

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

Geovani Nunes Grapiglia, Yurii Nesterov

REPRINT | 3237

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

<https://uclouvain.be/en/research-institutes/lidam/core/core-reprints.html>

TENSOR METHODS FOR MINIMIZING CONVEX 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 ($p \geq 2$) with ν -Hölder continuous p th derivatives. We propose tensor schemes with and without acceleration. For the schemes without acceleration, we establish 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 ν 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 ν is unknown, we present a universal accelerated tensor scheme with iteration complexity of $\mathcal{O}(\epsilon^{-p/[(p+1)(p+\nu-1)])}$. A lower complexity bound of $\mathcal{O}(\epsilon^{-2/[3(p+\nu)-2]})$ 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

DOI. 10.1137/19M1259432

1. Introduction.

1.1. Motivation. In [16], it was shown that a suitable cubic regularization of the Newton method (CNM) takes at most $\mathcal{O}(\epsilon^{-1/2})$ iterations to reduce the functional residual below a given precision $\epsilon > 0$ when the objective is a twice-differentiable convex function with a Lipschitz continuous Hessian. A better complexity bound of $\mathcal{O}(\epsilon^{-1/3})$ was shown in [13] 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, [17, 5]). 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 [14], 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 $\mathcal{O}(\epsilon^{-1/p})$ was proved for the method without acceleration, while an improved bound of $\mathcal{O}(\epsilon^{-1/(p+1)})$ was proved for the accelerated tensor method.

*Received by the editors May 1, 2019; accepted for publication (in revised form) April 14, 2020; published electronically October 5, 2020.

<https://doi.org/10.1137/19M1259432>

Funding: The first author's research 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). The second author's research was supported by the European Research Council (advanced grant 788368).

[†]Departamento de Matemática, Universidade Federal do Paraná, Centro Politécnico, 81531-980, Curitiba, Paraná, Brazil (grapiglia@ufpr.br).

[‡]Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 1348 Louvain-la-Neuve, Belgium (Yurii.Nesterov@uclouvain.be).

In the present paper, we study tensor methods (with and without acceleration) that can handle convex functions with ν -Hölder continuous p th derivatives ($p \geq 2$) and allow the inexact solution of auxiliary problems (in the sense of [4]). Specifically, our contribution is threefold:

1. For the schemes without acceleration, we establish iteration complexity bounds of $\mathcal{O}(\epsilon^{-1/(p+\nu-1)})$ for reducing the functional residual below a given $\epsilon \in (0, 1)$.
2. Assuming that ν 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 ν is unknown, we present a *universal* accelerated tensor scheme with iteration complexity of $\mathcal{O}(\epsilon^{-p/[(p+1)(p+\nu-1)])}$.
3. A lower complexity bound of $\mathcal{O}(\epsilon^{-2/[3(p+\nu)-2]})$ is also obtained, from which we conclude that our accelerated nonuniversal scheme is nearly optimal.

The methods and results presented here extend in a significant way the contributions in [2, 8, 9, 14]. Indeed, [8, 9] deal only with second-order schemes ($p = 2$) which require the exact solution of the auxiliary problems. On the other hand, the p -order methods proposed in [2, 14] are restricted to the Lipschitz case ($\nu = 1$), assuming that the Lipschitz constant is known and that the auxiliary problems are solved exactly. We believe that the development of p -order methods with affordable trial steps and automatic adjustment to the objective’s function class (universality) constitutes a fundamental step towards implementable high-order methods for convex optimization.

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 Appendix A.

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

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

For a smooth function $f : \text{dom } f \rightarrow \mathbb{R}$ with convex and open domain $\text{dom } f \subset \mathbb{E}$, 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 $\nabla f(x) \in \mathbb{E}^*$ and $\nabla^2 f(x)h \in \mathbb{E}^*$ for $x \in \text{dom } f$ and $h \in \mathbb{E}$.

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

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

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

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

For $h_1 = \dots = h_p = h \in \mathbb{E}$, 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:

$$(1.1) \quad f(x+h) = \Phi_{x,p}(x+h) + o(\|h\|^p), \quad x+h \in \text{dom } f,$$

Downloaded 06/27/23 to 130.104.59.42 . Redistribution subject to CCBY license

where

$$(1.2) \quad \Phi_{x,p}(y) \equiv f(x) + \sum_{i=1}^p \frac{1}{i!} D^i f(x)[y-x]^i, \quad y \in \mathbb{E}.$$

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

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

In fact, it can be shown that (see, e.g., [3])

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

Similarly, since $D^p f(x)[\cdot, \dots, \cdot] - D^p f(y)[\cdot, \dots, \cdot]$ 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:

$$(2.1) \quad \min_{x \in \mathbb{E}} f(x),$$

where $f : \mathbb{E} \rightarrow \mathbb{R}$ is a convex p -times differentiable function ($p \geq 2$). We assume that there exists at least one optimal solution $x^* \in \mathbb{E}$ for problem (2.1). Let us characterize the level of smoothness of the objective f by the system of Hölder constants

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

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

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

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

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

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

This property motivates the use of the following class of models of f around $x \in \mathbb{E}$:

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

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

$$(2.7) \quad f(y) \leq \Omega_{x,p,H}^{(\nu)}(y), \quad y \in \mathbb{E}.$$

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 ν is known, or ν is unknown. We cover both cases in a single framework by introducing parameter

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

Let $\epsilon \in (0, 1)$ 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

$$(3.2) \quad \min_{y \in \mathbb{E}} \Omega_{x_t, p, M_t}^{(\alpha)}(y),$$

with α given by (3.1). Similarly to [4], the trial point x_t^+ must satisfy the following conditions:

$$(3.3) \quad \Omega_{x_t, p, M_t}^{(\alpha)}(x_t^+) \leq f(x_t) \quad \text{and} \quad \|\nabla \Omega_{x_t, p, M_t}^{(\alpha)}(x_t^+)\|_* \leq \theta \|x_t^+ - x_t\|^{p+\alpha-1},$$

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

$$(3.4) \quad f(x_t) - f(x_t^+) \geq \frac{1}{8(p+1)! M_t^{\frac{1}{p+\alpha-1}}} \|\nabla f(x_t^+)\|_*^{\frac{p+\alpha}{p+\alpha-1}}$$

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 α 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.

Remark 3.1. 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. In [10], we investigated the possible use of gradient methods with Bregman distance for the case $p = 3$ and $\alpha = \nu$. Specifically, under assumptions H1 and H2 below, we showed that if $\{y_k\}_{k=0}^T$ is a sequence generated by these methods applied to $\min_{y \in \mathbb{E}} \Omega_{x_t, 3, M_t}^{(\nu)}(y)$ with

$$\|\nabla f(y_k)\|_* > \epsilon \quad \text{and} \quad \|\nabla \Omega_{x_t, 3, M_t}^{(\nu)}(y_k)\|_* > \theta \|y_k - x_0\|^{2+\nu} \quad \text{for } k = 0, \dots, T,$$

then

$$T \leq \begin{cases} \mathcal{O}(\epsilon^{-(3+\nu)}) & \text{if } M_t < [6/(3+\nu)]H_{f,3}(\nu) \text{ and } \nu \neq 0, \\ \mathcal{O}\left(\epsilon^{-\frac{(3+\nu)}{2}}\right) & \text{if } M_t = [6/(3+\nu)]H_{f,3}(\nu) \text{ and } \nu \neq 0, \\ \mathcal{O}(\log(\epsilon^{-1})) & \text{if } M_t > [6/(3+\nu)]H_{f,3}(\nu). \end{cases}$$

For more details, see Theorems 3.7 and 3.8 in [10].

