曲率方向也能带来实际益处。

1 引言

本文考虑如下形式的无约束非线性优化问题：$\min_{x\in\mathbb{R}^n} f(x)$，其中 $f\colon \mathbb{R}^n \to \mathbb{R}$ 是二阶连续可微且可能非凸的。我们研究无法获得精确目标函数及其相应导数（∇f 和 $\nabla^2 f$）的情形；取而代之的是，通过概率预言机获得零阶、一阶和二阶近似，即返回近似值的预言机

[†]密歇根大学工业与运筹工程系。（albertberahas@gmail.com，wanpingd@umich.edu）

[‡]德克萨斯大学奥斯汀分校运筹学与工业工程项目。（raghu.bollapragada@utexas.edu）

*通讯作者。

具有特定精度和可靠性。此类问题出现在仿真优化等应用中， [27], 机器学习 [9], 和决策 [28]。

尽管上述概率预言机设定中的大多数文献都聚焦于能够保证收敛到（近似）一阶驻点的算法，但我们的目标是开发、分析并实现能够以适当速率收敛到（近似）二阶驻点的方法。二阶驻点定义为满足以下条件的点 $x \in \mathbb{R}^n$ 及常数 $\bar{\epsilon}_g$， $\bar{\epsilon}_\lambda$， $\bar{\epsilon}_g > 0$：

$$\|\nabla f(x)\|_2 < \bar{\epsilon}_g, \quad \text{and} \quad \lambda_{\min}(\nabla f^2(x)) > -\max\{\bar{\epsilon}_\lambda, \bar{\epsilon}_H\}. \tag{1.1}$$

为此，我们开发了利用负曲率方向、且仅使用概率零阶、一阶和二阶预言机输出的自适应随机算法。所提出的方法采用两步结构，在梯度（或一般下降）步与负曲率步之间交替进行 [4,15], 并采用 Armijo 型步长搜索机制，在含噪评估下自适应地选择步长 [6,7,20],且具有高概率迭代复杂度保证。具体而言，我们证明，经过足够多次迭代后，算法以极高概率达到一个（$\bar{\epsilon}_g$， $\bar{\epsilon}_\lambda$， $\bar{\epsilon}_H$）驻点（1.1），且由此得到的收敛邻域大小随预言机噪声而变化，并且除噪声相关项外，与确定性情形相匹配。

1.1 相关文献

本节回顾与我们的问题设定及算法框架各组成部分最密切相关的文献。我们重点关注负曲率方法、各种不精确或随机预言机模型下的优化，以及随机线搜索/步长搜索策略方面的既有研究。这些领域构成了我们所提出方法的基础。

大量研究已证实，在确定性非凸优化中，负曲率（NC）对于逃离鞍点并实现二阶收敛具有重要作用 [25]。经典方法包括结合下降方向与负曲率方向的曲线方法和两步方案 [16,18,23],， 以及近年来在每次迭代中显式利用负曲率方向的变体 [4,15]。信赖域方法和三次正则化方法也通过其子问题求解器（例如截断共轭梯度法或 Lanczos）隐式利用负曲率信息，并具有清晰明确的全局收敛与复杂度保证 [13,24]。这些确定性研究成果为我们的算法结构提供了依据，但它们依赖于能够获取精确的函数与导数信息，或其确定性界；而这些假设在随机环境下不再成立。

基于负曲率的方法已在含噪环境中得到研究，例如，具有有界误差的确定性不精确预言机 [4,21,30], 以及在随机环境中以期望收敛的随机设置 [1,4,15]。然而，这些方法大多要求对精度作出严格假设，通常只能保证一阶收敛或期望收敛

而非高概率收敛。据我们所知，尚无先前研究在一般概率误差预言机设置下，为基于负曲率的方法提供高概率二阶复杂度保证。

在不精确预言机条件下研究优化方法，已逐步纳入确定性 [5, 12], 期望意义下 [4, 9, 26, 30], 以及概率性 [3, 7, 11, 20] 误差模型。尤其是，概率预言机允许返回的估计值以非平凡概率出现不准确或有偏的情况，因此越来越多地被用于模拟现实中的随机评估。然而，现有研究主要聚焦于零阶和一阶信息，而适用于在含噪环境中可靠识别负曲率方向的概率二阶预言机鲜少受到关注。我们的设定符合这一通用概率框架，但进一步考虑了函数评估的有界噪声和次指数噪声模型、以非零概率可能存在任意大误差的有偏梯度近似，以及用于检测负曲率的概率海森向量信息。

随机线搜索和步长搜索方法为我们的工作提供了另一项重要基础。传统的确定性线搜索程序严重依赖准确的函数与导数信息，因此无法直接应用于含噪环境 [25]。为解决这一问题，人们提出了若干自适应方案，包括结合样本量调整的随机回溯策略 [8, 17]。近年来，步长搜索方法（也称随机线搜索方法）得到了发展：它们依据概率充分下降条件自适应地增大或减小步长；若试探步未能通过检验，则在同一迭代点重新评估预言机；参见例如 [6, 14, 20, 26]。这些策略提升了方法在随机环境中的鲁棒性，但主要针对一阶方法，既未解决负曲率方向的处理问题，也未提供二阶收敛保证。

总而言之，尽管负曲率方法、不精确或概率预言机以及随机步长搜索技术都已分别得到研究，但如何结合这些要素，在一般噪声环境下获得二阶保证，仍存在空白。我们的工作通过开发一种面向一般概率环境、并具备高概率迭代复杂度保证的基于负曲率的自适应步长搜索框架，填补了这一空白。

1.2 贡献

本文针对随机非凸优化，在一般的概率预言机设定下提出了一种负曲率方法。在该设定中，零阶、一阶和二阶信息只能通过满足指定精度和可靠性的含噪近似获得。该框架适用于广泛的噪声模型：函数评估可能含有有界噪声或具有次指数尾部的噪声；梯度估计可能存在偏差，且以可控概率任意不准确；海森矩阵信息则通过专为在噪声下检测负曲率而设计的概率二阶预言机获得

噪声下的曲率。

我们的第一项贡献是针对这一问题构建了一个灵活的算法框架。该框架采用两步结构，在下降（与梯度相关）方向和负曲率方向之间交替进行，并放宽了传统的充分下降条件，以考虑噪声的影响。两类步长均采用 Armijo 型步长搜索过程，从而能够自适应地选择步长，并在需要时重新评估搜索方向。该框架还纳入了一种高效机制，仅通过两次函数评估且无需梯度评估即可选取负曲率方向，从而降低计算成本并增强大规模场景下的鲁棒性。总体而言，该方案将一大类随机优化方法涵盖为其特例，同时引入了处理噪声所必需的新组件。

我们的第二项贡献是对所提出方法的高概率二阶收敛性和复杂度进行分析。我们建立了显式尾界，表明达到一个 $(\bar{\epsilon}_g, \bar{\epsilon}_H, \bar{\epsilon}_\lambda)$-二阶驻点所需迭代次数超过 $\mathcal{O}(\max\{\bar{\epsilon}_g^{-2}, \bar{\epsilon}_H^{-3}, \bar{\epsilon}_\lambda^{-3}\})$ 的概率会随迭代次数增加而呈指数衰减。这些保证与确定性收敛速率相匹配，仅相差依赖于噪声的项；当噪声消失时，它们可作为特例还原确定性结果；此外，它们还在更广泛的概率预言机假设下进行分析，从而扩展了先前的随机分析。

最后，我们根据理论框架实现了一种实用方法：采用高效的共轭梯度子求解器检测负曲率，以取代理想化条件和二阶计算，同时保留分析所启发的步长选择与接受机制。在经典测试问题上的初步实验表明，该方法对噪声具有鲁棒性，在噪声环境下的表现往往优于基准随机方法。

1.3 论文结构

本文结构安排如下。本节最后将介绍全文使用的符号。我们在第2节介绍假设、概率预言机以及所提出的算法。在第3节，我们建立达到二阶驻点的高概率复杂度界。第4节报告数值实验，展示该算法的实际表现。最后，我们在第5节给出总结性评论。

1.4 符号

设 $\mathbb{N}$ 和 $\mathbb{R}$ 分别表示自然数集和实数集。对于 $n, m \in \mathbb{N}$，设 $\mathbb{R}^n$ 和 $\mathbb{R}^{n\times m}$ 分别表示 n维实向量集和 $n \times m$ 实矩阵集。对于矩阵 A，设 $\lambda_{\min}(A)$（$\lambda_{\max}(A)$）表示其最小（最大）特征值。我们考虑迭代点序列 $\{x_k, \hat{x}_k\}_{k\in\mathbb{N}}$，其中 $x_k, \hat{x}_k \in \mathbb{R}^n$。在迭代$k$时，算法可能生成下降方向 $d_k \in \mathbb{R}^n$ 以及负曲率

方向 $p_k \in \mathbb{R}^n$，其对应步长分别为 $\alpha_k > 0$ 和 $\beta_k > 0$。对于每个$k \in \mathbb{N}$，设 F_k表示 $f(x_k)$的估计值， $g_k \in \mathbb{R}^n$表示 $\nabla f(x_k)$的估计值， $H_k \in \mathbb{R}^{n\times n}$表示 $\nabla^2 f(\hat{} x_k)$的估计值。

2 假设、预言机与算法

本节将介绍假设、定义预言机，并描述算法框架。全文中，我们对目标函数作出如下假设。该假设在利用负曲率的方法以及更广泛的二阶方法文献中都很常见；例如，参见 [11,15,21]。

假设 2.1。 函数 f： $\mathbb{R}^n \to \mathbb{R}$ 是连续可微的， f 的梯度是 L_g-利普希茨连续的，并且 f 的海森矩阵是 L_H-利普希茨连续的，对所有$_x \in \mathbb{R}^n$均成立。此外，函数 f 由标量 $f \in \mathbb{R}$下有界。

我们的算法依赖于概率预言机提供的目标函数及其导数的近似值，这些概率预言机满足规定的精度和可靠性条件。下面给出零阶、一阶和二阶预言机的精确定义。

预言机 1。 给定一点 $x \in \mathbb{R}^n$，预言机计算 $F(x, \xi^{(0)})$，这是函数值 $f(x)$的一个（随机）估计值，其中 $\xi^{(0)}$是定义在概率空间 $(\Xi, \mathcal{F}, \mathbb{P})$上的随机变量，其分布可能依赖于 x。对于任意 $x \in \mathbb{R}^n$， $e(x, \xi^{(0)}) = |F(x, \xi^{(0)}) - f(x)|$至少满足以下两个条件之一：

1.（确定性有界噪声）存在常数 $\epsilon_f \geq 0$ ，使得对于 $\xi^{(0)}$的所有实现值， $e(x, \xi^{(0)}) \leq \epsilon_f$ 。

2.（独立次指数噪声）存在常数 $\epsilon_f \geq 0$和 $a > 0$，使得对于所有 $s > 0$， $\mathbb{P}_{\xi^{(0)}}\left[e(x, \xi^{(0)}) \geq s\right] \leq \exp(-a(s - \epsilon_f))$。

注 2.2. 当在点 $x \in \mathbb{R}^n$ 查询预言机 *1.1* 时，会返回一个估计值 $F(x, \xi^{(0)})$，其误差具有确定性上界 ϵ_f。在预言机 *1.2*下，函数噪声具有次指数尾部，且偏差容限为 ϵ_f。参数 $a > 0$ 和 $\epsilon_f \geq 0$是与 x无关的全局常数。关于零阶概率预言机的类似表述可见 *[11, 20]*。

预言机 2. 给定概率 $p_g \in \left(\frac{1}{2}, 1\right]$ 和点 $\underline{x} \in \mathbb{R}^n$，预言机计算 $g(x, \xi^{(1)})$，这是梯度 $\nabla f(x)$ 的一个（随机）估计值，并满足

$$\mathbb{P}_{\xi^{(1)}}\left[\|g(x, \xi^{(1)}) - \nabla f(x)\|_2 \leq \epsilon_g + \kappa_g \|\nabla f(x)\|_2\right] \geq p_g$$

其中， $\epsilon_g \geq 0$、 $\kappa_g \geq 0$和 $\xi^{(1)}$ 是在概率空间$(\Xi$、 $\mathcal{F}$、 $\mathbb{P})$上定义的随机变量，其分布可能依赖于 x。

备注 *2.3.* 预言机 *2* 返回一个梯度估计值 $g(x,\xi^{(1)})$，其满足精度条件的概率至少为$p_g > \frac{1}{2}$。右侧由绝对项和相对项组成，其中ϵ_g和κ_g是预言机固有的非负常数。当 $\|\nabla f(x)\|_2$较大或 $\epsilon_g = 0$时，相对项占主导地位，该条件要求满足范数相对精度 *[8, 10, 12]*。当 $\|\nabla f(x)\|_2$ 较小或 $\kappa_g = 0$时，绝对项占主导地位。常数ϵ_g 界定了可达到的最佳梯度精度（见 *(3.3)*）。若 $p_g = 1$，该条件为确定性的。这种结合绝对项与相对项的概率精度条件十分常见；例如参见 *[4, 7, 8, 20, 21]*。

在定义二阶概率预言机之前，我们先定义矩阵 $H \in \mathbb{R}^{n\times n}$ 的负曲率方向 $q \in \mathbb{R}^n$。

定义 1。考虑一个矩阵 $H \in \mathbb{R}^{n\times n}$，满足 $\lambda_{\min}(H) < 0$。如果

$$q^{\mathsf{T}} H q \le \gamma \lambda_{\min}(H)\|q\|_2^2 < 0, \quad \gamma \in (0,1] \tag{2.1a}$$
$$\|q\|_2 = \delta|\lambda_{\min}(H)|, \qquad \delta \in (0,\infty). \tag{2.1b}$$

二阶概率预言机定义如下。

预言机 3。给定一个概率 $p_H \in (0.5,\ 1]$ 和一个点 $x \in \mathbb{R}^n$，预言机计算出 $H(x,\xi^{(2)})$，即海森矩阵 $\nabla^2 f(x)$ 的一个（随机）估计值，使得

$$\mathbb{P}_{\xi^{(2)}}\Big[\|(\nabla^2 f(x) - H(x,\xi^{(2)}))q\|_2 \le \epsilon_H + \kappa_H|\lambda_{\min}(H(x,\xi^{(2)}))|\|q\|_2,$$
$$|\lambda_{\min}(\nabla^2 f(x)) - \lambda_{\min}(H(x,\xi^{(2)}))| \le \epsilon_\lambda + \kappa_\lambda|\lambda_{\min}(\nabla^2 f(x))|\Big] \ge p_H.$$

其中 $\epsilon_H \ge 0$, $\kappa_H \ge 0$, $\epsilon_\lambda \ge 0$, $\kappa_\lambda \ge 0$, $\xi^{(2)}$ 是定义在概率空间 $(\Xi,\mathcal{F},\mathbb{P})$ 上的随机变量，其分布可能取决于输入 x，而 q 是 $H(x,\xi^{(2)})$ 的负曲率方向。

注 2.4. 预言机*3* 包含两个非精确条件，仅在检测到负曲率时才施加。与预言机 *2*一样，每个条件都结合了相对项和绝对项，这些项由预言机固有的非负常数决定。第一个条件确保方向精度，要求海森矩阵估计值仅在负曲率方向上准确 *[4, 32]*。第二个条件控制对最小特征值的逼近。总之，这些条件严格弱于一致的谱范数界，即 $\|\nabla^2 f(x) - H(x,\ \xi^{(2)})\|_2$。若$p_H = 1$，则它们退化为非精确确定性情形。

为简化记号，我们将算法在每次迭代中生成的随机估计实现值记为 $F_k = F(x_k,\xi_k^{(0)})$、$F_k^+ = F(x_k+\alpha_k d_k,\xi_k^{(0+)})$、$\hat{F}_k = F(\hat{x}_k,\hat{\xi}_k^{(0)})$、$\hat{F}_k^{\pm} = F(\hat{x}_k\pm\beta_k q_k,\hat{\xi}_k^{(0\pm)})$、$g_k = g(x_k,\xi_k^{(1)})$、$H_k = H(\hat{x}_k,\hat{\xi}_k^{(2)})$ 和 $\lambda_k = \lambda_{\min}(H_k)$。算法框架见算法 2.1。

算法 2.1 带概率预言机的步长搜索负曲率方法

要求： $x_0 \in \mathbb{R}^n$, $\alpha_0, \beta_0 > 0$, $\tau \in (0,1)$, $c_d, c_p \in (0,1)$, $e_f, c_g, c_H \geq 0$, $\bar{\epsilon}_g, \bar{\epsilon}_H, \bar{\epsilon}_\lambda \geq 0$

