

COMPLEXITY OF PROJECTED GRADIENT METHODS FOR STRONGLY CONVEX OPTIMIZATION WITH HÖLDER CONTINUOUS GRADIENT TERMS*

XIAOJUN CHEN[†], C. T. KELLEY[‡], AND LEI WANG[§]

November 26, 2025

Abstract. This paper studies the complexity of projected gradient descent methods for a class of strongly convex constrained optimization problems where the objective function is expressed as a summation of m component functions, each possessing a gradient that is Hölder continuous with an exponent $\alpha_i \in (0, 1]$. Under this formulation, the gradient of the objective function may fail to be globally Hölder continuous, thereby existing complexity results inapplicable to this class of problems. Our theoretical analysis reveals that, in this setting, the complexity of projected gradient methods is determined by $\hat{\alpha} = \min_{i \in \{1, \dots, m\}} \alpha_i$. We first prove that, with an appropriately fixed stepsize, the complexity bound for finding an approximate minimizer with a distance to the true minimizer less than ε is $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+\hat{\alpha})})$, which extends the well-known complexity result for $\hat{\alpha} = 1$. Next we show that the complexity bound can be improved to $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+3\hat{\alpha})})$ if the stepsize is updated by the universal scheme. We illustrate our complexity results by numerical examples arising from elliptic equations with a non-Lipschitz term.

Key words. projected gradient descent, complexity, Hölder continuity

19 MSC codes. 90C25, 65L05, 65Y20

1. Introduction. Given a closed and convex set $\Omega \subseteq \mathbb{R}^n$, this paper considers the following optimization problem,

$$22 \quad (1.1) \qquad \min_{\mathbf{u} \in \Omega} f(\mathbf{u}) := \frac{1}{m} \sum_{i=1}^m f_i(\mathbf{u}),$$

23 where the objective function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies the following assumption.

ASSUMPTION 1.1.

25 1. The function f is μ -strongly convex with a parameter $\mu > 0$ on Ω , that is,

$$f(\mathbf{u}) \geq f(\mathbf{v}) + \langle \nabla f(\mathbf{v}), \mathbf{u} - \mathbf{v} \rangle + \frac{\mu}{2} \|\mathbf{u} - \mathbf{v}\|^2,$$

for all $\mathbf{u}, \mathbf{v} \in \Omega$.

28 2. For each $i \in [m] := \{1, 2, \dots, m\}$, the function $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously
 29 differentiable and the gradient ∇f_i is (globally) Hölder continuous with an
 30 exponent $\alpha_i \in (0, 1]$ on Ω , namely, there exists a constant $L_i > 0$ such that

$$(1.2) \quad \|\nabla f_i(\mathbf{u}) - \nabla f_i(\mathbf{v})\| \leq L_i \|\mathbf{u} - \mathbf{v}\|^{\alpha_i},$$

for all $\mathbf{u}, \mathbf{v} \in \Omega$.

*Submitted to the editors DATE.

Funding: We would like to acknowledge support for this project from RGC grant JLFS/P-501/24 for the CAS AMSS-PolyU Joint Laboratory in Applied Mathematics and Hong Kong Research Grant Council project PolyU15300024.

[†]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China (maxjchen@polyu.edu.hk).

[†]Department of Mathematics, Box 8205, North Carolina State University, Raleigh, NC 27695-8205, USA (Tim_Kelley@ncsu.edu).

[§]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China
(lei2wang@polyu.edu.hk).

33 Here, $\|\cdot\|$ is the ℓ_2 norm and $\langle \cdot, \cdot \rangle$ is the inner product on \mathbb{R}^n . We also denote by
 34 $\mathbf{u}^* \in \Omega$ and $f^* = f(\mathbf{u}^*)$ the global minimizer and the optimal value of problem (1.1),
 35 respectively.

36 Suppose that each ∇f_i is Lipschitz continuous, which corresponds to condition
 37 (1.2) with $\alpha_i = 1$ for all $\mathbf{u}, \mathbf{v} \in \Omega$. Then ∇f is also Lipschitz continuous and
 38 the associated Lipschitz constant is $L = \sum_{i=1}^m L_i/m$. Let $\Pi_\Omega(\cdot)$ be the projection
 39 operator onto the set Ω . It is well known that the classical projected gradient descent
 40 method

41 (1.3)
$$\mathbf{u}_{k+1} = \Pi_\Omega(\mathbf{u}_k - \tau \nabla f(\mathbf{u}_k)),$$

42 with any initial point $\mathbf{u}_0 \in \mathbb{R}^n$ and the stepsize $\tau \in (0, 2/(\mu + L)]$, achieves a linear
 43 rate of convergence [9, Theorem 2.2.14] as follows,

44
$$\|\mathbf{u}_k - \mathbf{u}^*\| \leq (1 - \mu\tau)^k \|\mathbf{u}_0 - \mathbf{u}^*\|.$$

45 Therefore, for a given $\varepsilon > 0$, method (1.3) is guaranteed to find a point $\mathbf{u}_k \in \Omega$
 46 satisfying $\|\mathbf{u}_k - \mathbf{u}^*\| \leq \varepsilon$ after at most $O(\log(\varepsilon^{-1}))$ iterations. Unfortunately, this
 47 analysis fails if there exists at least one index $i \in [m]$ such that $\alpha_i < 1$. We explain
 48 the failure of the convergence of method (1.3) to \mathbf{u}^* by the following example.

49 *Example 1.1.* [5, Example 1] Consider the following univariate optimization prob-
 50 lem,

51
$$\min_{x \in \mathbb{R}} f(x) = \frac{1}{2}x^2 + \frac{2}{3}|x|^{3/2},$$

52 which is a special instance of problem (1.1) with $f_1(x) = x^2/2$, $f_2(x) = 2|x|^{3/2}/3$, and
 53 $\Omega = \mathbb{R}$. It is easy to see that the global minimizer is $x^* = 0$. Method (1.3) with the
 54 fixed stepsize $\tau > 0$ starting from $x_0 \neq 0$ reads as follows,

55
$$x_{k+1} = x_k - \tau \nabla f(x_k) = (1 - \tau)x_k - \tau \text{sign}(x_k)|x_k|^{1/2},$$

56 where $\text{sign}(x) = 1$ if $x > 0$, 0 if $x = 0$, and -1 otherwise. A straightforward verification
 57 reveals that

58
$$|x_{k+1}|^2 - |x_k|^2 = -\tau(2 - \tau)|x_k|^2 - 2\tau(1 - \tau)|x_k|^{3/2} + \tau^2|x_k|.$$

59 It is evident that, when $|x_k|$ is sufficiently small, the last term in the right-hand side
 60 becomes dominant, resulting in that $|x_{k+1}|^2 - |x_k|^2 \geq 0$. Therefore, the distance to
 61 the global minimizer ceases to decrease once it achieves a certain level.

Moreover, in [5] we show that ∇f is locally Hölder continuous, but not globally
 Hölder continuous. In fact, from

$$\nabla f(x^* + |h|) - \nabla f(x^*) = |h| + |h|^{1/2} = (|h|^{1-\alpha} + |h|^{1/2-\alpha})|h|^\alpha =: (\hat{L}_1(h) + \hat{L}_2(h))|h|^\alpha,$$

62 we have $\hat{L}_1(h) \rightarrow \infty$ when $\alpha = (0, 1)$ and $|h| \rightarrow \infty$, while $\hat{L}_2(h) \rightarrow \infty$ when $\alpha = 1$
 63 and $|h| \rightarrow 0$.

64 This example demonstrates that a function f expressed as a sum of component
 65 functions f_i , each endowed with a Hölder continuous gradient, may itself fail to possess
 66 a Hölder continuous gradient. This phenomenon was revisited by Nesterov (see [10,
 67 Example 1]).

On the other hand, this example satisfies Assumption 1.1 (ii) as

$$|\nabla f_1(x) - \nabla f_1(y)| \leq L_1|x - y| \quad \text{and} \quad |\nabla f_2(x) - \nabla f_2(y)| \leq L_2|x - y|^{1/2}, \quad \forall x, y \in \mathbb{R}$$

68 with $L_1 = L_2 = 1$.

69 Since ∇f may not be globally Hölder continuous, most existing complexity results
70 are inapplicable to problem (1.1). For the special case where $m = 1$, namely, ∇f is
71 globally Hölder continuous with an exponent $\alpha \in (0, 1]$, Devolder et al. [6] presented
72 the following bound for method (1.3),

$$73 \quad f(\hat{\mathbf{u}}_N) - f(\mathbf{u}^*) \leq K(N) := \frac{L_\alpha \|\mathbf{u}_0 - \mathbf{u}^*\|^{1+\alpha}}{1 + \alpha} \left(\frac{2}{N} \right)^{\frac{1+\alpha}{2}},$$

74 where L_α is the Hölder constant and $\hat{\mathbf{u}}_N = \sum_{k=1}^N \mathbf{u}_k / N$. In the strongly convex case,
75 (51) in [6] comes to

$$76 \quad \|\hat{\mathbf{u}}_N - \mathbf{u}^*\|^2 \leq \frac{2}{\mu} K(N),$$

77 which implies that finding an N average of iterations $\hat{\mathbf{u}}_N$ satisfying $\|\hat{\mathbf{u}}_N - \mathbf{u}^*\| \leq \varepsilon$
78 requires $O(\varepsilon^{-4/(1+\alpha)})$ iterations.