To analyze the 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 3.2. *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

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

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

$$(3.6) \quad f(x_k) - f(x^*) \leq \frac{[24p(p+1)!]^{p+\alpha-1} M_\nu D_0^{p+\alpha}}{(k-m)^{p+\alpha-1}}.$$

Proof. By Step 1 in Algorithm 1, we have

$$(3.7) \quad M_k \leq M_\nu, \quad k = 0, \dots, T-1.$$

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

$$(3.8) \quad \begin{aligned} f(x_k) - f(x_{k+1}) &\geq \frac{1}{8(p+1)!} \left[\frac{1}{M_k} \right]^{\frac{1}{p+\alpha-1}} \|\nabla f(x_{k+1})\|_*^{\frac{p+\alpha}{p+\alpha-1}} \\ &\geq \frac{1}{8(p+1)!} \left[\frac{1}{M_\nu} \right]^{\frac{1}{p+\alpha-1}} \|\nabla f(x_{k+1})\|_*^{\frac{p+\alpha}{p+\alpha-1}} \\ &\geq \frac{1}{8(p+1)!} \left[\frac{1}{M_\nu D_0^{p+\alpha}} \right]^{\frac{1}{p+\alpha-1}} (\|\nabla f(x_{k+1})\|_* \|x_{k+1} - x^*\|)^{\frac{p+\alpha}{p+\alpha-1}} \\ (3.9) \quad &\geq \frac{1}{8(p+1)!} \left[\frac{1}{M_\nu D_0^{p+\alpha}} \right]^{\frac{1}{p+\alpha-1}} (f(x_{k+1}) - f(x^*))^{\frac{p+\alpha}{p+\alpha-1}}, \end{aligned}$$

where the last inequality is due to the convexity of f . Now, denoting

$$\delta_k = \frac{f(x_k) - f(x^*)}{[8(p+1)!]^{p+\alpha-1} M_\nu D_0^{p+\alpha}}$$

we see from (3.9) that this sequence satisfies condition (1.1) of Lemma 1.1 in [8] with $u = \frac{p+\alpha}{p+\alpha-1}$. Note that m is the first iteration for which $\delta_m \leq 2$. Using Lemma 1.1 in [8], this inequality allow us to obtain the upper bound (3.5) for m and also simplifies our final bound for the functional residual. Indeed, if $m > 0$, then $\delta_0 > 2$ and, in view of inequality (1.2) of Lemma 1.1 in [8], we have

$$\ln 2 \leq \ln \delta_{m-1} \leq \left(\frac{p+\alpha-1}{p+\alpha} \right)^{m-1} \ln \delta_0 \implies \left(\frac{p+\alpha}{p+\alpha-1} \right)^{m-1} \ln 2 \leq \ln \delta_0$$

$$\implies \left(\frac{p + \alpha}{p + \alpha - 1}\right)^{m-1} \leq \frac{\ln \delta_0}{\ln 2} = \log_2 \delta_0.$$

Thus, $m \leq \frac{\ln \delta_0}{\ln(\frac{p+\alpha}{p+\alpha-1})}$, and so (3.5) holds. Consequently, from inequality (1.3) of Lemma 1.1 in [8] we get the following rate of convergence:

$$\delta_k \leq \left[\frac{1 + \delta_m^{u-1}}{(u-1)(k-m)} \right]^{\frac{1}{u-1}},$$

that is,

$$\frac{f(x_k) - f(x^*)}{[8(p+1)!]^{p+\alpha-1} M_\nu D_0^{p+\alpha}} \leq \left[\frac{(p+\alpha-1)(1+2^{\frac{1}{p+\alpha-1}})}{k-m} \right]^{p+\alpha-1}.$$

Therefore,

$$\begin{aligned} f(x_k) - f(x^*) &\leq \frac{[8(1+2^{\frac{1}{p+\alpha-1}})(p+\alpha-1)(p+1)!]^{p+\alpha-1} M_\nu D_0^{p+\alpha}}{(k-m)^{p+\alpha-1}} \\ &\leq \frac{[24p(p+1)!]^{p+\alpha-1} M_\nu D_0^{p+\alpha}}{(k-m)^{p+\alpha-1}}. \end{aligned} \quad \square$$

If we assume that ν and $H_{f,p}(\nu)$ are known, by Lemma A.2, we can set

$$M_t = M_\nu \equiv \max \left\{ \frac{3H_{f,p}(\nu)}{2}, 3\theta(p-1)! \right\}.$$

Here, by (3.1) the corresponding version of Algorithm 1 takes at most $\mathcal{O}(\epsilon^{-1/(p+\nu-1)})$ 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 α by (3.1) and $t := 0$.

Step 1. Set $i := 0$.

Step 1.1 Compute an approximate solution $x_{t,i}^+$ to $\min_{y \in \mathbb{E}} \Omega_{x_t, p, 2^i H_t}^{(\alpha)}(y)$, such that

$$\Omega_{x_t, p, 2^i H_t}^{(\alpha)}(x_{t,i}^+) \leq f(x_t) \quad \text{and} \quad \|\nabla \Omega_{x_t, p, 2^i H_t}^{(\alpha)}(x_{t,i}^+)\|_* \leq \theta \|x_{t,i}^+ - x_t\|^{p+\alpha-1}.$$

Step 1.2. If

$$f(x_t) - f(x_{t,i}^+) \geq \frac{1}{8(p+1)!(2^i H_t)^{\frac{1}{p+\alpha-1}}} \|\nabla f(x_{t,i}^+)\|_*^{\frac{p+\alpha}{p+\alpha-1}}$$

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.

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

$$(3.10) \quad N_\nu(\epsilon) = \begin{cases} \max \left\{ \frac{3H_{f,p}(\nu)}{2}, 3\theta(p-1)! \right\} & \text{if } \alpha = \nu, \\ \max \left\{ \theta, \left(\frac{3H_{f,p}(\nu)}{2} \right)^{\frac{p}{p+\nu-1}} \left(\frac{4R(\epsilon)}{\epsilon} \right)^{\frac{1-\nu}{p+\nu-1}} \right\} & \text{if } \alpha = 1, \end{cases}$$

where

$$(3.11) \quad R(\epsilon) = \max_{x \in \mathbb{E}} \{ \|x - x^*\| : f(x) \leq f(x^*) + \epsilon \}.$$

The next lemma provides upper bounds on H_t and on the number of calls of the oracle.¹

LEMMA 3.3. *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*

$$(3.12) \quad f(x_0) - f(x^*) \geq \epsilon,$$

$$(3.13) \quad f(x_{t,i}^+) - f(x^*) \geq \epsilon, \quad i = 0, \dots, i_t \text{ and } t = 0, \dots, T.$$

Then

$$(3.14) \quad H_t \leq \max \{H_0, N_\nu(\epsilon)\} \quad \text{for } t = 0, \dots, T.$$

Moreover, the number O_T of calls of the oracle after T iterations is bounded as follows:

$$(3.15) \quad O_T \leq 2T + \log_2 \max \{H_0, N_\nu(\epsilon)\} - \log_2 H_0.$$

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 ν is known, then by (3.1) we have $\alpha = \nu$. Thus, by H1 and Lemma A.2 the final value of $2^{i_t} H_t$ 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,

$$H_{t+1} = \frac{1}{2} 2^{i_t} H_t \leq \max \left\{ \frac{3H_{f,p}(\nu)}{2}, 3\theta(p-1)! \right\} = N_\nu(\epsilon) \leq \max \{H_0, N_\nu(\epsilon)\};$$

that is, (3.14) holds for $t = t + 1$.

