TENSOR METHODS FOR MINIMIZING FUNCTIONS WITH HÖLDER CONTINUOUS HIGHER-ORDER DERIVATIVES

G.N. GRAPIGLIA* AND YU. NESTEROV†

Abstract. In this paper we study $p$-order methods for unconstrained minimization of convex functions that are $p$-times differentiable with $\nu$-Hölder continuous $p$th derivatives. We propose tensor schemes with and without acceleration. For the schemes without acceleration, we establish iteration complexity bounds of $O\left(\epsilon^{-1/(p+\nu-1)}\right)$ for reducing the functional residual below a given $\epsilon \in (0, 1)$. Assuming that $\nu$ is known, we obtain an improved complexity bound of $O\left(\epsilon^{-1/(p+\nu)}\right)$ for the corresponding accelerated scheme. For the case in which $\nu$ is unknown, we present a universal accelerated tensor scheme with iteration complexity of $O\left(\epsilon^{-2/\left[3(p+\nu)-2\right]}\right)$. A lower complexity bound of $O\left(\epsilon^{-2/\left[3(p+\nu)-2\right]}\right)$ is also obtained for this problem class.

Key words. unconstrained minimization, high-order methods, tensor methods, Hölder condition, worst-case global complexity bounds

AMS subject classifications. 49M15, 49M37, 58C15, 90C25, 90C30

1. Introduction.

1.1. Motivation. In [9], it was shown that a suitable cubic regularization of Newton method (CNM) takes at most $O\left(\epsilon^{-1/2}\right)$ iterations to reduce the functional residual below a given precision $\epsilon > 0$, when the objective is a twice-differentiable convex function with Lipschitz continuous Hessian. A better complexity bound of $O\left(\epsilon^{-1/3}\right)$ was shown in [10] for an accelerated version of CNM. Auxiliary problems in these methods consist in the minimization of a third-order regularization of the second-order Taylor approximation of the objective function around the current iterate. A natural generalization is to consider auxiliary problems in which one minimizes a $(p+1)$-order regularization of the $p$th order Taylor approximation of the objective function, resulting in tensor methods. Unconstrained optimization by Tensor methods is not a new subject (see, for example, [12, 4]). In the context of convex optimization, accelerated tensor methods (as described above) were first considered by Baes [2]. However, the author did not realize that under a proper choice of the regularization coefficient the auxiliary problems become convex. This important observation was done in a recent paper [11], where tensor methods with and without acceleration were proposed for unconstrained minimization of $p$-times differentiable convex functions with Lipschitz continuous $p$th derivatives. An iteration complexity bound of $O\left(\epsilon^{-1/p}\right)$ was proved for the method without acceleration, while an improved bound of $O\left(\epsilon^{-1/(p+1)}\right)$ was proved for the accelerated tensor method.

In the present paper, we study tensor methods (with and without acceleration) that can handle convex functions with $\nu$-Hölder continuous $p$th derivatives. For the schemes without acceleration, we establish iteration complexity bounds of $O\left(\epsilon^{-1/(p+\nu-1)}\right)$ for reducing the functional residual below a given $\epsilon \in (0, 1)$. Assum-

---

*Departamento de Matemática, Universidade Federal do Paraná, Centro Politécnico, Cx. postal 19.081, 81531-980, Curitiba, Paraná, Brazil (grapiglia@ufpr.br). This author was supported by the National Council for Scientific and Technological Development - Brazil (grants 401288/2014-5 and 406269/2016-5) and by the European Research Council Advanced Grant 788368.

†Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium (Yurii.Nesterov@uclouvain.be). This research was partially financed by the European Research Council Advanced Grant 788368. Generous support of Alibaba Group is also acknowledged.
ing that $\nu$ is known, we obtain an improved complexity bound of $O\left(\epsilon^{-1/(p+\nu)}\right)$ for the corresponding accelerated scheme. For the case in which $\nu$ is unknown, we present a universal accelerated tensor scheme with iteration complexity of $O\left(\epsilon^{-p/(\nu+1)(p+n-1)}\right)$. 

In all methods we allow inexact solution of the auxiliary problems by incorporating the acceptance conditions proposed in [3] in the context of nonconvex optimization. The complexity bounds established here generalize our previous results reported in [5, 6] for regularized Newton methods (i.e., case $p=2$). Finally, we also present a lower complexity bound for tensor methods under the Hölder condition.

1.2. Contents. The paper is organized as follows. In Section 2, we define our problem. In Section 3, we present tensor methods without acceleration and establish their convergence properties. In Section 4, we present complexity results for accelerated schemes. Finally, in Section 5 we obtain lower complexity bounds for tensor methods under the Hölder condition. All necessary auxiliary results are included in an appendix.

1.3. Notations and Generalities. In what follows, we denote by $E$ a finite-dimensional real vector space, and by $E^*$ its dual space, composed by linear functionals on $E$. The value of function $s \in E^*$ at point $x \in E$ is denoted by $\langle s, x \rangle$. Given a self-adjoint positive definite operator $B : E \to E^*$ (notation $B \succ 0$), we can endow these spaces with conjugate Euclidean norms:

$$
\|x\| = \langle Bx, x \rangle^{1/2}, \quad x \in E, \quad \|s\|_* = \langle s, B^{-1}s \rangle^{1/2}, \quad s \in E^*.
$$

For a smooth function $f : \text{dom} \ f \to \mathbb{R}$ with convex and open domain $\text{dom} \ f \subset E$, denote by $\nabla f(x)$ its gradient, and by $\nabla^2 f(x)$ its Hessian evaluated at point $x \in \text{dom} \ f$. Note that $\nabla f(x) \in E^*$ and $\nabla^2 f(x)h \in E^*$ for $x \in \text{dom} \ f$ and $h \in E$.

For any integer $p \geq 1$, denote by

$$
D^p f(x)[h_1, \ldots, h_p]
$$

the directional derivative of function $f$ at $x$ along directions $h_i \in E$, $i = 1, \ldots, p$. In particular, for any $x \in \text{dom} \ f$ and $h_1, h_2 \in E$ we have

$$
D f(x)[h_1] = \langle \nabla f(x), h_1 \rangle \quad \text{and} \quad D^2 f(x)[h_1, h_2] = \langle \nabla^2 f(x)h_1, h_2 \rangle.
$$

For $h_1 = \ldots = h_p = h \in E$, we use notation $D^p f(x)[h]^p$. Then the $p$th order Taylor approximation of function $f$ at $x \in \text{dom} \ f$ can be written as follows:

$$
f(x + h) = \Phi_{x,p}(x + h) + o(\|h\|^p), \quad x + h \in \text{dom} \ f,
$$

where

$$
\Phi_{x,p}(y) = f(x) + \sum_{i=1}^{p} \frac{1}{i!} D^i f(x)[y - x]^i, \quad y \in E.
$$

Note that $D^p f(x)[.]$ is a symmetric $p$-linear form. Its norm is defined by

$$
\|D^p f(x)\| = \max_{h_1, \ldots, h_p} \{ |D^p f(x)[h_1, \ldots, h_p]| : \|h_i\| \leq 1, \ i = 1, \ldots, p \}.
$$

In fact, it can be shown that (see, e.g., Appendix 1 in [8])

$$
\|D^p f(x)\| = \max_h \{ |D^p f(x)[h]^p| : \|h\| \leq 1 \}.
$$

Similarly, since $D^p f(x)[., \ldots, .] - D^p f(y)[., \ldots, .]$ is also a symmetric $p$-linear form for fixed $x, y \in \text{dom} \ f$, we can define

$$
\|D^p f(x) - D^p f(y)\| = \max_h \{ |D^p f(x)[h]^p - D^p f(y)[h]^p| : \|h\| \leq 1 \}.
$$
2. Problem Statement. In this paper we consider methods for solving the following minimization problem

\[(2.1) \min_{x \in \mathbb{R}} f(x),\]

where \(f : \mathbb{R} \to \mathbb{R}\) is a convex \(p\)-times differentiable function. We assume that there exists at least one optimal solution \(x^* \in \mathbb{R}\) for problem (2.1). Let us characterize the level of smoothness of the objective \(f\) by the system of Hölder constants

\[(2.2) H_{f,p}(\nu) \equiv \sup_{x,y \in \mathbb{R}} \left\{ \frac{\|D^p f(x) - D^p f(y)\|}{\|x - y\|^\nu} \right\}, \quad 0 \leq \nu \leq 1.\]

Then, from (2.2) and from the integral form of the mean-value theorem, it follows that

\[(2.3) |f(y) - \Phi_{x,p}(y)| \leq \frac{H_{f,p}(\nu)}{p!} \|y - x\|^{p+\nu},\]

\[(2.4) \|\nabla f(y) - \nabla \Phi_{x,p}(y)\|_* \leq \frac{H_{f,p}(\nu)}{(p-1)!} \|y - x\|^{p+\nu-1},\]

for all \(x, y \in \mathbb{R}\). Given \(x \in \mathbb{R}\), if \(H_{f,p}(\nu) < +\infty\) and \(H \geq H_{f,p}(\nu)\), by (2.3) we have

\[(2.5) f(y) \leq \Phi_{x,p}(y) + \frac{H}{p!} \|y - x\|^{p+\nu}, \quad y \in \mathbb{R}.\]

This property motivates the use of the following class of models of \(f\) around \(x \in \mathbb{R}\):

\[(2.6) \Omega^{(\alpha)}_{x,p,H}(y) = \Phi_{x,p}(y) + \frac{H}{p!} \|y - x\|^{p+\alpha}, \quad \alpha \in [0, 1].\]

In particular, as long as \(H \geq H_{f,p}(\nu)\), by (2.5) we have

\[(2.7) f(y) \leq \Omega^{(\nu)}_{x,p,H}(y), \quad y \in \mathbb{R}.\]

3. Tensor schemes without acceleration. If we assume that \(H_{f,p}(\nu) < +\infty\) for some \(\nu \in [0, 1]\), there are two possible situations: either \(\nu\) is known, or \(\nu\) is unknown. We cover both cases in a single framework by introducing parameter

\[(3.1) \alpha = \begin{cases} \nu, & \text{if } \nu \text{ is known}, \\ 1, & \text{if } \nu \text{ is unknown}. \end{cases}\]

Let \(\varepsilon \in (0, 1)\) be the target precision. At the beginning of the \(t\)th iteration one has an estimate \(x_t\) for the solution of (2.1) and a scaling coefficient \(M_t\). A trial point \(x_t^+\) is computed as an approximate solution to the auxiliary problem

\[(3.2) \min_{y \in \mathbb{R}} \Omega^{(\alpha)}_{x_t,p,M_t}(y),\]

with \(\alpha\) given by (3.1). Similarly to [3], the trial point \(x_t^+\) must satisfies the following conditions:

\[(3.3) \Omega^{(\alpha)}_{x_t,p,M_t}(x_t^+) \leq f(x_t) \quad \text{and} \quad \|\nabla \Omega^{(\alpha)}_{x_t,p,M_t}(x_t^+)\|_* \leq \theta \|x_t^+ - x_t\|^{p+\alpha-1},\]
where \( \theta \geq 0 \) is a user-defined parameter. When (3.2) is not convex, then \( x_t^+ \) is not necessarily an approximation of its global solution. If the descent condition

\[
(3.4) \quad f(x_t) - f(x_t^+) \geq \frac{1}{8(p+1)!M_t^{p+\alpha-1}}\|\nabla f(x_t^+)|^\frac{p+\alpha}{\alpha-1},
\]

holds, then \( x_t^+ \) is accepted and we define \( x_{t+1} = x_t^+ \). Otherwise, constant \( M_t \) is increased until the corresponding trial point \( x_t^+ \) is accepted. We will see that this process is well defined in the sense that there exists \( M_\nu > 0 \) such that \( M_t \leq M_\nu \) for all \( t \). This general scheme can be summarized in the following way.

**Algorithm 1. Tensor Method**

**Step 0.** Choose \( x_0 \in \mathbb{E} \) and \( \theta \geq 0 \). Set \( \alpha \) by (3.1) and \( t := 0 \).

**Step 1.** Find 0 < \( M_t \leq M_\nu \) such that (3.4) holds for an approximate solution \( x_t^+ \) to (3.2) satisfying conditions (3.3).