79 The contribution of this paper is to provide new complexity results of the pro-
80 jected gradient descent methods for problem (1.1), which are dictated by the parame-
81 ter $\hat{\alpha} = \min_{i \in [m]} \alpha_i \in (0, 1]$. We first show that, with an appropriately fixed stepsize,
82 the complexity bound for finding an iterate with a distance to the global minimizer
83 less than ε is $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+\hat{\alpha})})$, which extends the well-known complexity re-
84 sult for $\hat{\alpha} = 1$. Next, we demonstrate that this complexity bound can be improved to
85 $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+3\hat{\alpha})})$ if the stepsize is updated at each iteration using the uni-
86 versal scheme. Even in the special case where $m = 1$, our complexity bound is at least
87 $O(\varepsilon^{-1})$ lower than (51) in [6]. For example, when $\hat{\alpha} = 1$, our bound is $O(\log(\varepsilon^{-1}))$
88 but (51) in [6] is $O(\varepsilon^{-2})$.

89 Our study is motivated by elliptic equations with a non-Lipschitz term [2, 12],
90 as well as optimization problems with an ℓ_p -norm ($1 < p < 2$) regularization term
91 [1, 4]. We illustrate our complexity results by two numerical examples arising from
92 elliptic equations with a non-Lipschitz term in section 5, after we present complexity
93 of projected gradient methods with fixed stepsizes and updated stepsizes in sections 2
94 to 4, respectively.

95 **2. Vanilla Projected Gradient Descent Method with a Fixed Stepsize.**
96 In this section, we attempt to employ the vanilla projected gradient descent method
97 (1.3) with a fixed stepsize to solve problem (1.1), whose complexity bound is also
98 provided. Example 1.1 illustrates that the projected gradient descent method (1.3)
99 with a fixed stepsize will experience stagnation before reaching the global minimizer.

100 To obtain an approximate solution to problem (1.1), it is necessary to choose
101 a sufficiently small stepsize τ in the projected gradient descent method (1.3), the
102 magnitude of which depends on the desired level of accuracy. Let $M > 0$ be a
103 constant defined as

$$104 \quad (2.1) \quad M = \max_{i \in [m]} \left\{ \left[\frac{2(1 - \alpha_i)}{\mu(1 + \alpha_i)} \right]^{(1-\alpha_i)/(1+\alpha_i)} L_i^{2/(1+\alpha_i)} \right\}.$$

105 We select a specific stepsize $\tau = \varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}/M$ in the projected gradient descent
 106 method, whose complete framework is presented in Algorithm 1. Two sequences $\{\mathbf{v}_k\}$
 107 and $\{\mathbf{u}_k\}$ are maintained in Algorithm 1, where \mathbf{v}_k is generated by the projected
 108 gradient descent method and \mathbf{u}_k corresponds to the iterate achieving the smallest
 109 objective function value among the first k iterations.

Algorithm 1: Projected Gradient Descent Method (PGDM).

Input: $\varepsilon > 0$.

Initialize $\mathbf{u}_0 = \mathbf{v}_0 \in \Omega$.

Choose the stepsize $\tau = \varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}/M$.

for $k = 0, 1, 2, \dots$ **do**

Compute

$$\mathbf{v}_{k+1} = \Pi_{\Omega} (\mathbf{v}_k - \tau \nabla f(\mathbf{v}_k)).$$

Set

$$\mathbf{u}_{k+1} = \begin{cases} \mathbf{v}_{k+1}, & \text{if } f(\mathbf{v}_{k+1}) \leq f(\mathbf{u}_k), \\ \mathbf{u}_k, & \text{otherwise.} \end{cases}$$

Output: \mathbf{u}_{k+1} .

110 Our subsequent analysis is based on the inexact oracle [6] derived from the Hölder
 111 continuity condition of gradients, which is generalized to problem (1.1) and demon-
 112 strated in the following proposition.

113 **PROPOSITION 2.1.** *Suppose that Assumption 1.1 holds. Let $\delta > 0$ and*

$$114 \quad \rho \geq \max_{i \in [m]} \left\{ \left[\frac{1 - \alpha_i}{(1 + \alpha_i)\delta} \right]^{(1-\alpha_i)/(1+\alpha_i)} L_i^{2/(1+\alpha_i)} \right\}.$$

115 Then for all $\mathbf{u}, \mathbf{v} \in \Omega$, we have

$$116 \quad f(\mathbf{v}) \leq f(\mathbf{u}) + \langle \nabla f(\mathbf{u}), \mathbf{v} - \mathbf{u} \rangle + \frac{\rho}{2} \|\mathbf{v} - \mathbf{u}\|^2 + \frac{\delta}{2}.$$

117 *Proof.* Since ∇f_i is Hölder continuous with an exponent α_i , we can obtain that

$$118 \quad f_i(\mathbf{v}) \leq f_i(\mathbf{u}) + \langle \nabla f_i(\mathbf{u}), \mathbf{v} - \mathbf{u} \rangle + \frac{L_i}{1 + \alpha_i} \|\mathbf{v} - \mathbf{u}\|^{1+\alpha_i},$$

119 for all $\mathbf{u}, \mathbf{v} \in \Omega$. Then, for each i , it follows from [8, Lemma 2] that

$$120 \quad f_i(\mathbf{v}) \leq f_i(\mathbf{u}) + \langle \nabla f_i(\mathbf{u}), \mathbf{v} - \mathbf{u} \rangle + \frac{\rho}{2} \|\mathbf{v} - \mathbf{u}\|^2 + \frac{\delta}{2}.$$

121 Summing the above relationship over $i \in [m]$, we immediately arrive at the assertion
 122 of this proposition. The proof is completed. \square

123 Now, we are in the position to derive the complexity bound of Algorithm 1 in the
 124 following theorem.

125 THEOREM 2.2. Let $\varepsilon \in (0, 1)$ be a sufficiently small constant. Then after at most

126

$$O\left(\log\left(\frac{1}{\varepsilon}\right) \frac{1}{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}}\right)$$

127 iterations, Algorithm 1 will find an iterate $\mathbf{u}_k \in \Omega$ satisfying $\|\mathbf{u}_k - \mathbf{u}^*\| \leq \varepsilon$.

128 *Proof.* In view of Proposition 2.1, we take

129

$$\rho = \frac{1}{\tau} = \frac{M}{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}} \geq \max_{i \in [m]} \left\{ \left[\frac{2(1-\alpha_i)}{\mu(1+\alpha_i)\varepsilon^2} \right]^{(1-\alpha_i)/(1+\alpha_i)} L_i^{2/(1+\alpha_i)} \right\}.$$

130 Then it holds that

131

$$f(\mathbf{v}_{k+1}) \leq f(\mathbf{v}_k) + \langle \nabla f(\mathbf{v}_k), \mathbf{v}_{k+1} - \mathbf{v}_k \rangle + \frac{1}{2\tau} \|\mathbf{v}_{k+1} - \mathbf{v}_k\|^2 + \frac{\mu\varepsilon^2}{4},$$

132 which, after a suitable rearrangement, can be equivalently written as

133 (2.2)

$$\langle \nabla f(\mathbf{v}_k), \mathbf{v}_k - \mathbf{v}_{k+1} \rangle \leq f(\mathbf{v}_k) - f(\mathbf{v}_{k+1}) + \frac{\mu\varepsilon^2}{4} + \frac{1}{2\tau} \|\mathbf{v}_{k+1} - \mathbf{v}_k\|^2.$$

134 Recall that $f^* = f(\mathbf{u}^*)$. By virtue of the strong convexity of f , we can obtain that

135 (2.3)

$$\langle \nabla f(\mathbf{v}_k), \mathbf{u}^* - \mathbf{v}_k \rangle \leq f^* - f(\mathbf{v}_k) - \frac{\mu}{2} \|\mathbf{v}_k - \mathbf{u}^*\|^2.$$

136 The optimality condition of the projection problem defining \mathbf{v}_{k+1} yields that

137

$$\langle \mathbf{v}_{k+1} - \mathbf{v}_k + \tau \nabla f(\mathbf{v}_k), \mathbf{u} - \mathbf{v}_{k+1} \rangle \geq 0,$$

138 for all $\mathbf{u} \in \Omega$. Upon taking $\mathbf{u} = \mathbf{u}^*$, we have

139

$$\begin{aligned} \langle \mathbf{v}_{k+1} - \mathbf{v}_k, \mathbf{v}_{k+1} - \mathbf{u}^* \rangle &\leq \tau \langle \nabla f(\mathbf{v}_k), \mathbf{u}^* - \mathbf{v}_{k+1} \rangle \\ &= \tau \langle \nabla f(\mathbf{v}_k), \mathbf{u}^* - \mathbf{v}_k \rangle + \tau \langle \nabla f(\mathbf{v}_k), \mathbf{v}_k - \mathbf{v}_{k+1} \rangle, \end{aligned}$$

140 which together with (2.2) and (2.3) implies that

141

$$\begin{aligned} \langle \mathbf{v}_{k+1} - \mathbf{v}_k, \mathbf{v}_{k+1} - \mathbf{u}^* \rangle &\leq \tau \left(f^* - f(\mathbf{v}_{k+1}) + \frac{\mu\varepsilon^2}{4} \right) - \frac{\mu\tau}{2} \|\mathbf{v}_k - \mathbf{u}^*\|^2 \\ &\quad + \frac{1}{2} \|\mathbf{v}_{k+1} - \mathbf{v}_k\|^2. \end{aligned}$$