1: **对于所有** $k \in \{0,1,\dots\} =: \mathbb{N}$ **执行**
2: 生成 g_k 通过预言机 2
3: **如果** $\|g_k\|_2 \leq c_g \bar{\epsilon}_g$ **则** 设置 $\hat{}x_k \leftarrow x_k$, $\alpha_{k+1} \leftarrow \alpha_k$
4: **否则** 设置 $d_k \leftarrow -g_k$ 并生成 F_k 并 F_k^+ 通过预言机 1
5: **若** $F_k^+ \leq F_k + c_d \alpha_k d_k^\mathsf{T} g_k + e_f$ **则** 设定 $\hat{}x_k \leftarrow x_k + \alpha_k d_k$, $\alpha_{k+1} \leftarrow \tau^{-1}\alpha_k$
6: **否则** 设定 $\hat{}x_k \leftarrow x_k$, $\alpha_{k+1} \leftarrow \tau\alpha_k$
7: 生成 H_k 通过预言机 3并计算 λ_k
8: **若** $\lambda_k \geq -c_H \max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\}$ **则** 设定 $x_{k+1} \leftarrow \hat{}x_k$, $\beta_{k+1} \leftarrow \beta_k$
9: **否则** 计算满足条件 1 的负曲率方向 q_k 1
并生成 $\hat{F}_k$, $\hat{F}_k^+$ 且 $\hat{F}_k^-$ 通过预言机 1
10: **若** $\min\{\hat{F}_k^+, \hat{F}_k^-\} \leq \hat{F}_k + c_p \beta_k^2 q_k^\mathsf{T} H_k q_k + e_f$ **则**
11: 令 $p_k \leftarrow \arg\min_{\omega \in \{-1,1\}} \{F(\hat{}x_k + \omega \beta_k q_k, \hat{}\xi_k^{(0\pm)})\} q_k$
12: 设置 $x_{k+1} \leftarrow \hat{}x_k + \beta_k p_k$、 $\beta_{k+1} \leftarrow \tau^{-1}\beta_k$
13: **否则** 设置 $x_{k+1} \leftarrow \hat{}x_k$、 $\beta_{k+1} \leftarrow \tau\beta_k$

注 2.5。 我们对算法 *2.1*作几点说明。

- 算法*2.1* 在下降步与负曲率步之间交替进行，只有检测到负曲率时才计算后者。正如文献 $d_k \leftarrow -g_k$所述，尽管也允许采用更一般的形式 $d_k \leftarrow -W_k g_k$，其中 $w_1 I \preceq W_k \preceq w_2 I$ 对某个 $0 < w_1 \leq w_2$而言，不会影响理论保证。

- 下降步（第 *2–6* 行）的成本主要来自计算梯度估计 g_k 和两次函数评估，即 F_k 和 F_k^+。相比之下，负曲率步（第 *7–13* 行）的成本来自计算海森矩阵估计 H_k、其最小特征值 λ_k 以及可能的负曲率方向 q_k，此外还需要进行三次函数评估，即 $\hat{}F_k$ 和 $\hat{}F_k^\pm$。

- 提前终止规则（第*3* 行与第*8*行）仅接受那些预测下降量足够大的步；该预测下降量基于估计梯度范数或负曲率的幅度，从而过滤掉不太可能产生有意义进展的步。这些规则并不终止算法；一旦触发，便跳过相应的下降步或负曲率步。实施这些条件需要已知一阶和二阶预言机的绝对误差参数，或掌握其可靠估计值。提前终止是必要的，因为我们的框架涉及两种步类型，以及包含两个条件的二阶停止准则，因此需要在停止时间分析中分别控制成功事件与大步事件；这与先前工作形成对照，后者要么考虑由单一条件定义的停时 *[20]*，要么考虑每次迭代仅一步 *[11]*。

• 步长经由步长搜索过程确定，并依据放宽的 *Armijo* 条件接受（第*5–6* 行与第*10–13*行）。该随机线搜索在当前迭代点反复重新评估所需的预言机信息，直至条件满足，从而提升含噪函数评估下的鲁棒性。放宽后的条件包含噪声容忍参数e_f ，以确保在存在函数噪声时仍能得到可接受的步长。对于负曲率步，*Armijo* 条件被进一步修改，以计入曲率效应，如 *[18, 22, 23]*。

• 一个关键改动涉及负曲率方向的选择。由于q_k 和 $-q_k$ 都是有效的负曲率方向，目标是选择能够产生下降的符号。验证下降通常需要梯度信息，从而增加负曲率步的代价。先前的方法要么随机选择符号 *[21]*，要么使用近似梯度 *[4]*, ，从而在每次迭代代价与有效曲率利用之间进行权衡。相比之下，我们的方法通过比较两次试探函数求值并选择$\min\{\hat{} F_k^+, \hat{} F_k^-\}$（第*10*行）来确定符号，最多只需一次额外的函数求值。这提供了一种无需梯度信息的高效且可靠的符号选择机制，并能自然地与步长搜索过程中的 *Armijo* 条件相结合。

为分析算法 2.1 的收敛性，我们引入如下记号。设 $\{X_k\}$、 $\{X_k^+\}$、 $\{\hat{} X_k\}$、 $\{\hat{} X_k^+\}$和 $\{\hat{} X_k^-\}$ 为 $\mathbb{R}^n$ 中的随机向量，其实现值分别为 x_k、 $x_k + \alpha_k d_k$、$\hat{} x_k$、$\hat{} x_k + \beta_k q_k$ 和 $\hat{} x_k - \beta_k q_k$。我们将零阶预言机误差的幅度定义为随机变量 $\{E_k\}$、 $\{E_k^+\}$、$\{\hat{} E_k\}$、 $\{\hat{} E_k^+\}$和 $\{\hat{} E_k^-\}$，分别由 $E_k := |f(X_k) - F(X_k, \xi_k^{(0)})|$、 $E_k^+ := |f(X_k + \alpha_k d_k) - F(X_k + \alpha_k d_k, \xi_k^{(0+)})|$、 $\hat{} E_k := |f(\hat{} X_k) - F(\hat{} X_k, \hat{} \xi_k^{(0)})|$、 $\hat{} E_k^+ := |f(\hat{} X_k + \beta_k q_k) - F(\hat{} X_k + \beta_k q_k, \hat{} \xi_k^{(0+)})|$、 $\hat{} E_k^- := |f(\hat{} X_k - \beta_k q_k) - F(\hat{} X_k - \beta_k q_k, \hat{} \xi_k^{(0-)})|$给出，对应的实现值分别为 e_k、 e_k^+、$\hat{} e_k$、$\hat{} e_k^+$ 和 $\hat{} e_k^-$。为记号方便，我们用 $\{\alpha_k\}$ 和 $\{\beta_k\}$ 表示随机步长序列，其实现值仍使用相同符号书写。

在算法的迭代点 X_k 处，算法 2.1在以下操作中引入随机性 s:

1. **在 X_k 处的随机梯度：**给定 X_k ，梯度估计 $g_k = g(X_k, \xi_k^{(1)})$ 通过预言机 2 计算得到。然后，给定 X_k, α_k 和 g_k，中间点 X_k^+ 以确定性方式设定。

2. **在 X_k 和 X_k^+ 处的随机函数值：**给定 X_k 和 X_k^+ ，通过预言机 1 计算函数估计，并引入相应误差 E_k 和 E_k^+ 。这些估计值用于更新 α_{k+1} 并形成 X_k。

3. **在 $\hat{} X_k$ 处的随机海森矩阵：给定** X_k 海森矩阵近似 $H(\hat{} X_k, \hat{} \xi_k^{(2)})$ ，通过预言机 3计算，然后以确定性方式获取最小特征值 λ_k 及其对应的负曲率方向 q_k （如果存在）。如果 q_k 存在，则使用 $\hat{} X_k^+$ 和 $\hat{} X_k^-$ 以及 $\hat{} X_k$， q_k 和 β_k构造试探点 $\hat{} X_{k+1} \leftarrow X_k$ 和 $\beta_{k+1} \leftarrow \beta_k$。否则，算法将 $\hat{} X_{k+1} \leftarrow X_k$ 和 $\beta_{k+1} \leftarrow \beta_k$设为 。

4. **在 $\hat{}\, X_k$、$\hat{}\, X_k^+$ 和 $\hat{}\, X_k^-$ 处的随机函数值：** 当 q_k 存在时，给定 $\hat{}\, X_k$、$\hat{}\, X_k^+$ 和 $\hat{}\, X_k^-$，通过预言机 1计算函数估计值。这些估计值用于在两个方向之间进行选择，从而确定 X_{k+1} 和 β_{k+1}。

因此，在给定 X_k、α_k、β_k的条件下，迭代 k中的所有随机性均由 $\xi_k^{(0)}$、$\xi_k^{(0+)}$、$\xi_k^{(1)}$、$\hat{}\,\xi_k^{(0)}$、$\hat{}\,\xi_k^{(0\pm)}$ 和$\hat{}\,\xi_k^{(2)}$ 所刻画。令 $\mathcal{Z}_k := \left\{\xi_k^{(0)}、\xi_k^{(0+)}、\xi_k^{(1)}、\hat{}\,\xi_k^{(0)}、\hat{}\,\xi_k^{(0\pm)}、\hat{}\,\xi_k^{(2)}\right\}$ 和 $\mathcal{F}_k$ 表示截至迭代 k的算法历史所对应的σ-代数，即 $\mathcal{F}_k := \sigma\left(\cup_{i=1}^k \mathcal{Z}_i\right)$。为便于分析，我们引入增广滤过 $\mathcal{F}'_k := \sigma\left(\mathcal{F}_k \cup \sigma\left(\xi_{k+1}^{(1)}\right)\right)$和$\hat{}\,\mathcal{F}'_k := \sigma\left(\mathcal{F}_k \cup \sigma\left(\xi_{k+1}^{(0)}, \xi_{k+1}^{(0+)}, \xi_{k+1}^{(1)}, \xi_{k+1}^{(2)}\right)\right)$。

我们的分析依赖于将迭代 $k = 0,\mathbf{1},\ldots,t-1$（$t \geq 1$）划分为不同类型。这些类型可通过以下随机指示变量，按照估计准确度来定义：

$$
\begin{aligned}
I_k^g &:= \mathbb{1}\left\{\|g(X_k,\xi_k^{(1)}) - \nabla f(X_k)\|_2 \leq \epsilon_g + \kappa_g\|\nabla f(X_k)\|_2\right\},\\
I_k^H &:= \mathbb{1}\left\{\|(\nabla^2 f(X_k) - H(X_k,\xi_k^{(2)}))q_k\|_2 \leq \epsilon_H^2 + \kappa_H|\lambda_k|\|q_k\|_2,\right.\\
&\qquad\left.|\lambda_{\min}(H(X_k,\xi_k^{(2)})) - \lambda_{\min}(\nabla^2 f(X_k))| \leq \epsilon_\lambda + \kappa_\lambda|\lambda_{\min}(\nabla^2 f(X_k))|\right\},\\
\text{and}\quad I_k^f &:= \mathbb{1}\left\{E_k + E_k^+ \leq e_f\right\},\quad \hat{I}_k^f := \mathbb{1}\left\{\hat{E}_k + \max\{\hat{E}_k^+, \hat{E}_k^-\} \leq e_f\right\}.
\end{aligned}
$$

高概率下，生成了准确函数评估的迭代次数相对于 t 有一个下界。这对预言机1.1. 显然成立。

引理 2.6。 令 $p_f = 1 - 3\exp(-a(\frac{e_f}{2} - \epsilon_f))$，其中 a 是预言机 *1.2* 中的正常数。则指示变量 I_k^f 和 $\hat{}\, I_k^f$ 满足次鞅条件，即 $\mathbb{P}\left[I_k^f = 1|\mathcal{F}'_{k-1}\right] \geq p_f$ 和 $\mathbb{P}\left[\hat{}\, I_k^f = 1|\hat{}\,\mathcal{F}'_{k-1}\right] \geq p_f$。此外，若 $e_f > 2\epsilon_f + \frac{2}{a}\log 6$（等价地，$p_f > \frac{1}{2}$），则对于任意正整数 $t \geq 1$ 和任意 $\bar{p}_f \in [0, p_f]$，有 $\mathbb{P}\left[\sum_{k=0}^{t-1} I_k^f \geq \bar{p}_f t\right] \geq 1 - \exp\left(-\frac{(\hat{}\, p_f - p_f)^2}{2p_f^2}t\right)$。

证明。根据预言机1。2，在下降步中，对所有 $k \in \mathbb{N}$

$$
\mathbb{P}_{\xi_k^{(0)}}\left[E_k > s|\mathcal{F}'_{k-1}\right] \leq \exp(a(\epsilon_f - s)),\quad \mathbb{P}_{\xi_k^{(0+)}}\left[E_k^+ > s|\mathcal{F}'_{k-1}\right] \leq \exp(a(\epsilon_f - s)). \tag{2.2}
$$

因此，根据 I_k^f、预言机 1.2 和 p_f，可得

$$
\begin{aligned}
\mathbb{P}\left[I_k^f = 0|\mathcal{F}'_{k-1}\right] = \mathbb{P}\left[E_k + E_k^+ > e_f|\mathcal{F}'_{k-1}\right] &\leq \mathbb{P}\left[E_k > e_f/2 \text{ or } E_k^+ > e_f/2|\mathcal{F}'_{k-1}\right]\\
&\leq \mathbb{P}\left[E_k > e_f/2|\mathcal{F}'_{k-1}\right] + \mathbb{P}\left[E_k^+ > e_f/2|\mathcal{F}'_{k-1}\right]\\
&\leq 2\exp(a(\epsilon_f - e_f/2)) \leq 1 - p_f.
\end{aligned}
$$

因此，随机过程 $\left\{\sum_{k=0}^{t-1} I_k^f - \bar{p}_f t\right\}_{t=0}$，1,... 是一个次鞅。由于

$$\left|\left(\sum_{k=0}^{(t+1)-1} I_k^f - \bar{p}_f(t+1)\right) - \left(\sum_{k=0}^{t-1} I_k^f - \bar{p}_f t\right)\right| = |I_k^f - \bar{p}_f| \le \max\{|0-\bar{p}_f|, |1-\bar{p}_f|\} = \bar{p}_f,$$

对于任意 $t \ge 1$，根据 Azuma-Hoeffding不等式，对于任意正常数 c，$\mathbb{P}\left[\sum_{k=0}^{t-1} I_k^f - p_f t \le -c\right] \le \exp\left(-\frac{c^2}{2p_f^2 t}\right)$。令 $c = (p_f - \bar{p}_f)T$，并将两边同时减去 1，即得第二个结果。

类似地，在负曲率步中，对所有 $k \in \mathbb{N}$

$$\mathbb{P}_{\hat{\xi}_k^{(0)}}\left[\hat{E}_k > s|\hat{\mathcal{F}}'_{k-1}\right] \le \exp(a(\epsilon_f - s)), \quad \mathbb{P}_{\hat{\xi}_k^{(0\pm)}}\left[\hat{E}_k^{\pm} > s|\hat{\mathcal{F}}'_{k-1}\right] \le \exp(a(\epsilon_f - s))$$

因此，对所有 $k \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}_{\hat{\xi}_k^{(0\pm)}}\left[\max\{\hat{E}_k^+, \hat{E}_k^-\} > s|\hat{\mathcal{F}}'_{k-1}\right] &\le \mathbb{P}_{\hat{\xi}_k^{(0\pm)}}\left[\hat{E}_k^+ > s \text{ or } \hat{E}_k^- > s|\hat{\mathcal{F}}'_{k-1}\right] \\
&\le \mathbb{P}_{\hat{\xi}_k^{(0\pm)}}\left[\hat{E}_k^+ > s|\hat{\mathcal{F}}'_{k-1}\right] + \mathbb{P}_{\hat{\xi}_k^{(0\pm)}}\left[\hat{E}_k^- > s|\hat{\mathcal{F}}'_{k-1}\right] \\
&= \mathbb{P}_{\hat{\xi}_k^{(0+)}}\left[\hat{E}_k^+ > s|\hat{\mathcal{F}}'_{k-1}\right] + \mathbb{P}_{\hat{\xi}_k^{(0-)}}\left[\hat{E}_k^- > s|\hat{\mathcal{F}}'_{k-1}\right] \\
&\le 2\exp(a(\epsilon_f - s)) = \exp\left(a\left(\epsilon_f + \tfrac{1}{a}\log 2 - s\right)\right), \qquad (2.3)
\end{aligned}$$

并且因此，根据 I_k^f，预言机1。2，以及p_f的定义，可得

$$\mathbb{P}\left[\hat{I}_k^f = 0|\mathcal{F}'_{k-1}\right] = \mathbb{P}\left[\hat{E}_k + \max\{\hat{E}_k^+, \hat{E}_k^-\} > e_f|\hat{\mathcal{F}}'_{k-1}\right] \le 3\exp(a(\epsilon_f - \tfrac{e_f}{2})) = 1 - p_f.$$

□

次指数随机变量的分析借鉴了 [31, 第 2]章；然而，这些结果不足以满足我们的需求，因为此处考虑的分布由两个量 a 和ϵ_f参数化，而非一个。因此，我们采用来自 [11], 的一个更强结果，具体表述如下命题。

命题 2.7。 设 X 为一个随机变量，使得对于某些 $a > 0$ 和 $b \ge 0$，$\mathbb{P}[|X| \ge s] \le \exp(a(b-s))$，对所有 $s > 0$均成立。那么可得，对于所有 $\lambda \in [0, a)$，$\mathbb{E}\{\exp(\lambda|X|)\} \le \frac{1}{1-\lambda/a}\exp(\lambda b)$。

利用命题 2.7，我们接下来建立一个引理，为含噪函数评估导致的总误差提供一个高概率上界。

引理 2.8. 对于任意 $s \ge 0$和 $t \ge 1$, $\mathbb{P}\left[\sum_{k=0}^{t-1}(E_k + E_k^+) \ge t(\frac{5}{a} + 2\epsilon_f) + s\right] \le \exp\left(-\frac{a}{4}s\right)$ 和 $\mathbb{P}\left[\sum_{k=0}^{t-1}\left(\hat{E}_k + \max\{\hat{E}_k^+, \hat{E}_k^-\}\right) \ge t(\frac{5}{a} + 2\epsilon_f) + s\right] \le \exp\left(-\frac{a}{4}s\right)$.