On the other hand, if ν is unknown, we have $\alpha = 1$. In view of (3.11), (3.12), and H2, it follows that

$$(3.16) \quad R(\epsilon) \leq R(f(x_0) - f(x^*)) = \max_{x \in \mathcal{L}(x_0)} \|x - x^*\| \leq D_0 < +\infty.$$

Thus, by (3.13) and Lemma A.5 in [9] 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

$$2^{i_t} H_t \leq 2 \max \left\{ \theta, \left(\frac{3H_{f,p}(\nu)}{2} \right)^{\frac{p}{p+\nu-1}} \left(\frac{4R(\epsilon)}{\epsilon} \right)^{\frac{1-\nu}{p+\nu-1}} \right\} = 2N_\nu(\epsilon).$$

¹By calls of the oracle we mean the joint computation of f and its derivatives.

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 $H_{k+1} = 2^{i_k - 1} H_k$, it follows that $i_k - 1 = \log_2 H_{k+1} - \log_2 H_k$. Thus, by (3.14) we get

$$\begin{aligned} O_T &= \sum_{k=0}^{T-1} (i_k + 1) = \sum_{k=0}^{T-1} 2 + \log_2 H_{k+1} - \log_2 H_k = 2T + \log_2 H_T - \log_2 H_0 \\ &\leq 2T + \log_2 \max\{H_0, N_\nu(\epsilon)\} - \log_2 H_0. \end{aligned} \quad \square$$

From Lemma 3.3, we see that Algorithm 2 is a particular case of Algorithm 1 in which

$$M_t = 2^{i_t} H_t = 2H_{t+1} \quad \text{and} \quad M_\nu = 2 \max\{H_0, N_\nu(\epsilon)\}.$$

Thus, combining Theorem 3.2 and Lemma 3.3, we obtain the following result.

THEOREM 3.4. *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 2 such that*

$$(3.17) \quad f(x_0) - f(x^*) \geq \epsilon,$$

$$(3.18) \quad f(x_{t,i}^+) - f(x^*) \geq \epsilon, \quad i = 0, \dots, i_t \text{ and } t = 0, \dots, T.$$

Denote by m the first iteration number such that

$$f(x_m) - f(x^*) \leq [8(p + 1)!]^{p+\alpha-1} 4 \max\{H_0, N_\nu(\epsilon)\} D_0^{p+\alpha},$$

and assume that $m < T$. Then

$$(3.19) \quad m \leq \frac{1}{\ln\left(\frac{p+\alpha}{p+\alpha-1}\right)} \ln \max\left\{1, \log_2 \frac{f(x_0) - f(x^*)}{[8(p + 1)!]^{p+\alpha-1} 2 \max\{H_0, N_\nu(\epsilon)\} D_0^{p+\alpha}}\right\}$$

and

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

Consequently,

$$(3.21) \quad T \leq m + \kappa_1^{(\nu)} [24p(p + 1)!] \epsilon^{-\frac{1}{p+\nu-1}},$$

where

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

Proof. As mentioned above, by Lemma 3.3 we have

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

Then (3.19) and (3.20) follow directly from Theorem 3.2 with

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

Now, combining (3.18) and (3.20), we obtain

$$\epsilon \leq \frac{[24p(p+1)]^{p+\alpha-1} 2 \max \{H_0, N_\nu(\epsilon)\} D_0^{p+\alpha}}{(T-m)^{p+\alpha-1}},$$

and so

$$(3.22) \quad T \leq m + \frac{[24p(p+1)] (2 \max \{H_0, N_\nu(\epsilon)\} D_0^{p+\alpha})^{\frac{1}{p+\alpha-1}}}{\epsilon^{\frac{1}{p+\alpha-1}}}.$$

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

$$(3.23) \quad N_\nu(\epsilon) = \max \left\{ \frac{3H_{f,p}(\nu)}{2}, 3\theta(p-1)! \right\}.$$

Thus, combining (3.22) and (3.23), we get (3.21). On the other hand, if ν is unknown, then $\alpha = 1$ and, by (3.10), (3.16), and $\epsilon \in (0, 1)$, we have

$$(3.24) \quad \begin{aligned} N_\nu(\epsilon) &= \max \left\{ \theta, \left(\frac{3H_{f,p}(\nu)}{2} \right)^{\frac{p}{p+\nu-1}} \left(\frac{4R(\epsilon)}{\epsilon} \right)^{\frac{1-\nu}{p+\nu-1}} \right\} \\ &\leq \max \left\{ \theta, \left(\frac{3H_{f,p}(\nu)}{2} \right)^{\frac{p}{p+\nu-1}} (4D_0)^{\frac{1-\nu}{p+\nu-1}} \right\} \epsilon^{-\frac{1-\nu}{p+\nu-1}}. \end{aligned}$$

In this case, combining (3.22) and (3.24) we also get (3.21). □

Remark 3.5. Note that for any θ in the interval

$$0 < \theta < \begin{cases} \min \left\{ \frac{H_0}{3(p-1)!}, \frac{H_{f,p}(\nu)}{(p-1)!2} \right\} & \text{if } \nu \text{ is known,} \\ \min \left\{ H_0, \left(\frac{3H_{f,p}(\nu)}{2} \right)^{\frac{p}{p+\nu-1}} (4D_0)^{\frac{1-\nu}{p+\nu-1}} \right\} & \text{if } \nu \text{ is unknown,} \end{cases}$$

the corresponding right-hand side in (3.21) has the same value as for $\theta = 0$.

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 [8]. Looking at the efficiency bound (3.21), for ν known and ν unknown, we see that the universal scheme ensures the same dependence on the accuracy ϵ as the nonuniversal 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 section.

4. Accelerated tensor schemes. Similarly to section 3, we shall consider a general accelerated tensor method parametrized by the constant α 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 :

$$(4.1) \quad y_t = (1 - \gamma_t)x_t + \gamma_tv_t,$$

where

$$(4.2) \quad \gamma_t = \frac{a_t}{A_t + a_t},$$

with $a_t > 0$ being computed from the equation

$$(4.3) \quad a_t^{p+\alpha} = \frac{1}{2^{(3p-1)}} \left[\frac{(p-1)!}{M_t} \right] (A_t + a_t)^{p+\alpha-1}.$$

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

$$(4.4) \quad \min_{x \in \mathbb{E}} \Omega_{y_t, p, M_t}^{(\alpha)}(x),$$

such that

$$(4.5) \quad \Omega_{y_t, p, M_t}^{(\alpha)}(x_t^+) \leq f(y_t) \quad \text{and} \quad \|\nabla \Omega_{y_t, p, M_t}^{(\alpha)}(x_t^+)\|_* \leq \theta \|x_t^+ - y_t\|^{p+\alpha-1},$$

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

$$(4.6) \quad \langle \nabla f(x_t^+), y_t - x_t^+ \rangle \geq \frac{1}{4} \left[\frac{(p-1)!}{M_t} \right]^{\frac{1}{p+\alpha-1}} \|\nabla f(x_t^+)\|_*^{\frac{p+\alpha}{p+\alpha-1}}$$

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 $M_\nu > 0$ such that $M_t \leq M_\nu$ for all t . After obtaining x_{t+1} , we set $A_{t+1} = A_t + a_t$ and compute

$$(4.7) \quad v_{t+1} = \arg \min_{x \in \mathbb{E}} \psi_{t+1}(x),$$

where

$$(4.8) \quad \psi_{t+1}(x) = \psi_t(x) + a_t [f(x_{t+1}) + \langle \nabla f(x_{t+1}), x - x_{t+1} \rangle].$$

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 $x_0 \in \mathbb{E}$, $H_0 > 0$. Set α by (3.1), $v_0 = x_0$, $A_0 = 0$, and $t := 0$.

Step 1. Find $0 < M_t \leq M_\nu$ 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.