142 Moreover, it can be readily verified that

143 (2.4)

$$\begin{aligned} \|\mathbf{v}_{k+1} - \mathbf{u}^*\|^2 &= \|\mathbf{v}_{k+1} - \mathbf{v}_k + \mathbf{v}_k - \mathbf{u}^*\|^2 \\ &= \|\mathbf{v}_k - \mathbf{u}^*\|^2 + 2 \langle \mathbf{v}_{k+1} - \mathbf{v}_k, \mathbf{v}_k - \mathbf{u}^* \rangle + \|\mathbf{v}_{k+1} - \mathbf{v}_k\|^2 \\ &= \|\mathbf{v}_k - \mathbf{u}^*\|^2 + 2 \langle \mathbf{v}_{k+1} - \mathbf{v}_k, \mathbf{v}_{k+1} - \mathbf{u}^* \rangle - \|\mathbf{v}_{k+1} - \mathbf{v}_k\|^2. \end{aligned}$$

144 Collecting the above two relationships together, we arrive at

145

$$\|\mathbf{v}_{k+1} - \mathbf{u}^*\|^2 \leq (1 - \mu\tau) \|\mathbf{v}_k - \mathbf{u}^*\|^2 + 2\tau \left(f^* - f(\mathbf{v}_{k+1}) + \frac{\mu\varepsilon^2}{4} \right).$$

146 From the construction of \mathbf{u}_k in Algorithm 1, it then follows that $f(\mathbf{v}_l) \geq f(\mathbf{u}_k)$ for
 147 all $l \in \{1, 2, \dots, k\}$. Let $C_k = \sum_{l=1}^k (1 - \mu\tau)^{l-1}$ be a constant. Applying the above
 148 relationship recursively for k times leads to that

$$\begin{aligned} 149 \quad \|\mathbf{v}_k - \mathbf{u}^*\|^2 &\leq (1 - \mu\tau)^k \|\mathbf{u}_0 - \mathbf{u}^*\|^2 + 2\tau \sum_{l=1}^k (1 - \mu\tau)^{l-1} \left(f^* - f(\mathbf{v}_l) + \frac{\mu\varepsilon^2}{4} \right) \\ &\leq (1 - \mu\tau)^k \|\mathbf{u}_0 - \mathbf{u}^*\|^2 + 2\tau \left(f^* - f(\mathbf{u}_k) + \frac{\mu\varepsilon^2}{4} \right) C_k, \end{aligned}$$

150 which together with $\|\mathbf{v}_k - \mathbf{u}^*\| \geq 0$ and $C_k \geq 1$ implies that

$$151 \quad f(\mathbf{u}_k) - f^* \leq \frac{(1 - \mu\tau)^k}{2\tau C_k} \|\mathbf{u}_0 - \mathbf{u}^*\|^2 + \frac{\mu\varepsilon^2}{4} \leq \frac{(1 - \mu\tau)^k}{2\tau} \|\mathbf{u}_0 - \mathbf{u}^*\|^2 + \frac{\mu\varepsilon^2}{4}.$$

152 According to the strong convexity of f and the optimality condition of problem (1.1),
 153 we have

$$154 \quad (2.5) \quad f(\mathbf{u}_k) - f^* \geq \langle \nabla f(\mathbf{u}^*), \mathbf{u}_k - \mathbf{u}^* \rangle + \frac{\mu}{2} \|\mathbf{u}_k - \mathbf{u}^*\|^2 \geq \frac{\mu}{2} \|\mathbf{u}_k - \mathbf{u}^*\|^2.$$

155 Hence, it holds that

$$\begin{aligned} 156 \quad \|\mathbf{u}_k - \mathbf{u}^*\|^2 &\leq \frac{2}{\mu} (f(\mathbf{u}_k) - f^*) \leq \frac{(1 - \mu\tau)^k}{\mu\tau} \|\mathbf{u}_0 - \mathbf{u}^*\|^2 + \frac{\varepsilon^2}{2} \\ &\leq \frac{M \|\mathbf{u}_0 - \mathbf{u}^*\|^2}{\mu\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}} \left(1 - \frac{\mu}{M} \varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})} \right)^k + \frac{\varepsilon^2}{2}. \end{aligned}$$

157 We denote by K_ε^* the smallest iteration number k such that $\|\mathbf{u}_k - \mathbf{u}^*\| \leq \varepsilon$. Then
 158 solving the inequality $M \|\mathbf{u}_0 - \mathbf{u}^*\|^2 \varepsilon^{-2(1-\hat{\alpha})/(1+\hat{\alpha})} (1 - \mu\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}/M)^k / \mu \leq \varepsilon^2/2$
 159 indicates that

$$\begin{aligned} 160 \quad K_\varepsilon^* &\leq \frac{4 \log((2M \|\mathbf{u}_0 - \mathbf{u}^*\|^2 / \mu)^{(1+\hat{\alpha})/4} / \varepsilon)}{-\log(1 - \mu\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}/M)(1 + \hat{\alpha})} \\ &\leq \frac{4M \log((2M \|\mathbf{u}_0 - \mathbf{u}^*\|^2 / \mu)^{(1+\hat{\alpha})/4} / \varepsilon)}{\mu(1 + \hat{\alpha})\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}}. \end{aligned}$$

161 The proof is completed. \square

162 Theorem 2.2 demonstrates that the iteration complexity of Algorithm 1 with a
 163 fixed stepsize is $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+\hat{\alpha})})$ for problem (1.1). This complexity result
 164 generalizes the classical linear convergence when $\hat{\alpha} = 1$, which highlights the perfor-
 165 mance degradation incurred by non-Lipschitz gradients.

166 **3. Universal Primal Gradient Method.** The fixed stepsize τ chosen in Algo-
 167 rithm 1 depends on the parameters α_i and L_i for all $i \in [m]$, which are often unknown
 168 and hard to estimate in practice. To address this issue, we adopt the universal pri-
 169 mal gradient method (UPGM) proposed by Nesterov [8] to solve problem (1.1). This
 170 method incorporates a line-search procedure to adaptively determine the stepsize at
 171 each iteration, and its overall framework is outlined in Algorithm 2.

172 Next, we establish the iteration complexity of Algorithm 2, which remains on the
 173 same order as that of the projected gradient descent method with a fixed stepsize.

Algorithm 2: Universal Primal Gradient Method (UPGM).**Input:** $\varepsilon > 0$.Initialize $\mathbf{u}_0 = \mathbf{v}_0 \in \Omega$ and $\rho_0 > 0$.**for** $k = 0, 1, 2, \dots$ **do** **for** $j_k = 0, 1, 2, \dots$ **do**

Compute

$$\mathbf{v}_{k+1} = \Pi_{\Omega} \left(\mathbf{v}_k - \frac{1}{2^{j_k} \rho_k} \nabla f(\mathbf{v}_k) \right).$$

If \mathbf{v}_{k+1} satisfies the following line-search condition,

$$(3.1) \quad \begin{aligned} f(\mathbf{v}_{k+1}) &\leq f(\mathbf{v}_k) + \langle \nabla f(\mathbf{v}_k), \mathbf{v}_{k+1} - \mathbf{v}_k \rangle \\ &+ \frac{2^{j_k} \rho_k}{2} \|\mathbf{v}_{k+1} - \mathbf{v}_k\|^2 + \frac{\mu \varepsilon^2}{4}, \end{aligned}$$

then break. Update $\rho_{k+1} = 2^{j_k} \rho_k$.

Set

$$\mathbf{u}_{k+1} = \begin{cases} \mathbf{v}_{k+1}, & \text{if } f(\mathbf{v}_{k+1}) \leq f(\mathbf{u}_k), \\ \mathbf{u}_k, & \text{otherwise.} \end{cases}$$

Output: \mathbf{u}_{k+1} .174 THEOREM 3.1. Let $\varepsilon \in (0, 1)$ be a sufficiently small constant. Then after at most

175
$$O \left(\log \left(\frac{1}{\varepsilon} \right) \frac{1}{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}} \right)$$

176 iterations, Algorithm 2 will attain an iterate $\mathbf{u}_k \in \Omega$ satisfying that $\|\mathbf{u}_k - \mathbf{u}^*\| \leq \varepsilon$.177 *Proof.* Obviously, there exists $j_k \in \mathbb{N}$ such that

178
$$2^{j_k} \rho_k \geq \max_{i \in [m]} \left\{ \left[\frac{2(1-\alpha_i)}{\mu(1+\alpha_i)\varepsilon^2} \right]^{(1-\alpha_i)/(1+\alpha_i)} L_i^{2/(1+\alpha_i)} \right\}.$$

179 By invoking the results of Proposition 2.1, we know that condition (3.1) is satisfied.
180 Hence, the line-search step in Algorithm 2 can be terminated after a finite number of
181 trials and the required number of trials j_k satisfies

182 (3.2)
$$2^{j_k} \rho_k \leq 2 \max_{i \in [m]} \left\{ \left[\frac{2(1-\alpha_i)}{\mu(1+\alpha_i)\varepsilon^2} \right]^{(1-\alpha_i)/(1+\alpha_i)} L_i^{2/(1+\alpha_i)} \right\} \leq \frac{2M}{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}},$$

183 where $M > 0$ is a constant defined in (2.1). Moreover, the line-search condition (3.1)
184 directly yields that

185 (3.3)
$$\langle \nabla f(\mathbf{v}_k), \mathbf{v}_k - \mathbf{v}_{k+1} \rangle \leq f(\mathbf{v}_k) - f(\mathbf{v}_{k+1}) + \frac{2^{j_k} \rho_k}{2} \|\mathbf{v}_{k+1} - \mathbf{v}_k\|^2 + \frac{\mu \varepsilon^2}{4}.$$