**Step 2.** Set \( x_{t+1} = x_t^+ \).

**Step 3.** Set \( t := t + 1 \) and go back to Step 1.

To analyze convergence of Algorithm 1, we introduce the following assumptions:

- **H1** \( H_{f,p}(\nu) < +\infty \) for some \( \nu \in [0,1] \).
- **H2** The level sets of \( f \) are bounded, that is, \( \max_{x \in \mathcal{L}(x_0)} \|x - x^*\| \leq D_0 \in (1, +\infty) \) for \( \mathcal{L}(x_0) \equiv \{x \in \mathbb{E}: f(x) \leq f(x_0)\} \), with \( x_0 \) being the starting point.

The next theorem establishes global convergence rate for Algorithm 1.

**Theorem 3.1.** Suppose that H1 and H2 are true and let \( \{x_t\}_{t=0}^T \) be a sequence generated by Algorithm 1. Denote by \( m \) the first iteration number such that

\[
f(x_m) - f(x^*) \leq 2[8(p+1)!|p+\alpha-1]M_\nu D_0^p + \alpha,
\]

and assume that \( m < T \). Then

\[
(3.5) \quad m \leq \frac{1}{\ln \left( \frac{p+\alpha}{\alpha-1} \right)} \ln \max \left\{ 1, \log_2 \frac{f(x_0) - f(x^*)}{8(p+1)!|p+\alpha-1]M_\nu D_0^p + \alpha} \right\}
\]

and, for all \( k, m < k \leq T \), we have

\[
f(x_k) - f(x^*) \leq \frac{24p(p+1)!|p+\alpha-1]M_\nu D_0^p + \alpha}{(k-m)(p+\alpha-1)}.
\]

**Proof.** By Step 1 in Algorithm 1, we have

\[
M_k \leq M_\nu, \quad k = 0, \ldots, T - 1.
\]

Thus, in view of (3.4), (3.7) and H2, for \( k = 0, \ldots, T - 1 \) we have

\[
f(x_k) - f(x_{k+1}) \geq \frac{1}{8(p+1)!} \left[ \frac{1}{M_k} \right]^\frac{p+\alpha}{\alpha-1} \|\nabla f(x_{k+1})\|^\frac{p+\alpha}{\alpha-1}
\]

\[
\geq \frac{1}{8(p+1)!} \left[ \frac{1}{M_\nu} \right]^\frac{p+\alpha}{\alpha-1} \|\nabla f(x_{k+1})\|^\frac{p+\alpha}{\alpha-1}
\]

\[
\geq \frac{1}{8(p+1)!} \left[ \frac{1}{M_\nu D_0^p + \alpha} \right]^\frac{p+\alpha}{\alpha-1} \left( \|\nabla f(x_{k+1})\| \|x_{k+1} - x^*\| \right)^\frac{p+\alpha}{\alpha-1}
\]

\[
\geq \frac{1}{8(p+1)!} \left[ \frac{1}{M_\nu D_0^p + \alpha} \right]^\frac{p+\alpha}{\alpha-1} \left( f(x_{k+1}) - f(x^*) \right)^\frac{p+\alpha}{\alpha-1},
\]

for all \( k = 0, \ldots, T - 1 \).
where the last inequality is due to the convexity of $f$. Now, denoting
\[ \delta_k = \frac{f(x_k) - f(x^*)}{[8(p + 1)!!]^{p+\alpha-1} M_{\nu} D_0^p} \]
we see from (3.9) that this sequence satisfies condition (1.1) of Lemma 1.1 in [5] with $u = \frac{p+\alpha}{p+\alpha-1}$. Note that $m$ is the first iteration for which $\delta_m \leq 2$. If $m > 0$, then $\delta_0 > 2$ and, in view of inequality (1.2) of Lemma 1.1 in [5], we have
\[ \ln 2 \leq \ln \delta_m \leq \ln \left( \frac{p + \alpha - 1}{p + \alpha} \right)^{m-1} \ln 2 \leq \ln \delta_0 \]
\[ \Rightarrow \left( \frac{p + \alpha - 1}{p + \alpha} \right)^{m-1} \leq \frac{\ln \delta_0}{\ln 2} = \log_2 \delta_0. \]
Thus, $m \leq \frac{\ln \delta_0}{\ln \left( \frac{p + \alpha}{p + \alpha - 1} \right)}$, and so, (3.5) holds. Consequently, from inequality (1.3) of Lemma 1.1 in [5] we get the following rate of convergence:
\[ \delta_k \leq \left[ \frac{1 + \delta_{m-1}^{u-1}}{(u-1)(k-m)} \right]^{\frac{1}{u-1}}, \]
that is,
\[ \frac{f(x_k) - f(x^*)}{[8(p + 1)!!]^{p+\alpha-1} M_{\nu} D_0^p} \leq \left[ \frac{(p + \alpha - 1)(1 + 2^{\frac{1}{p+\alpha-1}})}{k-m} \right]^{p+\alpha-1}. \]
Therefore,
\[ f(x_k) - f(x^*) \leq \frac{[8(1 + 2^{\frac{1}{p+\alpha-1}})(p + \alpha - 1)(p + 1)!!]^{p+\alpha-1} M_{\nu} D_0^{p+\alpha}}{(k-m)^{p+\alpha-1}} \]
\[ \leq \frac{[24p(p + 1)!!]^{p+\alpha-1} M_{\nu} D_0^{p+\alpha}}{(k-m)^{p+\alpha-1}}. \]

If we assume that $\nu$ and $H_f(\nu)$ are known, by Lemma A.4, we can set
\[ M_t = M_{\nu} \equiv \max \left\{ \frac{3H_{f,p}(\nu)}{2}, 3\theta(p-1)! \right\}. \]
In this case, by (3.1), the corresponding version of Algorithm 1 takes at most $O(\epsilon^{-1/(p+\nu-1)})$ iterations to generate $x_k$ such that $f(x_k) - f(x^*) \leq \epsilon$, for a given $\epsilon \in (0, 1)$. However, in most practical problems, $H_{f,p}(\nu)$ is not known. To deal with this situation, we can consider the following adaptive version of Algorithm 1:
Algorithm 2. Adaptive Tensor Method

Step 0. Choose \( x_0 \in \mathbb{E}, H_0 > 0 \) and \( \theta \geq 0 \). Set \( \alpha \) by (3.1) and \( t := 0 \).

Step 1. Set \( i := 0 \).

Step 1.1 Compute an approximate solution \( x^+_{t,i} \) to \( \min_{y \in \mathbb{E}} \Omega^{(\alpha)}_{x_t,p,2^i H_t}(y) \), such that
\[
\Omega^{(\alpha)}_{x_t,p,2^i H_t}(x^+_{t,i}) \leq f(x_t) \quad \text{and} \quad \|\nabla \Omega^{(\alpha)}_{x_t,p,2^i H_t}(x^+_{t,i})\|_* \leq \theta \|x^+_{t,i} - x_t\|^{p+\alpha-1}.
\]

Step 1.2. If
\[
f(x_t) - f(x^+_{t,i}) \geq \frac{1}{8(p+1)!} \frac{\|\nabla f(x^+_{t,i})\|_*^{p+\alpha-1}}{2^{p+\alpha}}
\]
holds, set \( i_t := i \) and go to Step 2. Otherwise, set \( i := i + 1 \) and go to Step 1.1.

Step 2. Set \( x_{t+1} = x^+_{t,i_t} \) and \( H_{t+1} = 2^{i_t-1} H_t \).

Step 3. Set \( t := t + 1 \) and go to Step 1.

Note that Algorithm 2 is a particular case of Algorithm 1 in which
\[
M_t = 2^{i_t} H_t, \quad \forall t \geq 0.
\]

Let us define the following function of \( \epsilon > 0 \):
\[
N_\nu(\epsilon) = \begin{cases} 
\max \left\{ \frac{3H_f(p(\nu))}{2}, 3\theta(p-1)! \right\}, & \text{if } \alpha = \nu, \\
\max \left\{ \theta, \left( \frac{3H_f(p(\nu))}{2} \right)^{\frac{p}{p+\alpha}} \left( \frac{4R(\epsilon)}{\epsilon} \right)^{\frac{1}{p+\alpha}} \right\}, & \text{if } \alpha = 1,
\end{cases}
\]
where
\[
R(\epsilon) = \max_{x \in \mathbb{E}} \{\|x - x^*\| : f(x) \leq f(x^*) + \epsilon\}.
\]

The next lemma provides upper bounds on \( H_t \) and on the number of calls of oracle.

**Lemma 3.2.** Suppose that \( H1 \) and \( H2 \) are true. Given \( \epsilon > 0 \), assume that \( \{x_t\}_{t=0}^T \) is a sequence generated by Algorithm 2 such that
\[
f(x_0) - f(x^*) \geq \epsilon,
\]
\[
f(x^+_{t,i}) - f(x^*) \geq \epsilon, \quad i = 0, \ldots, i_t \text{ and } t = 0, \ldots, T.
\]
Then,
\[
H_t \leq \max \{H_0, N_\nu(\epsilon)\}, \quad \text{for } t = 0, \ldots, T.
\]
Moreover, the number \( O_T \) of calls of the oracle after \( T \) iterations is bounded as follows:
\[
O_T \leq 2T + \log_2 \max \{H_0, N_\nu(\epsilon)\} - \log_2 H_0.
\]

**Proof.** Let us prove (3.14) by induction. Clearly it holds for \( t = 0 \). Assume that (3.14) is true for some \( t, 0 \leq t \leq T - 1 \). If \( \nu \) is known, then by (3.1) we have \( \alpha = \nu \). Thus, by \( H1 \) and Lemma A.2, the final value of \( 2^i H_t \) cannot exceed
\[
2 \max \left\{ \frac{3H_f(p(\nu))}{2}, 3\theta(p-1)! \right\},
\]
since otherwise we should stop the line search earlier. Therefore,
\[
H_{t+1} = \frac{1}{2} \left( \frac{3H_{f,p}(\nu)}{2}, 3\theta(p-1)! \right) = N_\nu(\epsilon) \leq \max \{ H_0, N_\nu(\epsilon) \},
\]
that is, (3.14) holds for \( t = t + 1 \).

On the other hand, if \( \nu \) is unknown, we have \( \alpha = 1 \). In view of (3.11), (3.12) and H2, it follows that
\[
(3.16) \quad R(\epsilon) \leq R(f(x_0) - f(x^*)) = \max_{x \in L(x_0)} \| x - x^* \| \leq D_0 < +\infty.
\]

Thus, by (3.13) and Lemma A.5 in [6] we have \( \| \nabla f(x_{t+1}) \| \geq \frac{\epsilon}{R(\epsilon)} \). In this case, it follows from Corollary A.5 with \( \delta = \epsilon/R(\epsilon) \) that
\[
2^{i_t} H_t \leq 2 \max \left\{ \frac{3H_{f,p}(\nu)}{2}, \frac{\nu}{\nu - 1} \left( \frac{4R(\epsilon)}{\epsilon} \right)^{\frac{1}{\nu - 1}} \right\} = 2N_\nu(\epsilon).
\]

Consequently, we also have
\[
H_{t+1} = \frac{1}{2} 2^{i_t} H_t \leq N_\nu(\epsilon) \leq \max \{ H_0, N_\nu(\epsilon) \},
\]
that is, (3.14) holds for \( t + 1 \). This completes the induction argument.

Finally, note that at the \( k \)th iteration of Algorithm 2, the oracle is called \( i_k + 1 \) times. Since \( H_{k+1} = 2^{i_k} H_k \), it follows that \( i_k + 1 = \log_2 H_{k+1} - \log_2 H_k \). Thus, by (3.14) we get
\[
O_T = \sum_{k=0}^{T-1} (i_k + 1) = \sum_{k=0}^{T-1} 2 + \log_2 H_{k+1} - \log_2 H_k = 2T + \log_2 H_T - \log_2 H_0
\]
\[
\leq 2T + \log_2 \max \{ H_0, N_\nu(\epsilon) \} - \log_2 H_0.
\]

Combining Theorem 3.1 and Lemma 3.2, we obtain the following result.