证明。 对于下降步，根据命题2.7，由于（2.2）对两种函数估计误差均成立 E_k 和 E_k^+ ，因此可得

$$\mathbb{E}\left[\exp(2\lambda E_k)\right] \leq \frac{1}{1-2\lambda/a}\exp(2\lambda\epsilon_f) \quad \text{and} \quad \mathbb{E}\left[\exp(2\lambda E_k^+)\right] \leq \frac{1}{1-2\lambda/a}\exp(2\lambda\epsilon_f),$$

对所有 $\lambda \in \left[0, \frac{a}{2}\right)$ 均成立。于是，由柯西-施瓦茨不等式可知，对所有 $\lambda \in \left[0, \frac{a}{2}\right)$,

$$\begin{aligned}\mathbb{E}\left[\exp(\lambda E_k + \lambda E_k^+)|\mathcal{F}'_{k-1}\right] &\leq \sqrt{\mathbb{E}\left[\exp(2\lambda E_k)|\mathcal{F}'_{k-1}\right]\mathbb{E}\left[\exp(2\lambda E_k^+)|\mathcal{F}'_{k-1}\right]} \\ &\leq \frac{1}{1-2\lambda/a}\exp(2\lambda\epsilon_f).\end{aligned} \tag{2.4}$$

由马尔可夫不等式可知，对于任意 $\lambda \in \left[0, \frac{a}{2}\right)$, $s \geq 0$, 和正整数 t，有

$$\begin{aligned}\mathbb{P}\left[\sum_{k=0}^{t-1}(E_k + E_k^+) \geq s\right] &= \mathbb{P}\left[\exp\left(\lambda\sum_{k=0}^{t-1}(E_k + E_k^+)\right) \geq \exp(\lambda s)\right] \\ &\leq e^{-\lambda s}\mathbb{E}\left[\exp\left(\lambda\sum_{k=0}^{t-1}(E_k + E_k^+)\right)\right] \\ &\leq e^{-\lambda s}\left(\frac{1}{1-2\lambda/a}\exp(2\lambda\epsilon_f)\right)^t,\end{aligned} \tag{2.5}$$

其中，最后一个不等式可通过以下数学归纳法证明。首先，由于（ $t = 1$ 2.4），该不等式对 t 成立。现在，假设它对任意正整数 t都成立，那么对于 $t+1$,

$$\begin{aligned}&e^{-\lambda s}\mathbb{E}\left[\exp\left(\lambda\sum_{k=0}^{(t+1)-1}(E_k + E_k^+)\right)\right] \\ &= e^{-\lambda s}\mathbb{E}\left[\exp\left(\lambda\sum_{k=0}^{t-1}(E_k + E_k^+)\right)\mathbb{E}_{\xi_t^{(0)},\xi_t^{(0+)}}\left[\exp\left(\lambda(E_t + E_t^+)\right)|\mathcal{F}'_{t-1}\right]\right] \\ &\leq e^{-\lambda s}\mathbb{E}\left[\exp\left(\lambda\sum_{k=0}^{t-1}(E_k + E_k^+)\right)\frac{1}{1-2\lambda/a}\exp(2\lambda\epsilon_f)\right] \leq e^{-\lambda s}\left(\frac{1}{1-2\lambda/a}\exp(2\lambda\epsilon_f)\right)^t\end{aligned}$$

其中，第一个不等式由（2.4）得出，第二个不等式则由归纳假设得出。至此，数学归纳法证明完成。为便于说明，我们对所有 $1/(1-x)$ $\leq\exp(2x)$ 使用 $x \in [0, 1/2]$ 来简化结果

$$\mathbb{P}\left[\sum_{k=0}^{t-1}(E_k + E_k^+) \geq s\right] \leq e^{-\lambda s}\left[\exp(\tfrac{4\lambda}{a})\exp(2\lambda\epsilon_f)\right]^t \leq \exp(\lambda[t(\tfrac{5}{a} + 2\epsilon_f) - s]),$$

其中 $\lambda \in \left[0, \frac{a}{4}\right]$。 仅当 $s \geq t(\frac{5}{a} + 2\epsilon_f)$ 时，右侧才小于或等于 1。将 s替换为 $t(\frac{5}{a} + 2\epsilon_f)$ $+ s$并选择 $\lambda = \frac{a}{4}$, -即可得到第一个结果。 -

对于负曲率步，根据命题2.7 以及不等式 (2.2)-(2.3),

$$\mathbb{E}\left[\exp(2\lambda \hat{E}_k)\right] \le \tfrac{1}{1-2\lambda/a}\exp(2\lambda\epsilon_f),$$
$$\text{and}\quad \mathbb{E}\left[\exp(2\lambda \max\{\hat{E}_k^+, \hat{E}_k^-\})\right] \le \tfrac{1}{1-2\lambda/a}\exp(2\lambda(\epsilon_f + \tfrac{1}{a}\log 2)),$$

对所有 $\lambda \in [0, \frac{a}{2})$。按照（2.5）中的相同步骤,

$$\mathbb{P}\left[\sum_{k=0}^{t-1}\left(\hat{E}_k + \max\{\hat{E}_k^+, \hat{E}_k^-\}\right) \ge s\right] \le e^{-\lambda s}\left(\tfrac{1}{1-2\lambda/a}\exp\left(\lambda(2\epsilon_f + \tfrac{1}{a}\log 2)\right)\right)^t$$
$$\le \exp(\lambda[t(\tfrac{5}{a} + 2\epsilon_f) - s]).$$

仅当 $s \ge t(\frac{5}{a} + 2\epsilon_f)$ 时，右侧才小于或等于 1。将 s 替换为 $t(\frac{5}{a} + 2\epsilon_f) + s$ 并选择 $\lambda = \frac{a}{4}$，即可得到第二个结果。 □

我们引入模型精度假设（关于预言机成功概率的条件），以推导高概率迭代复杂度界。

假设 2.9。 我们对概率预言机中的概率（p_f、 p_g 和 p_H）作出以下假设:

(*i*) $\mathbb{P}\left[I_k^f I_k^g = 1|\mathcal{F}_{k-1}\right] \ge p_f p_g$ 且 $\mathbb{P}\left[\hat{I}_k^f I_k^H = 1|\mathcal{F}_{k-1},\ \hat{x}_k\right] \ge p_f p_H$, (*ii*) $\mathbb{P}\left[I_k^f I_k^g \hat{I}_k^f I_k^H = 1|\mathcal{F}_{k-1}\right] \ge p_f^2 p_g p_H$, 且 (*iii*) $p_f^2 p_g p_H + p_f p_g + p_f p_H - 2 > 0$。

注 *2.10. (i)-(ii)* 中的界表明，过程 $\sum_{k=0}^{t-1} I_k^f I_k^g - p_f p_g t$, $\sum_{k=0}^{t-1} \hat{I}_k^f I_k^H - p_f p_H t$ 和 $\sum_{k=0}^{t-1} I_k^f I_k^g \hat{I}_k^f I_k^H - p_f^2 p_g p_H t$ 是次鞅。当 $p_f = 1$ 时，我们几乎必然有 $I_k^f = \hat{I}_k^f = 1$，并且假设 *2.9* 简化为有界噪声情形中使用的相应条件。此外，如果算法仅执行下降步，则可以省略 $\hat{I}_k^f$ 和 I_k^H，假设 *2.9* 可简化为 $\mathbb{P}\left[I_k^f I_k^g = 1|\mathcal{F}_{k-1}\right] \ge p_f p_g =: p_{fg} > \frac{1}{2}$，这与 *[20]* 中的假设一致。 *[11]* 中也出现了形式如 *(ii)* 的相关联合概率条件。

以下引理可由假设 2.9(ii) 及 Azuma–Hoeffding 不等式 [2]推出。其证明与引理 2.6 的证明类似，为简洁起见，此处略去。

引理 2.11。 对于所有正整数 t，以及任意的 $\bar{p}_f \in [0, p_f)$、 $\bar{p}_g \in [0, p_g)$ 和 $\bar{p}_H \in [0, p_H)$, $\mathbb{P}\left[\sum_{k=0}^{t-1} I_k^f I_k^g > \bar{p}_f \bar{p}_g t\right] \ge 1 - \exp\left(-\frac{(p_f p_g - \bar{p}_f \bar{p}_g)^2}{2p_f^2 p_g^2} t\right)$、 $\mathbb{P}\left[\sum_{k=0}^{t-1} \hat{I}_k^f I_k^H > \bar{p}_f \bar{p}_H t\right] \ge 1 - \exp\left(-\frac{(p_f p_H - \bar{p}_f \bar{p}_H)^2}{2p_f^2 p_H^2} t\right)$, 以及 $\mathbb{P}\left[\sum_{k=0}^{t-1} I_k^f I_k^g \hat{I}_k^f I_k^H > \bar{p}_f^2 \bar{p}_g \bar{p}_H t\right] \ge 1 - \exp\left(-\frac{(p_f^2 p_g p_H - \bar{p}_f^2 \bar{p}_g \bar{p}_H)^2}{2p_f^4 p_g^2 p_H^2} t\right)$。

引理 2.6、2.8和引理 2.11的结果是第 3.2 节次指数噪声分析的核心构件。

3 收敛性分析

本节将对算法 2.1进行全面的收敛性与复杂度分析。为便于分析，我们引入以下指示随机变量:

$$\begin{aligned} \Omega_k^g &:= \mathbb{1}\left\{\|g_k\|_2 \geq c_g\bar{\epsilon}_g\right\},\ \Omega_k^H := \mathbb{1}\left\{\lambda_k < -c_H \max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\}\right\}, \\ \Theta_k^g &:= \mathbb{1}\{\text{descent step is successful, i.e., } \alpha_{k+1} = \tau^{-1}\alpha_k\}, \\ \text{and} \quad \Theta_k^H &:= \mathbb{1}\{\text{negative curvature step is successful, i.e., } \beta_{k+1} = \tau^{-1}\beta_k\}. \end{aligned}$$

为形式化定义迭代复杂度，我们定义停时

$$N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} := \min\left\{k : \|\nabla f(x_k)\|_2 \leq \bar{\epsilon}_g,\ \lambda_{\min}(\nabla^2 f(x_k)) \geq -\max\{\bar{\epsilon}_\lambda, \bar{\epsilon}_H\}\right\}, \tag{3.1}$$

即达到一个 $(\bar{\epsilon}_g, \bar{\epsilon}_\lambda, \bar{\epsilon}_H)$-驻点所需的迭代次数 .

3.1 收敛性分析：有界噪声情形

本小节分析预言机1.1（有界函数评估预言机）下的收敛性。具体而言，函数估计误差的实现值满足

$$e_k \leq \epsilon_f,\ e_k^+ \leq \epsilon_f,\ \hat{e}_k \leq \epsilon_f,\ \hat{e}_k^\pm \leq \epsilon_f. \tag{3.2}$$

线搜索 c_d、 c_p 以及提前终止 c_g 和 c_H 参数设定为满足

$$\begin{aligned} &0 < c_d < \tfrac{1}{2} + \tfrac{(1-\eta)(1-\kappa_g)}{2(1+\kappa_g)}, \quad 0 < c_g < 1 - \kappa_g \leq 1, \quad 0 < \eta < 1, \\ \text{and} \quad &0 < c_p < \tfrac{\gamma-\kappa_H}{2\gamma} = \tfrac{1}{2} - \tfrac{\kappa_H}{2\gamma}, \quad 0 \leq \kappa_H < \gamma \leq 1, \quad 0 < c_H < 1 - \kappa_\lambda \leq 1. \end{aligned}$$

这些界量化了预言机精度如何影响可接受的参数范围。对于梯度步，随着 $\kappa_g \to 1$， c_d 和 c_g 的上界会减小，因此在梯度估计精度较低时，需要采用更保守的接受阈值和终止阈值。对于负曲率步，增大 κ_λ 会使 c_H 的允许范围趋近于零，从而对基于曲率的终止提出更严格的要求；而 $\kappa_H \to 1$ 则迫使 $\gamma \to 1$，使接受负曲率方向的条件更加严格（极端情况下$\gamma = 1$，该方向必须与特征向量重合）。在无噪声情形下，即 $\epsilon_g = \kappa_g = 0$和 $\epsilon_H = \kappa_H = \epsilon_\lambda = \kappa_\lambda = 0$，这些要求简化为 $0 < c_d < 1 - \frac{\eta}{2}$、 $0 < c_g,\ c_H < 1$、
$0 < c_p < \frac{1}{2}$ $0 < \gamma \leq 1$ 以及
，与确定性情形下的界一致。

接下来，我们建立由算法2.1 所生成的随机过程的性质，此时函数、梯度与Hessian估计分别通过预言机1.1、2和3计算得到。

引理 3.1. 设 $e_f \geq 2\epsilon_f$，且

$$\bar{\epsilon}_g \geq \frac{2\epsilon_g}{\min\{\eta c_g(1-\kappa_g),1-\kappa_g-c_g\}},\ \bar{\epsilon}_H \geq \frac{\epsilon_H}{c_H}\sqrt{\frac{2}{\delta(\gamma-\kappa_H-2c_p\gamma)}},\quad and\quad \bar{\epsilon}_\lambda \geq \frac{\epsilon_\lambda}{1-\kappa_\lambda-c_H}. \tag{3.3}$$

那么，存在常数 $\bar{\alpha}$、$\beta > 0$

$$\bar{\alpha} := \frac{2\big(1/(1+\kappa_g')-c_d\big)}{L_g},\quad and\quad \bar{\beta} := \frac{3(\gamma-\kappa_H-2c_p\gamma)}{2\delta L_H}, \tag{3.4}$$

其中 $0 \leq \kappa_g' < 1$，以及非递减函数 $h_d(\cdot)$ $h_p(\cdot)$: $\mathbb{R} \to \mathbb{R}$，满足 $h_d(\alpha)$ $h_p(\beta)$ >0 对任意 α，$\beta > 0$，使得对于算法 2.1 的任意实现，下列性质对所有 $k < N_{\bar{\epsilon}_g,\ \bar{\epsilon}_H,\ \bar{\epsilon}_\lambda}$ 成立：

(i) 若迭代 k 的下降步处的梯度估计是准确的（即 $I_k^g = 1$）且$\|g_k\|_2 \geq c_g\bar{\epsilon}_g$ （即 $\Omega_k^g = 1$），则对于 $\alpha_k \leq \bar{\alpha}$，该迭代（下降步）是成功的（$\Theta_k^g = 1$），这意味着 $\alpha_{k+1} = \tau^{-1}\alpha_k$;

(ii) 若迭代 k 的负曲率步处的海森矩阵估计是准确的（即$I_k^H = 1$）且 $\lambda_k \leq -c_H\max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\}$（即 $\Omega_k^H = 1$），则对于 $\beta_k \leq \bar{\beta}$，该迭代（负曲率）是成功的（$\Theta_k^H = 1$），这意味着 $\beta_{k+1} = \tau^{-1}\beta_k$；

(iii) 当 $I_k^g\Omega_k^g\Theta_k^g = 1$时，则 $f(x_{k+1}) \leq f(x_k) - h_d(\alpha_k) + 4e_f$;

(iv) 当 $I_k^H\Omega_k^H\Theta_k^H = 1$时，则 $f(x_{k+1}) \leq f(x_k) - h_p(\beta_k) + 4e_f$;

(v) $I_k^g(1-\Omega_k^g)I_k^H(1-\Omega_k^H) = 0$。

总之，

$$f(x_{k+1}) \leq \begin{cases} f(x_k) - h_d(\alpha_k) + 4e_f, & if\ I_k^g\Omega_k^g\Theta_k^g = 1, \\ f(x_k) - h_p(\beta_k) + 4e_f, & if\ I_k^H\Omega_k^H\Theta_k^H = 1, \\ f(x_k) + 4e_f, & otherwise. \end{cases} \tag{3.5}$$

证明。我们从下降步开始，并考虑 I_k^g $\Omega_k^g = 1$ （梯度准确且 $\|g_k\|_2 > c_g\bar{\epsilon}_g$）的情形。根据预言机 2、(3.3) 以及算法 2.1 中的终止条件，我们有 $\|\nabla f(x_k) - g_k\|_2 \leq \epsilon_g + \kappa_g\|\nabla f(x_k)\|_2$且$\|g_k\|_2 \geq c_g\bar{\epsilon}_g \geq \frac{2c_g\epsilon_g}{\eta c_g}{}_{(1-\kappa_g)} = \eta^{2\epsilon_g}{}_{(1-\kappa_g)}$，因此 $\|\nabla f(x_k) - g_k\|_2 \leq \kappa_g'\|\nabla f(x_k)\|_2$，其中 $\kappa_g' := 1 - {}^{2}{}_{(1-\eta)(1-\kappa_g)1+\kappa_g+(1-\eta)(1-\kappa_g)} < 1$，由此可得

$$-\nabla f(x_k)^\mathsf{T} g_k \leq -\frac{1}{1+\kappa_g'}\|g_k\|_2^2,\quad -\nabla f(x_k)^\mathsf{T} g_k \leq -(1-\kappa_g')\|\nabla f(x_k)\|_2^2. \tag{3.6}$$