186 According to the optimality condition of the projection problem defining \mathbf{v}_{k+1} , we
 187 have

$$188 \quad \left\langle \mathbf{v}_{k+1} - \mathbf{v}_k + \frac{1}{2^{j_k} \rho_k} \nabla f(\mathbf{v}_k), \mathbf{u}^* - \mathbf{v}_{k+1} \right\rangle \geq 0,$$

189 which further implies that

$$190 \quad \begin{aligned} \langle \mathbf{v}_{k+1} - \mathbf{v}_k, \mathbf{v}_{k+1} - \mathbf{u}^* \rangle &\leq \frac{1}{2^{j_k} \rho_k} \langle \nabla f(\mathbf{v}_k), \mathbf{u}^* - \mathbf{v}_{k+1} \rangle \\ &\leq \frac{1}{2^{j_k} \rho_k} \langle \nabla f(\mathbf{v}_k), \mathbf{u}^* - \mathbf{v}_k \rangle + \frac{1}{2^{j_k} \rho_k} \langle \nabla f(\mathbf{v}_k), \mathbf{v}_k - \mathbf{v}_{k+1} \rangle. \end{aligned}$$

191 Substituting (2.3) and (3.3) into the above relationship leads to that

$$192 \quad \begin{aligned} \langle \mathbf{v}_{k+1} - \mathbf{v}_k, \mathbf{v}_{k+1} - \mathbf{u}^* \rangle &\leq \frac{1}{2^{j_k} \rho_k} \left(f^* - f(\mathbf{v}_{k+1}) + \frac{\mu \varepsilon^2}{4} \right) \\ &\quad + \frac{1}{2} \|\mathbf{v}_{k+1} - \mathbf{v}_k\|^2 - \frac{\mu}{2^{j_k+1} \rho_k} \|\mathbf{v}_k - \mathbf{u}^*\|^2, \end{aligned}$$

193 Thus, it follows from relationship (2.4) that

$$194 \quad \begin{aligned} \|\mathbf{v}_{k+1} - \mathbf{u}^*\|^2 &\leq \left(1 - \frac{\mu}{2^{j_k} \rho_k} \right) \|\mathbf{v}_k - \mathbf{u}^*\|^2 + \frac{2}{2^{j_k} \rho_k} \left(f^* - f(\mathbf{v}_{k+1}) + \frac{\mu \varepsilon^2}{4} \right) \\ &\leq \left(1 - \frac{\mu \varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}}{2M} \right) \|\mathbf{v}_k - \mathbf{u}^*\|^2 + \frac{2}{\rho_0} \left(f^* - f(\mathbf{v}_{k+1}) + \frac{\mu \varepsilon^2}{4} \right), \end{aligned}$$

195 where the last inequality comes from (3.2) and $2^{j_k} \rho_k \geq \rho_0$. The remaining part of
 196 the proof follows the same line of reasoning as that of Theorem 2.2 and is therefore
 197 omitted here for the sake of brevity. \square

198 We end this section by estimating the total number of line-search steps required
 199 by Algorithm 2.

200 COROLLARY 3.2. *Let $\varepsilon \in (0, 1)$ be a sufficiently small constant. Then Algorithm 2
 201 requires at most*

$$202 \quad O \left(\log \left(\frac{1}{\varepsilon} \right) \frac{1}{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}} \right)$$

203 line-search steps for the generated sequence $\{\mathbf{u}_k\}$ to satisfy $\|\mathbf{u}_k - \mathbf{u}^*\| \leq \varepsilon$.

204 Proof. Let N_k be the total number of line-search steps after k iterations in Algo-
 205 rithm 2. From the update rule $\rho_{k+1} = 2^{j_k} \rho_k$, we can obtain that $j_k = \log \rho_{k+1} - \log \rho_k$.
 206 Then a straightforward verification reveals that

$$207 \quad (3.4) \quad N_k = \sum_{l=0}^k (j_l + 1) = k + 1 + \log \rho_{k+1} - \log \rho_0,$$

208 which together with relationship (3.2) implies that

$$209 \quad \begin{aligned} N_k &\leq k + 1 + \log \left(\frac{2M}{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}} \right) - \log \rho_0 \\ &\leq k + \frac{2(1-\hat{\alpha})}{1+\hat{\alpha}} \log \left(\frac{1}{\varepsilon} \right) + \log \left(\frac{2M}{\rho_0} \right) + 1. \end{aligned}$$

210 By invoking the results of Theorem 3.1, we conclude that Algorithm 2 requires at
 211 most $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+\hat{\alpha})})$ line-search steps, which completes the proof. \square

212 At each iteration of Algorithm 2, we evaluate both the function value and the
 213 gradient at \mathbf{v}_k . In addition, an extra function evaluation at \mathbf{v}_{k+1,j_k} is involved during
 214 each line-search step. Therefore, Theorem 3.1 and Corollary 3.2 together reveal that
 215 the total number of function and gradient evaluations required by Algorithm 2 is
 216 $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+\hat{\alpha})})$.

217 **4. Universal Fast Gradient Method.** To obtain a sharper complexity bound,
 218 we devise in this section a universal fast gradient method (UFGM) tailored to prob-
 219 lem (1.1). The proposed scheme, summarized in Algorithm 3, exhibits slight but
 220 essential differences from the algorithm introduced by Nesterov [8] to exploit the
 221 strong convexity of the objective function.

Algorithm 3: Universal Fast Gradient Method (UFGM).

Input: $\varepsilon > 0$.

Initialize $\mathbf{u}_0 = \mathbf{w}_0 \in \Omega$, $\rho_0 \geq \mu$, and $\sigma_0 = 1$.

for $k = 0, 1, 2, \dots$ **do**

for $j_k = 0, 1, 2, \dots$ **do**

Set $\nu_k = \sqrt{\mu/(2^{j_k}\rho_k)}$ and $\eta_k = \nu_k/(1 + \nu_k)$.
 Compute

$$(4.1) \quad \mathbf{v}_k = (1 - \eta_k)\mathbf{u}_k + \eta_k \Pi_{\Omega}(\mathbf{w}_k),$$

and

$$(4.2) \quad \mathbf{z}_k = \Pi_{\Omega} \left(\Pi_{\Omega}(\mathbf{w}_k) - \frac{\nu_k}{\mu} \nabla f(\mathbf{v}_k) \right).$$

Set

$$(4.3) \quad \mathbf{u}_{k+1} = (1 - \eta_k)\mathbf{u}_k + \eta_k \mathbf{z}_k.$$

If \mathbf{u}_{k+1} satisfies the following line-search condition,

$$(4.4) \quad \begin{aligned} f(\mathbf{u}_{k+1}) &\leq f(\mathbf{v}_k) + \langle \nabla f(\mathbf{v}_k), \mathbf{u}_{k+1} - \mathbf{v}_k \rangle \\ &+ \frac{\mu}{2\nu_k^2} \|\mathbf{u}_{k+1} - \mathbf{v}_k\|^2 + \frac{\eta_k \mu \varepsilon^2}{4}, \end{aligned}$$

then break.

Set $\rho_{k+1} = 2^{j_k} \rho_k$ and update \mathbf{w}_{k+1} by

$$(4.5) \quad \mathbf{w}_{k+1} = (1 - \eta_k)\mathbf{w}_k + \eta_k \mathbf{v}_k - \frac{\eta_k}{\mu} \nabla f(\mathbf{v}_k).$$

Output: \mathbf{u}_{k+1} .

222 The following lemma illustrates that the line-search process in (4.4) is well-defined,
 223 which is guaranteed to terminate in a finite number of trials.

224 LEMMA 4.1. *There exists an integer $j_k \in \mathbb{N}$ such that the line-search condition
 225 (4.4) is satisfied in Algorithm 3.*

226 *Proof.* It follows from the definition of η_k and $\nu_k \leq 1$ that

$$227 \quad \eta_k = \frac{\nu_k}{1 + \nu_k} \geq \frac{\nu_k}{2}, \quad \text{and} \quad \frac{\mu}{\nu_k^2} = 2^{j_k} \rho_k.$$

228 Recall that $\hat{\alpha} = \min_{i \in [m]} \alpha_i \in (0, 1]$. Then we have

$$\begin{aligned} 229 \quad \frac{\mu}{\nu_k^2} \eta_k^{(1-\hat{\alpha})/(1+\hat{\alpha})} &\geq \frac{2^{j_k} \rho_k}{2^{(1-\hat{\alpha})/(1+\hat{\alpha})}} \nu_k^{(1-\hat{\alpha})/(1+\hat{\alpha})} \\ &= \frac{2^{j_k} \rho_k}{2^{(1-\hat{\alpha})/(1+\hat{\alpha})}} \left[\frac{\mu}{2^{j_k} \rho_k} \right]^{(1-\hat{\alpha})/(2(1+\hat{\alpha}))} \\ &= \frac{\mu^{(1-\hat{\alpha})/(2(1+\hat{\alpha}))}}{2^{(1-\hat{\alpha})/(1+\hat{\alpha})}} [2^{j_k} \rho_k]^{(1+3\hat{\alpha})/(2(1+\hat{\alpha}))}, \end{aligned}$$

230 where the first equality comes from the definition of ν_k . Now it is clear that

$$231 \quad \frac{\mu}{\nu_k^2} \eta_k^{(1-\hat{\alpha})/(1+\hat{\alpha})} \rightarrow \infty,$$