THEOREM 3.3. Suppose that H1 and H2 are true. Given \( \epsilon \in (0, 1) \), assume that \( \{ x_t \}_{t=0}^T \) is a sequence generated by Algorithm 2 such that (3.12) and (3.13) hold. Denote by \( m \) the first iteration number such that
\[
f(x_m) - f(x^*) \leq [8(p + 1)!]^{p+\alpha-1} \max \{ H_0, N_\nu(\epsilon) \} D_0^{p+\alpha},
\]
and assume that \( m < T \). Then,
\[
(3.17) \quad m \leq \frac{1}{\ln \left( \frac{p+\alpha-1}{p+\alpha-1} \right)} \ln \max \left\{ 1, \log_2 \frac{f(x_0) - f(x^*)}{[8(p + 1)!]^{p+\alpha-1} \max \{ H_0, N_\nu(\epsilon) \} D_0^{p+\alpha}} \right\}
\]
and, for all \( k, m < k \leq T \), we have
\[
(3.18) \quad f(x_k) - f(x^*) \leq \frac{[24p(p + 1)!]^{p+\alpha-1} \max \{ H_0, N_\nu(\epsilon) \} D_0^{p+\alpha}}{(k - m)^{p+\alpha-1}}.
\]

Consequently,
\[
(3.19) \quad T \leq m + \kappa_1^{(\nu)} [24p(p + 1)!] \epsilon^{-\frac{1}{p+\alpha-1}},
\]
where

\[
\kappa_1^{(\nu)} = \begin{cases} 
\left(2 \max \left\{ H_0, \frac{3H_f(\nu)}{2}, 3\theta(p-1)! \right\} D_0^{p+\nu} \right)^{\frac{1}{p+\nu-1}}, & \text{if } \nu \text{ is known}, \\
\left(2 \max \left\{ H_0, \theta, \frac{3H_f(\nu)}{2} \right\} \frac{p}{p+\nu-1} (4D_0)^{\frac{1}{p+\nu-1}} \right)^{\frac{1}{p}}, & \text{if } \nu \text{ is unknown}.
\end{cases}
\]

Proof. By Lemma 3.2, we have

\[2^t H_t = 2^{(2^t-1)H_t} = 2H_{t+1} \leq 2 \max \{ H_0, N_\nu(\epsilon) \}, \quad t = 0, \ldots, T - 1.\]

Then, (3.17) and (3.18) follow directly from Theorem 3.2 with

\[M_\nu = 2 \max \{ H_0, N_\nu(\epsilon) \}.\]

Now, combining (3.13) and (3.18) for \(k = T\), we obtain

\[
\epsilon \leq \frac{[24p(p + 1)!]^{p+\alpha-1} \max \{ H_0, N_\nu(\epsilon) \} D_0^{p+\alpha}}{(T - m)^{p+\alpha-1}}.
\]

and so,

\[
(3.20) \quad T \leq m + \frac{[24p(p + 1)!] \left(2 \max \{ H_0, N_\nu(\epsilon) \} D_0^{p+\alpha}\right)^{\frac{1}{p+\alpha-1}}}{\epsilon^{p+\alpha-1}}.
\]

If \(\nu\) is known, then \(\alpha = \nu\) and, by (3.10), we have

\[
(3.21) \quad N_\nu(\epsilon) = \max \left\{ \frac{3H_{f,p}(\nu)}{2}, 3\theta(p-1)! \right\}.
\]

Thus, combining (3.20) and (3.21), we get (3.19). On the other hand, if \(\nu\) is unknown, then \(\alpha = 1\) and, by (3.10), (3.16) and \(\epsilon \in (0, 1)\), we have

\[
N_\nu(\epsilon) = \max \left\{ \theta, \left(\frac{3H_{f,p}(\nu)}{2}\right)^{\frac{p}{p+\nu-1}} \left(4R(\epsilon)\varepsilon\right)^{\frac{1-\nu}{p+\nu-1}} \right\}
\]

\[
\leq \max \left\{ \theta, \left(\frac{3H_{f,p}(\nu)}{2}\right)^{\frac{p}{p+\nu-1}} (4D_0)^{\frac{1-\nu}{p+\nu-1}} \right\} \epsilon^{-\frac{1-\nu}{p+\nu-1}}.
\]

(3.22)

In this case, combining (3.20) and (3.22) we also get (3.19). 

Note that Algorithm 2 with \(\alpha = 1\) is a universal scheme: it works for any Hölder parameter \(\nu \in [0, 1]\) without using it explicitly. This algorithm can be viewed as a generalization of the universal method (6.10) in [5]. Looking at the efficiency bound (3.19), for \(\nu\) known and \(\nu\) unknown, we see that the universal scheme ensures the same dependence on the accuracy \(\epsilon\) as the non-universal scheme (\(\alpha = \nu \neq 1\)). Remarkably, this is not true for the accelerated schemes obtained from the standard estimating sequences technique, as we will see in the next session.
4. Accelerated tensor schemes. Similarly to Section 3, we shall consider a general accelerated tensor method parametrized by the constant $\alpha$ given in (3.1). Specifically, at the beginning of the $t$th iteration ($t > 0$) one has an estimate $x_t$ for the solution of (2.1), an auxiliary vector $v_t$ and constants $A_t, M_t > 0$. A new vector $y_t$ is computed as a convex combination of $x_t$ and $v_t$:

$$y_t = (1 - \gamma_t)x_t + \gamma_tv_t,$$

where

$$\gamma_t = \frac{a_t}{A_t + a_t}$$

with $a_t > 0$ being computed from the equation

$$a_t^{p+\alpha} = \frac{1}{2(3p-1)!} \left[ \frac{(p-1)!}{M_t} \right] (A_t + a_t)^{p+\alpha-1}$$

Then, a trial point $x_t^+$ is computed as an approximate solution to the auxiliary problem

$$\min_{x \in \mathbb{R}} \Omega^{(\alpha)}_{y_t,p,M_t}(x),$$

such that

$$\Omega^{(\alpha)}_{y_t,p,M_t}(x^+_t) \leq f(y_t) \quad \text{and} \quad \|\nabla \Omega^{(\alpha)}_{y_t,p,M_t}(x^+_t)\|_* \leq \theta\|x^+_t - y_t\|^{p+\alpha-1},$$

where $\theta \geq 0$ is a user-defined parameter. If the descent condition

$$\langle \nabla f(x^+_t), y_t - x^+_t \rangle \geq \frac{1}{4} \left[ \frac{(p-1)!}{M_t} \right]^{1/p+\alpha-1} \|\nabla f(x^+_t)\|_*^{p+\alpha-1}$$

is satisfied, then $x^+_t$ is accepted, and we define $x_{t+1} = x^+_t$. Otherwise, constant $M_t$ is increased until the corresponding trial point $x^+_t$ is accepted. As in Algorithm 1, we assume that there exists $M_0 > 0$ such that $M_t \leq M_0$ for all $t$. After obtaining $x_{t+1}$, we set $A_{t+1} = A_t + a_t$ and compute

$$v_{t+1} = \arg \min_{x \in \mathbb{R}} \psi_{t+1}(x),$$

where

$$\psi_{t+1}(x) = \psi_t(x) + a_t [f(x_{t+1}) + \langle \nabla f(x_{t+1}), x - x_{t+1} \rangle].$$

To initialize, we choose $x_0$ and we set $v_0 = x_0$, $A_0 = 0$ and $\psi_0(x) = \frac{1}{p+\alpha} \|x-x_0\|^{p+\alpha}$. This general scheme can be summarized in the following way.

**Algorithm 3. Accelerated Tensor Method**

**Step 0.** Choose $x_0 \in \mathbb{R}$, $H_0 > 0$. Set $\alpha$ by (3.1), $v_0 = x_0$, $A_0 = 0$ and $t := 0$.

**Step 1.** Find $0 < M_t \leq M_0$ such that (4.6) holds for an approximate solution $x^+_t$ to (4.4) satisfying (4.5), with $y_t$ being defined by (4.1)-(4.3).

**Step 2.** Set $x_{t+1} = x^+_t$ and $A_{t+1} = A_t + a_t$ with $a_t > 0$ obtained from (4.3).

**Step 3.** Define $\psi_{t+1}(\cdot)$ by (4.8) and compute $v_{t+1}$ by (4.7).

**Step 4.** Set $t := t + 1$ and go back to Step 1.
The next result establishes the relationship between the estimating functions $\psi_t(\cdot)$ and the objective function $f(\cdot)$.

**Lemma 4.1.** For all $t \geq 0$,

$$
(4.9) \quad \psi_t(x) \leq A_t f(x) + \frac{1}{(p+\alpha)} \|x - x_0\|^{p+\alpha}, \forall x \in \mathbb{E}.
$$

*Proof.* We prove this result by induction in $t$. Since $A_0 = 0$, for all $x \in \mathbb{E}$

$$
\psi_0(x) = \frac{1}{(p+\alpha)} \|x - x_0\|^{p+\alpha} = A_0 f(x) + \frac{1}{(p+\alpha)} \|x - x_0\|^{p+\alpha},
$$

that is, (4.9) is true for $t = 0$. Suppose that (4.9) is true for some $t \geq 0$. Then, (4.8) and convexity of $f$ imply that, for all $x \in \mathbb{E},$

$$
\begin{align*}
\psi_{t+1}(x) &= \psi_t(x) + a_t \left[ f(x_{t+1}) + (\nabla f(x_{t+1}), x - x_{t+1}) \right] \\
&\leq \psi_t(x) + a_t f(x) \\
&\leq (A_t + a_t) f(x) + \frac{(p+\alpha) \|x - x_0\|^{p+\alpha}}{(p+\alpha)} \\
&= A_{t+1} f(x) + \frac{(p+\alpha) \|x - x_0\|^{p+\alpha}}{(p+\alpha)}.
\end{align*}
$$

Thus, (4.9) is also true for $t + 1$, and the proof is completed. \(\square\)

The theorem below establishes the global convergence rate for Algorithm 3.

**Theorem 4.2.** Assume that $H1$ is true and let the sequence $\{x_t\}_{t=0}^T$ be generated by Algorithm 3. Then, for $t = 2, \ldots, T$,

$$
(4.10) \quad f(x_t) - f(x^*) \leq \frac{2^{3p-1} M_r (p + \alpha)^{p+\alpha-1} \|x_0 - x^*\|^{p+\alpha}}{(p-1)! (t-1)^{p+\alpha}},
$$

*Proof.* Let us prove by induction that

$$
(4.11) \quad A_t f(x_t) \leq \psi_t^* \equiv \min_{x \in \mathbb{E}} \psi_t(x).
$$

Since $A_0 = 0$, we have $A_0 f(x_0) = 0 = \min_{x \in \mathbb{E}} \psi_0(x)$. Thus, (4.11) is true for $t = 0$. Assume that it is true for some $t \geq 0$. Note that, for any $x \in \mathbb{E}$ we have

$$
\psi_t(x) = \sum_{i=0}^{t-1} a_i \left[ f(x_{i+1}) + (\nabla f(x_{i+1}), x - x_{i+1}) \right] + \frac{(p+\alpha) \|x - x_0\|^{p+\alpha}}{p+\alpha} \\
\equiv \ell_t(x) + \frac{(p+\alpha) \|x - x_0\|^{p+\alpha}}{p+\alpha}, \forall t \geq 1.
$$

Note that $\ell_t(x)$ is a linear function. Moreover, by Lemma 4 in [10], function $\frac{1}{(p+\alpha)} \|x - x_0\|^{p+\alpha}$ is uniformly convex of degree $p+\alpha$ with parameter $2^{-(p+\alpha-2)}$. Thus, $\psi_t(x)$ is also a uniformly convex function of degree $p + \alpha$ with parameter $2^{-(p+\alpha-2)}$. Therefore, Lemma A.2 in [6] and the induction assumption imply that

$$
\psi_t(x) \geq \psi_t^* + \frac{2^{-(p+\alpha-1)}}{p+\alpha} \|x - v_t\|^{p+\alpha} \geq A_t f(x_t) + \frac{2^{-(p+\alpha-1)}}{p+\alpha} \|x - v_t\|^{p+\alpha}.
$$

Thus,

$$
\begin{align*}
\psi_{t+1}^* &= \min_{x \in \mathbb{E}} \{ \psi_t(x) + a_t \left[ f(x_{t+1}) + (\nabla f(x_{t+1}), x - x_{t+1}) \right] \} \\
&\geq \min_{x \in \mathbb{E}} \{ A_t f(x_t) + \frac{2^{-(p+\alpha-2)}}{(p+\alpha)} \|x - v_t\|^{p+\alpha} \\
&\quad + a_t \left[ f(x_{t+1}) + (\nabla f(x_{t+1}), x - x_{t+1}) \right] \}.
\end{align*}
$$
Since $f$ is convex and differentiable, we have
\[ f(x_t) \geq f(x_{t+1}) + \langle \nabla f(x_{t+1}), x_t - x_{t+1} \rangle. \]