那么，根据假设2.1， $d_k = -g_k$以及（3.6），

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \alpha_k\nabla f(x_k)^\mathsf{T} d_k + \frac{L_g}{2}\alpha_k^2\|d_k\|_2^2$$

$$\leq f(x_k) - \tfrac{\alpha_k}{1+\kappa_g'}\|g_k\|_2^2 + \tfrac{L_g}{2}\alpha_k^2\|g_k\|_2^2.$$

因此，只要 $\alpha_k \leq \bar{\alpha} = {}^{2}(1/(1+\kappa_g')-c_d)\ L_g$ 该不等式就可以进一步被界为

$$f(x_k+\alpha_k d_k) \leq f(x_k) - c_d\alpha_k\|g_k\|_2^2 = f(x_k) + c_d\alpha_k g_k^{\mathsf{T}} d_k. \tag{3.7}$$

利用有界噪声条件（3.2）以及 $e_f \geq 2\epsilon_f$,

$$\begin{aligned} F(x_k+\alpha_k d_k,\ \xi_k^{(0+)}) &\leq f(x_k+\alpha_k d_k) + e_k^+ \leq f(x_k) + c_d\alpha_k g_k^{\mathsf{T}} d_k + \epsilon_f \\ &\leq F\ (x_k, \xi_k^{(0)}\ ) + c_d\alpha_k g_k^{\mathsf{T}} d_k + e_f, \end{aligned} \tag{3.8}$$

这表明迭代点 k 处的下降步是成功的（$\Theta_k^g = 1$），只要步长足够小（第（i）部分，引理3.1）。

当该步同样成功时，即 $I_k^g\ \Omega_k^g\Theta_k^g = 1$，根据算法2.1 中的流程以及噪声界（3.2），可得

$$F(x_k+\alpha_k d_k, \xi_k^{(0+)}) \leq F(x_k, \xi_k^{(0)}) - c_d\alpha_k\|g_k\|_2^2 + e_f \leq F(x_k, \xi_k^{(0)}) - c_d\alpha_k c_g^2\bar{\epsilon}_g^2 + e_f,$$

w此处 $h_d(\alpha) := c_g^2\bar{\epsilon}_g^2\alpha$。再次利用有界噪声条件以及 $e_f \geq 2\epsilon$ $_f$,

$$\begin{aligned} f(\hat{x}_k) = f(x_k+\alpha_k d_k) &\leq F(x_k+\alpha_k d_k, \xi_k^{(0+)}) + e_k^+ \\ &\leq F(x_k, \xi_k^{(0)}) - h_d(\alpha_k) + e_f + \epsilon_f \\ &\leq f(x_k) - h_d(\alpha_k) + 2e_f, \end{aligned} \tag{3.9}$$

从而使真实函数值下降。

若相反，$I_k^g(1-\Omega_k^g) = 1$，则 $\|\nabla f(x_k)\|_2 \leq \|g_k\|_2 + \|\nabla f(x_k) - g_k\|_2 < c_g\bar{\epsilon}_g + \epsilon_g + \kappa_g\|\nabla f(x_k)\|_2$，由此可得

$$\|\nabla f(x_k)\|_2 \leq \tfrac{c_g\bar{\epsilon}_g+\epsilon_g}{1-\kappa_g} \leq \tfrac{2c_g+1-\kappa_g-c_g}{2(1-\kappa_g)}\bar{\epsilon}_g < \bar{\epsilon}_g, \tag{3.10}$$

因此，停止时间定义的第一部分（3.1）成立。

当 $\Theta_k^g = 0$（下降步未成功）时，试探点会被拒绝，且 $\hat{}x_k = x_k, f(\hat{}x_k) = f(x_k)$。否则，$\Theta_k^g = 1$ 确保函数值的潜在增幅始终有界

$$\begin{aligned} f(x_k+\alpha_k d_k) &\leq F(x_k+\alpha_k d_k, \xi_k^{(0+)}) + e_k^+ \\ &\leq F(x_k, \xi_k^{(0)}) + c_d\alpha_k g_k^{\mathsf{T}} d_k + e_f + \epsilon_f \\ &\leq F(x_k, \xi_k^{(0)}) + e_f + \epsilon_f \leq f(x_k) + 2e_f \end{aligned} \tag{3.11}$$

需要指出的是，这个界也适用于 $\Theta_k^g = 0$ 的情形。

接下来，在 $\hat{x}_k \leftarrow x_k + \alpha_k d_k$，我们考虑负曲率步。当 $I_k^H \Omega_k^H = 1$（海森矩阵准确且 $\lambda_k < -c_H$ 最大值$\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\}$）时，根据（3.3）和（2.1b）

$$\begin{aligned} \|(\nabla^2 f(\hat{x}_k) - H_k) q_k\|_2 &\le \epsilon_H^2 + \kappa_H |\lambda_k| \|q_k\|_2 \\ &\le \tfrac{\delta(\gamma - \kappa_H - 2c_p\gamma)(c_H \bar{\epsilon}_H)^2}{2} + \kappa_H \delta |\lambda_k|^2 \\ &< \tfrac{\delta(\gamma + \kappa_H - 2c_p\gamma)|\lambda_k|^2}{2}, \end{aligned} \tag{3.12}$$

根据 $\nabla^2 f$ （假设2.1）的利普希茨连续性，

$$f(\hat{x}_k \pm \beta_k q_k) \le f(\hat{x}_k) \pm \beta_k \nabla f(\hat{x}_k)^\mathsf{T} q_k + \tfrac{\beta_k^2}{2} q_k^\mathsf{T} \nabla^2 f(\hat{x}_k) q_k + \tfrac{L_H}{6} \beta_k^3 \|q_k\|_2^3.$$

由于 $\min\{\nabla f(\hat{x}_k)^\mathsf{T} q_k,\ -\nabla f(\hat{x}_k)^\mathsf{T} q_k\} \le 0$，取上述

$$\min\{f(\hat{x}_k + \beta_k q_k), f(\hat{x}_k - \beta_k q_k)\} \le f(\hat{x}_k) + \tfrac{\beta_k^2}{2} q_k^\mathsf{T} \nabla^2 f(\hat{x}_k) q_k + \tfrac{L_H}{6} \beta_k^3 \|q_k\|_2^3. \tag{3.13}$$

注意到 $q_k^\mathsf{T} \nabla^2 f(\hat{x}_k) q_k = q_k^\mathsf{T} (\nabla^2 f(\hat{x}_k) - H_k) q_k + q_k^\mathsf{T} H_k q_k \le \|q_k\|_2 \|(\nabla^2 f(\hat{x}_k) - H_k) q_k\|_2 + q_k^\mathsf{T} H_k q_k$。由（3.12），对任意 $\beta_k \le \bar{\beta}$，可得

$$\tfrac{\beta_k^2}{2} \|q_k\|_2 \|(\nabla^2 f(\hat{x}_k) - H_k) q_k\|_2 + \tfrac{\beta_k^2}{2} q_k^\mathsf{T} H_k q_k + \tfrac{L_H}{6} \beta_k^3 \|q_k\|_2^3 \le c_p \beta_k^2 q_k^\mathsf{T} H_k q_k.$$

将上述界代入（3.13），当 $I_k^g \, \Omega_k^g = 1$ 且 $\beta_k \le \bar{\beta}$时，

$$\min\{f(\hat{x}_k + \beta_k q_k), f(\hat{x}_k - \beta_k q_k)\} \le f(\hat{x}_k) + c_p \beta_k^2 q_k^\mathsf{T} H_k q_k,$$

因此，由（3.2）和 $e_f \ge 2\epsilon_f$，可得

$$\begin{aligned} &\min\left\{F(\hat{x}_k + \beta_k q_k, \hat{\xi}_k^{(0+)}), F(\hat{x}_k - \beta_k q_k, \hat{\xi}_k^{(0-)})\right\} \\ \le\ &\min\left\{f(\hat{x}_k + \beta_k q_k) + \hat{e}_k^+, f(\hat{x}_k - \beta_k q_k) + \hat{e}_k^-\right\} \\ \le\ &\min\{f(\hat{x}_k + \beta_k q_k), f(\hat{x}_k - \beta_k q_k)\} + \max\{\hat{e}_k^+, \hat{e}_k^-\} \\ \le\ &f(\hat{x}_k) + c_p \beta_k^2 q_k^\mathsf{T} H_k q_k + \epsilon_f \le F(\hat{x}_k, \hat{\xi}_k^{(0)}) + c_p \beta_k^2 q_k^\mathsf{T} H_k q_k + e_f. \end{aligned}$$

因此，$\Theta_k^H = 1$ 当步长足够小时（引理 ii 的第 3.1 部分）。当 $I_k^H \Omega_k^H \Theta_k^H = 1$ 时，该步是成功的，且 $\lambda_k < -c_H \max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\}$，并且

$$\begin{aligned} &\min\left\{F(\hat{x}_k + \beta_k q_k, \hat{\xi}_k^{(0+)}), F(\hat{x}_k - \beta_k q_k, \hat{\xi}_k^{(0-)})\right\} \\ \le\ &F(\hat{x}_k, \hat{\xi}_k^{(0)}) + c_p \beta_k^2 p_k^\mathsf{T} H_k p_k + e_f \\ \le\ &F(\hat{x}_k, \hat{\xi}_k^{(0)}) - c_p \gamma \delta^2 \beta_k^2 |\lambda_k|^3 + e_f \\ \le\ &F(\hat{x}_k, \hat{\xi}_k^{(0)}) - \beta_k^2 c_p \gamma \delta^2 c_H^3 (\max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\})^3 + e_f, \end{aligned} \tag{3.14}$$

其中 $h_p(\beta) := c_p\gamma\delta^2 c_H^3(\max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\})^3\beta^2$。若 $F(\hat{x}_k \pm \beta_k q_k, \hat{\xi}_k^{(0+)}) = \min\{F(\hat{x}_k + \beta_k q_k, \hat{\xi}_k^{(0+)}) \; F(\hat{x}_k - \beta_k q_k, \hat{\xi}_k^{(0-)})\}$，根据 p_k 的定义和 (3.14)

$$\begin{aligned} f(\hat{x}_k + \beta_k p_k) &= f(\hat{x}_k \pm \beta_k q_k) \\ &\le F(\hat{x}_k \pm \beta_k q_k, \hat{\xi}_k^{(0+)}) + \hat{e}_k^+ \\ &= \min\{F(\hat{x}_k + \beta_k q_k, \hat{\xi}_k^{(0+)}), F(\hat{x}_k - \beta_k q_k, \hat{\xi}_k^{(0-)})\} + \hat{e}_k^+ \\ &\le F(\hat{x}_k, \hat{\xi}_k^{(0)}) - h_p(\beta_k) + e_f + \max\{\hat{e}_k^+, \hat{e}_k^-\}. \end{aligned} \tag{3.15}$$

综合这两种情形

$$\begin{aligned} f(\hat{x}_k + \beta_k p_k) &\le F(\hat{x}_k, \hat{\xi}_k^{(0)}) - h_p(\beta_k) + e_f + \max\{\hat{e}_k^+, \hat{e}_k^-\} \\ &\le f(\hat{x}_k) + \epsilon_f - h_p(\beta_k) + e_f + \epsilon_f \le f(\hat{x}_k) - h_p(\beta_k) + 2e_f. \end{aligned} \tag{3.16}$$

当 $I_k^H(1 - \Omega_k^H) = 1$时，我们有 $\lambda_k \ge -c_H \max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\}$ 并由（3.3）

$$\begin{aligned} |\lambda_{\min}(\nabla^2 f(\hat{x}_k)) - \lambda_k| &\le \epsilon_\lambda + \kappa_\lambda |\lambda_{\min}(\nabla^2 f(\hat{x}_k))| \\ &\le (1 - \kappa_\lambda - c_H)\bar{\epsilon}_\lambda + \kappa_\lambda |\lambda_{\min}(\nabla^2 f(\hat{x}_k))| \\ &\le (1 - \kappa_\lambda - c_H)\max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\} + \kappa_\lambda |\lambda_{\min}(\nabla^2 f(\hat{x}_k))|, \end{aligned}$$

由此可得

$$\begin{aligned} |\lambda_{\min}(\nabla^2 f(\hat{x}_k)) - \lambda_k| &\ge \lambda_k - \lambda_{\min}(\nabla^2 f(\hat{x}_k)) \\ &\ge -c_H \max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\} - \lambda_{\min}(\nabla^2 f(\hat{x}_k)), \end{aligned} \tag{3.17}$$

且 $\lambda_{\min}(\nabla^2 f(\hat{x}_k)) \ge -\max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\}$。

当 $\Theta_k^H = 0$（负曲率步不成功）时，试探点被拒绝，且 $x_{k+1} = \hat{x}_k$ 以及 $f(x_{k+1}) = f(\hat{x}_k)$。否则，$\Theta_k^H = 1$ 保证

$$\begin{aligned} f(x_{k+1}) = f(\hat{x}_k + \beta_k p_k) &\le F(\hat{x}_k + \beta_k p_k, \hat{\xi}_k^{(0\pm)}) + \max\{\hat{e}_k^+, \hat{e}_k^-\} \\ &\le F(\hat{x}_k, \hat{\xi}_k^{(0)}) + c_p\beta_k^2 p_k^{\mathsf{T}} H_k p_k + e_f + \max\{\hat{e}_k^+, \hat{e}_k^-\} \\ &\le F(\hat{x}_k, \hat{\xi}_k^{(0)}) - \gamma c_p \beta_k^2 |\lambda_k| \|p_k\|_2^2 + e_f + \max\{\hat{e}_k^+, \hat{e}_k^-\} \\ &\le F(\hat{x}_k, \hat{\xi}_k^{(0)}) + e_f + \max\{\hat{e}_k^+, \hat{e}_k^-\} \\ &\le f(\hat{x}_k) + e_f + 2\epsilon_f \le f(\hat{x}_k) + 2e_f. \end{aligned} \tag{3.18}$$

在这两种情形下，$\Theta_k^H = 0$与 $\Theta_k^H = 1$，我们有 $f(x_{k+1}) \le f(\hat{x}_k) + 2e_f$。

结合（3.9）、（3.18）和$I_k^g \; \Omega_k^g \Theta_k^g = 1$ 可得（3.5）中的第一个结果（引理 iii）的第3.1部分）。结合（3.11）、（3.16）和 $I_k^H \Omega_k^H \Theta_k^H = 1$ 可得（3.5）中的第二个结果（引理 iv）的第3.1部分）。最后一种情况可由（3.11）和（3.18）推得。

对于（v），反证法可知，若 $I_k^g(1-\Omega_k^g)I_k^H(1-\Omega_k^H)=1$，则 $\hat{}x_k=x_k$，并利用（3.10）和（3.17），可以得出

$$\|\nabla f(x_k)\|_2<\bar{\epsilon}_g,\quad \lambda_{\min}(\nabla^2 f(x_k))=\lambda_{\min}(\nabla^2 f(\hat{x}_k))\geq-\max\{\bar{\epsilon}_H,\bar{\epsilon}_\lambda\},$$

因此 $N_{\bar{\epsilon}_g,\ \bar{\epsilon}_H,\ \bar{\epsilon}_\lambda}\leq k$ 这与假设 $k<N_{\bar{\epsilon}_g,\ \bar{\epsilon}_H,\ \bar{\epsilon}_\lambda}$ 相矛盾。 □

引理3.1 给出了迭代过程中目标函数值变化的基本界，并考虑了算法中可能出现的不同情况。给定¯步长界 $\bar{\alpha}$ 和 β 引理中指定的参数，我们定义以下随机指示变量，用于指示第 k 次迭代是大步还是小步。

$$U_k^g=\begin{cases}1, & \min\{\alpha_k,\alpha_{k+1}\}\geq\bar{\alpha},\Omega_k^g=1\\ 0, & \max\{\alpha_k,\alpha_{k+1}\}\leq\bar{\alpha},\Omega_k^g=1\\ 0, & \text{if }\Omega_k^g=0,\end{cases},\quad U_k^H=\begin{cases}1, & \min\{\beta_k,\beta_{k+1}\}\geq\bar{\beta},\Omega_k^H=1\\ 0, & \max\{\beta_k,\beta_{k+1}\}\leq\bar{\beta},\Omega_k^H=1\\ 0, & \Omega_k^H=0.\end{cases}$$

¯不失一般性，假设 $\bar{\alpha}=\alpha_0\tau^{m_\alpha}$ 且 $\beta=\beta_0\tau^{m_\beta}$ 对某些正整数 m_α，m_β。在这种情况下，可以证明，对于每一步，α_k 和 β_k 要么是大步，要么是小步，要么是稳定步（例如，Ω_k^g 或 $\Omega_k^H=0$）。由引理3.1中的性质（i）-（ii）可直接推出，

$$I_k^g\Omega_k^g(1-U_k^g)(1-\Theta_k^g)=0,\quad\text{and}\quad I_k^H\Omega_k^H(1-U_k^H)(1-\Theta_k^H)=0.\tag{3.19}$$