232 as $j_k \rightarrow \infty$. Thus, there exists $j_k \in \mathbb{N}$ such that

$$233 \quad (4.6) \quad \frac{\mu}{\nu_k^2} \eta_k^{(1-\hat{\alpha})/(1+\hat{\alpha})} \geq \max_{i \in [m]} \left\{ \left[\frac{2(1 - \alpha_i)}{\mu(1 + \alpha_i)\varepsilon^2} \right]^{(1-\alpha_i)/(1+\alpha_i)} L_i^{2/(1+\alpha_i)} \right\},$$

234 which further implies that

$$\begin{aligned} 235 \quad \frac{\mu}{\nu_k^2} &\geq \frac{1}{\eta_k^{(1-\hat{\alpha})/(1+\hat{\alpha})}} \max_{i \in [m]} \left\{ \left[\frac{2(1 - \alpha_i)}{\mu(1 + \alpha_i)\varepsilon^2} \right]^{(1-\alpha_i)/(1+\alpha_i)} L_i^{2/(1+\alpha_i)} \right\} \\ &\geq \max_{i \in [m]} \left\{ \left[\frac{2(1 - \alpha_i)}{\eta_k \mu(1 + \alpha_i)\varepsilon^2} \right]^{(1-\alpha_i)/(1+\alpha_i)} L_i^{2/(1+\alpha_i)} \right\}. \end{aligned}$$

236 As a direct consequence of Proposition 2.1, we can proceed to show that the line-search
237 condition (4.4) is satisfied, which completes the proof. \square

238 **LEMMA 4.2.** *Let $\{\sigma_k\}$ be a sequence of positive constants defined recursively by*

$$239 \quad (4.7) \quad \sigma_{k+1} = (1 + \nu_k) \sigma_k,$$

240 *with $\sigma_0 = 1$. And let $\{\phi_k\}$ be a sequence of functions defined recursively by*

$$\begin{aligned} 241 \quad (4.8) \quad \phi_{k+1}(\mathbf{u}) &= \phi_k(\mathbf{u}) - \nu_k \sigma_k f^* + \nu_k \sigma_k f(\mathbf{v}_k) + \nu_k \sigma_k \langle \nabla f(\mathbf{v}_k), \mathbf{u} - \mathbf{v}_k \rangle \\ &\quad + \frac{\nu_k \sigma_k \mu}{2} \|\mathbf{u} - \mathbf{v}_k\|^2, \end{aligned}$$

242 *with $\phi_0(\mathbf{u}) = c_0 + \sigma_0 \mu \|\mathbf{u} - \mathbf{w}_0\|^2 / 2$ for $c_0 = f(\mathbf{u}_0) - f^* - \mu \varepsilon^2 / 4$ and $\mathbf{w}_0 \in \Omega$. Then,
243 for all $k \in \mathbb{N}$, the function ϕ_k preserves the following canonical form,*

$$244 \quad (4.9) \quad \phi_k(\mathbf{u}) = c_k + \frac{\sigma_k \mu}{2} \|\mathbf{u} - \mathbf{w}_k\|^2,$$

245 *where $\{c_k\}$ is a sequence of real numbers and $\{\mathbf{w}_k\}$ is defined recursively by (4.5).*

246 *Proof.* We first prove that $\nabla^2\phi_k = \sigma_k\mu I$ for all $k \in \mathbb{N}$ by induction. It is evident
 247 that $\nabla^2\phi_0 = \sigma_0\mu I$. Now we assume that $\nabla^2\phi_k = \sigma_k\mu I$ for some k . Then relationships
 248 (4.7) and (4.8) imply that

$$249 \quad \nabla^2\phi_{k+1} = \nabla^2\phi_k + \nu_k\sigma_k\mu I = \sigma_k\mu I + \nu_k\sigma_k\mu I = \sigma_{k+1}\mu I.$$

250 Thus, we know that $\nabla^2\phi_k = \sigma_k\mu I$ for all $k \in \mathbb{N}$, which, in turn, justifies the canonical
 251 form of ϕ_k in (4.9).

252 Next, by combining two relationships (4.8) and (4.9) together, we can obtain that

$$253 \quad \begin{aligned} \phi_{k+1}(\mathbf{u}) &= c_k + \frac{\sigma_k\mu}{2} \|\mathbf{u} - \mathbf{w}_k\|^2 - \nu_k\sigma_k f^* + \nu_k\sigma_k f(\mathbf{v}_k) \\ &\quad + \nu_k\sigma_k \langle \nabla f(\mathbf{v}_k), \mathbf{u} - \mathbf{v}_k \rangle + \frac{\nu_k\sigma_k\mu}{2} \|\mathbf{u} - \mathbf{v}_k\|^2. \end{aligned}$$

254 Since \mathbf{w}_{k+1} is a global minimizer of ϕ_{k+1} over \mathbb{R}^n , the first-order optimality condition
 255 yields that

$$256 \quad \begin{aligned} 0 &= \nabla\phi_{k+1}(\mathbf{w}_{k+1}) = \sigma_k\mu(\mathbf{w}_{k+1} - \mathbf{w}_k) + \nu_k\sigma_k\nabla f(\mathbf{v}_k) + \nu_k\sigma_k\mu(\mathbf{w}_{k+1} - \mathbf{v}_k) \\ &= (1 + \nu_k)\sigma_k\mu\mathbf{w}_{k+1} - \sigma_k\mu\mathbf{w}_k - \nu_k\sigma_k\mu\mathbf{v}_k + \nu_k\sigma_k\nabla f(\mathbf{v}_k), \end{aligned}$$

257 from which the closed-form expression of \mathbf{w}_{k+1} in (4.5) can be derived. The proof is
 258 completed. \square

259 LEMMA 4.3. *Let σ_k and $\{\phi_k\}$ be the sequences defined in Lemma 4.2. Then we
 260 have*

$$261 \quad (4.10) \quad \phi_k(\mathbf{u}) \leq \sigma_k(f(\mathbf{u}) - f^*) + \phi_0(\mathbf{u}),$$

262 for all $\mathbf{u} \in \Omega$ and $k \in \mathbb{N}$.

263 *Proof.* We prove that $\{\phi_k\}$ and $\{\sigma_k\}$ satisfy relationship (4.10) by induction. It
 264 is obvious that (4.10) holds for $k = 0$ since $f(\mathbf{u}) \geq f^*$ for any $\mathbf{u} \in \Omega$. Now we assume
 265 that (4.10) holds for some $k \in \mathbb{N}$. It follows from the strong convexity of f that

$$266 \quad f(\mathbf{u}) \geq f(\mathbf{v}_k) + \langle \nabla f(\mathbf{v}_k), \mathbf{u} - \mathbf{v}_k \rangle + \frac{\mu}{2} \|\mathbf{u} - \mathbf{v}_k\|^2,$$

267 for all $\mathbf{u} \in \Omega$. Then substituting the above relationship into (4.8) leads to that

$$268 \quad \begin{aligned} \phi_{k+1}(\mathbf{u}) &\leq \phi_k(\mathbf{u}) - \nu_k\sigma_k f^* + \nu_k\sigma_k f(\mathbf{u}) \\ &\leq \sigma_k(f(\mathbf{u}) - f^*) + \phi_0(\mathbf{u}) + \nu_k\sigma_k(f(\mathbf{u}) - f^*) \\ &= \sigma_{k+1}(f(\mathbf{u}) - f^*) + \phi_0(\mathbf{u}), \end{aligned}$$

269 which indicates that (4.10) also holds for $k + 1$. We complete the proof. \square

270 LEMMA 4.4. *Let $\{\sigma_k\}$ and $\{\phi_k\}$ be the sequences defined in Lemma 4.2. Then,
 271 for the sequence $\{\mathbf{u}_k\}$ generated by Algorithm 3, it holds that*

$$272 \quad (4.11) \quad \sigma_k \left(f(\mathbf{u}_k) - f^* - \frac{\mu\varepsilon^2}{4} \right) \leq \phi_k^* := \min_{\mathbf{u} \in \Omega} \phi_k(\mathbf{u}),$$

273 for all $k \in \mathbb{N}$.

274 *Proof.* We aim to prove the assertion of this lemma by induction. It is clear that
 275 (4.11) holds for $k = 0$ since $\sigma_0 = 1$ and $\phi_0^* = \phi_0(\mathbf{w}_0) = f(\mathbf{u}_0) - f^* - \mu\varepsilon^2/4$. Now we
 276 assume that (4.11) holds for some $k \in \mathbb{N}$ and investigate the situation for $k + 1$.