Then, substituting this inequality above, we obtain
\[
\psi^*_{t+1} \geq \min_{x \in \mathbb{E}} \{ A_{t+1} f(x_{t+1}) + \langle \nabla f(x_{t+1}), A_t x_t - A_t x_{t+1} \rangle + a_t \langle \nabla f(x_{t+1}), x - x_{t+1} \rangle + 2^{-\frac{p+\alpha-1}{p+\alpha}} \| x - v_t \|^\frac{p+\alpha}{p+\alpha}. \]

Note that $y_t = (1 - \gamma_t)x_t + \gamma_t v_t = \frac{A_t}{A_{t+1}} x_t + \frac{\alpha}{A_{t+1}} v_t$. Therefore, $A_t x_t = A_{t+1} y_t - a_t v_t$, and
\[
\psi^*_{t+1} \geq \min_{x \in \mathbb{E}} \{ A_{t+1} f(x_{t+1}) + \langle \nabla f(x_{t+1}), A_{t+1} y_t - a_t v_t - A_t x_{t+1} \rangle + a_t \langle \nabla f(x_{t+1}), x - x_{t+1} \rangle + 2^{-\frac{p+\alpha-1}{p+\alpha}} \| x - v_t \|^\frac{p+\alpha}{p+\alpha}. \]

Moreover, $A_{t+1} x_{t+1} = A_t x_t + a_t x_{t+1}$, and so
\[
\psi^*_{t+1} \geq \min_{x \in \mathbb{E}} \{ A_{t+1} f(x_{t+1}) + A_{t+1} \langle \nabla f(x_{t+1}), y_t - x_{t+1} \rangle + a_t \langle \nabla f(x_{t+1}), x - v_t \rangle + 2^{-\frac{p+\alpha-1}{p+\alpha}} \| x - v_t \|^\frac{p+\alpha}{p+\alpha} \},
\]
where the last inequality is due to (4.6). Thus, to prove that (4.11) is true for $t + 1$, it is enough to show that
\[
A_{t+1} \frac{1}{4} \left[ \frac{(p-1)!}{M_t} \right]^{\frac{1}{p-1}} \| \nabla f(x_{t+1}) \|_{\frac{1}{p-1}}^{\frac{p+\alpha}{p-1}} + a_t \langle \nabla f(x_{t+1}), x - v_t \rangle + 2^{-\frac{p+\alpha-1}{p+\alpha}} \| x - v_t \|^\frac{p+\alpha}{p+\alpha} \geq 0
\]
for all $x \in \mathbb{E}$. Using Lemma 2 in [10] with $r = p + \nu$, $s = a_t \nabla f(x_{t+1})$ and $\omega = 2^{-\frac{p+\alpha-1}{p+\alpha}}$, we see that a sufficient condition for (4.12) is
\[
A_{t+1} \frac{1}{4} \left[ \frac{(p-1)!}{M_t} \right]^{\frac{1}{p-1}} \| \nabla f(x_{t+1}) \|_{\frac{1}{p-1}}^{\frac{p+\alpha}{p-1}} \geq \left( \frac{p+\alpha-1}{p+\alpha} \right) \frac{2}{p+\alpha} a_t^{\frac{p+\alpha-2}{p+\alpha}} \| \nabla f(x_{t+1}) \|_{\frac{1}{p+\alpha-1}}^{\frac{p+\alpha}{p+\alpha-1}},
\]
which is equivalent to
\[
a_t^{p+\alpha} \leq 2 \left( \frac{p+\alpha}{p+\alpha-1} \right)^{p+\alpha-1} \left( \frac{1}{8} \right)^{p+\alpha-1} \left[ \frac{(p-1)!}{M_t} \right] A_{t+1}^{p+\alpha-1}.
\]

Note that, $2 \left( \frac{p+\alpha}{p+\alpha-1} \right)^{p+\alpha-1} \left( \frac{1}{8} \right)^{p+\alpha-1} \geq \frac{1}{2(p+\alpha-1)}$. Therefore, by (4.3) we have
\[
a_t^{p+\alpha} = \frac{1}{2(p+\alpha-1)} \left[ \frac{(p-1)!}{M_t} \right] (A_t + a_t)^{p+\alpha-1}
\]
\[
\leq 2 \left( \frac{p+\alpha}{p+\alpha-1} \right)^{p+\alpha-1} \left( \frac{1}{8} \right)^{p+\alpha-1} \left[ \frac{(p-1)!}{M_t} \right] A_{t+1}^{p+\alpha-1}.
\]
Thus (4.11) is true for \( t + 1 \), completing the induction argument.

Let us now estimate the growth of the coefficients \( A_t \). Since \( M_t \leq M_\nu \) for all \( t = 0, \ldots, T \), by (4.3) we get \( a_t^{p+\alpha} \geq \frac{1}{M} (A_t + a_t)^{p+\alpha-1} \) with

\[
\tilde{M} = \frac{2^{(3p-1)} M_\nu}{(p-1)!}.
\]

Consequently,

\[
A_{t+1} - A_t = a_t \geq \left( \frac{1}{\tilde{M}} \right)^{\frac{p+\alpha-1}{p+\alpha}} A_t^{\frac{p+\alpha-1}{p+\alpha}+\alpha}. \tag{4.14}
\]

Now, denoting \( B_t = \tilde{M} A_t \) for all \( t \geq 0 \), it follows from (4.14) that,

\[
B_{t+1} - B_t \geq B_t^{\frac{p+\alpha-1}{p+\alpha}}. \tag{4.14}
\]

Then, by Lemma A.4 in [6], we have

\[
B_t \geq \left[ \left( \frac{1}{p+\alpha} \right) \left( \frac{B_t^{\frac{1}{p+\alpha}}}{B_t^{\frac{1}{p+\alpha}+1}} \right) \right]^{p+\alpha} (t-1)^{p+\alpha} \quad \forall t \geq 2.
\]

Note that \( A_1 \geq \frac{1}{2 \tilde{M}}. \) Thus, \( B_1 \geq 1 \) and consequently

\[
B_t \geq \left[ \left( \frac{1}{p+\alpha} \right) \left( \frac{1}{2} \right) \right]^{p+\alpha} (t-1)^{p+\alpha}.
\]

Therefore, for all \( t \geq 2 \), we have

\[
A_t \geq \frac{1}{M} \left[ \left( \frac{1}{p+\alpha} \right) \left( \frac{1}{2} \right) \right]^{p+\alpha} (t-1)^{p+\alpha}. \tag{4.15}
\]

Finally, by (4.11) and Lemma 4.1, for \( t \geq 0 \), we have

\[
A_t f(x_t) \leq \psi^* \leq A_t f(x^*) + \frac{1}{p+\alpha} \| x^* - x_0 \|^{p+\alpha}.
\]

Hence, \( A_t (f(x_t) - f(x^*)) \leq \frac{1}{2p+1} \| x^* - x_0 \|^{2+\nu} \), and (4.10) follows immediately from (4.13) and (4.15).

If we assume that \( \nu \) and \( H_{f,p}(\nu) \) are known, then, by Lemma A.6, we can set

\[
M_t = M_\nu \equiv (p + \nu - 1)(H_{f,p}(\nu) + \theta(p-1)!).
\]

In this case, by (3.1), the corresponding version of Algorithm 3 takes at most \( O(\epsilon^{-1/(p+\nu)}) \) iterations to generate \( x_t \) such that \( f(x_t) - f(x^*) \leq \epsilon. \) For problems in which \( H_{f,p}(\nu) \) is not known, let us consider the following adaptive version of Algorithm 3:
Algorithm 4. Adaptive Accelerated Tensor Method

Step 0. Choose \( x_0 \in \mathbb{R} \), \( H_0 > 0 \) and \( \theta \geq 0 \). Set \( \alpha \) by (3.1) and define function \( \psi_0(x) = \frac{1}{p+\alpha} \| x - x_0 \|^{p+\alpha} \). Set \( v_0 = x_0 \), \( A_0 = 0 \) and \( t := 0 \).

Step 1. Set \( i := 0 \).

Step 1.1. Compute the coefficient \( a_{t,i} > 0 \) by solving equation

\[
a_{t,i}^{p+\alpha} = \frac{1}{2(3p-1)} \left( \frac{(p-1)!}{2^{p-1}H_t} \right) (A_t + a_{t,i})^{p+\alpha-1}.
\]

Step 1.2. Set \( \gamma_{t,i} = \frac{a_{t,i}}{A_t + a_{t,i}} \) and compute vector \( y_{t,i} = (1 - \gamma_{t,i})x_t + \gamma_{t,i}v_t \).

Step 1.3 Compute an approximate solution \( x_{t,i}^+ \) to \( \min_{x \in \mathbb{R}} \Omega^{(\alpha)}_{y_{t,i},p,2^iH_t}(x) \), such that

\[
\Omega^{(\alpha)}_{y_{t,i},p,2^iH_t}(x_{t,i}^+) \leq f(y_{t,i}) \quad \text{and} \quad \| \nabla \Omega^{(\alpha)}_{y_{t,i},p,2^iH_t}(x_{t,i}^+) \|_* \leq \theta \| x_{t,i}^+ - y_{t,i} \|^{p+\alpha-1}.
\]

Step 1.4. If condition

\[
\langle \nabla f(x_{t,i}^+), y_{t,i} - x_{t,i}^+ \rangle \geq \frac{1}{4} \left[ \frac{(p-1)!}{2^{p-1}H_t} \right] \frac{1}{p+\alpha} \| \nabla f(x_{t,i}^+) \|_e^{p+\alpha},
\]

set \( i_t := i \) and go to Step 2. Otherwise, set \( i := i + 1 \) and go back to Step 1.1.

Step 2. Set \( x_{t+1} = x_{i_t,i_t}^+ \), \( y_{t+1} = y_{i_t,i_t} \), \( a_t = a_{i_t,i_t} \) and \( \gamma_t = \gamma_{i_t,i_t} \). Define \( A_{t+1} = A_t + a_t \) and \( H_{t+1} = 2^{i_t-1}H_t \).

Step 3. Define \( \psi_{t+1}(\cdot) \) by (4.8) and compute \( v_{t+1} \) by (4.7).

Step 4. Set \( t := t + 1 \) and go back to Step 1.

Note that Algorithm 4 is a particular case of Algorithm 3 in which

\[ M_t = 2^i H_t, \quad \forall t \geq 0. \]

Let us define the following function of \( \epsilon > 0 \):

\[
(4.16) \quad \tilde{N}_\nu(\epsilon) = \begin{cases} 
(p + \nu - 1)(H_{f,p}(\nu) + \theta(p - 1)!), & \text{if } \alpha = \nu, \\
\max \left\{ \theta(p - 1)!, \frac{4R(\epsilon)}{\epsilon} \right\}^{\frac{\nu - \alpha}{\nu + 1}}, & \text{if } \alpha = 1.
\end{cases}
\]

The next lemma provides upper bounds on \( H_t \) and on the number of calls of the oracle in Algorithm 4.