上述结论表明，若梯度或负曲率信息足够准确且模足够大，则任意低于某一阈值的步长都会产生成功步。在引理3.2-3.4中，我们针对不同情形给出了迭代次数的界。

引理 3.2. 对任意正整数 t,

$$\sum_{k=0}^{t-1}I_k^g\Omega_k^g\Theta_k^gU_k^g+\sum_{k=0}^{t-1}I_k^H\Omega_k^H\Theta_k^HU_k^H\leq\frac{f(x_0)-f(x_t)}{c_{\bar{\alpha},\bar{\beta}}}+\frac{4e_f}{c_{\bar{\alpha},\bar{\beta}}}t.$$

¯ ¯其中 $c_{\bar{\alpha},\ \bar{\beta}}:=\min\{h_d(\bar{\alpha})\ h_p(\beta)\}$与 $\bar{\alpha}$，β由 (3.4)给出。

证明。 考虑步长指示变量（U_k^g和 U_k^H），可由引理3.5对引理3.1中的式子作轻微修改，得到:

$$f(x_{k+1})\leq\begin{cases}f(x_k)-h_d(\bar{\alpha})+4e_f, & \text{if } I_k^g\Omega_k^g\Theta_k^gU_k^g=1\\ f(x_k)-h_p(\bar{\beta})+4e_f, & \text{if } I_k^H\Omega_k^H\Theta_k^HU_k^H=1\\ f(x_k)+4e_f, & \text{otherwise,}\end{cases}$$

上述结论成立，是因为 $h_d(\cdot)$ 和 $h_p(\cdot)$ 在 $\mathbb{R}_{\geq 0}$ 上是非递减函数。将从 $k=0$ 到 $t-1$ 的不等式相加，可得

$$f(x_t) \leq f(x_0) - h_d(\bar{\alpha}) \sum_{k=0}^{t-1} I_k^g \Omega_k^g \Theta_k^g U_k^g - h_p(\bar{\beta}) \sum_{k=0}^{t-1} I_k^H \Omega_k^H \Theta_k^H U_k^H + 4te_f.$$

重新整理上述内容并利用 $c_{\bar{\alpha},\bar{\beta}}$ 的定义，即完成证明。 □

引理 3.2 由引理 3.1 推出，并给出了“良好”迭代（精确、成功、步长较大）的数量界。下一个引理给出了步长 α_k 和 β_k 发生变化的迭代次数界。当 $\|g_k\|_2$ 足够大（$\Omega_k^g = 1$）时，α_k在成功迭代（$\Theta_k^g = 1$）中增大，否则减小。因此，在满足 $U_k^g = 1$ 且 α_k 减小的迭代次数，受其增大的迭代次数以及将 α_k 从 α_0 减小到 $\bar{\alpha}$所需迭代次数之和所界定。类似地，在 $\Omega_k^g(1-U_k^g) = 1$ 且 α_k增大的迭代次数，受其减小的迭代次数所界定。下面的引理将这些观察形式化。

引理 3.3。 对于任意正整数 t,

$$\sum_{k=0}^{t-1} \Omega_k^g (1-U_k^g)\Theta_k^g + \sum_{k=0}^{t-1} \Omega_k^g U_k^g (1-\Theta_k^g) \leq \tfrac{1}{2}\left(\sum_{k=0}^{t-1} \Omega_k^g + c_\tau\right), \tag{3.20}$$

$$\sum_{k=0}^{t-1} \Omega_k^H (1-U_k^H)\Theta_k^H + \sum_{k=0}^{t-1} \Omega_k^H U_k^H (1-\Theta_k^H) \leq \tfrac{1}{2}\left(\sum_{k=0}^{t-1} \Omega_k^H + c_\tau\right), \tag{3.21}$$

其中 $c_\tau :=$ 最大值$\{$对数$_\tau \frac{\bar{\alpha}}{\alpha_0}$, 对数$_\tau \frac{\bar{\beta}}{\beta_0}, 0\}$ 以及 $\bar{\alpha}, \bar{\beta}$ 见 (3.4)。

证明。 我们首先考虑下降步的界（3.20）。对于所有 $\alpha_t, t \in \mathbb{N}$，步长可表示为

$$\alpha_t = \alpha_0 \tau^{\sum_{k=0,\,\Omega_k^g=1}^{t-1} ((1-\Theta_k^g)-\Theta_k^g)} = \alpha_0 \tau^{\sum_{k=0}^{t-1} \Omega_k^g ((1-\Theta_k^g)-\Theta_k^g)}. \tag{3.22}$$

令 $s_{-1} := 0$, $s_0 :=$ 最小值$\{k \in \mathbb{N} :\ \alpha_k = \bar{\alpha}\}$，以及 $s_i :=$ 最小值$\{k :\ k > s_{i-1}, \alpha_k = \bar{\alpha}\}$ 对于任意$i \geq 1$。由（3.22）以及 s_i 的定义，对于所有 $i \geq 0$,

$$\bar{\alpha} = \alpha_{s_{i+1}} = \alpha_{s_i} \tau^{\sum_{k=s_i}^{s_{i+1}-1} \Omega_k^g ((1-\Theta_k^g)-\Theta_k^g)} = \bar{\alpha} \tau^{\sum_{k=s_i}^{s_{i+1}-1} \Omega_k^g ((1-\Theta_k^g)-\Theta_k^g)},$$

由此可得 $0 = \sum_{k=s_i}^{s_{i+1}-1} \Omega_k^g ((1-\Theta_k^g)-\Theta_k^g)$。根据 s_i 的定义，我们有：(i) $\alpha_k > \bar{\alpha}$ 且 $U_k^g = 1$ 对任意 $0 \leq k < s_0$ 和 $\Omega_k^g = 1$；并且，(ii) $\{k :\ s_i \leq k < s_{i+1},\ \Omega_k^g = 1\}$是以下任一集合的子集：$\{k :\ s_i \leq k < s_{i+1},\ \Omega_k^g = 1,\ U_k^g = 1\}$ 或 $\{k :\ s_i \leq k < s_{i+1},\ \Omega_k^g = 1,\ U_k^g =$

0}。因此，对于 $i \in \mathbb{N}$,

$$\begin{aligned} 0 &= \sum_{k=s_i}^{s_{i+1}-1} \Omega_k^g((1-\Theta_k^g)-\Theta_k^g) \\ &= \sum_{k=s_i}^{s_{i+1}-1} \Omega_k^g U_k^g((1-\Theta_k^g)-\Theta_k^g) = \sum_{k=s_i}^{s_{i+1}-1} \Omega_k^g(1-U_k^g)((1-\Theta_k^g)-\Theta_k^g). \end{aligned} \tag{3.23}$$

令 $i_t \coloneqq \max\{i \in \mathbb{N} : s_i \le t\}$ 对任意 $t \in \mathbb{N}$。我们考虑两种情况。

情况 *1*（$\alpha_t \ge \bar{\alpha}$）：对于任意 $s_{i_t} \le k < t$ 使得 $\Omega_k^g = 1$，可得 $\alpha_k \ge \bar{\alpha}$ 且 $U_k^g = 1$。因此，步长方程（3.22）中的指数可表示为

$$\begin{aligned} \sum_{k=0}^{t-1} \Omega_k^g((1-\Theta_k^g)-\Theta_k^g) &= \sum_{i=-1}^{i_t} \sum_{k=s_i}^{\min\{s_{i+1},t\}-1} \Omega_k^g((1-\Theta_k^g)-\Theta_k^g) \\ &= \sum_{i=-1}^{i_t} \sum_{k=s_i}^{\min\{s_{i+1},t\}-1} \Omega_k^g U_k^g((1-\Theta_k^g)-\Theta_k^g) \\ &= \sum_{i=0}^{t-1} \Omega_k^g U_k^g((1-\Theta_k^g)-\Theta_k^g), \end{aligned} \tag{3.24}$$

由（3.23）以及 $U_k^g = 1$ 对于 $k = 0, \ldots\ldots, s_0 - 1$，只要 $\Omega_k^g = 1$，即可得 $\alpha_t = \alpha_0 \tau^{\sum_{k=0}^{t-1} \Omega_k^g U_k^g \left((1-\Theta_k^g)-\Theta_k^g\right)} \ge \bar{\alpha}$。因此，由（3.24）以及 c_τ 的定义，

$$\sum_{k=0}^{t-1} \Omega_k^g U_k^g(1-\Theta_k^g) \le \sum_{k=0}^{t-1} \Omega_k^g U_k^g \Theta_k^g + c_\tau. \tag{3.25}$$

情况 *2*（$\alpha_t \le \bar{\alpha}$）：在这种情况下，对于任意 $s_{i_t} \le k < t$，可得 $\alpha_k \le \bar{\alpha}$ 且 $U_k^g = 0$。步长 α_t 可用 α_{s_0} 处的信息表示如下

$$\alpha_t = \alpha_{s_0} \tau^{\sum_{k=s_0}^{t-1} \Omega_k^g((1-\Theta_k^g)-\Theta_k^g)} = \bar{\alpha} \tau^{\sum_{k=s_0}^{t-1} \Omega_k^g((1-\Theta_k^g)-\Theta_k^g)}. \tag{3.26}$$

按照与式（3.24）中相同的方法，式（3.26）中的指数可表示为

$$\sum_{k=s_0}^{t-1} \Omega_k^g((1-\Theta_k^g)-\Theta_k^g) = \sum_{i=0}^{t-1} \Omega_k^g(1-U_k^g)((1-\Theta_k^g)-\Theta_k^g),$$

由（3.23）以及$1-U_k^g = 0$对于 $k = 0, \ldots\ldots, s_0 - 1$在$\Omega_k^g = 1$时成立这一事实。因此，$\alpha_t = \bar{\alpha} \cdot \tau^{\sum_{i=0}^{t-1} \Omega_k^g(1-U_k^g)((1-\Theta_k^g)-\Theta_k^g)} \le \bar{\alpha}$，由此可得

$$\sum_{i=0}^{t-1} \Omega_k^g(1-U_k^g)(1-\Theta_k^g) \ge \sum_{i=0}^{t-1} \Omega_k^g(1-U_k^g)\Theta_k^g. \tag{3.27}$$

（3.20）中的界通过对（3.25）与（3.27）求和得到。负曲率步长的相应结果（3.21）可类似地得到。

□

引理 3.3 界定了对应于两种“糟糕”情形的迭代次数：步长过小的成功步和步长过大的不成功步。现在我们给出一个刻画“良好”迭代的引理。

引理 3.4。对于所有正整数 t，以及满足 $\bar{p}_g\bar{p}_H+\bar{p}_g+\bar{p}_H-2>0$的任意 $0<\bar{p}_g<p_g$和 $0<\bar{p}_H<p_H$，如果 $N_{\bar{\epsilon}_g,\,\bar{\epsilon}_H,\,\bar{\epsilon}_\lambda}>t$和 $\sum_{k=0}^{t-1}I_k^g\geq\bar{p}_g t$，$\sum_{k=0}^{t-1}I_k^H\geq\bar{p}_H t$，且 $\sum_{k=0}^{t-1}I_k^gI_k^H\geq\bar{p}_g\bar{p}_H t$，则有：

$$\sum_{k=0}^{t-1}I_k^g\Omega_k^g\Theta_k^gU_k^g+\sum_{k=0}^{t-1}I_k^H\Omega_k^H\Theta_k^HU_k^H\geq\tfrac{1}{2}c_{gH}t-c_\tau,$$

其中 $c_{gH}:=\bar{p}_g\bar{p}_H+\bar{p}_g+\bar{p}_H-2$，$c_\tau:=$最大值{对数$_\tau\,\frac{\bar{\alpha}}{\alpha_0}$，对数$_\tau\,\frac{\bar{\beta}}{\beta_0}$，0}以及 $\bar{\alpha}$，$\bar{\beta}$均在（3.4）中给出。因此，

$$\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda}>t,\quad\sum_{k=0}^{t-1}I_k^g\geq\bar{p}_g t,\quad\sum_{k=0}^{t-1}I_k^H\geq\bar{p}_H t,\quad\sum_{k=0}^{t-1}I_k^gI_k^H\geq\bar{p}_g\bar{p}_H t,\right.\tag{3.28}$$
$$\left.\text{and}\quad\sum_{k=0}^{t-1}I_k^g\Omega_k^g\Theta_k^gU_k^g+\sum_{k=0}^{t-1}I_k^H\Omega_k^H\Theta_k^HU_k^H<\tfrac{1}{2}c_{gH}t-c_\tau\right]=0.$$

证明。若$N_{\bar{\epsilon}_g,\,\bar{\epsilon}_H,\,\bar{\epsilon}_\lambda}>t$且 $\sum_{k=0}^{t-1}I_k^g\geq\bar{p}_g t$，$\sum_{k=0}^{t-1}I_k^H\geq\bar{p}_H t$，且 $\sum_{k=0}^{t-1}I_k^gI_k^H\geq\bar{p}_g\bar{p}_H t$，则对于下降步，由（3.19）和（3.20），

$$\begin{aligned}&\sum_{k=0}^{t-1}I_k^g\Omega_k^g\Theta_k^gU_k^g\\&=\sum_{k=0}^{t-1}I_k^g\Omega_k^g-\sum_{k=0}^{t-1}I_k^g\Omega_k^g(1-\Theta_k^g)U_k^g-\sum_{k=0}^{t-1}I_k^g\Omega_k^g\Theta_k^g(1-U_k^g)-\sum_{k=0}^{t-1}I_k^g\Omega_k^g(1-\Theta_k^g)(1-U_k^g)\\&\geq\sum_{k=0}^{t-1}I_k^g\Omega_k^g-\tfrac{1}{2}\left(\sum_{k=0}^{t-1}\Omega_k^g+c_\tau\right)=\tfrac{1}{2}\sum_{k=0}^{t-1}I_k^g\Omega_k^g-\tfrac{1}{2}\sum_{k=0}^{t-1}(1-I_k^g)\Omega_k^g-\tfrac{c_\tau}{2}.\end{aligned}$$

同样，对于负曲率步，由 (3.19) 和 (3.21) 可得

$$\sum_{k=0}^{t-1}I_k^H\Omega_k^H\Theta_k^HU_k^H\geq\tfrac{1}{2}\sum_{k=0}^{t-1}I_k^H\Omega_k^H-\tfrac{1}{2}\sum_{k=0}^{t-1}(1-I_k^H)\Omega_k^H-\tfrac{c_\tau}{2}.$$

综合上述两个不等式，可得

$$\begin{aligned}&\sum_{k=0}^{t-1} I_k^g \Omega_k^g \Theta_k^g U_k^g + \sum_{k=0}^{t-1} I_k^H \Omega_k^H \Theta_k^H U_k^H \\ \geq & \frac{1}{2}\sum_{k=0}^{t-1} I_k^g \Omega_k^g + \frac{1}{2}\sum_{k=0}^{t-1} I_k^H \Omega_k^H - \frac{1}{2}\sum_{k=0}^{t-1}(1-I_k^g)\Omega_k^g - \frac{1}{2}\sum_{k=0}^{t-1}(1-I_k^H)\Omega_k^H - c_\tau.\end{aligned} \tag{3.29}$$

根据引理 v 的 (3.1 以及 $I_k^g(1-\Omega_k^g)I_k^H(1-\Omega_k^H) = 0$，对于任意 $k \leq t < N_{\bar{\epsilon}_g,\ \bar{\epsilon}_H,\ \bar{\epsilon}_\lambda}$,

$$\begin{aligned}\sum_{k=0}^{t-1} I_k^g I_k^H &= \sum_{k=0}^{t-1} I_k^g \Omega_k^g I_k^H \Omega_k^H + \sum_{k=0}^{t-1} I_k^g (1-\Omega_k^g) I_k^H \Omega_k^H + \sum_{k=0}^{t-1} I_k^g \Omega_k^g I_k^H (1-\Omega_k^H) + \sum_{k=0}^{t-1} I_k^g (1-\Omega_k^g) I_k^H (1-\Omega_k^H) \\ &= \sum_{k=0}^{t-1} I_k^g I_k^H \Omega_k^H + \sum_{k=0}^{t-1} I_k^g \Omega_k^g I_k^H (1-\Omega_k^H) \\ &\leq \sum_{k=0}^{t-1} I_k^H \Omega_k^H + \sum_{k=0}^{t-1} I_k^g \Omega_k^g.\end{aligned}$$

将上述结果代入（3.29）并利用c_{gH} 的定义即可完成证明,

$$\begin{aligned}\sum_{k=0}^{t-1} I_k^g \Omega_k^g \Theta_k^g U_k^g + \sum_{k=0}^{t-1} I_k^H \Omega_k^H \Theta_k^H U_k^H &\geq \frac{1}{2}\sum_{k=0}^{t-1} I_k^g I_k^H - \frac{1}{2}\sum_{k=0}^{t-1}(1-I_k^g) - \frac{1}{2}\sum_{k=0}^{t-1}(1-I_k^H) - c_\tau \\ &\geq \frac{1}{2}\bar{p}_g\bar{p}_H t - \frac{1}{2}(1-\bar{p}_g)t - \frac{1}{2}(1-\bar{p}_H)t - c_\tau.\end{aligned}$$