277 From the canonical form (4.9), it follows that ϕ_k is a strongly convex function
 278 and $\Pi_\Omega(\mathbf{w}_k) = \arg \min_{\mathbf{u} \in \Omega} \phi_k(\mathbf{u})$. By invoking the result of [9, Corollary 2.2.1], we
 279 have

$$\begin{aligned} 280 \quad \phi_k(\mathbf{u}) &\geq \phi_k^* + \frac{\sigma_k \mu}{2} \|\mathbf{u} - \Pi_\Omega(\mathbf{w}_k)\|^2 \\ &\geq \sigma_k \left(f(\mathbf{u}_k) - f^* - \frac{\mu \varepsilon^2}{4} \right) + \frac{\sigma_k \mu}{2} \|\mathbf{u} - \Pi_\Omega(\mathbf{w}_k)\|^2, \end{aligned}$$

281 for all $\mathbf{u} \in \Omega$. Then relationship (4.8) yields that

$$\begin{aligned} 282 \quad \phi_{k+1}(\mathbf{u}) &\geq \sigma_k \left(f(\mathbf{u}_k) - f^* - \frac{\mu \varepsilon^2}{4} \right) + \frac{\sigma_k \mu}{2} \|\mathbf{u} - \Pi_\Omega(\mathbf{w}_k)\|^2 - \nu_k \sigma_k f^* \\ &\quad + \nu_k \sigma_k f(\mathbf{v}_k) + \nu_k \sigma_k \langle \nabla f(\mathbf{v}_k), \mathbf{u} - \mathbf{v}_k \rangle + \frac{\nu_k \sigma_k \mu}{2} \|\mathbf{u} - \mathbf{v}_k\|^2 \\ &\geq \sigma_{k+1} (f(\mathbf{v}_k) - f^*) - \frac{\sigma_k \mu \varepsilon^2}{4} + \langle \nabla f(\mathbf{v}_k), \sigma_k \mathbf{u}_k - \sigma_{k+1} \mathbf{v}_k \rangle \\ &\quad + \nu_k \sigma_k \langle \nabla f(\mathbf{v}_k), \mathbf{u} \rangle + \frac{\sigma_k \mu}{2} \|\mathbf{u} - \Pi_\Omega(\mathbf{w}_k)\|^2 \\ &= \sigma_{k+1} (f(\mathbf{v}_k) - f^*) - \frac{\sigma_k \mu \varepsilon^2}{4} + \nu_k \sigma_k \langle \nabla f(\mathbf{v}_k), \mathbf{u} - \Pi_\Omega(\mathbf{w}_k) \rangle \\ &\quad + \frac{\sigma_k \mu}{2} \|\mathbf{u} - \Pi_\Omega(\mathbf{w}_k)\|^2, \end{aligned}$$

283 where the second inequality comes from the strong convexity of f and (4.7), and the
 284 last equality holds due to the definition of \mathbf{v}_k in (4.1). According to the definition of
 285 \mathbf{z}_k in (4.2), we can obtain that

$$\begin{aligned} 286 \quad &\nu_k \sigma_k \langle \nabla f(\mathbf{v}_k), \mathbf{u} - \Pi_\Omega(\mathbf{w}_k) \rangle + \frac{\sigma_k \mu}{2} \|\mathbf{u} - \Pi_\Omega(\mathbf{w}_k)\|^2 \\ &= \frac{\sigma_k \mu}{2} \left\| \mathbf{u} - \left(\Pi_\Omega(\mathbf{w}_k) - \frac{\nu_k}{\mu} \nabla f(\mathbf{v}_k) \right) \right\|^2 - \frac{\nu_k^2 \sigma_k}{2\mu} \|\nabla f(\mathbf{v}_k)\|^2 \\ &\geq \frac{\sigma_k \mu}{2} \left\| \mathbf{z}_k - \left(\Pi_\Omega(\mathbf{w}_k) - \frac{\nu_k}{\mu} \nabla f(\mathbf{v}_k) \right) \right\|^2 - \frac{\nu_k^2 \sigma_k}{2\mu} \|\nabla f(\mathbf{v}_k)\|^2 \\ &= \nu_k \sigma_k \langle \nabla f(\mathbf{v}_k), \mathbf{z}_k - \Pi_\Omega(\mathbf{w}_k) \rangle + \frac{\sigma_k \mu}{2} \|\mathbf{z}_k - \Pi_\Omega(\mathbf{w}_k)\|^2. \end{aligned}$$

287 As a result, it holds that

$$\begin{aligned} 288 \quad (4.12) \quad \phi_{k+1}(\mathbf{u}) &\geq \sigma_{k+1} (f(\mathbf{v}_k) - f^*) - \frac{\sigma_k \mu \varepsilon^2}{4} + \nu_k \sigma_k \langle \nabla f(\mathbf{v}_k), \mathbf{z}_k - \Pi_\Omega(\mathbf{w}_k) \rangle \\ &\quad + \frac{\sigma_k \mu}{2} \|\mathbf{z}_k - \Pi_\Omega(\mathbf{w}_k)\|^2, \end{aligned}$$

289 for all $\mathbf{u} \in \Omega$. From the definitions of \mathbf{v}_k and \mathbf{u}_{k+1} in (4.1) and (4.3), it can be derived
 290 that $\mathbf{z}_k - \Pi_\Omega(\mathbf{w}_k) = (\mathbf{u}_{k+1} - \mathbf{v}_k)/\eta_k$. Substituting this relationship into (4.12) and
 291 taking $\mathbf{u} = \Pi_\Omega(\mathbf{w}_{k+1})$, we arrive at

$$292 \quad \frac{\phi_{k+1}^*}{\sigma_{k+1}} \geq f(\mathbf{v}_k) - f^* + \langle \nabla f(\mathbf{v}_k), \mathbf{u}_{k+1} - \mathbf{v}_k \rangle + \frac{\mu}{2\nu_k^2} \|\mathbf{u}_{k+1} - \mathbf{v}_k\|^2 - \frac{(1 - \eta_k)\mu \varepsilon^2}{4},$$

293 which together with the line-search condition (4.4) implies that

$$294 \quad \frac{\phi_{k+1}^*}{\sigma_{k+1}} \geq f(\mathbf{u}_{k+1}) - f^* - \frac{\eta_k \mu \varepsilon^2}{4} - \frac{(1 - \eta_k)\mu \varepsilon^2}{4} = f(\mathbf{u}_{k+1}) - f^* - \frac{\mu \varepsilon^2}{4}.$$

295 Therefore, relationship (4.11) also holds for $k + 1$. The proof is completed. \square

296 COROLLARY 4.5. *Let $\{\sigma_k\}$ and $\{\phi_k\}$ be the sequences defined in Lemma 4.2. Then
297 the sequence $\{\mathbf{u}_k\}$ generated by Algorithm 3 satisfies*

298 (4.13)
$$f(\mathbf{u}_k) - f^* \leq \frac{1}{\sigma_k} \phi_0(\mathbf{u}^*) + \frac{\mu \varepsilon^2}{4},$$

299 for any $k \in \mathbb{N}$.

300 *Proof.* Collecting two relationships (4.10) and (4.11) together, we can obtain that

301
$$\begin{aligned} \sigma_k \left(f(\mathbf{u}_k) - f^* - \frac{\mu \varepsilon^2}{4} \right) &\leq \min_{\mathbf{u} \in \Omega} \phi_k(\mathbf{u}) \leq \min_{\mathbf{u} \in \Omega} \{ \sigma_k(f(\mathbf{u}) - f^*) + \phi_0(\mathbf{u}) \} \\ &\leq \sigma_k(f(\mathbf{u}^*) - f^*) + \phi_0(\mathbf{u}^*) \\ &= \phi_0(\mathbf{u}^*), \end{aligned}$$

302 which completes the proof. \square

303 We proceed to establish the iteration complexity of Algorithm 3, as articulated
304 in the theorem below.

305 THEOREM 4.6. *Let $\varepsilon \in (0, 1)$ be a sufficiently small constant. Then after at most*

306
$$O \left(\log \left(\frac{1}{\varepsilon} \right) \frac{1}{\varepsilon^{2(1-\hat{\alpha})/(1+3\hat{\alpha})}} \right)$$

307 iterations, Algorithm 3 will reach an iterate \mathbf{u}_k satisfying $\|\mathbf{u}_k - \mathbf{u}^*\| \leq \varepsilon$.

308 *Proof.* In view of relationship (4.6), the number of line-search steps j_k in (4.4)
309 satisfies

310
$$\frac{\mu}{\nu_k^2} \eta_k^{(1-\hat{\alpha})/(1+\hat{\alpha})} \leq 2 \max_{i \in [m]} \left\{ \left[\frac{2(1-\alpha_i)}{\mu(1+\alpha_i)\varepsilon^2} \right]^{(1-\alpha_i)/(1+\alpha_i)} L_i^{2/(1+\alpha_i)} \right\} \leq \frac{2M}{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}},$$

311 where $M > 0$ is a constant defined in (2.1). Since $\eta_k = \nu_k/(1+\nu_k) \geq \nu_k/2$, we arrive
312 at

313 (4.14)
$$\frac{\nu_k^2}{\mu} \geq \frac{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}}{2M} \eta_k^{(1-\hat{\alpha})/(1+\hat{\alpha})} \geq \frac{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}}{2^{2/(1+\hat{\alpha})} M} \nu_k^{(1-\hat{\alpha})/(1+\hat{\alpha})}.$$

314 Let $\omega > 0$ be a constant defined as

315
$$\omega = \frac{1}{2^{2/(1+3\hat{\alpha})}} \left[\frac{\mu}{M} \right]^{(1+\hat{\alpha})/(1+3\hat{\alpha})}.$$

316 Then it follows from relationship (4.14) that

317 (4.15)
$$\nu_k \geq \omega \varepsilon^{2(1-\hat{\alpha})/(1+3\hat{\alpha})},$$

318 which further infers that

319
$$\sigma_{k+1} = (1 + \nu_k) \sigma_k \geq \left(1 + \omega \varepsilon^{2(1-\hat{\alpha})/(1+3\hat{\alpha})} \right) \sigma_k.$$

320 Applying the above inequality for k times recursively yields that

321
$$\sigma_k \geq \left(1 + \omega \varepsilon^{2(1-\hat{\alpha})/(1+3\hat{\alpha})} \right)^k.$$

322 As a direct consequence of (2.5) and (4.13), we can show that

$$\begin{aligned} 323 \quad \|\mathbf{u}_k - \mathbf{u}^*\|^2 &\leq \frac{2}{\mu} (f(\mathbf{u}_k) - f^*) \leq \frac{2}{\mu} \left(\frac{1}{\sigma_k} \phi_0(\mathbf{u}^*) + \frac{\mu \varepsilon^2}{4} \right) \\ &\leq \chi \left(1 + \omega \varepsilon^{2(1-\hat{\alpha})/(1+3\hat{\alpha})} \right)^{-k} + \frac{\varepsilon^2}{2}, \end{aligned}$$