**Lemma 4.3.** Suppose that H1 and H2 are true. Given \( \epsilon > 0 \), assume that \( \{x_t\}_{t=0}^T \) is a sequence generated by Algorithm 4 such that

\[
(4.17) \quad f(x_0) - f(x^*) \geq \epsilon,
\]

and

\[
(4.18) \quad f(x_{i_t,i_t}^+) - f(x^*) \geq \epsilon, \quad i = 0, \ldots, i_t \text{ and } t = 0, \ldots, T.
\]

Then,

\[
(4.19) \quad H_t \leq \max \left\{ H_0, \tilde{N}_\nu(\epsilon) \right\}, \quad \text{for } t = 0, \ldots, T.
\]
Moreover, the number $O_T$ of calls of the oracle after $T$ iterations is bounded as follows:

\begin{equation}
O_T \leq 2T + \log_2 \max \left\{ H_0, \tilde{N}_\nu(\epsilon) \right\} - \log_2 H_0.
\end{equation}

Proof. Let us prove by induction that the scaling coefficients $H_t$ in Algorithm 4 satisfy (4.19). This is obvious for $t = 0$. Assume that (4.19) is true for some $t \geq 0$. If $\alpha = \nu$, it follows from Lemma A.6 that the final value $2^\nu H_t$ cannot be bigger than

$$2 \left[ (p + \nu - 1)(H_{f,p}(\nu) + \theta(p - 1)! \right],$$

since otherwise we should stop the line-search earlier. Thus,

$$H_{t+1} = \frac{1}{2} 2^\nu H_t \leq (p + \nu - 1)(H_{f,p}(\nu) + \theta(p - 1)! \leq \max \left\{ \tilde{N}_\nu(\epsilon), H_0 \right\},$$

that is, (4.20) holds for $t + 1$. On the other hand, suppose that $\alpha = 1$. In view of Lemma A.8, at any trial point $t_{i,t}$ we have

$$\| \nabla f(x_{i,t}) \|_* \geq \frac{\epsilon}{R(\epsilon)}.$$ 

Thus, it follows from Lemma A.7 that

$$2^\nu H_t \leq 2 \max \left\{ 4\theta(p - 1)!, (4H_{f,p}(\nu))^{p+\nu} \left( \frac{4R(\epsilon)}{\epsilon} \right)^{\frac{1}{p+\nu}} \right\} \leq 2 \max \left\{ \tilde{N}_\nu(\epsilon), H_0 \right\}.$$ 

Consequently, we also have $H_{t+1} \leq \max \{ M_\nu(\epsilon), H_0 \}$, i.e., (4.19) holds for $t + 1$. This completes the induction argument. Finally, as in the proof of Lemma 3.2, from (4.19) we get (4.20). \hspace{1cm} \Box

Now, we can prove the following convergence result for Algorithm 4.

Theorem 4.4. Suppose that H1 and H2 are true. Given $\epsilon \in (0, 1)$, assume that $\{x_t\}_{t=0}^T$ is a sequence generated by Algorithm such that (4.17) and (4.18) hold. Then, (4.21)

$$f(x_t) - f(x^*) \leq 2^{3p} \max \left\{ \tilde{N}_\nu(\epsilon), H_0 \right\} \frac{(p + \nu)p^{p+\nu-1} \| x_0 - x^* \|^{p+\nu} \| x_0 - x^* \|^{p+\nu}}{(p - 1)! (t - 1)^{p+\nu}}, \hspace{1cm} 2 \leq t \leq T.$$ 

Consequently,

\begin{equation}
T \leq 1 + \left[ \frac{2^{3p} \max \{ H_0, (p + \nu - 1)(H_{f,p}(\nu) + \theta(p - 1)!) \} (p + \nu)^{p+\nu-1} \| x_0 - x^* \|^{p+\nu}}{(p - 1)!} \right] \frac{\epsilon}{\nu + 1} \left( \frac{1}{\epsilon} \right)^{\frac{1}{p+\nu}}
\end{equation}

if $\nu$ is known (i.e., $\alpha = \nu$), and

\begin{equation}
T \leq 1 + \left[ \frac{2^{3p} \max \{ H_0, 4\theta(p - 1)!, (4H_{f,p}(\nu))^{p+\nu} (4D_0)^{\frac{1}{p+\nu}} \} (p + 1)^p \| x_0 - x^* \|^{p+1}}{(p - 1)!} \right] \frac{\epsilon}{\nu + 1} \left( \frac{1}{\epsilon} \right)^{\frac{p}{p+\nu+1}}
\end{equation}

if $\nu$ is unknown (i.e., $\alpha = 1$).

Proof. By Lemma 4.3, we have

$$2^\nu H_t = 2(2^\nu H_t) = 2H_{t+1} \leq 2 \max \left\{ H_0, \tilde{N}_\nu(\epsilon) \right\}, \hspace{1cm} t = 0, \ldots, T - 1.$$
Then, (4.21) follows directly from Theorem 4.2 with

\[ \nu \in 2 \max \left\{ H_0, \tilde{N}_\nu (\epsilon) \right\}. \]

Now, combining (4.21) and (4.18) for \( k = T \), we obtain

\[ \epsilon \leq \frac{2^p \max \left\{ H_0, \tilde{N}_\nu (\epsilon) \right\} (p + \alpha)^{p+\alpha-1} \| x_0 - x^* \|^{p+\alpha}}{(p-1)! (T-1)^{p+\alpha}} \]

and so,

\[ (4.24) \quad T \leq 1 + \left[ \frac{2^p \max \left\{ H_0, \tilde{N}_\nu (\epsilon) \right\} (p + \alpha)^{p+\alpha-1} \| x_0 - x^* \|^{p+\alpha}}{\epsilon (p-1)!} \right]^{\frac{1}{p+\alpha}}. \]

If \( \nu \) is known, then \( \alpha = \nu \) and, by (4.16), we have

\[ (4.25) \quad \tilde{N}_\nu (\epsilon) = (p + \nu - 1)(H_{f,p}(\nu) + \theta (p-1)!). \]

Thus, combining (4.24) and (4.25), we get (4.22). On the other hand, if \( \nu \) is unknown, then \( \alpha = 1 \) and, by (4.16), (3.16) and \( \epsilon \in (0,1) \), we have

\[ \tilde{N}_\nu (\epsilon) = \max \left\{ 4\theta (p-1)!, (4H_{f,p}(\nu))^{\frac{p}{p^\nu-1}} \left( \frac{4R(\epsilon)}{\epsilon} \right)^{\frac{1}{p^\nu-1}} \right\} \]

\[ (4.26) \quad \leq \max \left\{ 4\theta (p-1)!, (4H_{f,p}(\nu))^{\frac{p}{p^\nu-1}} (4D_0)^{\frac{1}{p^\nu-1}} \right\} \epsilon^{-\frac{1}{p+\nu}}. \]

In this case, combining (4.24) and (4.26) we get (4.23). \( \square \)

5. Lower complexity bounds under Hölder condition. In this section we investigate how much the convergence rates of our tensor methods can be improved with respect to problems satisfying H1. Specifically, we derive lower complexity bounds for \( p \)-order tensor methods applied to the problem (2.1), where the objective \( f \) is convex and \( H_{f,p}(\nu) < +\infty \) for some \( \nu \in [0,1] \).

5.1. Hard functions and Lower Complexity Bounds. For simplicity, let us consider \( E = \mathbb{R}^n \) and \( B = I_n \). Given an approximation \( \tilde{x} \) for the solution of (2.1), \( p \)-order methods usually compute the next test point as \( x^+ = \tilde{x} + h \), where the search direction \( h \) is the solution of an auxiliary problem of the form

\[ (5.1) \quad \min_{h \in \mathbb{R}_+^p} \phi_{a,\gamma,m}(h) \equiv \sum_{i=1}^{p} a^{(i)} D^i f(\tilde{x}) [h]^i + \gamma \| h \|^m, \]

with \( a \in \mathbb{R}^p, \gamma > 0 \) and \( m > 1 \). Denote by \( \Gamma_{\tilde{x},f}(a,\gamma,m) \) the set of all stationary points of function \( \phi_{a,\gamma,m}(\cdot) \) and define the linear subspace

\[ (5.2) \quad S_f(\tilde{x}) = \text{Lin} (\Gamma_{\tilde{x},f}(a,\gamma,m) \mid a \in \mathbb{R}^p, \gamma > 0, m > 1). \]
With this notation, we can characterize the class of $p$-order tensor methods by the following assumption.

**Assumption 1.** Given $x_0 \in \mathbb{R}^n$, the method generates a sequence of test points \( \{x_k\}_{k \geq 0} \) such that

\[
(5.3) \quad x_{k+1} \in x_0 + \sum_{i=0}^{k} S_f(x_i), \quad k \geq 0.
\]

Given \( \nu \in [0, 1] \), our parametric family of difficult functions for $p$-order tensor methods is defined as

\[
(5.4) \quad f_k(x) = \frac{1}{p + \nu} \left[ \sum_{i=1}^{k-1} |x^{(i)} - x^{(i+1)}|^{p+\nu} + \sum_{i=k}^{n} |x^{(i)}|^{p+\nu} \right] - x^{(1)}, \quad 2 \leq k \leq n.
\]

The next lemma establishes that for each \( f_k(.) \) we have \( H_{f_k,p}(\nu) < +\infty \).

**Lemma 5.1.** Given an integer \( k \in [2, n] \), the \( p \)th derivative of \( f_k(.) \) is \( \nu \)-Hölder continuous with

\[
(5.5) \quad H_{f_k,p}(\nu) = 2^{\frac{2+\nu}{p-\nu}} \prod_{i=1}^{p-1} (p + \nu - i).
\]

**Proof.** In view of (5.4) we have

\[
(5.6) \quad f_k(x) = \eta_{p+\nu}(A_k x) - \langle e_1, x \rangle,
\]

where

\[
(5.7) \quad \eta_{p+\nu}(u) = \frac{1}{p + \nu} \sum_{i=1}^{n} |u^{(i)}|^{p+\nu},
\]

\[
(5.8) \quad A_k = \begin{pmatrix} U_k & 0 \\ 0 & I_{n-k} \end{pmatrix}, \quad \text{with} \quad U_k = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -1 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}.
\]

It can be shown that (see page 13 in [11]):

\[
(5.9) \quad \|A_k\| \leq 2.
\]

On the other hand, for any \( x, h \in \mathbb{R}^n \), we have

\[
D^\ell \eta_{p+\nu}(x)[h]^\ell = \begin{cases} \left( \frac{\Pi_{i=0}^{\ell-1}(p + \nu - i)}{p + \nu} \right) \sum_{i=1}^{n} |x^{(i)}|^{p+\nu-\ell}(h^{(i)})^\ell, & \text{if } \ell \text{ is even}, \\ \left( \frac{\Pi_{i=0}^{\ell-1}(p + \nu - i)}{p + \nu} \right) \sum_{i=1}^{n} |x^{(i)}|^{p+\nu-1-\ell}x^{(i)}(h^{(i)})^\ell, & \text{if } \ell \text{ is odd}. \end{cases}
\]
Therefore, for all \( x, y, h \in \mathbb{R}^n \), it follows that
\[
|D^p\eta_{p+\nu}(x)[h]^p - D^p\eta_{p+\nu}(y)[h]^p| \leq \left( \Pi_{i=1}^{p-1}(p + \nu - i) \right) \sum_{i=1}^{n} |x^{(i)} - y^{(i)}|^\nu (h^{(i)})^p
\leq \left( \Pi_{i=1}^{p-1}(p + \nu - i) \right) \|x - y\|_\infty \sum_{i=1}^{n} (h^{(i)})^p
\leq \left( \Pi_{i=1}^{p-1}(p + \nu - i) \right) \|x - y\|_\infty \sum_{i=1}^{n} \left[ (h^{(i)})^2 \right]^\frac{\nu}{2}
\leq \left( \Pi_{i=1}^{p-1}(p + \nu - i) \right) \|x - y\|_\infty \|h\|^p.
\]

Consequently, for all \( x, d, h \in \mathbb{R}^n \), we have
\[
|D^p f_k(x + d)[h]^p - D^p f_k(x)[h]^p| = |D^p\eta_{p+\nu}(A_k(x + d))[A_k h]^p - D^p\eta_{p+\nu}(A_k x)[A_k h]^p|
\leq \Pi_{i=1}^{p-1}(p + \nu - i) \|A_k d\|_\infty \|A_k h\|^p.
(5.10)
\]

Note that
\[
\|A_k d\|_\infty = \max_{1 \leq i \leq n} |(A_k d)^{(i)}| \leq \max_{1 \leq i \leq n-1} \left( |d^{(i)}| + |d^{(i+1)}| \right)
\leq \max_{1 \leq i \leq n-1} \sqrt{2[(d^{(i)})^2 + (d^{(i+1)})^2]}
\leq 2 \sqrt{\frac{2}{\max_i |d^{(i)}|^2}},
\] and, by (5.9),
\[
\|A_k h\| \leq \|A\| \|h\| \leq 2 \|h\|.
(5.12)
\]

Thus, combining (5.10)-(5.12), we get
\[
\|D^p f_k(x + d) - D^p f_k(x)\| \leq 2 \sqrt{\frac{2}{\max_i |d^{(i)}|^2}} \Pi_{i=1}^{p-1}(p + \nu - i) \|d\|^{\nu}.
\]

The next lemma provides additional properties of \( f_k(\cdot) \).