Remark 4.1. From the expression of $\psi_{t+1}(\cdot)$ we can see that $\min_{x \in \mathbb{E}} \psi_{t+1}(x)$ admits a closed form solution, namely,

$$v_{t+1} = x_0 - \frac{B^{-1}s_{t+1}}{\|s_{t+1}\|^{\frac{p+\alpha-2}{p+\alpha-1}}}, \quad s_{t+1} = \sum_{i=0}^t a_i \nabla f(x_{i+1}).$$

Regarding the computation of a_t , for $t = 0$, (4.3) gives

$$a_0 = \frac{1}{2^{(3p-1)}} \left[\frac{(p-1)!}{M_t} \right].$$

For $t > 0$, we have $A_t > 0$. Thus, the computation of a_t requires the solution of a univariate nonlinear equation of the form

$$x^{p+\alpha} - B(A+x)^{p+\alpha-1} = 0, \quad B, A > 0.$$

Denoting $g(x) = x^{p+\alpha} - B(A+x)^{p+\alpha-1}$, it is easy to see that

$$g(0) = -BA^{p+\alpha-1} < 0 \leq g(c),$$

where $c = \max\{A, 2^{p+\alpha-1}B\}$. Since $g(\cdot)$ is continuous, we can use the bisection method to compute an approximation to $a^* \in (0, c]$ such that $g(a^*) = 0$. As can be seen in the proof of Theorem 4.3, our convergence results only require

$$0 < a_t^{p+\alpha} \leq \frac{1}{2^{(3p-1)}} \left[\frac{(p-1)!}{M_t} \right] (A_t + a_t)^{p+\alpha-1}.$$

The next result establishes the relationship between the estimating functions $\psi_t(\cdot)$ and the objective function $f(\cdot)$.

LEMMA 4.2. *For all $t \geq 0$,*

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

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 the convexity of f imply that, for all $x \in \mathbb{E}$,

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

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

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

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

$$\begin{aligned} \psi_t(x) &= \sum_{i=0}^{t-1} a_i [f(x_{i+1}) + \langle \nabla f(x_{i+1}), x - x_{i+1} \rangle] + \frac{\|x - x_0\|^{p+\alpha}}{p + \alpha} \\ &\equiv \ell_t(x) + \frac{1}{p+\alpha} \|x - x_0\|^{p+\alpha} \quad \forall t \geq 1. \end{aligned}$$

Note that $\ell_t(x)$ is a linear function. Moreover, by Lemma 4 in [13], the 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 [9] 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,

$$\begin{aligned} \psi_{t+1}^* &= \min_{x \in \mathbb{E}} \{ \psi_t(x) + a_t [f(x_{t+1}) + \langle \nabla f(x_{t+1}), x - x_{t+1} \rangle] \} \\ &\geq \min_{x \in \mathbb{E}} \{ A_t f(x_t) + \frac{2^{-(p+\alpha-2)}}{(p+\alpha)} \|x - v_t\|^{p+\alpha} \\ &\quad + a_t [f(x_{t+1}) + \langle \nabla f(x_{t+1}), x - x_{t+1} \rangle] \}. \end{aligned}$$

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

$$\begin{aligned} \psi_{t+1}^* &\geq \min_{x \in \mathbb{E}} \{ A_{t+1} f(x_{t+1}) + \langle \nabla f(x_{t+1}), A_t x_t - A_t x_{t+1} \rangle \\ &\quad + a_t \langle \nabla f(x_{t+1}), x - x_{t+1} \rangle + \frac{2^{-(p+\alpha-1)}}{(p+\alpha)} \|x - v_t\|^{p+\alpha} \}. \end{aligned}$$

Note that $y_t = (1 - \gamma_t)x_t + \gamma_t v_t = \frac{A_t}{A_{t+1}}x_t + \frac{a_t}{A_{t+1}}v_t$. Therefore, $A_t x_t = A_{t+1}y_t - a_t v_t$, and

$$\begin{aligned} \psi_{t+1}^* &\geq \min_{x \in \mathbb{E}} \{ A_{t+1} f(x_{t+1}) + \langle \nabla f(x_{t+1}), A_{t+1}y_t - a_t v_t - A_t x_{t+1} \rangle \\ &\quad + a_t \langle \nabla f(x_{t+1}), x - x_{t+1} \rangle + \frac{2^{-(p+\alpha-1)}}{(p+\alpha)} \|x - v_t\|^{p+\alpha} \}. \end{aligned}$$

Moreover, $A_{t+1}x_{t+1} = A_t x_{t+1} + a_t x_{t+1}$, and so

$$\begin{aligned} \psi_{t+1}^* &\geq \min_{x \in \mathbb{E}} \{ A_{t+1} f(x_{t+1}) + A_{t+1} \langle \nabla f(x_{t+1}), y_t - x_{t+1} \rangle \\ &\quad + a_t \langle \nabla f(x_{t+1}), x - v_t \rangle + \frac{2^{-(p+\alpha-1)}}{(p+\alpha)} \|x - v_t\|^{p+\alpha} \} \\ &\geq A_{t+1} f(x_{t+1}) + \min_{x \in \mathbb{E}} \{ A_{t+1} \frac{1}{4} \left[\frac{(p-1)!}{M_t} \right]^{\frac{1}{p+\alpha-1}} \|\nabla f(x_{t+1})\|_*^{\frac{p+\alpha}{p+\alpha-1}} \\ &\quad + a_t \langle \nabla f(x_{t+1}), x - v_t \rangle + \frac{2^{-(p+\alpha-1)}}{(p+\alpha)} \|x - v_t\|^{p+\alpha} \}, \end{aligned}$$

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

$$(4.12) \quad A_{t+1} \frac{1}{4} \left[\frac{(p-1)!}{M_t} \right]^{\frac{1}{p+\alpha-1}} \left\{ \|\nabla f(x_{t+1})\|_*^{\frac{p+\alpha}{p+\alpha-1}} + a_t \langle \nabla f(x_{t+1}), x - v_t \rangle + \frac{2^{-(p+\alpha-1)}}{(p+\alpha)} \|x - v_t\|^{p+\alpha} \right\} \geq 0$$

for all $x \in \mathbb{E}$. Using Lemma 2 in [13] with $r = p + \nu$, $s = a_t \nabla f(x_{t+1})$, and $\omega = 2^{-(p+\alpha-1)}$, we see that a sufficient condition for (4.12) is

$$A_{t+1} \frac{1}{4} \left[\frac{(p-1)!}{M_t} \right]^{\frac{1}{p+\alpha-1}} \|\nabla f(x_{t+1})\|_*^{\frac{p+\alpha}{p+\alpha-1}} \geq \left(\frac{p+\alpha-1}{p+\alpha} \right) 2^{\frac{p+\alpha-2}{p+\alpha-1}} a_t^{\frac{p+\alpha}{p+\alpha-1}} \|\nabla f(x_{t+1})\|_*^{\frac{p+\alpha}{p+\alpha-1}},$$

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 $2 \left(\frac{p+\alpha}{p+\alpha-1} \right)^{p+\alpha-1} \left(\frac{1}{8} \right)^{p+\alpha-1} \geq \frac{1}{2^{(3p-1)}}$. Therefore, by (4.3) we have

$$\begin{aligned} a_t^{p+\alpha} &= \frac{1}{2^{(3p-1)}} \left[\frac{(p-1)!}{M_t} \right] (A_t + a_t)^{p+\alpha-1} \\ &\leq 2 \left(\frac{p+\alpha}{p+\alpha-1} \right)^{p+\alpha-1} \left(\frac{1}{8} \right)^{p+\alpha-1} \left[\frac{(p-1)!}{M_t} \right] A_{t+1}^{p+\alpha-1}. \end{aligned}$$