324 where $\chi = 2(f(\mathbf{u}_0) - f^*)/\mu + \|\mathbf{u}_0 - \mathbf{u}^*\|^2 > 0$ is a constant. Let K_ε^* be the small-
325 est iteration number k such that $\|\mathbf{u}_k - \mathbf{u}^*\| \leq \varepsilon$. By solving the inequality $\chi(1 +$
326 $\omega \varepsilon^{2(1-\hat{\alpha})/(1+3\hat{\alpha})})^{-k} \leq \varepsilon^2/2$, we have

$$327 \quad K_\varepsilon^* \leq \log \left(\frac{\sqrt{2\chi}}{\varepsilon} \right) \frac{2}{\log(1 + \omega \varepsilon^{2(1-\hat{\alpha})/(1+3\hat{\alpha})})} \leq \log \left(\frac{\sqrt{2\chi}}{\varepsilon} \right) \frac{4}{\omega \varepsilon^{2(1-\hat{\alpha})/(1+3\hat{\alpha})}}.$$

328 The proof is completed. \square

329 The complexity bound established in Theorem 4.6 is markedly lower than those
330 presented in Theorems 2.2 and 3.1, thereby highlighting the acceleration effect at-
331 tained by Algorithm 3. Finally, we demonstrate that the number of line-search steps
332 required by Algorithm 3 is also $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+3\hat{\alpha})})$.

333 **COROLLARY 4.7.** *Let $\varepsilon \in (0, 1)$ be a sufficiently small constant. Then, to achieve
334 an iterate \mathbf{u}_k satisfying $\|\mathbf{u}_k - \mathbf{u}^*\| \leq \varepsilon$, Algorithm 3 requires at most*

$$335 \quad O \left(\log \left(\frac{1}{\varepsilon} \right) \frac{1}{\varepsilon^{2(1-\hat{\alpha})/(1+3\hat{\alpha})}} \right)$$

336 *line-search steps.*

337 *Proof.* It follows from relationship (4.14) that

$$338 \quad \rho_{k+1} = 2^{j_k} \rho_k = \frac{\mu}{\nu_k^2} \leq \frac{2^{2/(1+\hat{\alpha})} M}{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}} \left[\frac{1}{\nu_k} \right]^{(1-\hat{\alpha})/(1+\hat{\alpha})},$$

339 which together with (4.15) implies that

$$340 \quad \rho_{k+1} \leq \frac{2^{2/(1+\hat{\alpha})} M}{\varepsilon^{2(1-\hat{\alpha})/(1+\hat{\alpha})}} \left[\frac{1}{\omega \varepsilon^{2(1-\hat{\alpha})/(1+3\hat{\alpha})}} \right]^{(1-\hat{\alpha})/(1+\hat{\alpha})} = \frac{2^{2/(1+\hat{\alpha})} M}{\omega^{(1-\hat{\alpha})/(1+\hat{\alpha})} \varepsilon^{4(1-\hat{\alpha})/(1+3\hat{\alpha})}}.$$

341 Let N_k be the total number of line-search steps after k iterations in Algorithm 3. In
342 view of (3.4), we have

$$\begin{aligned} 343 \quad N_k &\leq k + 1 + \log \left(\frac{2^{2/(1+\hat{\alpha})} M}{\omega^{(1-\hat{\alpha})/(1+\hat{\alpha})} \varepsilon^{4(1-\hat{\alpha})/(1+3\hat{\alpha})}} \right) - \log \rho_0 \\ &\leq k + \frac{4(1-\hat{\alpha})}{1+3\hat{\alpha}} \log \left(\frac{1}{\varepsilon} \right) + \log \left(\frac{2^{2/(1+\hat{\alpha})} M}{\omega^{(1-\hat{\alpha})/(1+\hat{\alpha})} \rho_0} \right) + 1. \end{aligned}$$

344 Consequently, Theorem 4.6 indicates that the total number of line-search steps in
345 Algorithm 3 is at most $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+3\hat{\alpha})})$, which completes the proof. \square

346 **5. Numerical Experiments.** Preliminary numerical results are presented in
347 this section to provide additional insights into the performance guarantees of the
348 gradient descent method (1.3). We aim to elucidate that the final error attained by

349 the gradient descent method (1.3) is influenced by both the stepsize τ and the Hölder
 350 exponent α .

351 We generated the results using Julia [3] version 1.12 on an Apple Macintosh Mini
 352 with a M2 processor, 8 performance cores, and 32GB of memory.

353 We have placed the Julia codes for the results in the GitHub repository https://github.com/ctkelley/Grad_Des_CKW.jl with instructions for reproducing the figures.
 354

355 **5.1. Two-dimensional PDE with a non-Lipschitz term.** Hölder continuous
 356 gradients arise naturally in partial differential equations (PDEs) involving non-
 357 Lipschitz nonlinearity [2, 12]. In this subsection, we introduce a numerical example
 358 from [2]. This problem is to solve the following two-dimensional PDE,

359 (5.1)
$$\mathcal{F}(u) = -\Delta u + \nu u_+^p = 0,$$

360 where $p \in (0, 1)$, $\nu > 0$ is a constant and $u_+ = \max\{u, 0\}$. It should be noted that \mathcal{F}
 361 is the gradient of the following energy functional,

362
$$\hat{f}(u) = \frac{1}{2} \|\nabla u\|^2 + \frac{\nu}{p+1} \int_D u_+^{p+1}(y) dy.$$

363 Discretizing (5.1) with the standard five point difference scheme [7] leads to the
 364 following nonlinear system,

365 (5.2)
$$\mathbf{F}(\mathbf{u}) = \mathbf{A}\mathbf{u} + \nu \mathbf{u}_+^{1/2} - \mathbf{b} = 0,$$

366 where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the discretization of $-\Delta$ with zero boundary conditions, $\mathbf{b} \in$
 367 \mathbb{R}^n encodes the boundary conditions, and $\mathbf{u}_+^{1/2} = \max\{\mathbf{u}, 0\}^{1/2}$ is understood as a
 368 component-wise operation. Problem (5.2) is equivalent to optimization problem (1.1)
 369 with $\Omega = \mathbb{R}^n$, and

370
$$f(\mathbf{u}) = \frac{1}{2}(f_1(\mathbf{u}) + f_2(\mathbf{u})) \quad \text{with} \quad f_1(\mathbf{u}) = \mathbf{u}^\top \mathbf{A}\mathbf{u} - 2\mathbf{b}^\top \mathbf{u}, \quad f_2(\mathbf{u}) = \frac{\nu}{p+1} \mathbf{e}^\top \mathbf{u}_+^{1+p},$$

371 where $\mathbf{e} \in \mathbb{R}^n$ is the vector of all ones.

372 It is clear that ∇f_1 is Lipschitz continuous with the Lipschitz constant $L_1 = \|\mathbf{A}\|$,
 373 and ∇f_2 is locally Hölder continuous with $\alpha = 1/2$ and $L_2 = \nu n^{1/4}$ from

374
$$\|\nabla f_2(\mathbf{u}) - \nabla f_2(\mathbf{v})\| = \nu \left\| \mathbf{u}_+^{1/2} - \mathbf{v}_+^{1/2} \right\| \leq \nu n^{1/4} \|\mathbf{u} - \mathbf{v}\|^{1/2},$$

375 for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. The function f is $\lambda(\mathbf{A})$ -strongly convex, where $\lambda(\mathbf{A})$ is the smallest
 376 eigenvalue of the symmetric positive definite matrix \mathbf{A} .

377 We now modify the problem to enable direct computation of the errors in the
 378 iteration. To this end we follow Example 4.4 in [11] and take as the exact solution
 379 the function

380
$$u^*(x, y) = \left(\frac{3r-1}{2} \right)^2 \max(0, r-1/3)$$

381 where $r = \sqrt{x^2 + y^2}$, and let \mathbf{u}^* be u^* evaluated at the interior grid points. We
 382 enforce the boundary conditions

383
$$u(x, 1) = u^*(x, 1), u(x, 0) = u^*(x, 0), u(1, y) = u^*(1, y), u(0, y) = u^*(0, y)$$

384 for $0 < x, y < 1$ and encode this into \mathbf{b} . Letting $\mathbf{c}^* = \mathbf{F}(\mathbf{u}^*)$ our modified equation is

385 (5.3)
$$\mathbf{F}(\mathbf{u}) - \mathbf{c}^* = 0.$$



FIG. 1. Numerical performance of gradient descent method (1.3) for problem ??.

386 In the iteration we use the solution of $\mathbf{A}\mathbf{u}_0 = -\mathbf{b}$ as the initial iterate. This is the
 387 discretization of Laplace's equation with the problem boundary conditions. In this
 388 way we ensure that the entire iteration satisfies the boundary conditions.

389 In the first experiment, we scrutinize the performance of the gradient descent
 390 method (1.3) under different stepsizes. Specifically, with the parameter p fixed at 0.5,
 391 the algorithm is tested for stepsizes chosen from the set $\{0.01, 0.005, 0.001, 0.0005\}$.
 392 The corresponding numerical results, presented in Figure 1(a), illustrate the decay
 393 of the distance between the iterates and the global minimizer over iterations. It can
 394 be observed that a larger stepsize facilitates a more rapid descent in the early stage
 395 of iterations, albeit at the expense of a greater asymptotic error. This phenomenon
 396 corroborates our theoretical predictions.