**Lemma 5.2.** Given an integer \( k \in [2, n] \), let function \( f_k(\cdot) \) be defined by (5.4). Then, \( f_k(\cdot) \) has a unique global minimizer \( x_k^* \). Moreover,
\[
f_k^* = -\frac{(p + \nu - 1)k}{p + \nu} \quad \text{and} \quad \|x_k^*\| < \frac{(k + 1)^{\frac{\nu}{2}}}{\sqrt{3}}.
(5.13)
\]

**Proof.** The existence and uniqueness of \( x_k^* \) follows from the fact that \( f_k(\cdot) \) is uniformly convex. In view of (5.6), it follows from the first order optimality condition that
\[
A_k^T \nabla\eta_{p+\nu}(A_k x_k^*) = e_1.
\]
Therefore, \( A_k x_k^* = y_k^* \), where \( y_k^* \) satisfies
\[
\nabla\eta_{p+\nu}(y_k^*) = A_k^T e_1 = \hat{e}_k = \begin{bmatrix} \bar{e}_k \\ 0_{n-k} \end{bmatrix}
(5.14)
\]
with $\hat{e}_k \in \mathbb{R}^k$ being the vector of all ones, and $0_{n-k}$ being the origin in $\mathbb{R}^{n-k}$. Note that

$$\frac{\partial \eta_{p+\nu}}{\partial y_i}(y) = |y^{(i)}|^{p+\nu-2} y^{(i)}; \quad i = 1, \ldots, n.$$  
(5.15)

Consequently, (5.14) is equivalent to

$$|(y_k^*)^{(i)}|^{p+\nu-2}(y_k^*)^{(i)} = \begin{cases} 1, & \text{for } i = 1, \ldots, k, \\ 0, & \text{for } i = k+1, \ldots, n. \end{cases}$$

Thus,

$$(5.16) \quad A_k x_k^* = y_k^* = \hat{e}_k,$$

and so

$$(5.17) \quad (x_k)^{(i)} = (A_k^{-1} y_k^*)^{(i)} = (A_k^{-1} \hat{e}_k)^{(i)} = (k-i+1)_{+}, \quad i = 1, \ldots, n,$$

where $(\tau)_+ = \max \{0, \tau\}$. Finally, combining (5.6), (5.7), (5.16) and (5.17) we get

$$f_k^* = \eta_{p+\nu}(A_k x_k^*) - \langle e_1, x_k^* \rangle = \eta_{p+\nu}(\hat{e}_k) - (x_k^*)^{(1)} = \frac{1}{p+\nu} \sum_{i=1}^{n} |(\hat{e}_k)^{(i)}|^{p+\nu} - k = \frac{k}{p+\nu} - k = -\frac{(p+\nu-1)k}{p+\nu},$$

$$\|x_k^*\|^2 = \sum_{i=1}^{n} [(x_k^*)^{(i)}]^2 = k^2 + (k-1)^2 + \ldots + 2^2 + 1^2$$

$$= \sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6} < \frac{(k+1)^3}{3}.$$

Our goal is to understand the behavior of the tensor methods specified by Assumption 1 when applied to the minimization of $f_k(\cdot)$ with a suitable $k$. For that, let us consider the following subspaces:

$$\mathbb{R}_{k}^{n} = \left\{ x \in \mathbb{R}^n \mid x^{(i)} = 0, \quad i = k + 1, \ldots, n \right\}, \quad 1 \leq k \leq n - 1.$$

**Lemma 5.3.** For any $q \geq 0$ and $x \in \mathbb{R}_{k}^{n}$, $f_{k+q}(x) = f_k(x)$.

*Proof.* It follows directly from (5.4). $\square$

**Lemma 5.4.** Let $\mathcal{M}$ be a $p$-order tensor method satisfying Assumption 1. If $\mathcal{M}$ is applied to the minimization of $f_k(\cdot)$ starting from $x_0 = 0$, then the sequence $\{x_k\}_{k \geq 0}$ of test points generated by $\mathcal{M}$ satisfies

$$x_{k+1} \in \bigcap_{i=0}^{k} S_{f_i}(x_i) \subset \mathbb{R}_{k+1}^{n}, \quad 0 \leq k \leq t - 1.$$ 

*Proof.* See Lemma 2 in [11]. $\square$

Now, we can prove the lower complexity bound for $p$-order tensor methods applied to the minimization of functions with $\nu$-Hölder continuous $p$th derivatives.
\textbf{Theorem 5.5.} Let $\mathcal{M}$ be a $p$-order tensor method satisfying Assumption 1. Assume that for any function $f$ with $H_{f,p}(\nu) < +\infty$ this method ensures the rate of convergence:

\begin{equation}
\min_{0 \leq k \leq t} f(x_k) - f^* \leq \frac{H_{f,p}(\nu)\|x_0 - x^*\|^{p+\nu}}{\kappa(t)}, \quad t \geq 1,
\end{equation}

where $\{x_k\}_{k \geq 0}$ is the sequence generated by method $\mathcal{M}$ and $x^*$ is a global minimizer of $f$. Then, for all $t \geq 1$ such that $2t + 1 \leq n$ we have

\begin{equation}
\kappa(t) \leq C_{p,\nu}(t + 1)^{\frac{3(p+\nu)-2}{2}},
\end{equation}

where

\begin{equation}
C_{p,\nu} = \frac{2^{3\nu+2} \Pi_{i=0}^{p-1}(p + \nu - i)}{3^{\frac{p+\nu}{2}}(p + \nu - 1)}.
\end{equation}

\textbf{Proof.} Let us apply method $\mathcal{M}$ for minimizing function $f_{2t+1}(\cdot)$ starting from point $x_0 = 0$. By Lemma 5.4 we have $x_i \in \mathbb{R}^n_i$ for all $i$, $0 \leq i \leq t$. Moreover, by Lemma 5.3 we have

\begin{equation}
f_{2t+1}(x) = f_t(x), \quad \forall x \in \mathbb{R}^n.
\end{equation}

Thus, from (5.18), (5.21), Lemma 5.1 and Lemma 5.2 we get

\begin{align*}
\kappa(t) &\leq \frac{H_{f_{2t+1},p}(\nu)\|x_0 - x^*\|^{p+\nu}}{\min_{0 \leq k \leq t} f_{2t+1}(x_k) - f^*_{2t+1}} = \frac{2^{2\nu} \Pi_{i=1}^{p-1}(p + \nu - i)\|x^*_t\|^{p+\nu}}{\min_{0 \leq k \leq t} f_t(x_k) - f^*_t} \\
&\leq \frac{2^{2\nu} \Pi_{i=1}^{p-1}(p + \nu - i)\|x^*_t\|^{p+\nu}}{f_t - f^*_t} \leq \frac{2^{2\nu} \Pi_{i=1}^{p-1}(p + \nu - i)(2(t + 1))^{\frac{1}{2}(p+\nu)}}{3^{\frac{p+\nu}{2}}(f_t^* - f^*_t)} \\
&= \frac{2^{3\nu+2} \Pi_{i=0}^{p-1}(p + \nu - i)(2(t + 1))^{\frac{3(p+\nu)}{2}}}{3^{\frac{p+\nu}{2}}(p + \nu - 1)(t + 1)} = C_{p,\nu}(t + 1)^{\frac{3(p+\nu)-2}{2}},
\end{align*}

where constant $C_{p,\nu}$ is given by (5.20). \hfill \Box

\textbf{5.2. Discussion.} Theorem 5.5 establishes that the lower bound for the rate of convergence of tensor methods applied to functions with $\nu$-Hölder continuous $p$th derivatives is of $\mathcal{O}\left(\left(\frac{1}{\kappa}\right)^{\frac{3(p+\nu)-2}{2}}\right)$. In the Lipschitz case (i.e., $\nu = 1$) we have $\mathcal{O}\left(\left(\frac{1}{\kappa}\right)^{\frac{3+1}{2}}\right)$, which coincide with the bounds in [1, 11]. On the other hand, for first-order methods (i.e., $p = 1$) we have $\mathcal{O}\left(\left(\frac{1}{\kappa}\right)^{\frac{1+1}{2}}\right)$, which is the bound in [7].

The rate of $\mathcal{O}\left(\left(\frac{1}{\kappa}\right)^{\frac{3(p+\nu)-2}{2}}\right)$ corresponds to a worst-case complexity bound of $\mathcal{O}(\varepsilon^{-2/[(3(p+\nu)-2)]})$ iterations necessary to ensure $f(x_k) - f(x^*) \leq \varepsilon$. This means that the non-universal accelerated schemes proposed in this paper (e.g., Algorithm 4 with
\(\alpha = \nu\) are nearly optimal tensor methods. In fact, for \(\epsilon \in (0, 1)\), note that
\[
\left(\frac{1}{\epsilon}\right)^{\frac{1}{p-\nu}} = \left(\frac{1}{\epsilon}\right)^{\frac{p+\nu-2}{(p+\nu)(\beta+\nu)-2}} \left(\frac{1}{\epsilon}\right)^{\frac{2}{\beta(p+\nu)-2}} \leq \left(\frac{1}{\epsilon}\right)^{\frac{p-1}{3p^2-2p}} \left(\frac{1}{\epsilon}\right)^{\frac{2}{3(p+\nu)-2}}
\]
In particular, if \(\epsilon = 10^{-6}\), we have \(\left(\frac{1}{\epsilon}\right)^{\frac{1}{p-\nu}} \leq 6 \left(\frac{1}{\epsilon}\right)^{\frac{2}{3(p+\nu)-2}}\). Thus, in practice, the complexity bounds of our accelerated non-universal methods differ from the lower bound just by a small constant factor.

6. Conclusion. In this paper, we presented \(p\)-order methods for unconstrained minimization of convex functions that are \(p\)-times differentiable with \(\nu\)-Hölder continuous \(p\)th derivatives. For the universal and the non-universal schemes without acceleration, we established iteration complexity bounds of \(\mathcal{O}(\epsilon^{-1/(p+\nu-1)})\) for reducing the functional residual below a given \(\epsilon \in (0, 1)\). Assuming that \(\nu\) is known, we obtain an improved complexity bound of \(\mathcal{O}(\epsilon^{-1/(p+\nu)})\) for the corresponding accelerated scheme. For the case in which \(\nu\) is unknown, we present a accelerated universal tensor scheme with iteration complexity of \(\mathcal{O}(\epsilon^{-\nu/(p(\nu+1)/2)})\). Regarding the approximate solution of the auxiliary problems, it is easy to see that \(x_t^+\) satisfying (3.3) can be computed by any monotone optimization scheme that drives the gradient of the objective to zero. Moreover, if \(H_{f,p}(\alpha) < +\infty\) and \(M \geq (p-1)H_{f,p}(\alpha)\), we can show that \(\Omega^{(\alpha)}_{x,p,M}(\cdot)\) is convex for any \(x \in \mathbb{E}\) (as in the proof of Theorem 1 in [11]). Therefore, when \(M_t\) is sufficiently large, the computation of \(x_t^+\) satisfying (3.3) will be very fast, since the corresponding auxiliary optimization process can converge with a linear rate defined by an absolute constant.

Finally, a lower complexity bound of \(\mathcal{O}(\epsilon^{-2/(3(p+\nu)-2)})\) was also obtained for the referred problem class. This means that, in practice, our accelerated non-universal schemes are nearly optimal. Remarkably, the complexity bound obtained for the accelerated universal schemes is slightly worse than the bound obtained for the non-universal accelerated schemes. Up to now, it is not clear whether the estimating sequences technique can be modified to provide an accelerated universal \(p\)-order method with a complexity bound of \(\mathcal{O}(\epsilon^{-1/(p+\nu)})\).

REFERENCES

Appendix A. Auxiliary Results.