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 $t = 0, \dots, T$, by (4.3) we get $a_t^{p+\alpha} \geq \frac{1}{M} (A_t + a_t)^{p+\alpha-1}$, with

$$(4.13) \quad \tilde{M} = \frac{2^{(3p-1)} M_\nu}{(p-1)!}.$$

Consequently,

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

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

$$B_{t+1} - B_t \geq B_{t+1}^{\frac{p+\alpha-1}{p+\alpha}}.$$

Then, by Lemma A.4 in [9], we have

$$B_t \geq \left[\left(\frac{1}{p+\alpha} \right) \left(\frac{B_1^{\frac{1}{p+\alpha}}}{B_1^{\frac{1}{p+\alpha}} + 1} \right)^{\frac{p+\alpha-1}{p+\alpha}} \right]^{p+\alpha} (t-1)^{p+\alpha} \quad \forall t \geq 2.$$

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

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

Therefore, for all $t \geq 2$, we have

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

Finally, by (4.11) and Lemma 4.2, for $t \geq 0$, we have

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

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

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

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

Here, by (3.1) the corresponding version of Algorithm 3 takes at most $\mathcal{O}(\epsilon^{-1/(p+\nu)})$ 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 $x_0 \in \mathbb{E}$, $H_0 > 0$, and $\theta \geq 0$. Set α 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

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

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 $\min_{x \in \mathbb{E}} \Omega_{y_{t,i}, p, 2^i H_t}^{(\alpha)}(x)$, such that

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

Step 1.4. If condition

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

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

Step 2. Set $x_{t+1} = x_{t,i_t}^+$, $y_t = y_{t,i_t}$, $a_t = a_{t,i_t}$, and $\gamma_t = \gamma_{t,i_t}$. 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 $\epsilon > 0$:

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

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

LEMMA 4.4. *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

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

Then

$$(4.19) \quad H_t \leq \max \{H_0, \tilde{N}_\nu(\epsilon)\} \quad \text{for } t = 0, \dots, T.$$

Moreover, the number O_T of calls of the oracle after T iterations is bounded as follows:

$$(4.20) \quad O_T \leq 2T + \log_2 \max \{H_0, \tilde{N}_\nu(\epsilon)\} - \log_2 H_0.$$

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 $2^{i_t} H_t$ cannot be bigger than

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

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

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

that is, (4.20) holds for $t + 1$. On the other hand, suppose that $\alpha = 1$. In view of Lemma A.5 in [9], at any trial point $x_{t,i}^+$ we have

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

Thus, it follows from Lemma A.7 that

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

Consequently, we also have $H_{t+1} \leq \max \{ \tilde{N}_\nu(\epsilon), H_0 \}$; i.e., (4.19) holds for $t + 1$. This completes the induction argument. Finally, as in the proof of Lemma 3.3, from (4.19) we get (4.20). \square

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

THEOREM 4.5. *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 4 such that (4.17) and (4.18) hold. Then*

$$(4.21) \quad f(x_t) - f(x^*) \leq \frac{2^{3p} \max \{ \tilde{N}_\nu(\epsilon), H_0 \} (p + \alpha)^{p+\alpha-1} \|x_0 - x^*\|^{p+\alpha}}{(p-1)!(t-1)^{p+\alpha}}, \quad 2 \leq t \leq T.$$

Consequently,

$$(4.22) \quad T \leq 1 + \left[\frac{2^{3p} \max \{ H_0, (p + \nu - 1)(H_{f,p}(\nu) + \theta(p-1)!) \} (p + \nu)^{p+\nu-1} \|x_0 - x^*\|^{p+\nu}}{(p-1)!} \right]^{\frac{1}{p+\nu}} \left(\frac{1}{\epsilon} \right)^{\frac{1}{p+\nu}}$$

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

$$(4.23) \quad T \leq 1 + \left[\frac{2^{3p} \max \left\{ H_0, 4\theta(p-1)!, (4H_{f,p}(\nu))^{\frac{p}{p+\nu-1}} (4D_0)^{\frac{1-\nu}{p+\nu-1}} \right\} (p+1)^p \|x_0 - x^*\|^{p+1}}{(p-1)!} \right]^{\frac{1}{p+1}} \left(\frac{1}{\epsilon} \right)^{\frac{p}{(p+1)(p+\nu-1)}}$$

if ν is unknown (i.e., $\alpha = 1$).

Proof. By Lemma 4.4, we have

$$2^{it} H_t = 2(2^{it-1} H_t) = 2H_{t+1} \leq 2 \max \{ H_0, \tilde{N}_\nu(\epsilon) \}, \quad t = 0, \dots, T-1.$$

Then (4.21) follows directly from Theorem 4.3 with

$$M_\nu = 2 \max \{ H_0, \tilde{N}_\nu(\epsilon) \}.$$

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

$$\epsilon \leq \frac{2^{3p} \max \{ H_0, \tilde{N}_\nu(\epsilon) \} (p + \alpha)^{p+\alpha-1} \|x_0 - x^*\|^{p+\alpha}}{(p-1)!(T-1)^{p+\alpha}},$$

and so

$$(4.24) \quad T \leq 1 + \left[\frac{2^{3p} \max \{ H_0, \tilde{N}_\nu(\epsilon) \} (p + \alpha)^{p+\alpha-1} \|x_0 - x^*\|^{p+\alpha}}{\epsilon (p-1)!} \right]^{\frac{1}{p+\alpha}}.$$

If ν 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 ν is unknown, then $\alpha = 1$ and, by (4.16), (3.16), and $\epsilon \in (0, 1)$, we have

$$(4.26) \quad \begin{aligned} \tilde{N}_\nu(\epsilon) &= \max \left\{ 4\theta(p-1)!, (4H_{f,p}(\nu))^{\frac{p}{p+\nu-1}} \left(\frac{4R(\epsilon)}{\epsilon} \right)^{\frac{1-\nu}{p+\nu-1}} \right\} \\ &\leq \max \left\{ 4\theta(p-1)!, (4H_{f,p}(\nu))^{\frac{p}{p+\nu-1}} (4D_0)^{\frac{1-\nu}{p+\nu-1}} \right\} \epsilon^{-\frac{1-\nu}{p+\nu-1}}. \end{aligned}$$

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

When $\nu = 1$, bounds (4.22) and (4.23) have the same dependence on ϵ . However, when $\nu \neq 1$, the bound of $\mathcal{O}(\epsilon^{-p/(p+1)(p+\nu-1)})$ obtained for the universal scheme (i.e., Algorithm 4 with $\alpha = 1$) is worse than the bound of $\mathcal{O}(\epsilon^{-1/(p+\nu)})$ obtained for the nonuniversal scheme ($\alpha = \nu$). For high-order methods ($p \geq 2$), to the best of our knowledge, there is no simple procedure by which one can identify the level of smoothness ν of the p th derivatives (in general). Therefore, despite this gap in the complexity bounds, we believe that the automatic choice of the best function subclass in the universal scheme is a very attractive feature. Moreover, in the nonuniversal scheme, for any $\theta > 0$ with $(p + \nu - 1)(H_{f,p}(\nu) + \theta(p - 1)!) > H_0$, the corresponding right-hand side of (4.22) has an additional term

$$\left[\frac{2^{3p}\theta(p-1)!(p+\nu)^{p+\nu-1}\|x_0 - x^*\|^{p+\nu}}{(p-1)!} \right]^{\frac{1}{p+\nu}} \left(\frac{1}{\epsilon} \right)^{\frac{1}{p+\nu}}$$

in comparison to its value when $\theta = 0$. In contrast, in the accelerated universal scheme, for any θ in the interval

$$0 < \theta < \min \left\{ \frac{H_0}{4(p-1)!}, \frac{H_{f,p}(\nu)^{\frac{p}{p+\nu-1}} D_0^{\frac{1-\nu}{p+\nu-1}}}{(p-1)!} \right\},$$

the corresponding right-hand side in (4.23) is the same as for $\theta = 0$. In this sense, it appears that the accelerated universal scheme is more robust than the accelerated nonuniversal scheme in terms of the inexact solution of the auxiliary problems.