□

上述引理表明：当准确迭代的次数足够大时，“好”迭代（准确、成功、大步长）的数量存在一个随 t增大的下界。主要定理如下。

定理 3.5. 假设 2.1 与 2.9 成立且 $c_{gH} \geq 4\epsilon_f$。则对任意满足 $\bar{p}_g\bar{p}_H + \bar{p}_g + \bar{p}_H - 2 =: c_{gH} > 8\epsilon_f$ 以及 $0 < \bar{p}_g < p_g,\quad 0 < \bar{p}_H < p_H$，

$$\mathbb{P}[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} \leq t] \geq 1 - \exp\left(-\frac{(p_g-\bar{p}_g)^2}{2p_g^2}t\right) - \exp\left(-\frac{(p_H-\bar{p}_H)^2}{2p_H^2}t\right) - \exp\left(-\frac{(p_g p_H - \bar{p}_g\bar{p}_H)^2}{2p_g^2 p_H^2}t\right),$$

其中 $R = \frac{f_{(x_0)} - f^*}{c_{\bar{\alpha},\ \bar{\beta}}} + c_\tau$，$c_{\bar{\alpha},\ \bar{\beta}} := \min\left\{c_d\bar{\alpha}c_g^2\bar{\epsilon}_g^2, c_p\bar{\beta}^2 c_H^3(\max\{\bar{\epsilon}_H,\bar{\epsilon}_\lambda\})^3\right\}$，$\bar{\alpha}$，$\bar{\beta}$ 由(3.4)给出，c_τ 由引理 3.3给出，$\epsilon_c = 16\epsilon_f$，并且

$$\bar{\epsilon}_g > \max\left\{\left(\frac{\epsilon_c}{(\bar{p}_g\bar{p}_H+\bar{p}_g+\bar{p}_H-2)c_d\bar{\alpha}c_g^2}\right)^{1/2}, \frac{2\epsilon_g}{\min\{\eta c_g(1-\kappa_g), 1-\kappa_g-c_g\}}\right\},$$

$$\bar{\epsilon}_H > \max\left\{\left(\frac{\epsilon_c}{(\bar{p}_g\bar{p}_H+\bar{p}_g+\bar{p}_H-2)c_p\bar{\beta}^2c_H^3}\right)^{1/3}, \frac{\epsilon_H}{c_H}\sqrt{\frac{2}{\delta(\gamma-\kappa_H-2c_p\gamma)}}\right\}, \tag{3.30}$$
$$\bar{\epsilon}_\lambda > \max\left\{\left(\frac{\epsilon_c}{(\bar{p}_g\bar{p}_H+\bar{p}_g+\bar{p}_H-2)c_p\bar{\beta}^2c_H^3}\right)^{1/3}, \frac{\epsilon_\lambda}{1-\kappa_\lambda-c_H}\right\}.$$

证明。 当 $N_{\bar{\epsilon}_g,\ \bar{\epsilon}_H,\ \bar{\epsilon}_\lambda} > t \geq T$，根据引理 3.2 以及 T的定义

$$\sum_{k=0}^{t-1} I_k^g\Omega_k^g\Theta_k^gU_k^g + \sum_{k=0}^{t-1} I_k^H\Omega_k^H\Theta_k^HU_k^H \leq \frac{f(x_0)-f(x_t)}{c_{\bar{\alpha},\bar{\beta}}} + \frac{4e_f}{c_{\bar{\alpha},\bar{\beta}}}t < \frac{f(x_0)-f^*}{c_{\bar{\alpha},\bar{\beta}}} + \frac{1}{2}c_{gH}t - R$$
$$= \frac{1}{2}c_{gH}t - c_\tau.$$

我们将这些事件记为

$$E_g \coloneqq \left\{\sum_{k=0}^{t-1} I_k^g \geq \bar{p}_g t\right\}, \quad E_H \coloneqq \left\{\sum_{k=0}^{t-1} I_k^H \geq \bar{p}_H t\right\}, \quad E_{gH} \coloneqq \left\{\sum_{k=0}^{t-1} I_k^gI_k^H \geq \bar{p}_g\bar{p}_H t\right\}.$$

经过 t 次迭代后仍未到达二阶驻点的概率为

$$\begin{aligned}
&\mathbb{P}[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t]\\
=&\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t,\ \sum_{k=0}^{t-1} I_k^g\Omega_k^g\Theta_k^gU_k^g + \sum_{k=0}^{t-1} I_k^H\Omega_k^H\Theta_k^HU_k^H < \frac{1}{2}c_{gH}t - c_\tau\right]\\
=&\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t,\ \sum_{k=0}^{t-1} I_k^g\Omega_k^g\Theta_k^gU_k^g + \sum_{k=0}^{t-1} I_k^H\Omega_k^H\Theta_k^HU_k^H < \frac{1}{2}c_{gH}t - c_\tau,\ E_g\cap E_H\cap E_{gH}\right]\\
&+\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t,\ \sum_{k=0}^{t-1} I_k^g\Omega_k^g\Theta_k^gU_k^g + \sum_{k=0}^{t-1} I_k^H\Omega_k^H\Theta_k^HU_k^H < \frac{1}{2}c_{gH}t - c_\tau,\ E_g^c\cup E_H^c\cup E_{gH}^c\right]\\
\leq&0 + \mathbb{P}\left\{E_g^c\cup E_H^c\cup E_{gH}^c\right\}\\
\leq&\mathbb{P}[E_g^c] + \mathbb{P}[E_H^c] + \mathbb{P}[E_{gH}^c]\\
\leq&\exp\left(-\frac{(p_g-\bar{p}_g)^2}{2p_g^2}t\right) + \exp\left(-\frac{(p_H-\bar{p}_H)^2}{2p_H^2}t\right) + \exp\left(-\frac{(p_gp_H-\bar{p}_g\bar{p}_H)^2}{2p_g^2p_H^2}t\right),
\end{aligned}$$

其中，第一个不等式由（3.28）得出，最后一个不等式则由引理 2.11 的一个特例得出，其中 $p_f = \bar{p}_f = 1$ 且 $I_k^f = \hat{}I_k^f = 1$ 对所有正整数 k。 □

注释 3.6。 我们对定理 *3.5*作如下评论。

- 复杂度界对 $\bar{\epsilon}_g$ 和 $\bar{\epsilon}_H$ 的依赖关系，与确定性情形*[4, 15, 21]*,相呼应：迭代点收敛到驻点的一个邻域。邻域参数 $\bar{\epsilon}_g$、$\bar{\epsilon}_H$和 $\bar{\epsilon}_\lambda$ 以不同阶数依赖于 ϵ_f，满足 $\bar{\epsilon}_g > \mathcal{O}(\epsilon_f^{1/2}+\epsilon_g)$、$\bar{\epsilon}_H > \mathcal{O}(\epsilon_f^{1/3}+\epsilon_H)$和 $\bar{\epsilon}_\lambda > \mathcal{O}(\epsilon_f^{1/3}+\epsilon_\lambda)$。因此，算法 *2.1* 可达到的二阶精度为 $\left(\mathcal{O}(\epsilon_f^{1/2}+\epsilon_g)\ \mathcal{O}(\epsilon_f^{1/3}+\epsilon_H)\ \mathcal{O}(\epsilon_f^{1/3}+\epsilon_\lambda)\right)$。

- 当 $p_g = p_H = 1$ 且 $\epsilon_f = \epsilon_g = \epsilon_H = \epsilon_\lambda = 0$时，结果退化为确定性非精确情形 *[4, 定理 2。10]*，其中，$\bar{\epsilon}_g$ 和 $\bar{\epsilon}_H$ 的下界均为零。此外，若 $p_H = 1$，对 p_g 的要求变为 $p_g > 1/2$，这与 *[20]*中的一阶条件相符。

- $N_{\bar{\epsilon}_g,\ \bar{\epsilon}_H,\ \bar{\epsilon}_\lambda}$ 超过 t 的概率对于 $t \geq T$随 t 呈指数衰减，其中 $T = \mathcal{O}\Big(\epsilon_f^{-1} + \epsilon_g^{-2} + (\max\{\epsilon_H, \epsilon_\lambda\})^{-3}\Big)$。

3.2 收敛性分析：次指数噪声情形

我们现在将分析扩展到函数估计中的噪声具有次指数尾的情形，对应于预言机 1.2。有界噪声情形中的引理在将 $I_k^g(p_g)$ 替换为 $I_k^f I_k^g(p_f p_g)$，并将 $I_k^H(p_H)$ 替换为 $\hat{}I_k^f I_k^H(p_f p_H)$ 之后仍然成立。我们首先陈述引理3.1 在该情形下的对应结论。

引理 3.7. 设 $e_f \geq 4/a + 2\epsilon_f$以及邻域参数（$\bar{\epsilon}_g$，$\bar{\epsilon}_H$，$\bar{\epsilon}_\lambda$）按（3.3）中预先¯规定的方式选取。则存在常数 $\bar{\alpha}$、$\beta > 0$ （3.4）以及非减函数$h_d(\cdot)$ $h_p(\cdot)$ ： $\mathbb{R} \to \mathbb{R}$，使得对任意 α、$\beta > 0$满足 h_d（α） h_p（β） > 0，从而对算法*2.1* 的任意实现，下列结论对所有$k < N_{\bar{\epsilon}_g,\ \bar{\epsilon}_H,\ \bar{\epsilon}_\lambda}$成立：

(i) 若第 k 次迭代的下降步中的函数和梯度估计值准确（即 $I_k^f I_k^g = 1$），且 $\|g_k\|_2 \geq c_g\bar{\epsilon}_g$（即 $\Omega_k^g = 1$），则对于 $\alpha_k \leq \bar{\alpha}$，该迭代（下降步）是成功的（$\Theta_k^g = 1$），这意味着 $\alpha_{k+1} = \tau^{-1}\alpha_k$。

(ii) 若第 k 次迭代的负曲率步中的函数和海森矩阵估计值准确（即 $\hat{}I_k^f I_k^H = 1$），且 $\lambda_k \leq -c_H \max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\}$（即 $\Omega_k^H = 1$），则对于 $\beta_k \leq \bar{\beta}$，该迭代（负曲率步）是成功的（$\Theta_k^H = 1$），这意味着 $\beta_{k+1} = \tau^{-1}\beta_k$。

(iii) 当 $I_k^f I_k^g \Omega_k^g \Theta_k^g = 1$，$f(x_{k+1}) \leq f(x_k) - h_d(\alpha_k) + 3e_f + \hat{}e_k + \max\{\hat{}e_k^+,\ \hat{}e_k^-\}$。

(iv) 当 $\hat{}I_k^f I_k^H \Omega_k^H \Theta_k^H = 1$, $f(x_{k+1}) \leq f(x_k) - h_p(\beta_k) + 3e_f + e_k + e_k^+$。

(v) $I_k^g(1 - \Omega_k^g) I_k^H(1 - \Omega_k^H) = 0.$

综上所述，

$$f(x_{k+1}) \leq \begin{cases} f(x_k) - h_d(\alpha_k) + 3e_f + \hat{e}_k + \max\{\hat{e}_k^+, \hat{e}_k^-\}, & \text{if } I_k^f I_k^g \Omega_k^g \Theta_k^g = 1, \\ f(x_k) - h_p(\beta_k) + 3e_f + e_k + e_k^+, & \text{if } \hat{I}_k^f I_k^H \Omega_k^H \Theta_k^H = 1, \\ f(x_k) + 2e_f + e_k + e_k^+ + \hat{e}_k + \max\{\hat{e}_k^+, \hat{e}_k^-\}, & \text{otherwise.} \end{cases} \tag{3.31}$$

证明。 证明沿用引理3.1中的相同论证。具体而言，部分 (v) 在陈述和证明中完全相同，因此略去。下面重述其余论证，并着重说明差异。

对于下降步，当 $I_k^f I_k^g \Omega_k^g = 1$ 时，该情形与引理3.1 中的 (i) 相同。由 (3.7) 和 (3.8), 只要满足 ($\alpha_k \le \bar{\alpha}$) 中定义的条件 (3.4，便可得

$$F(x_k + \alpha_k d_k, \xi_k^{(0+)}) \le F(x_k, \xi_k^{(0)}) + c_d \alpha_k g_k^{\mathsf{T}} d_k + e_k + e_k^+ \le F(x_k, \xi_k^{(0)}) + c_d \alpha_k g_k^{\mathsf{T}} d_k + e_f,$$

这表明该下降步是成功的（$\Theta_k^g = 1$）。当 $I_k^f I_k^g \Omega_k^g \Theta_k^g = 1$时，该情形类似于（$iii$），即引理3.1中的相应情形，并且

$$f(x_k + \alpha_k d_k) \le f(x_k) - h_d(\alpha_k) + e_f + e_k + e_k^+ \le f(x_k) - h_d(\alpha_k) + 2e_f,$$

其中 h_d（α）$= c_g^2 \bar{\epsilon}_g^2 \alpha$。对 f（$\hat{}x_k$）的最坏情况界，类似于（3.11），为

$$f(\hat{x}_k) = f(x_k + \alpha_k d_k) \le f(x_k) + e_f + e_k + e_k^+.$$

对于负曲率步，当$\hat{}I_k^f I_k^H \Omega_k^H = 1$时，对于 $\beta_k \le \bar{\beta}$（定义见（3.4））,

$$\begin{aligned}
&\min\{F(\hat{x}_k + \beta_k q_k, \hat{\xi}_k^{(0+)}), F(\hat{x}_k - \beta_k q_k, \hat{\xi}_k^{(0-)})\} \\
\le &\min\{f(\hat{x}_k + \beta_k q_k), f(\hat{x}_k - \beta_k q_k)\} + \max\{\hat{e}_k^+, \hat{e}_k^-\} \\
\le &f(\hat{x}_k) + c_p \beta_k^2 q_k^{\mathsf{T}} H_k q_k + \max\{\hat{e}_k^+, \hat{e}_k^-\} \\
\le &F(\hat{x}_k, \hat{\xi}_k^{(0)}) + c_p \beta_k^2 q_k^{\mathsf{T}} H_k q_k + \hat{e}_k + \max\{\hat{e}_k^+, \hat{e}_k^-\} \le F(\hat{x}_k, \hat{\xi}_k^{(0)}) + c_p \beta_k^2 q_k^{\mathsf{T}} H_k q_k + e_f
\end{aligned}$$

并且该步是成功的（$\Theta_k^H = 1$）。当$\hat{}I_k^f I_k^H \Omega_k^H \Theta_k^H = 1$时，

$$\begin{aligned}
f(\hat{x}_k + \beta_k p_k) &\le \min\{F(\hat{x}_k + \beta_k q_k, \hat{\xi}_k^{(0+)}), F(\hat{x}_k - \beta_k q_k, \hat{\xi}_k^{(0-)})\} + \max\{\hat{e}_k^+, \hat{e}_k^-\} \\
&\le F(\hat{x}_k, \hat{\xi}_k^{(0)}) + c_p \beta_k^2 q_k^{\mathsf{T}} H_k q_k + e_f + \max\{\hat{e}_k^+, \hat{e}_k^-\} \\
&\le f(\hat{x}_k) + \hat{e}_k - h_p(\beta_k) + e_f + \max\{\hat{e}_k^+, \hat{e}_k^-\} \le f(\hat{x}_k) - h_p(\beta_k) + 2e_f,
\end{aligned}$$

其中 h_p（β）$= c_p \gamma \delta^2 c_H^3$（$\max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\}$）$^3 \beta^2$。类似于（3.18）的界为

$$f(x_{k+1}) = f(\hat{x}_k + \beta_k p_k) \le f(\hat{x}_k) + e_f + \hat{e}_k + \max\{\hat{e}_k^+, \hat{e}_k^-\}.$$

综合上述不同情形的不等式，即可得到所需结果 . □

下面的引理与引理 3.2类似，它给出了“好”迭代次数的界，并可直接由引理 3.7推出。

引理 3.8。对于任意正整数 t，我们有

$$\begin{aligned}
\sum_{k=0}^{t-1} I_k^f I_k^g \Omega_k^g \Theta_k^g U_k^g + \sum_{k=0}^{t-1} \hat{I}_k^f I_k^H \Omega_k^H \Theta_k^H U_k^H \le &\frac{f(x_0) - f(x_t)}{c_{\bar{\alpha},\bar{\beta}}} + \frac{3e_f}{c_{\bar{\alpha},\bar{\beta}}} t \\
&+ \frac{1}{c_{\bar{\alpha},\bar{\beta}}} \sum_{k=0}^{t-1} \left(e_k + e_k^+ + \hat{e}_k + \max\{\hat{e}_k^+, \hat{e}_k^-\}\right),
\end{aligned}$$

其中$c_{\bar{\alpha},\bar{\beta}} := \min\{h_d(\bar{\alpha}), h_p(\bar{\beta})\}$以及 $\bar{\alpha}$，$\bar{\beta}$在（3.4）中给出。