397 In the second experiment, the stepsize τ is fixed at 0.001, while the parameter p
 398 is varied over the values $\{0.2, 0.4, 0.6, 0.8\}$. Figure 1(b) similarly tracks the decay of
 399 the distance to the global minimizer over iterations. It is evident that, as the value
 400 of p decreases, the final error attained by the algorithm increases under the same
 401 stepsize. Therefore, the associated optimization problems become increasingly ill-
 402 conditioned and thus more challenging to solve for smaller values of p . These findings
 403 offer empirical support for our theoretical analysis.

404 **5.2. Example 2.** We consider a second numerical example motivated by a semi-
 405 linear elliptic problem with a constraint on the solution in a certain set [12]. Let
 406 $D = (0, 1)^3$ and

407 (5.4)
$$\mathcal{H}(u) = -\Delta u + \lambda|u|^\nu - |u|^{p-1}u$$

408 on D with the boundary condition $u = 1$ on the boundary ∂D , where $p > 1$, $\nu \in (0, 1)$
 409 and $\lambda > p/\nu$ are constants. We consider the variational inequality that is to find
 410 $u^* \in [-1, 1]$ such that for any $u \in [-1, 1]$,

411
$$\mathcal{H}(u^*)(u - u^*) \geq 0.$$

412 This problem is equivalent to the nonlinear equation

413 (5.5)
$$0 = \mathcal{F}(u) := \begin{cases} \mathcal{H}(u) & \text{if } u - \mathcal{H}(u) \in [-1, 1], \\ u - 1 & \text{if } u - \mathcal{H}(u) \geq 1, \\ u + 1 & \text{otherwise.} \end{cases}$$

414 Discretizing (5.4) with the standard five point difference scheme [7], problem (5.5)
 415 leads to the following system of nonlinear equations

416 (5.6)
$$\mathbf{F}(\mathbf{u}) = \mathbf{u} - \Pi_{\mathbf{U}} \left(\mathbf{u} - \tau (\mathbf{A}\mathbf{u} + \lambda|\mathbf{u}|^\nu - |\mathbf{u}|^{p-1}\mathbf{u} - \mathbf{b}) \right) = 0,$$

417 where $\mathbf{U} = [-1, 1]^n$, $\tau > 0$ is a constant, $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix and $\mathbf{b} \in \mathbb{R}^n$. Note that (5.6) is the first-order optimal condition of the
 418 minimization problem
 419

420 (5.7)
$$\min_{\mathbf{u} \in [-1, 1]^n} f(\mathbf{u}) := \frac{1}{2} \mathbf{u}^\top \mathbf{A} \mathbf{u} + \frac{\lambda}{1+\nu} \mathbf{e}^\top |\mathbf{u}|^{\nu+1} - \frac{1}{1+p} \mathbf{e}^\top \max(\mathbf{u}, -\mathbf{u})^{p+1} + \mathbf{b}^\top \mathbf{u}.$$

The Hessian matrix of f at \mathbf{u} with $\mathbf{u}_i \neq 0$, $i = 1, \dots, n$ has the form

$$\nabla^2 f(\mathbf{u}) = \mathbf{A} + \lambda\nu |\mathbf{u}|^{\nu-1} - p \text{diag} \left(\max(-\mathbf{u}, \mathbf{u})^{p-1} \right),$$

421 Since $\lambda\nu > p$, $\nabla^2 f(\mathbf{u})$ is symmetric positive definite for any $\mathbf{u} \in [-1, 1]^n$ with $\mathbf{u}_i \neq 0$,
 422 $i = 1, \dots, n$. Hence f is μ -strongly convex in $[-1, 1]^n$ with $\mu = \lambda_{\min}(\mathbf{A})$ and the
 423 system (5.6) has a unique solution in $[-1, 1]^n$. However, ∇f is not Lipschitz continuous
 424 in $[-1, 1]^n$.

Let

$$f_1(\mathbf{u}) = \frac{1}{2} \mathbf{u}^\top \mathbf{A} \mathbf{u} + \mathbf{b}^\top \mathbf{u}, f_2(\mathbf{u}) = \frac{\lambda}{1+\nu} \mathbf{e}^\top |\mathbf{u}|^{\nu+1}, f_3(\mathbf{u}) = -\frac{1}{1+p} \mathbf{e}^\top \max(\mathbf{u}, -\mathbf{u})^{p+1}$$

425 This example satisfies Assumption 1.1 (ii) with $L_1 = \lambda_{\max}(\mathbf{A})$, $L_2 = \lambda\nu$, $L_3 =$
 426 $pn^{\frac{1}{2}}$, $\alpha_1 = \alpha_3 = 1$, $\alpha_2 = 1 - \nu$.

427 **6. Conclusion.** In this paper, we consider a class of strongly convex constrained
 428 optimization problems of the form (1.1).

429 This example satisfies Assumption 1.1 (ii) with $L_1 = \lambda_{\max}(\mathbf{A})$, $L_2 = \lambda\nu$, $L_3 =$
 430 $pn^{\frac{1}{2}}$, $\alpha_1 = \alpha_3 = 1$, $\alpha_2 = 1 - \nu$. Example 1.1 shows that although each component
 431 function f_i of the objective function f admits a Hölder continuous gradient with an
 432 component $\alpha_i \in (0, 1]$, the gradient of f is not necessarily Hölder continuous. To
 433 establish the iteration complexity of the projected gradient descent methods for this
 434 class of problems, we use the parameter $\hat{\alpha} = \min_{i \in [m]} \alpha_i$ to determine the complex-
 435 ity bound. Algorithm 1 is a new version of projected gradient method for problem
 436 (1.1) with an appropriately fixed stepsize. Theorem 2.2 shows that Algorithm 1 can
 437 find an iterate in the feasible set Ω with a distance to the global minimizer less than
 438 ε at most $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+\hat{\alpha})})$ iterations. This recovers the classical complex-
 439 ity result when $\hat{\alpha} = 1$ and reveals the additional difficulty imposed by the weaker
 440 smoothness of the objective function for $\hat{\alpha} < 1$. Algorithm 2 is a modification of
 441 Algorithm 1 for problems where the parameters α_i and L_i are difficult to estimate
 442 for the stepsize. In Algorithm 3, the stepsize is updated by the universal scheme at
 443 each iteration, which improves the complexity bound to $O(\log(\varepsilon^{-1})\varepsilon^{2(\hat{\alpha}-1)/(1+3\hat{\alpha})})$.

444 Numerical experiments are conducted to validate our theoretical findings, demon-
 445 strating the expected behavior of projected gradient descent methods under different
 446 stepsizes and Hölder exponents. These results offer new insights into the performance
 447 guarantees of the classic projected gradient descent methods for a broader class of
 448 optimization problems with non-Lipschitz gradients.

449 REFERENCES

- 450 [1] J.-C. BARITAUX, K. HASSLER, AND M. UNSER, *An efficient numerical method for general L_p*
 451 *regularization in fluorescence molecular tomography*, IEEE Trans. Med. Imaging, 29 (2010),
 452 pp. 1075–1087.
- 453 [2] J. W. BARRETT AND R. M. SHANAHAN, *Finite element approximation of a model reaction-*
 454 *diffusion problem with a non-lipschitz nonlinearity*, Numer. Math., 59 (1991), pp. 217–242.
- 455 [3] J. BEZANSON, A. EDELMAN, S. KARPINSKI, AND V. B. SHAH, *Julia: A fresh approach to nu-*
 456 *merical computing*, SIAM Review, 59 (2017), pp. 65–98.
- 457 [4] L. S. BORGES, F. S. V. BAZÁN, AND L. BEDIN, *A projection-based algorithm for ℓ_2 - ℓ_p Tikhonov*
 458 *regularization*, Math. Methods Appl. Sci., 41 (2018), pp. 5919–5938.
- 459 [5] X. CHEN, C. T. KELLEY, AND L. WANG, *A new complexity result for strongly convex optimiza-*
 460 *tion with locally α -hölder continuous gradients*, arXiv:2505.03506v1, (2025).
- 461 [6] O. DEVOLDER, F. GLINEUR, AND Y. NESTEROV, *First-order methods of smooth convex opti-*
 462 *mization with inexact oracle*, Math. Program., 146 (2014), pp. 37–75.
- 463 [7] R. J. LEVEQUE, *Finite Difference Methods for Ordinary and Partial Differential Equations:*
 464 *Steady-State and Time-Dependent Problems*, Society for Industrial and Applied Mathe-
 465 *matics*, 2007.
- 466 [8] Y. NESTEROV, *Universal gradient methods for convex optimization problems*, Math. Program.,
 467 152 (2015), pp. 381–404.
- 468 [9] Y. NESTEROV, *Lectures on Convex Optimization*, Springer, 2018.
- 469 [10] Y. NESTEROV, *Universal complexity bounds for universal gradient methods in nonlinear opti-*
 470 *mization*, arXiv:2509.20902, (2025).
- 471 [11] X. QU, W. BIAN, AND X. CHEN, *An extra gradient Anderson-accelerated algorithm for*
 472 *pseudomonotone variational inequalities*, 2024, <https://arxiv.org/abs/2408.06606>, <https://arxiv.org/abs/2408.06606>.
- 474 [12] M. TANG, *Uniqueness of bound states to $\Delta u - u + |u|^{p-1}u = 0$ in \mathbb{R}^n , $n \geq 3$* , Invent. Math.,
 475 (2025), pp. 1–47.