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 $\mathbb{E} = \mathbb{R}^n$ and $B = I_n$. Given an approximation \bar{x} for the solution of (2.1), p -order methods usually compute the next test point as $x^+ = \bar{x} + \bar{h}$, where the search direction \bar{h} is the solution of an auxiliary problem of the form

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

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

$$(5.2) \quad S_f(\bar{x}) = \text{Lin}(\Gamma_{\bar{x},f}(a, \gamma, m) \mid a \in \mathbb{R}^p, \gamma > 0, m > 1).$$

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 $f_k(\cdot)$ we have $H_{f_k,p}(\nu) < +\infty$.

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

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

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

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

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

It can be shown that (see page 13 in [14])

$$(5.9) \quad \|A_k\| \leq 2.$$

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

$$\begin{aligned} |D^p \eta_{p+\nu}(x)[h]^p - D^p \eta_{p+\nu}(y)[h]^p| &\leq \left(\Pi_{i=1}^{p-1} (p + \nu - i) \right) \sum_{i=1}^n |x^{(i)} - y^{(i)}|^\nu (h^{(i)})^p \\ &\leq \left(\Pi_{i=1}^{p-1} (p + \nu - i) \right) \|x - y\|_\infty^\nu \sum_{i=1}^n (h^{(i)})^p \\ &\leq \left(\Pi_{i=1}^{p-1} (p + \nu - i) \right) \|x - y\|_\infty^\nu \sum_{i=1}^n \left[(h^{(i)})^2 \right]^{\frac{p}{2}} \\ &\leq \left(\Pi_{i=1}^{p-1} (p + \nu - i) \right) \|x - y\|_\infty^\nu \|h\|^p. \end{aligned}$$

Downloaded 06/27/23 to 130.104.59.42 . Redistribution subject to CCBY license

Consequently, for all $x, d, h \in \mathbb{R}^n$, we have

$$(5.10) \quad \begin{aligned} |D^p f_k(x+d)[h]^p - D^p f_k(x)[h]^p| &= |D^p \eta_{p+\nu}(A_k(x+d))[A_k h]^p - D^p \eta_{p+\nu}(A_k x)[A_k h]^p| \\ &\leq \Pi_{i=1}^{p-1} (p+\nu-i) \|A_k d\|_\infty^\nu \|A_k h\|^p. \end{aligned}$$

Note that

$$(5.11) \quad \begin{aligned} \|A_k d\|_\infty &= \max_{1 \leq i \leq n} |(A_k d)^{(i)}| \leq \max_{1 \leq i \leq n-1} (|d^{(i)}| + |d^{(i+1)}|) \\ &\leq \max_{1 \leq i \leq n-1} \sqrt{2[(d^{(i)})^2 + (d^{(i+1)})^2]} \leq 2^{\frac{1}{2}} \|d\| \end{aligned}$$

and, by (5.9), that

$$(5.12) \quad \|A_k h\| \leq \|A\| \|h\| \leq 2 \|h\|.$$

Thus, combining (5.10)–(5.12), we get

$$\|D^p f_k(x+d) - D^p f_k(x)\| \leq 2^{\frac{2+\nu}{2}} \Pi_{i=1}^{p-1} (p+\nu-i) \|d\|^\nu. \quad \square$$

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,*

$$(5.13) \quad f_k^* = -\frac{(p+\nu-1)k}{p+\nu} \quad \text{and} \quad \|x_k^*\| < \frac{(k+1)^{\frac{3}{2}}}{\sqrt{3}}.$$

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

$$(5.14) \quad \nabla \eta_{p+\nu}(y_k^*) = A_k^T e_1 = \hat{e}_k = \begin{bmatrix} \bar{e}_k \\ 0_{n-k} \end{bmatrix},$$

with $\bar{e}_k \in \mathbb{R}^k$ being the vector of all ones and 0_{n-k} being the origin in \mathbb{R}^{n-k} . Note that

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

Consequently, (5.14) is equivalent to

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

Thus,

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

and so

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

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

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

$$\begin{aligned} \|x_k^*\|^2 &= \sum_{i=1}^n \left[(x_k^*)^{(i)} \right]^2 = k^2 + (k-1)^2 + \dots + 2^2 + 1^2 \\ &= \sum_{i=1}^k i^2 = \frac{k(k+1)(2k+1)}{6} < \frac{(k+1)^3}{3}. \end{aligned} \quad \square$$

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:

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

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). □

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

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

Proof. See Lemma 2 in [14]. □

Now we can prove the lower complexity bound for p -order tensor methods applied to the minimization of functions with ν -Hölder continuous p th derivatives.

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:

$$(5.18) \quad \min_{0 \leq k \leq t} f(x_k) - f^* \leq \frac{H_{f,p}(\nu) \|x_0 - x^*\|^{p+\nu}}{\kappa(t)}, \quad t \geq 1,$$

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

$$(5.19) \quad \kappa(t) \leq C_{p,\nu} (t+1)^{\frac{3(p+\nu)-2}{2}},$$

where

$$(5.20) \quad C_{p,\nu} = \frac{2^{\frac{3p+4\nu+2}{2}} \prod_{i=0}^{p-1} (p+\nu-i)}{3^{\frac{p+\nu}{2}} (p+\nu-1)}.$$

Downloaded 06/27/23 to 130.104.59.42 . Redistribution subject to CCBY license

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 $x_i \in \mathbb{R}_t^n$ for all i , $0 \leq i \leq t$. Moreover, by Lemma 5.3 we have

$$(5.21) \quad f_{2t+1}(x) = f_t(x) \quad \forall x \in \mathbb{R}_t^n.$$

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

$$\begin{aligned} \kappa(t) &\leq \frac{H_{f_{2t+1}, p}(\nu) \|x_0 - x_{2t+1}^*\|^{p+\nu}}{\min_{0 \leq k \leq t} f_{2t+1}(x_k) - f_{2t+1}^*} = \frac{2^{\frac{2+\nu}{2}} \prod_{i=1}^{p-1} (p + \nu - i) \|x_{2t+1}^*\|^{p+\nu}}{\min_{0 \leq k \leq t} f_t(x_k) - f_{2t+1}^*} \\ &\leq \frac{2^{\frac{2+\nu}{2}} \prod_{i=1}^{p-1} (p + \nu - i) \|x_{2t+1}^*\|^{p+\nu}}{f_t^* - f_{2t+1}^*} < \frac{2^{\frac{2+\nu}{2}} \prod_{i=1}^{p-1} (p + \nu - i) [2(t+1)]^{\frac{3}{2}(p+\nu)}}{3^{\frac{p+\nu}{2}} (f_t^* - f_{2t+1}^*)} \\ &= \frac{2^{\frac{3p+4\nu+2}{2}} \prod_{i=1}^{p-1} (p + \nu - i) (t+1)^{\frac{3(p+\nu)}{2}}}{3^{\frac{p+\nu}{2}} \left[\frac{-(p + \nu - 1)t + (p + \nu - 1)(2t+1)}{p + \nu} \right]} \\ &= \frac{2^{\frac{3p+4\nu+2}{2}} \prod_{i=0}^{p-1} (p + \nu - i) (t+1)^{\frac{3(p+\nu)}{2}}}{3^{\frac{p+\nu}{2}} (p + \nu - 1) (t+1)} = C_{p,\nu} (t+1)^{\frac{3(p+\nu)-2}{2}}, \end{aligned}$$

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

5.2. Discussion. Theorem 5.5 establishes that the lower bound for the rate of convergence of tensor methods applied to functions with ν -Hölder continuous p th derivatives is of $\mathcal{O}\left(\left(\frac{1}{k}\right)^{\frac{3(p+\nu)-2}{2}}\right)$. In the Lipschitz case (i.e., $\nu = 1$), we have $\mathcal{O}\left(\left(\frac{1}{k}\right)^{\frac{3p+1}{2}}\right)$, which coincides with the bounds in [1, 14]. On the other hand, for first-order methods (i.e., $p = 1$) we have $\mathcal{O}\left(\left(\frac{1}{k}\right)^{\frac{1+3\nu}{2}}\right)$, which is the bound in [12].