证明。类似于引理3.2的证明，通过引入步长指示变量U_k^g和 U_k^H，可将（3.31）改写为

$$f(x_{k+1}) \leq \begin{cases} f(x_k) - h_d(\bar{\alpha}) + 3e_f + \hat{e}_k + \max\{\hat{e}_k^+, \hat{e}_k^-\}, & \text{if } I_k^f I_k^g \Omega_k^g \Theta_k^g U_k^g = 1, \\ f(x_k) - h_p(\bar{\beta}) + 3e_f + e_k + e_k^+, & \text{if } \hat{I}_k^f I_k^H \Omega_k^H \Theta_k^H U_k^H = 1, \\ f(x_k) + 3e_f + e_k + e_k^+ + \hat{e}_k + \max\{\hat{e}_k^+, \hat{e}_k^-\}, & \text{otherwise.} \end{cases}$$

因为 $h_d(\cdot)$ 和 $h_p(\cdot)$ 在 $\mathbb{R}_{\geq 0}$ 上均为非递减函数。将上述不等式从 $k=0$ 到 $t-1$ 求和，重新排列各项，并根据 $c_{\bar{\alpha},\bar{\beta}}$ 的定义，即可完成证明。

□

引理3.3可直接应用于此情形。下一个引理与引理3.4类似，它给出了“良好”迭代（准确、成功且步长较大）的数量下界，这意味着随着 t 增加，其占比仍有正的下界。

引理 3.9. 对于所有正整数 t，以及任意满足 $\bar{p}_{fg}\cdot\bar{p}_{fH}+\bar{p}_{fg}+\bar{p}_{fH}-2>0$ 的 $0<\bar{p}_{fg}<p_f p_g$ 和 $0<\bar{p}_{fH}<p_f p_H$，若 $N_{\bar{\epsilon}_g,\ \bar{\epsilon}_H,\ \bar{\epsilon}_\lambda} > t$ 且 $\sum_{k=0}^{t-1} I_k^f I_k^g \geq \bar{p}_g t$, $\sum_{k=0}^{t-1} \hat{} I_k^f I_k^H \geq \bar{p}_{fH} t$，并且 $\sum_{k=0}^{t-1} I_k^f I_k^g \hat{} I_k^f I_k^H \geq \bar{p}_{fg}\bar{p}_{fH} t$，则，

$$\sum_{k=0}^{t-1} I_k^f I_k^g \Omega_k^g \Theta_k^g U_k^g + \sum_{k=0}^{t-1} \hat{I}_k^f I_k^H \Omega_k^H \Theta_k^H U_k^H \geq \tfrac{1}{2}(\bar{p}_{fg}\bar{p}_{fH} + \bar{p}_{fg} + \bar{p}_{fH} - 2)t - c_\tau = \tfrac{1}{2} c_{gH} t - c_\tau,$$

其中 $c_{gH} :=\ \bar{p}_{fg}\bar{p}_{fH} + \bar{p}_{fg} + \bar{p}_{fH} - 2$，$c_\tau :=$ 最大值{对数$_\tau$ $\frac{\bar{\alpha}}{\alpha_0}$，对数$_\tau$ $\frac{\bar{\beta}}{\beta_0}$，0} 以及 $\bar{\alpha}$，$\bar{\beta}$ 均在 (3.4)中给出。因此，

$$\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t, \quad \sum_{k=0}^{t-1} I_k^f I_k^g \geq \bar{p}_{fg} t, \quad \sum_{k=0}^{t-1} \hat{I}_k^f I_k^H \geq \bar{p}_{fH} t, \quad \sum_{k=0}^{t-1} I_k^f I_k^g \hat{I}_k^f I_k^H \geq \bar{p}_{fg}\bar{p}_{fH} t, \right.$$
$$\left. and \quad \sum_{k=0}^{t-1} I_k^f I_k^g \Omega_k^g \Theta_k^g U_k^g + \sum_{k=0}^{t-1} \hat{I}_k^f I_k^H \Omega_k^H \Theta_k^H U_k^H < \tfrac{1}{2} c_{gH} t - c_\tau \right] = 0.$$

证明。 证明过程与引理 3.4的证明类似，并结合 I_k^f 和 $\hat{} I_k^f$ 及其对应的概率，因此略去。

□

主要结果由此得出，其论证与定理 3.5类似 .

定理 3.10. 假设 *2.1* 和 *2.9* 成立，且 $e_f \geq 2\epsilon_f + 5/a$。那么，对于任意满足 $\bar{p}_{fg}\bar{p}_{fH}+\bar{p}_{fg}+\bar{p}_{fH}-2 =:\ c_{gH} > \frac{10e_f+4s}{c_{\bar{\alpha}\ \bar{\beta}}}$、$t \geq T :=\ _{c_{gH}}\ \frac{R}{5e_f+2s}$、$-^2$ $c_{\bar{\alpha},\bar{\beta}}$ 的 ~~$s \geq 0$~~、$0<\bar{p}_{fg}<p_f\underline{p_g}$、~~$0<\bar{p}_{fH}<p_f p_H$~~

$$\mathbb{P}\{N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} \leq t\} \geq 1 - \exp\left(-\tfrac{(p_f p_g - \bar{p}_{fg})^2}{2p_f^2 p_g^2} t\right) - \exp\left(-\tfrac{(p_f p_H - \bar{p}_{fH})^2}{2p_f^2 p_H^2} t\right)$$

$$- \exp\left(-\frac{(p_f^2 p_g p_H - \bar{p}_{fg}\bar{p}_{fH})^2}{2p_f^4 p_g^2 p_H^2} t\right) - 2\exp\left(-\frac{a}{4}st\right)$$

其中 $R = {}^{f}(x_0)c_{\bar{\alpha}}^{\ -f^*}$、$_{\bar{\beta}} \quad + c_\tau$、$c_{\bar{\alpha}}$、$\bar{\beta}$:=*min* $\left\{c_d\bar{\alpha}c_g^2\bar{\epsilon}_g^2,\ c_p\bar{\beta}^2c_H^3(max\{\bar{\epsilon}_H,\bar{\epsilon}_\lambda\})^3\right\}$、$\bar{\alpha}$、$\bar{\beta}$ 在 *(3.4)* 中给出，c_τ 见引理 *3.3*，而 $\bar{\epsilon}_g$、$\bar{\epsilon}_H$ 和 $\bar{\epsilon}_\lambda$ 的界与 *(3.30)* 中的形式相同，因子为 $\epsilon_c = 16\epsilon_f + 32/a + 4s$。

证明。我们定义以下两个事件：$A_t := \left\{\frac{1}{t}\sum_{k=0}^{t-1}(e_k + e_k^+) \le e_f + s\right\}$ 以及 $\hat{} A_t := \left\{\frac{1}{t}\sum_{k=0}^{t-1}(\hat{}e_k + \max\{\hat{}e_k^+, \hat{}e_k^-\}) \le e_f + s\right\}$。事件 $\{N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t\}$ 随后可利用 A_t 和 A_t 划分为两个事件，

$$\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t\right] = \mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t,\ A_t \cap \hat{A}_t\right] + \mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t,\ A_t^c \cup \hat{A}_t^c\right]. \tag{3.32}$$

我们首先对后一个项取界。由 $e_f \ge 2\epsilon_f + 5/a$ 和引理 2.8，可得

$$\begin{aligned}
&\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t,\ A_t^c \cup \hat{A}_t^c\right] \qquad (3.33)\\
&\le \mathbb{P}\left[A_t^c\right] + \mathbb{P}\left[\hat{A}_t^c\right]\\
&= \mathbb{P}\left[\frac{1}{t}\sum_{k=0}^{t-1}(e_k + e_k^+) > e_f + s\right] + \mathbb{P}\left[\frac{1}{t}\sum_{k=0}^{t-1}(\hat{e}_k + \max\{\hat{e}_k^+, \hat{e}_k^-\}) > e_f + s\right]\\
&\le \mathbb{P}\left[\frac{1}{t}\sum_{k=0}^{t-1}(e_k + e_k^+) > 2\epsilon_f + \frac{5}{a} + s\right] + \mathbb{P}\left[\frac{1}{t}\sum_{k=0}^{t-1}(\hat{e}_k + \max\{\hat{e}_k^+, \hat{e}_k^-\}) > 2\epsilon_f + \frac{5}{a} + s\right]\\
&\le 2\exp\left(-\frac{a}{4}st\right)
\end{aligned}$$

(3.32)中的前一个项可分解如下，

$$\begin{aligned}
&\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t,\ A_t \cap \hat{A}_t\right]\\
&= \mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t,\ A_t \cap \hat{A}_t, \sum_{k=0}^{t-1} I_k^f I_k^g \Omega_k^g \Theta_k^g U_k^g + \sum_{k=0}^{t-1} \hat{I}_k^f I_k^H \Omega_k^H \Theta_k^H U_k^H < \frac{1}{2}c_{gH}t - c_\tau\right]\\
&\quad + \mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda} > t,\ A_t \cap \hat{A}_t, \sum_{k=0}^{t-1} I_k^f I_k^g \Omega_k^g \Theta_k^g U_k^g + \sum_{k=0}^{t-1} \hat{I}_k^f I_k^H \Omega_k^H \Theta_k^H U_k^H \ge \frac{1}{2}c_{gH}t - c_\tau\right]
\end{aligned} \tag{3.34}$$

$\hat{}$ 根据引理 3.8，当 $A_t \cap A_t$ 为真且 $N_{\bar{\epsilon}_g、\ \bar{\epsilon}_H、\ \bar{\epsilon}_\lambda} > t \ge T$ 时，可得

$$\sum_{k=0}^{t-1} I_k^f I_k^g \Omega_k^g \Theta_k^g U_k^g + \sum_{k=0}^{t-1} \hat{I}_k^f I_k^H \Omega_k^H \Theta_k^H U_k^H \le \frac{f(x_0) - f(x_t)}{c_{\bar{\alpha},\bar{\beta}}} + \frac{3e_f}{c_{\bar{\alpha},\bar{\beta}}}t$$

$$+\frac{1}{c_{\bar{\alpha},\bar{\beta}}}\sum_{k=0}^{t-1}\left(e_k+e_k^+ +\hat{e}_k+\max\{\hat{e}_k^+,\hat{e}_k^-\}\right)$$
$$\leq \frac{f(x_0)-f^*}{c_{\bar{\alpha},\bar{\beta}}}+\frac{5e_f+2s}{c_{\bar{\alpha},\bar{\beta}}}t<\frac{c_{gH}}{2}t-c_\tau.$$

因此，式（3.34）中的第二项为 0，我们只需对第一项求界，

$$\begin{aligned}&\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda}>t,\ A_t\cap\hat{A}_t,\ \sum_{k=0}^{t-1}I_k^fI_k^g\Omega_k^g\Theta_k^gU_k^g+\sum_{k=0}^{t-1}\hat{I}_k^fI_k^H\Omega_k^H\Theta_k^HU_k^H<\tfrac{1}{2}c_{gH}t-c_\tau\right]\\ \leq&\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda}>t,\ \sum_{k=0}^{t-1}I_k^fI_k^g\Omega_k^g\Theta_k^gU_k^g+\sum_{k=0}^{t-1}\hat{I}_k^fI_k^H\Omega_k^H\Theta_k^HU_k^H<\tfrac{1}{2}c_{gH}t-c_\tau\right].\end{aligned}\tag{3.35}$$

证明的其余部分与有界噪声情形下的证明类似。我们将这些事件记为

$$E_{fg}=\left\{\sum_{k=0}^{t-1}I_k^fI_k^g\geq\bar{p}_{fg}t\right\},\ E_{fH}=\left\{\sum_{k=0}^{t-1}\hat{I}_k^fI_k^H\geq\bar{p}_{fH}t\right\},\ E_{fgH}=\left\{\sum_{k=0}^{t-1}I_k^fI_k^g\hat{I}_k^fI_k^H\geq\bar{p}_{fg}\bar{p}_{fH}t\right\}.$$

那么，式（3.35）中的概率可作如下界定

$$\begin{aligned}&\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda}>t,\ \sum_{k=0}^{t-1}I_k^fI_k^g\Omega_k^g\Theta_k^gU_k^g+\sum_{k=0}^{t-1}\hat{I}_k^fI_k^H\Omega_k^H\Theta_k^HU_k^H<\tfrac{1}{2}c_{gH}t-c_\tau\right]\\ =&\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda}>t,\ \sum_{k=0}^{t-1}I_k^fI_k^g\Omega_k^g\Theta_k^gU_k^g+\sum_{k=0}^{t-1}\hat{I}_k^fI_k^H\Omega_k^H\Theta_k^HU_k^H<\tfrac{1}{2}c_{gH}t-c_\tau,\ E_{fg}\cap E_{fH}\cap E_{fgH}\right]\\ &+\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda}>t,\ \sum_{k=0}^{t-1}I_k^fI_k^g\Omega_k^g\Theta_k^gU_k^g+\sum_{k=0}^{t-1}\hat{I}_k^fI_k^H\Omega_k^H\Theta_k^HU_k^H<\tfrac{1}{2}c_{gH}t-c_\tau,\ E_{fg}^c\cup E_{fH}^c\cup E_{fgH}^c\right]\\ =&0+\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda}>t,\ \sum_{k=0}^{t-1}I_k^fI_k^g\Omega_k^g\Theta_k^gU_k^g+\sum_{k=0}^{t-1}\hat{I}_k^fI_k^H\Omega_k^H\Theta_k^HU_k^H<\tfrac{1}{2}c_{gH}t-c_\tau,\ E_{fg}^c\cup E_{fH}^c\cup E_{fgH}^c\right]\\ \leq&\mathbb{P}\left[E_{fg}^c\cup E_{fH}^c\cup E_{fgH}^c\right]\\ \leq&\mathbb{P}\left[E_{fg}^c\right]+\mathbb{P}\left[E_{fH}^c\right]+\mathbb{P}\left[E_{fgH}^c\right]\\ \leq&\exp\left(-\frac{(p_fp_g-\bar{p}_{fg})^2}{2p_f^2p_g^2}t\right)+\exp\left(-\frac{(p_fp_H-\bar{p}_{fH})^2}{2p_f^2p_H^2}t\right)+\exp\left(-\frac{(p_f^2p_gp_H-\bar{p}_{fg}\bar{p}_{fH})^2}{2p_f^4p_g^2p_H^2}t\right).\end{aligned}$$

其中第四行由引理 3.9 得出，而最后一个不等式由引理2.11得出。结合（3.33）和（3.36），可得

$$\mathbb{P}\left[N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda}>t\right]\leq 2\exp\left(-\tfrac{a}{4}st\right)+\exp\left(-\frac{(p_fp_g-\bar{p}_{fg})^2}{2p_f^2p_g^2}t\right)+\exp\left(-\frac{(p_fp_H-\bar{p}_{fH})^2}{2p_f^2p_H^2}t\right)$$

$$+\exp\left(-\frac{(p_f^2 p_g p_H - \bar{p}_{fg}\bar{p}_{fH})^2}{2p_f^4 p_g^2 p_H^2}t\right).$$

重新整理上述内容即可完成证明。 □

注 3.11。 该定理与有界噪声情况下的定理 *3.5*相类似，但有两处修改，以反映函数预言机的次指数尾部：

- 噪声控制参数 ^{e_f} 变为 $2\epsilon_f + 5/a > 2\epsilon_f$，其中 $a > 0$ 和 $\epsilon_f \geq 0$ 是预言机参数。邻域参数 $\bar{\epsilon}_g$、 $\bar{\epsilon}_H$和 $\bar{\epsilon}_\lambda$ 的下界 e_f $s \geq 0$ $\bar{\epsilon}_g > \mathcal{O}\ \epsilon_f + a^{-1} + s^{\ 1/2} + \epsilon_g$ $\bar{\epsilon}_H >$ 现 $\mathcal{O}\ \epsilon_f + a^{-1} + s^{\ 1/3}\ \epsilon_H$ $\bar{\epsilon}_\lambda > \mathcal{O}\ \epsilon_f + a^{-1} + s^{\ 1/3} + \epsilon_\lambda$ 在取决于 以及一个满足条件的自由参数：(( )) (( ),)(())。因此，算法 *2.1*可达到的二阶精度为 $\big(\mathcal{O}((\epsilon_f + a^{-1} + s)^{1/2} + \epsilon_g)\ \mathcal{O}((\epsilon_f + a^{-1} + s)^{1/3} + \epsilon_H)$, $\mathcal{O}((\epsilon_f + a^-$
- 除了有界噪声分析中出现的三个指数项（由于线搜索中的函数预言机，此处涉及 $p_f p_g$ 和 $p_f p_H$ ）外，还出现了一个额外因子 $2\exp(-\frac{a}{4}st)$，用于刻画函数噪声的次指数尾部。选择 $s > 0$ 会以稍大的邻域为代价，换取失败概率更快衰减。与之前一样，$N_{\bar{\epsilon}_g,\bar{\epsilon}_H,\bar{\epsilon}_\lambda}$ 超过 t的概率，对于 $t \geq T$，会随 t呈指数衰减，其中 $T = \mathcal{O}\Big(\epsilon_f^{-1} + a + s^{-1} + \epsilon_g^{-2} + (\max\{\epsilon_H, \epsilon_\lambda\})^{-3}\Big)$。