Lemma A.1. Let $H_{f,p}(\nu) < +\infty$ for some $\nu \in [0,1]$ and assume that $x^+$ satisfies

(A.1) $\Omega^{(\nu)}_{x,p,H}(x^+) \leq f(\bar{x}),$

for some $\bar{x} \in \mathbb{E}$ and $H > 0$. If $H \geq \frac{3}{2}H_{f,p}(\nu)$, then

(A.2) $f(\bar{x}) - f(x^+) \geq \frac{H}{(p+1)!} \|x^+ - \bar{x}\|^{p+\nu}.$

Proof. In view of (2.7) and (A.1), we have

\[
\begin{align*}
f(x^+) &\leq \Omega^{(\nu)}_{x,p,H}(x^+) \\
&= \Phi_{x,p}(x^+) + \frac{H}{p!}\|x^+ - \bar{x}\|^{p+\nu} - \frac{(H - H_{f,p}(\nu))}{p!}\|x^+ - \bar{x}\|^{p+\nu} \\
&= \Omega^{(\nu)}_{x,p,H}(x^+) - \frac{(H - H_{f,p}(\nu))}{p!}\|x^+ - \bar{x}\|^{p+\nu} \\
&\leq f(\bar{x}) - \frac{(H - H_{f,p}(\nu))}{p!}\|x^+ - \bar{x}\|^{p+\nu},
\end{align*}
\]

which gives

\[
f(\bar{x}) - f(x^+) \geq \frac{H - H_{f,p}(\nu)}{p!}\|x^+ - \bar{x}\|^{p+\nu}.
\]

Since $H \geq \frac{3}{2}H_{f,p}(\nu) \geq \frac{p+1}{p}H_{f,p}(\nu)$ for all $p \geq 2$, it follows that

\[
f(\bar{x}) - f(x^+) \geq \left(1 - \frac{p}{p+1}\right)\frac{H}{p!}\|x^+ - \bar{x}\|^{p+\nu} = \frac{H}{(p+1)!}\|x^+ - \bar{x}\|^{p+\nu}.
\]

Lemma A.2. Let $H_{f,p}(\nu) < +\infty$ for some $\nu \in [0,1]$ and assume that $x^+ \in \mathbb{E}$ satisfies (A.1) and

(A.3) $\|\nabla\Omega^{(\nu)}_{x,p,H}(x^+)\|_* \leq \theta \|x^+ - \bar{x}\|^{p+\nu-1},$

for some $\bar{x} \in \mathbb{E}$, $H > 0$ and $\theta \geq 0$. If

(A.4) $H \geq \max\left\{\frac{3H_{f,p}(\nu)}{2}, 3\theta(p-1)!\right\},$

then...
\[ f(\bar{x}) - f(x^+) \geq \frac{1}{8(p+1)!H^{\frac{1}{p+1}}} \| \nabla f(x^+) \|_*^{\frac{p+\nu}{p+1}}. \]

**Proof.** By (2.4), (2.6), (A.3) and (A.4), we have

\[
\| \nabla f(x^+) \|_* \leq \| \nabla f(x^+) - \nabla \Phi_{x,p}(x^+) \|_* + \| \nabla \Phi_{x,p}(x^+) - \nabla \Omega_{x,p,H}(x^+) \|_* \\
+ \| \nabla \Omega_{x,p,H}(x^+) \|_* \\
\leq \left[ \frac{H_{f,p}(\nu)}{(p-1)!} + \frac{H(p+\nu)}{p!} + \theta \right] |x^+ - \bar{x}|^{p+\nu-1} \\
\leq 2H |x^+ - \bar{x}|^{p+\nu-1}.
\]

Thus,

\[ \|x^+ - \bar{x}\|^{p+\nu} \geq \left( \frac{1}{2H} \right)^{\frac{p+\nu}{p+1}} \| \nabla f(x^+) \|_*^{\frac{p+\nu}{p+1}}. \]

On the other hand, by (A.1) and (A.4), it follows from Lemma A.1 that

\[ f(\bar{x}) - f(x^+) \geq \frac{H}{(p+1)!} |x^+ - \bar{x}|^{p+\nu}. \]

Then, combining (A.6) and (A.7) we get (A.5). \(\Box\)

**Lemma A.3.** Let \( H_{f,p}(\nu) < +\infty \) for some \( \nu \in [0,1] \) and assume that \( x^+ \) satisfies

\[ \| \nabla \Omega_{x,p,H}^{(1)}(x^+) \|_* \leq \theta |x^+ - \bar{x}|^p, \]

for some \( \bar{x} \in E, H > 0 \) and \( \theta > 0 \). If for some \( \delta > 0 \) we have

\[ \| \nabla f(x^+) \|_* \geq \delta \quad \text{and} \quad H \geq \max \left\{ \theta, (CH_{f,p}(\nu))^{\frac{p}{p+\nu}} \left( \frac{4}{\delta} \right)^{\frac{1}{p+\nu}} \right\}, \]

with constant \( C \geq 1 \), then

\[ |x^+ - \bar{x}|^{1-\nu} \geq \frac{CH_{f,p}(\nu)}{H} \]

and, consequently,

\[ 4H |x^+ - \bar{x}|^p \geq \| \nabla f(x^+) \|_* . \]

**Proof.** For \( \nu = 1 \), (A.10) is obvious. Thus, assume that \( \nu \in [0,1) \) and denote \( r = |x^+ - \bar{x}| \). Then, by (2.4), (2.6) and (A.8), we have

\[
\delta < \| \nabla f(x^+) \|_* \leq \| \nabla f(x^+) - \nabla \Phi_{x,p}(x^+) \|_* + \| \nabla \Phi_{x,p}(x^+) - \nabla \Omega_{x,p,H}(x^+) \|_* \\
+ \| \nabla \Omega_{x,p,H}(x^+) \|_* \\
\leq \frac{H_{f,p}(\nu)}{(p-1)!} r^{p+\nu-1} + \left( \frac{H(p+1)}{p!} + \theta \right) r^p \\
= r^{p+\nu-1} \left[ \frac{H_{f,p}(\nu)}{(p-1)!} + \left( \frac{p+1}{p!} + \frac{\theta}{H} \right) H r^{1-\nu} \right].
\]
Assume by contradiction that (A.10) is not true, i.e., $H_{r,\nu}^{1-\nu} < CH_{f,p}(\nu)$. Since $H \geq \theta$ and $C \geq 1$, it follows that
\[
\delta < r^{p+\nu-1} \left[ \frac{H_{f,p}(\nu)}{(p-1)!} + \left( \frac{p+1}{p!} + 1 \right) CH_{f,p}(\nu) \right]
= \frac{r^{p+\nu-1}H_{f,p}(\nu)}{(p-1)!} \left[ 1 + \frac{C(p+1)}{p} + C(p-1)! \right]
\leq 4CH_{f,p}(\nu)r^{p+\nu-1} < 4CH_{f,p}(\nu) \left( \frac{CH_{f,p}(\nu)}{H} \right)^{\frac{p+\nu-1}{1-\nu}}
= 4(CH_{f,p}(\nu))^{\frac{1}{1-\nu}} \left( \frac{1}{H} \right)^{\frac{p+\nu-1}{1-\nu}}.
\]
This implies that $H < (CH_{f,p}(\nu))^{\frac{1}{1-\nu}} \left( \frac{4}{\delta} \right)^{\frac{1}{1-\nu}}$, contradicting the second inequality in (A.9). Therefore, (A.10) holds.

Finally, let us prove (A.11). In view of inequality (A.10) we have
\[
H_{f,p}(\nu) \leq \frac{H}{C(p-1)!} r^{1-\nu}.
\]
Thus, it follows from (A.12) that
\[
\|\nabla f(x^+)\|_* \leq r^{p+\nu-1} \left[ \frac{H}{C(p-1)!} r^{1-\nu} + \left( \frac{p+1}{p!} + \frac{\theta}{H} \right) H r^{1-\nu} \right]
= r^p H \left[ \frac{1}{Cp!} + \left( \frac{p+1}{p!} + \frac{\theta}{H} \right) H \right] \leq 4r^p H.
\]

\[\square\]

**Lemma A.4.** Let $H_{f,p}(\nu) < +\infty$ for some $\nu \in [0,1]$ and assume that $x^+ \in E$ satisfies
\[
(A.13) \quad \Omega_{x^+,H}^{(1)}(x^+) < f(x)
\]
and
\[
(A.14) \quad \|\nabla \Omega_{x^+,H}^{(1)}(x^+)\|_* \leq \theta \|x^+ - \bar{x}\|^p
\]
for some $\bar{x} \in \mathbb{E}$, $H > 0$ and $\theta \geq 0$. If for some $\delta > 0$ we have
\[
(A.15) \quad \|\nabla f(x^+)\|_* \geq \delta \quad \text{and} \quad H \geq \max \left\{ \theta, (CH_{f,p}(\nu))^{\frac{1}{1-\nu}} \left( \frac{4}{\delta} \right)^{\frac{1}{1-\nu}} \right\},
\]
with constant $C \geq \frac{3}{2}$, then
\[
(A.16) \quad f(\bar{x}) - f(x^+) \geq \frac{H}{(p+1)!} \|x^+ - \bar{x}\|^{p+1}.
\]
Proof. In view of (2.3), (2.6) and (A.13), we have

\[ f(x^+) \leq \Omega^{(\nu)}_{x,p,H_f,p(\nu)}(x^+) \]

\[ = \Phi_{x,p}(x^+) + \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} \]

\[ = \Phi_{\bar{x},p}(x^+) + \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} + \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} \]

\[ = \Omega^{(1)}_{\bar{x},p,H}(x^+) - \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} + \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} \]

\[ < f(\bar{x}) - \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} + \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} \]

and so

(A.17) \[ f(\bar{x}) - f(x^+) \geq \frac{H}{p!} \|x^+ - \bar{x}\|^{p+1} - \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu}. \]

Assume by contradiction that (A.16) is not true, i.e.,

(A.18) \[ f(\bar{x}) - f(x^+) < \frac{H}{(p+1)!} \|x^+ - \bar{x}\|^{p+1}. \]

Then, combining (A.17) and (A.18) we obtain

\[ \frac{H}{p!} \|x^+ - \bar{x}\|^{p+1} - \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} < \frac{H}{(p+1)!} \|x^+ - \bar{x}\|^{p+1} \]

which implies that

(A.19) \[ H \left( 1 - \frac{1}{p+1} \right) < H_{f,p}(\nu) \|x^+ - \bar{x}\|^{p+1}. \]

By (A.14) and (A.15), the conclusions of Lemma A.3 hold. In particular, we have

\[ 4H \|x^+ - \bar{x}\|^p \geq \|\nabla f(x^+)\|_* \geq \delta \]

and so

(A.20) \[ \|x^+ - \bar{x}\|^{p+1} \leq \left( \frac{\delta}{4H} \right)^{\frac{p+1}{p}}. \]

Then, it follows from (A.19) and (A.20) that

\[ \frac{H}{p+1} < H_{f,p}(\nu) \left( \frac{\delta}{4H} \right)^{\frac{p+1}{p}}, \]

\[ \Rightarrow H < \left( \frac{3}{2} H_{f,p}(\nu) \right)^{\frac{p}{p+1}} \left( \frac{4}{\delta} \right)^{\frac{1-p}{p+\nu-1}}, \]

contradicting the second inequality in (A.15). Therefore, (A.16) is true. \( \square \)

**Corollary A.5.** Let \( H_{f,p}(\nu) < +\infty \) for some \( \nu \in [0,1] \) and assume that \( x^+ \in E \) satisfies (A.13) and (A.14) for some \( \bar{x} \in E \), \( H > 0 \) and \( \theta \geq 0 \). Given \( \delta > 0 \), define

(A.21) \[ \xi_\nu(\delta) \equiv \max \left\{ \theta, \left( \frac{3H_{f,p}(\nu)}{2} \right)^{\frac{p}{p+\nu-1}} \left( \frac{4}{\delta} \right)^{\frac{1-p}{p+\nu-1}} \right\}. \]
If \( \| \nabla f(x^+) \|_* \geq \delta \) and \( H \geq \xi_\nu(\delta) \), then

\[
f(\bar{x}) - f(x^+) \geq \frac{1}{8(p + 1)! H^p} \| \nabla f(x^+) \|_*^{p+1}.
\]

Proof. From inequality (A.11) in Lemma A.3 we have

\[
\| x^+ - \bar{x} \|_p \geq \frac{1}{4H} \| \nabla f(x^+) \|_*
\]

which implies that

\[
\| x^+ - \bar{x} \|^{p+1} \geq \left( \frac{1}{4H} \right)^{\frac{p+1}{p}} \| \nabla f(x^+) \|_*^{\frac{p+1}{p}}.
\]

Then, it follows from inequality (A.16) in Lemma A.4 that

\[
f(\bar{x}) - f(x^+) \geq \frac{H}{(p + 1)!} \| x^+ - \bar{x} \|^{p+1} \geq \frac{H}{(p + 1)!} \frac{1}{4^{\frac{p+1}{p}} H^{\frac{p+1}{p}}} \| \nabla f(x^+) \|_*^{\frac{p+1}{p}}
\]

\[
\geq \frac{1}{8(p + 1)! H^p} \| \nabla f(x^+) \|_*^{\frac{p+1}{p}}.
\]