The rate of $\mathcal{O}\left(\left(\frac{1}{k}\right)^{\frac{3(p+\nu)-2}{2}}\right)$ corresponds to a worst-case complexity bound of $\mathcal{O}\left(\epsilon^{-2/\lceil 3(p+\nu)-2 \rceil}\right)$ iterations necessary to ensure $f(x_k) - f(x^*) \leq \epsilon$. This means that the nonuniversal 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

$$\begin{aligned} \left(\frac{1}{\epsilon}\right)^{\frac{1}{p+\nu}} &= \left(\frac{1}{\epsilon}\right)^{\frac{p+\nu-2}{(p+\nu)\lceil 3(p+\nu)-2 \rceil}} \left(\frac{1}{\epsilon}\right)^{\frac{2}{3(p+\nu)-2}} \leq \left(\frac{1}{\epsilon}\right)^{\frac{p-1}{3p^2-2p}} \left(\frac{1}{\epsilon}\right)^{\frac{2}{3(p+\nu)-2}} \\ &\leq \left(\frac{1}{\epsilon}\right)^{\frac{1}{8}} \left(\frac{1}{\epsilon}\right)^{\frac{2}{3(p+\nu)-2}}. \end{aligned}$$

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 nonuniversal methods differ from the lower bound just by a small constant factor.

Notice that the lower-bound described in Theorem 5.5 is only valid while the iteration counter t satisfies $t < \frac{1}{2}(n-1)$, where n is the dimension of the domain of the objective. The same condition on t appears in other lower bounds in the literature for the case $p = 1$ and $\nu = 1$ (see, e.g., Theorem 2.1.7 in [15]).

6. Conclusion. In this paper, we presented p -order methods for unconstrained minimization of convex functions that are p -times differentiable with ν -Hölder continuous p th derivatives. For the universal and the nonuniversal schemes without acceleration, we established iteration complexity bounds of $\mathcal{O}\left(\epsilon^{-1/(p+\nu-1)}\right)$ for reducing the functional residual below a given $\epsilon \in (0, 1)$. Assuming that ν is known,

we obtained an improved complexity bound of $\mathcal{O}(\epsilon^{-1/(p+\nu)})$ for the corresponding accelerated scheme. For the case in which ν is unknown, we presented an accelerated universal tensor scheme with an iteration complexity of $\mathcal{O}(\epsilon^{-p/[(p+1)(p+\nu-1)])}$.

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 nonuniversal schemes are nearly optimal. Remarkably, the complexity bound obtained for the accelerated universal schemes is slightly worse than the bound obtained for the nonuniversal 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 $\mathcal{O}(\epsilon^{-1/(p+\nu)})$.

It is worth mentioning that the study of high-order methods is still at its early stages, with the majority of recent works in this area focusing on the derivation of global complexity bounds (see, e.g., [2, 4, 6, 7, 11, 14]). These bounds predict that high-order methods with $p \geq 3$ may require significantly fewer calls of the oracle than second-order methods. As pointed out in [7, 14], the computation of high-order derivatives may be affordable for structured objectives (such as separable functions). Moreover, at least for $p = 3$ and $\alpha = \nu$, the auxiliary problems can be solved using Bregman gradient methods that also take into account their particular structure [10, 14]. Nevertheless, the practical impact of high-order methods is yet to be seen.

Appendix A. Auxiliary results. In all algorithms described in this paper, the acceptance of new points is conditioned to the achievement of a sufficient decrease of the objective function value. In the nonaccelerated schemes, the sufficient decrease condition is specified by (3.4), while for accelerated schemes, it is specified by (4.6). In this appendix, we present auxiliary results from which we conclude that these conditions are satisfied when the regularization parameter is sufficiently large.

A.1. Results for schemes without acceleration. Our first lemma gives a lower bound for the functional decrease in terms of a suitable power of the norm of the displacement when ν is known.

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

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

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

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

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

$$\begin{aligned} f(x^+) &\leq \Omega_{\bar{x},p,H_{f,p}(\nu)}^{(\nu)}(x^+) \\ &= \Phi_{\bar{x},p}(x^+) + \frac{H}{p!} \|x^+ - \bar{x}\|^{p+\nu} - \frac{(H - H_{f,p}(\nu))}{p!} \|x^+ - \bar{x}\|^{p+\nu} \\ &= \Omega_{\bar{x},p,H}^{(\nu)}(x^+) - \frac{(H - H_{f,p}(\nu))}{p!} \|x^+ - \bar{x}\|^{p+\nu} \\ &\leq f(\bar{x}) - \frac{(H - H_{f,p}(\nu))}{p!} \|x^+ - \bar{x}\|^{p+\nu}, \end{aligned}$$

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

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

The next lemma provides a lower bound for the functional decrease in terms of a suitable power of the norm of the gradient when ν is known.

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*

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

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

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

then

$$(A.5) \quad f(\bar{x}) - f(x^+) \geq \frac{1}{8(p+1)!H^{\frac{1}{p+\nu-1}}} \|\nabla f(x^+)\|_*^{\frac{p+\nu}{p+\nu-1}}.$$

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

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

Thus,

$$(A.6) \quad \|x^+ - \bar{x}\|^{p+\nu} \geq \left(\frac{1}{2H} \right)^{\frac{p+\nu}{p+\nu-1}} \|\nabla f(x^+)\|_*^{\frac{p+\nu}{p+\nu-1}}.$$

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

$$(A.7) \quad f(\bar{x}) - f(x^+) \geq \frac{H}{(p+1)!} \|x^+ - \bar{x}\|^{p+\nu}.$$

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

The lemma below gives lower bounds for powers of the norm of the displacement when ν is unknown.

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

$$(A.8) \quad \|\nabla\Omega_{\bar{x},p,H}^{(1)}(x^+)\|_* \leq \theta \|x^+ - \bar{x}\|^p$$

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

$$(A.9) \quad \|\nabla f(x^+)\|_* \geq \delta \quad \text{and} \quad H \geq \max \left\{ \theta, (CH_{f,p}(\nu))^{\frac{p}{p+\nu-1}} \left(\frac{4}{\delta} \right)^{\frac{1-\nu}{p+\nu-1}} \right\},$$

with constant $C \geq 1$, then

$$(A.10) \quad \|x^+ - \bar{x}\|^{1-\nu} \geq \frac{CH_{f,p}(\nu)}{H}$$

and, consequently,

$$(A.11) \quad 4H\|x^+ - \bar{x}\|^p \geq \|\nabla f(x^+)\|_*.$$

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

$$(A.12) \quad \begin{aligned} \delta &< \|\nabla f(x^+)\|_* \\ &\leq \|\nabla f(x^+) - \nabla \Phi_{\bar{x},p}(x^+)\|_* + \|\nabla \Phi_{\bar{x},p}(x^+) - \nabla \Omega_{\bar{x},p,H}^{(1)}(x^+)\|_* \\ &\quad + \|\nabla \Omega_{\bar{x},p,H}^{(1)}(x^+)\|_* \\ &\leq \frac{H_{f,p}(\nu)}{(p-1)!} r^{p+\nu-1} + \left(\frac{H(p+1)}{p!} + \theta\right) r^p \\ &= r^{p+\nu-1} \left[\frac{H_{f,p}(\nu)}{(p-1)!} + \left(\frac{p+1}{p!} + \frac{\theta}{H}\right) Hr^{1-\nu}\right]. \end{aligned}$$