4 数值实验

本节报告数值实验，以评估所提出的两步法（算法 2.1，记为 SS2-NC-G）的实际性能。首先，我们考察该方法对关键参数的敏感性，即函数噪声水平 ϵ_f （图 1）和 Armijo 噪声容忍参数 e_f （图 2），随后将其与基于预言机估计的自适应线搜索（记为SS-G） [20] 以及一种共轭梯度变体（记为 SS-NC-CG） [29] （图 3）进行比较。我们使用 Rosenbrock 函数 [19] 作为测试案例，并根据迭代次数和函数评估次数，评估目标函数值、梯度范数、最小海森矩阵特征值和步长等指标。

为模拟预言机输出，我们对精确的函数值、梯度值和海森矩阵值施加受控的有界噪声。这对应于预言机 1。1、预言机 2 （其中 $p_g = 1$ 和$\kappa_g = 0$），以及预言机 3 （其中 $p_H = 1$ 和 $\kappa_H = \kappa_\lambda = 0$。为简化实验设置并使其与理论缩放保持一致，我们施加 $\epsilon_f = \epsilon_g^2 = \epsilon_H^3$。由此得到的扰动模型如下所述。

- **函数估计：** 含噪函数评估通过向精确的目标函数值 $F \leftarrow f(x) + \epsilon_f \mathcal{U}(-1, 1)$ 添加有界随机噪声 ϵ_f生成。
- **梯度估计：**为生成范数有界的梯度扰动，我们首先抽取 $r \sim \mathcal{N}(0, I_n)$，并将其归一化为 $u \leftarrow r/\|r\|_2$，从而得到在单位球面上均匀分布的随机方向。然后我们抽取 $U \sim \mathcal{U}(0, 1)$，并设定半径

$\rho \leftarrow \epsilon_g U^{1/n}$。噪声梯度估计为 $g \leftarrow \nabla f(x) + \rho u$。该构造保证扰动满足 $\|g - \nabla f(x)\|_2 \leq \epsilon_g$。

- **Hessian估计：**同样，我们先抽取一个随机矩阵 $R \sim \mathcal{N}(0,1)^{n\times n}$，并将其归一化为 $u \leftarrow R/\|R\|_2$，从而生成海森矩阵扰动。然后，我们抽取 $U \sim$Uniform$(0,1)$，并设置 $\rho \leftarrow \epsilon_H U^{1/n^2}$。噪声Hessian估计为 $H \leftarrow \nabla^2 f(x) + \rho u$。在此构造下，扰动满足 $\|H - \nabla^2 f(x)\|_2 \leq \epsilon_H$。

对于所有方法，我们均使用相同的步长搜索超参数 $\alpha_0 = \beta_0 = 1$, $\tau = 0.5$,$c_d = c_p = 0.2$，并设置 $c_g = 0$。对于负曲率方法，我们将提前终止阈值设为 $c_H \max\{\bar{\epsilon}_H, \bar{\epsilon}_\lambda\} = 10^{-3}$。所有实验均在 MATLAB R2021b 中进行。

我们首先研究 SS2-NC-G 对噪声水平 $\epsilon_f \in \{10^{-2}, 10^{-3}, 10^{-5}, 10^{-8}, 0$ 的敏感性 $\}$. 如图1所示，噪声幅度决定了收敛邻域的大小。较小的 ϵ_f (例如 0 和 10^{-8}) 会产生更平滑的轨迹，并收敛到更小的邻域，但初始进展较慢；较大的噪声则会带来更大的变化和更大的收敛邻域，初始下降通常也更快。这些趋势与理论缩放 $\epsilon_f = \epsilon_g^2 = \epsilon_H^3$一致；在此条件下，进入高概率区间之前的迭代复杂度为 $\mathcal{O}(\epsilon_f^{-1})$。尽管较高噪声的运行在早期阶段可能进展更快，但较低噪声的运行最终能达到更好的目标函数值和平稳性。等高线图进一步展示了不同噪声水平下轨迹行为的差异。

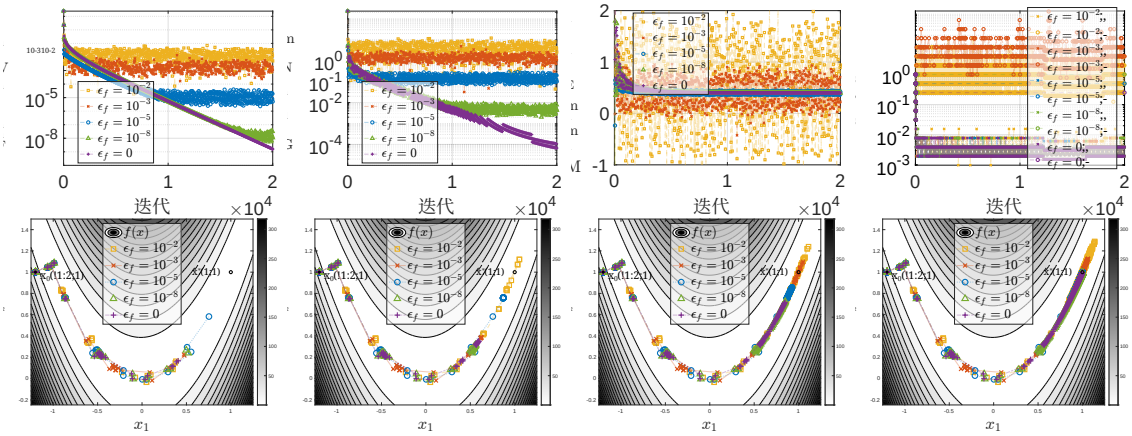

图 1：算法 2.1 在 Rosenbrock问题上的敏感性（针对 $\epsilon_f \in \{10^{-2}, 10^{-3}, 10^{-5}, 10^{-8}, 0\}$，其中 $e_f = 2\epsilon_f$。上排：指标随迭代变化。下排：带有轨迹的等高线图（分别经过 100、200、2000 和 20000 次迭代）。

接下来，我们考察错误设定e_f的影响，该量在含噪函数评估下控制步接受。我们测试 $e_f \in \{0.25\epsilon_f, 2\epsilon_f, 16\epsilon_f, 128\epsilon_f\}$。在有界噪声情形（预言机1.1）中，理论要求 $e_f \geq 2\epsilon_f$ 才能收敛；尽管如此，我们还

考虑更小的取值，以评估噪声被低估情形下的行为。如图2所示，较大的e_f 会收敛到更大的邻域，这与精度界对e_f的依赖相一致，同时允许更大的步长和更快的初期进展。选择 $e_f = 2\epsilon_f$由理论所启发，似乎在早期进展与最终精度之间取得了平衡。

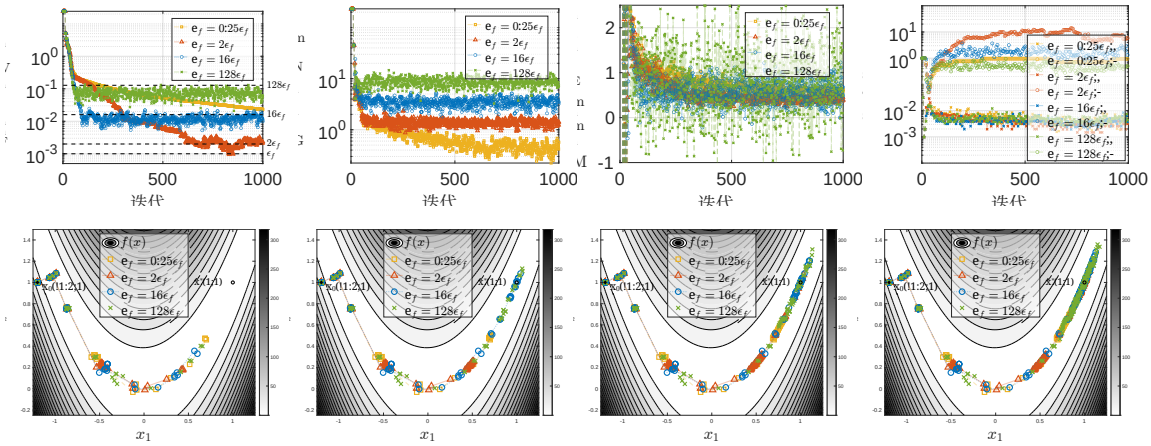

图 2: 算法 2.1 在 Rosenbrock问题上的敏感性，适用于 $\epsilon_f = 10^{-3}$ 和 $e_f \in \{0.25\epsilon_f, 2\epsilon_f, 16\epsilon_f, 128\epsilon_f\}$。结果为 10 次运行的平均值。上排：指标随迭代次数的变化。下排：包含轨迹的等高线图（分别经过 100、200、1000 和 5000 次迭代后）。

在最后的实验中，我们比较SS2-NC-G与一阶方法SS-G和SS-NC-CG。后者遵循 [29] 中框架的简化版本，并计算牛顿型步或负曲率步；由于问题的随机性，步长由步长搜索确定。我们报告目标函数值、梯度范数、最小特征值和步长随迭代和函数评估的变化。Rosenbrock问题在 $\epsilon_f = 10^{-3}$ 和 $e_f = 2\epsilon_f$ 下的结果如图3所示。引入负曲率的方法（SS2-NC-G 和SS-NC-CG）在负曲率区域更有效地降低目标函数，而SS-G 在类鞍点区域附近进展较慢。轨迹图反映了这一行为。

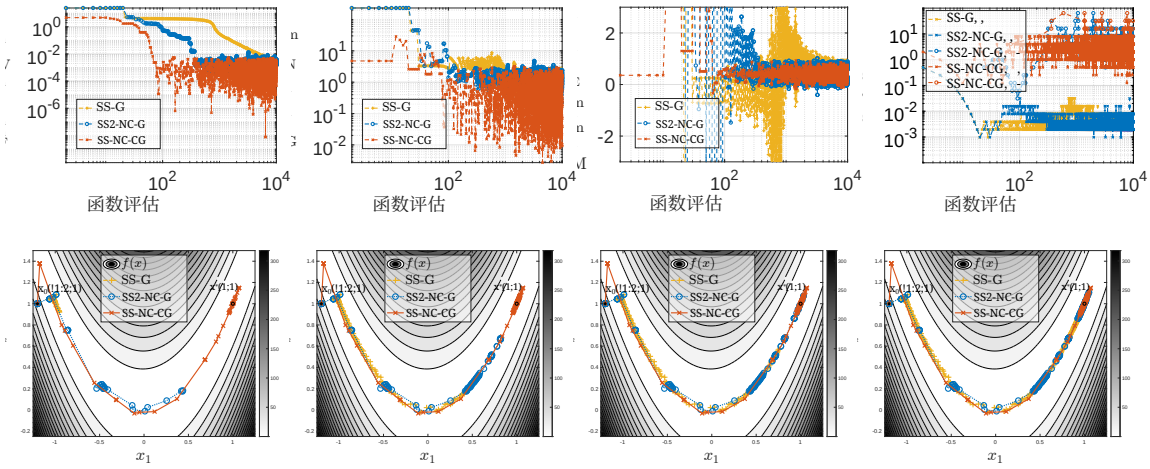

图 3： 算法 2.1 (SS2-NC-G)、 SS-G和 SS-NC-CG 在 Rosenbrock问题上的比较，其中 $\epsilon_f = 10^{-3}$ 和 $e_f = 2\epsilon_f$。上排：函数评估指标。下排：等高线图及迭代点轨迹（经过 100、500、1000 和 5000 次迭代后）。

5 最后总结

我们考虑目标函数及其导数受到噪声干扰、并通过概率预言机访问的优化问题。在该框架内，我们提出了一种两步负曲率方法，其中包含带有放松的Armijo型充分下降条件的步长搜索过程，以及一种用于选择负曲率方向的新机制。在关于概率预言机精度的温和假设下，所提方法具有高概率二阶收敛和迭代复杂度保证。在一个标准非凸问题上的数值实验说明了所提框架的效率和鲁棒性。

参考文献

[1] Zeyuan Allen-Zhu。Natasha 2：比 SGD 更快的非凸优化。神经信息处理系统进展，31，2018。[2] Kazuoki Azuma。若干相依随机变量的加权和。东北数学杂志，第二辑，19(3):357–367，1967。[3] Stefania Bellavia、Eugenio Fabrizi 和 Benedetta Morini。用于含噪凸优化的线搜索牛顿-CG 方法。费拉拉大学年刊，68(2):483–504，2022。[4] Albert S Berahas、Raghu Bollapragada 和 Wanping Dong。结合自适应采样利用负曲率：理论结果与实用算法。计算优化与应用，第 1–37 页，2026。[5] Albert S Berahas、Richard H Byrd 和 Jorge Nocedal。通过拟牛顿方法对含噪函数进行无导数优化。*SIAM* 优化杂志，29(2):965–993，2019。[6] Albert S Berahas、Liyuan Cao 和 Katya Scheinberg。含噪通用线搜索算法的全局收敛速率分析。*SIAM* 优化杂志，31(2):1489–1518，2021。[7] Albert S Berahas、Miaolan Xie 和 Baoyu Zhou。一种用于非线性等式约束随机优化、具有高概率复杂度界的序列二次规划方法。*SIAM* 优化杂志，35(1):240–269，2025。[8] Raghu Bollapragada、Richard Byrd 和 Jorge Nocedal。随机优化的自适应采样策略。*SIAM* 优化杂志，28(4):3312–3343，2018。[9] L´eon Bottou、Frank E Curtis 和 Jorge Nocedal。大规模机器学习的优化方法。*SIAM* 综述，60(2):223–311，2018。[10] Richard H Byrd、Gillian M Chin、Jorge Nocedal 和 Yuchen Wu。机器学习优化方法中的样本量选择。数学规划，134(1):127–155，2012。[11] Liyuan Cao、Albert S Berahas 和 Katya Scheinberg。带含噪预言机的信赖域方法的一阶和二阶高概率复杂度界。数学规划，第 1–52 页，2023。[12] Richard G Carter。关于使用非精确梯度信息的信赖域算法的全局收敛性。*SIAM* 数值分析杂志，28(1):251–265，1991。

[13] Coralia Cartis、Nicholas IM Gould 和 Philippe L Toint。无约束优化的自适应三次正则化方法。第一部分：动机、收敛性与数值结果。数学规划，127(2):245–295，2011。[14] Coralia Cartis 和 Katya Scheinberg。基于概率模型的无约束优化方法的全局收敛速率分析。数学规划，169:337–375，2018。[15] Frank E Curtis 和 Daniel P Robinson。在确定性与随机优化中利用负曲率。数学规划，176(1):69–94，2019。[16] Anders Forsgren、Philip E Gill 和 Walter Murray。使用部分 Cholesky 分解计算修正牛顿方向。*SIAM* 科学计算杂志，16(1):139–150，1995。[17] Michael P Friedlander 和 Mark Schmidt。用于数据拟合的混合确定性-随机方法。*SIAM* 科学计算杂志，34(3):A1380–A1405，2012。[18] Donald Goldfarb。利用负曲率方向的最小化曲线路径步长算法。数学规划，18(1):31–40，1980。[19] Serge Gratton 和 Ph L Toint。用于 MATLAB、Python 和 Julia 的 S2mpj 与 cutest 优化问题。*Optimization Methods and Software*，第 1–33 页，2025。[20] Billy Jin、Katya Scheinberg 和 Miaolan Xie。基于随机预言机的线搜索高概率复杂度界。神经信息处理系统，34:9193–9203，2021。[21] Shuyao Li 和 Stephen J Wright。一种用于非凸最小化的随机化算法，具有非精确评估和复杂度保证。*JOTA*，207(3):66，2025。[22] Garth P McCormick。针对负曲率修正 Armijo 步长规则。数学规划，13(1):111–115，1977。[23] Jorge J Mor´e 和 Danny C Sorensen。关于在修正牛顿法中使用负曲率方向。数学规划，16:1–20，1979。[24] Yurii Nesterov 和 Boris T Polyak。牛顿法的三次正则化及其全局性能。数学规划，108(1):177–205，2006。[25] Jorge Nocedal 和 Stephen J Wright。数值优化。Springer，1999。[26] Courtney Paquette 和 Katya Scheinberg。一种具有期望复杂度分析的随机线搜索方法。*SIAM* 优化杂志，30(1):349–376，2020。

[27] Raghu Pasupathy、Peter Glynn、Soumyadip Ghosh 和 Fatemeh S Hashemi。基于仿真的递归中的采样率。*SIAM* 优化杂志，28(1)：45–73，2018。[28] R Tyrrell Rockafellar、Stanislav Uryasev 等。条件风险价值的优化。风险杂志，2：21–42，2000。[29] Cl´ement W Royer、Michael O'Neill 和 Stephen J Wright。一种具有复杂度保证的光滑无约束优化牛顿共轭梯度算法。数学规划，180：451–488，2020。[30] Cl´ement W Royer 和 Stephen J Wright。光滑非凸优化二阶线搜索算法的复杂度分析。*SIAM* 优化杂志，28(2)：1448–1477，2018。[31] Roman Vershynin。高维概率：介绍及其在数据科学中的应用，第47卷。剑桥大学出版社，2018。[32] Peng Xu、Fred Roosta 和 Michael W Mahoney。非精确海森信息下非凸优化的牛顿型方法。数学规划，184(1)：35–70，2020。