**Lemma A.6.** Let \( H_{f,p}(\nu) < +\infty \) for some \( \nu \in [0, 1] \) and assume that \( x^+ \) satisfies

(A.22) \[ \| \nabla \Omega^{(\nu)}_{x,p,H}(x^+) \|_* \leq \theta \| x^+ - \bar{x} \|^{p+\nu-1}, \]

for some \( \bar{x} \in \mathbb{E}, H > 0 \) and \( \theta \geq 0 \). If

(A.23) \[ H \geq (p + \nu - 1) (H_{f,p}(\nu) + \theta (p - 1)!), \]

then

(A.24) \[ \langle \nabla f(x^+), \bar{x} - x^+ \rangle \geq \frac{1}{3} \left[ \frac{(p - 1)!}{H} \right]^{\frac{p+\nu-1}{p}} \| \nabla f(x^+) \|_*^{\frac{p+\nu-1}{p}}. \]

Proof. Denote \( r = \| x^+ - \bar{x} \| \). Then, by (2.4), (2.6) and (A.22), we have

\[
\| \nabla f(x^+) + \frac{H(p + \nu)}{p!} r^{p+\nu-2} B(x^+ - \bar{x}) \|_* = \| \nabla f(x^+) - \nabla \Phi_{x,p}(x^+) + \nabla \Omega^{(\nu)}_{x,p,H}(x^+) \|_*
\]

\[
\leq \| \nabla f(x^+) - \nabla \Phi_{x,p}(x^+) \|_* + \| \nabla \Omega^{(\nu)}_{x,p,H}(x^+) \|_*
\]

\[
\leq \left( \frac{H_{f,p}(\nu)}{(p - 1)!} + \theta \right) r^{p+\nu-1}.
\]

Thus, we obtain

\[
\left( \frac{H_{f,p}(\nu)}{(p - 1)!} + \theta \right) r^{2(p+\nu-1)} \geq \| \nabla f(x^+) + \frac{H(p + \nu)}{p!} r^{p+\nu-2} B(x^+ - \bar{x}) \|_*^2
\]

\[
= \| \nabla f(x^+) \|_*^2 + \frac{2(p + \nu)}{p!} H r^{p+\nu-2} (\nabla f(x^+), x^+ - \bar{x})
\]

\[
+ \frac{H^2 (p + \nu)^2}{(p!)^2} r^{2(p+\nu-1)}
\]
Thus, (A.25)
\[
\langle \nabla f(x^+), \bar{x} - x^+ \rangle \geq \frac{p!}{2(p + \nu)H^{p+\nu-2}} \| \nabla f(x^+) \|_*^2 + \frac{p^2 H}{2(p + \nu)p!} \left[ 1 - \frac{H f_\nu(v) + \theta(p - 1)!}{H} \right]^2 r^{p + \nu}.
\]

For \( \nu = 0 \), (A.25) leads to the desired relation. Let us assume that \( \nu > 0 \). Denote \( g = \| \nabla f(x^+) \| \) and \( \Delta^2 = 1 - \left( \frac{H f_\nu(v) + \theta(p - 1)!}{H} \right)^2 \). By (A.23), we have

(A.26) \[
\Delta^2 \geq 1 - \frac{1}{(p + \nu - 1)^2} = \frac{(p + \nu - 1)^2 - 1}{(p + \nu - 1)^2} = \frac{(p + \nu - 2)(p + \nu)}{(p + \nu - 1)^2} > 0.
\]

Consider the right-hand side of inequality (A.25) as a function of \( r \):

\[
h(r) = \frac{p!}{2(p + \nu)H^{p+\nu-2}g^2} + \frac{H^2 \Delta^2 r^{p+\nu}}{2(p + \nu)p!}.
\]

Since \( \Delta^2 > 0 \), \( h \) is a convex function for \( r > 0 \). Thus, let us find the optimal \( r_* \) as a solution to the first-order optimality condition for function \( h \):

\[
g^2(p + \nu - 2)p! \frac{H^2 \Delta^2 r_*^{p+\nu-1}}{(p + \nu)H r_*^{p+\nu-1}} = \frac{H^2 \Delta^2 r_*^{p+\nu-1}}{p!}.
\]

Solving this equation for \( r_* \), we obtain \( r_*^{p+\nu-1} = \frac{g(p - 1)!}{H \Delta} \sqrt{\frac{p + \nu - 2}{p + \nu}} \). Consequently,

\[
h(r_*) = \frac{r_*}{2H(p + \nu)} \left[ \frac{g^2p!}{r_*^{p+\nu-1}} + \frac{H^2 \Delta^2 r_*^{p+\nu-1}}{p!} \right] = \frac{(p + \nu - 1)p H^{\frac{p+\nu-2}{p+\nu}}}{(p + \nu)\sqrt{(p + \nu - 2)(p + \nu)}} \left( \sqrt{\frac{p + \nu - 2}{p + \nu}} \right)^{\frac{p+\nu-1}{p+\nu}} \left[ \frac{(p - 1)!}{H} \right]^{\frac{1}{p+\nu-1}} g^{\frac{p+\nu}{p+\nu-1}}.
\]

Now, using (A.26) we obtain

\[
h(r_*) \geq \frac{(p + \nu - 1)p}{(p + \nu)\sqrt{(p + \nu - 2)(p + \nu)}} \left( \sqrt{\frac{p + \nu - 2}{p + \nu}} \right)^{\frac{p+\nu-2}{p+\nu}} \left( \sqrt{\frac{p + \nu - 2}{p + \nu}} \right)^{\frac{1}{p+\nu-1}} \left[ \frac{(p - 1)!}{H} \right]^{\frac{1}{p+\nu-1}} g^{\frac{p+\nu}{p+\nu-1}}.
\]

Note that

\[
\frac{(p + \nu - 1)p}{(p + \nu)\sqrt{(p + \nu - 2)(p + \nu)}} = \frac{p(p + \nu - 1)}{(p + \nu)(p + \nu)\sqrt{p + \nu - 1}} \geq \left( \frac{p}{p + 1} \right) \left( \frac{p - 1}{p + 1} \right) \geq \frac{1}{3}.
\]

Thus, \( h(r_*) \geq \frac{1}{3} \left[ \frac{(p - 1)!}{H} \right]^{\frac{1}{p+\nu-1}} g^{\frac{p+\nu}{p+\nu-1}} \) and so, by (A.25), we get (A.24).
\textbf{Lemma A.7.} Let $H_{f,p}(\nu) < +\infty$ for some $\nu \in [0, 1]$ and assume that $x^+$ satisfies
\begin{equation}
(A.27) \quad \|\nabla \Omega^{(1)}_{x,p,H}(x^+)\|_* \leq \theta \|x^+ - \bar{x}\|^p,
\end{equation}
for some $\bar{x} \in \mathbb{E}$, $H > 0$ and $\theta \geq 0$. If for some $\delta > 0$ we have
\begin{equation}
(A.28) \quad \|\nabla f(x^+)\|_* \geq \delta \quad \text{and} \quad H \geq \max \left\{C\theta(p-1)!, (CH_{f,p}(\nu))^{\frac{p}{p+1}} \left(\frac{4}{\delta}\right)^{\frac{p}{p+1}} \right\}
\end{equation}
with $C \geq 4$, then
\begin{equation}
(A.29) \quad \langle \nabla f(x^+), \bar{x} - x^+ \rangle \geq \frac{1}{4} \left(\frac{p!}{H}\right) \|\nabla f(x^+)\|_{x^+,p}^{\frac{p}{p+1}}.
\end{equation}

\textbf{Proof.} Denote $r = \|x^+ - \bar{x}\|$. Then, by (2.4), (2.6) and (A.27) we have
\begin{align*}
\|\nabla f(x^+) + \frac{H(p+1)}{p!}r^{p-1}B(x^+ - \bar{x})\|_* &= \|\nabla f(x^+) - \nabla \Phi_{x,p}(x^+) + \nabla \Omega^{(1)}_{x,p,H}(x^+)\|_* \\
&\leq \|\nabla f(x^+) - \nabla \Phi_{x,p}(x^+)\|_* + \|\nabla \Omega_{x,p,H}(x^+)\|_* \\
&\leq \frac{H_{f,p}(\nu)}{(p-1)!}r^{p+\nu-1} + \theta r^p \\
&\leq \left(\frac{H}{C(p-1)!} + \theta\right)r^p
\end{align*}
Therefore,
\begin{align*}
\left(\frac{H}{C(p-1)!} + \theta\right)^2 r^{2p} &\geq \|\nabla f(x^+) + \frac{H}{p!}r^{p-1}B(x^+ - \bar{x})\|^2 \\
&= \|\nabla f(x^+)\|^2 + \frac{2(p+1)}{p!}Hr^{p-1}\langle f^{(1)}(x^+), x^+ - \bar{x}\rangle + \frac{H^2(p+1)^2}{(p!)^2}r^{2p},
\end{align*}
which gives
\begin{equation}
(A.30) \quad \langle \nabla f(x^+), \bar{x} - x^+ \rangle \geq \frac{p!}{2(p+1)Hr^{p-1}}\|\nabla f(x^+)\|^2 + \frac{p!}{2H(p+1)} \left[\left(\frac{H(p+1)}{p!}\right)^2 - \left(\frac{H}{C(p-1)!} + \theta\right)^2\right]r^{p+1}.
\end{equation}
Since $H \geq C\theta(p-1)!$, it follows that
\begin{equation}
\left(\frac{H(p+1)}{p!}\right)^2 - \left(\frac{H}{C(p-1)!} + \theta\right)^2 \geq \frac{H^2}{2(p+1)} \left[\left(\frac{p+1}{p!}\right)^2 - \left(\frac{1}{2(p-1)!}\right)^2\right]^2.
\end{equation}
Because $C \geq 4$, we have
\begin{equation}
-\left(\frac{2}{C(p-1)!}\right)^2 \geq -\left(\frac{1}{2(p-1)!}\right)^2
\end{equation}
and so,
\begin{align*}
\left(\frac{H(p+1)}{p!}\right)^2 - \left(\frac{H}{C(p-1)!} + \theta\right)^2 &\geq \frac{H^2}{2(p+1)} \left[\left(\frac{p+1}{p!}\right)^2 - \left(\frac{1}{2(p-1)!}\right)^2\right]^2 \\
&\geq \frac{3H^2}{8(p+1)!}.
\end{align*}
Therefore,

\begin{equation}
\langle \nabla f(x^+), \bar{x} - x^+ \rangle \geq \frac{p!}{2H(p + 1)r^{p-1}} \|\nabla f(x^+)\|^2 + \frac{3Hr^2}{8(p + 1)!} r^{p+1}.
\end{equation}

Denote \( g = \|\nabla f(x^+)\| \) and consider the right-hand side of (A.30) as a function of \( r \):

\[
h(r) = \frac{p!}{2(p + 1)Hr^{p-1}} g^2 + \frac{3Hr^2r^{p+1}}{8(p + 1)!}.
\]

Let us find the optimal \( r^* \) as a solution to the first-order optimality condition for function \( h \):

\[
\frac{(p - 1)g^2p!}{2(p + 1)Hr^p} = \frac{3Hp^2(p + 1)r^p}{8(p + 1)!} = \frac{3Hr^p}{8p!}.
\]

Solving this equation for \( r^* \), we obtain

\[
r^p = \frac{g(p - 1)!}{H \Delta} \sqrt{\frac{8(p - 1)}{3(p + 1)}}.
\]

Consequently,

\[
h(r^*) = \left[ \frac{g p}{2(p + 1)} \frac{\sqrt{6(p + 1)}}{8(p - 1)} + \frac{3gp}{8(p + 1)} \sqrt{\frac{8(p - 1)}{6(p + 1)}} \right] \left[ \frac{g(p - 1)!}{H} \sqrt{\frac{8(p - 1)}{6(p + 1)}} \right]^{\frac{1}{p}} \geq \frac{1}{4} \left[ \frac{(p - 1)!}{H} \right]^{\frac{1}{p}} g^{\frac{p+1}{p}}.
\]

Therefore, (A.29) holds. \( \square \)