Assume by contradiction that (A.10) is not true, i.e., $Hr^{1-\nu} < CH_{f,p}(\nu)$. Since $H \geq \theta$ and $C \geq 1$, it follows that

$$\begin{aligned} \delta &< r^{p+\nu-1} \left[\frac{H_{f,p}(\nu)}{(p-1)!} + \left(\frac{p+1}{p!} + 1\right) CH_{f,p}(\nu)\right] \\ &= \frac{r^{p+\nu-1} H_{f,p}(\nu)}{(p-1)!} \left[1 + \frac{C(p+1)}{p} + C(p-1)!\right] \\ &\leq 4CH_{f,p}(\nu)r^{p+\nu-1} < 4CH_{f,p}(\nu) \left(\frac{CH_{f,p}(\nu)}{H}\right)^{\frac{p+\nu-1}{1-\nu}} \\ &= 4(CH_{f,p}(\nu))^{\frac{p}{1-\nu}} \left(\frac{1}{H}\right)^{\frac{p+\nu-1}{1-\nu}}. \end{aligned}$$

This implies that $H < (CH_{f,p}(\nu))^{\frac{p}{p+\nu-1}} \left(\frac{4}{\delta}\right)^{\frac{1-\nu}{p+\nu-1}}$, 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

$$\frac{H_{f,p}(\nu)}{(p-1)!} \leq \frac{H}{C(p-1)!} r^{1-\nu}.$$

Thus, it follows from (A.12) that

$$\begin{aligned} \|\nabla f(x^+)\|_* &\leq r^{p+\nu-1} \left[\frac{H}{C(p-1)!} r^{1-\nu} + \left(\frac{p+1}{p!} + \frac{\theta}{H}\right) Hr^{1-\nu}\right] \\ &= r^p H \left[\frac{1}{Cp!} + \frac{(p+1)}{p!} + \frac{\theta}{H}\right] \leq 4r^p H. \quad \square \end{aligned}$$

Now, using Lemma A.3, we obtain a lower bound for the functional decrease in terms of a computable power of the norm of the displacement when ν is unknown.

LEMMA A.4. Let $H_{f,p}(\nu) < +\infty$ for some $\nu \in [0, 1]$, and assume that $x^+ \in \mathbb{E}$ satisfies

$$(A.13) \quad \Omega_{\bar{x},p,H}^{(1)}(x^+) < f(\bar{x})$$

and

$$(A.14) \quad \|\nabla \Omega_{\bar{x},p,H}^{(1)}(x^+)\|_* \leq \theta \|x^+ - \bar{x}\|^p$$

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

$$(A.15) \quad \|\nabla f(x^+)\|_* \geq \delta \quad \text{and} \quad H \geq \max \left\{ \theta, (CH_{f,p}(\nu))^{\frac{p}{p+\nu-1}} \left(\frac{4}{\delta} \right)^{\frac{1-\nu}{p+\nu-1}} \right\},$$

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

$$\begin{aligned} f(x^+) &\leq \Omega_{\bar{x},p,H_{f,p}(\nu)}^{(\nu)}(x^+) \\ &= \Phi_{\bar{x},p}(x^+) + \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} \\ &= \Phi_{\bar{x},p}(x^+) + \frac{H}{p!} \|x^+ - \bar{x}\|^{p+1} - \frac{H}{p!} \|x^+ - \bar{x}\|^{p+1} + \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} \\ &= \Omega_{\bar{x},p,H}^{(1)}(x^+) - \frac{H}{p!} \|x^+ - \bar{x}\|^{p+1} + \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu} \\ &< f(\bar{x}) - \frac{H}{p!} \|x^+ - \bar{x}\|^{p+1} + \frac{H_{f,p}(\nu)}{p!} \|x^+ - \bar{x}\|^{p+\nu}, \end{aligned}$$

and so

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

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

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

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

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

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

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

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

$$\frac{Hp}{p+1} < H_{f,p}(\nu) \left(\frac{\delta}{4H}\right)^{\frac{\nu-1}{p}} \implies H < \left(\frac{3}{2}H_{f,p}(\nu)\right)^{\frac{p}{p+\nu-1}} \left(\frac{4}{\delta}\right)^{\frac{1-\nu}{p+\nu-1}},$$

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

Finally, the next lemma gives a lower bound for the functional decrease in terms of a computable power of the norm of the gradient when ν is unknown.

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

$$(A.21) \quad \xi_\nu(\delta) \equiv \max \left\{ \theta, \left(\frac{3H_{f,p}(\nu)}{2}\right)^{\frac{p}{p+\nu-1}} \left(\frac{4}{\delta}\right)^{\frac{1-\nu}{p+\nu-1}} \right\}.$$

If $\|\nabla f(x^+)\|_* \geq \delta$ and $H \geq \xi_\nu(\delta)$, then

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

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

$$\begin{aligned} f(\bar{x}) - f(x^+) &\geq \frac{H}{(p+1)!} \|x^+ - \bar{x}\|^{p+1} \geq \frac{H}{(p+1)!} \frac{1}{4^{\frac{p+1}{p}} H^{\frac{p+1}{p}}} \|\nabla f(x^+)\|_*^{\frac{p+1}{p}} \\ &\geq \frac{1}{8(p+1)!H^{\frac{1}{p}}} \|\nabla f(x^+)\|_*^{\frac{p+1}{p}}. \end{aligned} \quad \square$$

A.2. Results for accelerated schemes. For the case in which ν is known, the lemma below establishes that (4.6) is achievable when the regularization parameter is sufficiently large.

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

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

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

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

then

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

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

$$\begin{aligned} \|\nabla f(x^+) + \frac{H(p+\nu)}{p!} r^{p+\nu-2} B(x^+ - \bar{x})\|_* &= \|\nabla f(x^+) - \nabla \Phi_{\bar{x},p}(x^+) + \nabla \Omega_{\bar{x},p,H}^{(\nu)}(x^+)\|_* \\ &\leq \|\nabla f(x^+) - \nabla \Phi_{\bar{x},p}(x^+)\|_* + \|\nabla \Omega_{\bar{x},p,H}^{(\nu)}(x^+)\|_* \\ &\leq \left(\frac{H_{f,p}(\nu)}{(p-1)!} + \theta\right) r^{p+\nu-1}. \end{aligned}$$

Thus, we obtain

$$\begin{aligned} \left(\frac{H_{f,p}(\nu)}{(p-1)!} + \theta\right) r^{2(p+\nu-1)} &\geq \|\nabla f(x^+) + \frac{H(p+\nu)}{p!} r^{p+\nu-2} B(x^+ - \bar{x})\|_*^2 \\ &= \|\nabla f(x^+)\|_*^2 + \frac{2(p+\nu)}{p!} H r^{p+\nu-2} \langle \nabla f(x^+), x^+ - \bar{x} \rangle \\ &\quad + \frac{H^2(p+\nu)^2}{(p!)^2} r^{2(p+\nu-1)}, \end{aligned}$$

which implies that
(A.25)

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

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

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

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

$$h(r) = \frac{p!}{2(p+\nu)Hr^{p+\nu-2}} g^2 + \frac{Hp^2\Delta^2 r^{p+\nu}}{2(p+\nu)p!}.$$

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 :

$$\frac{g^2(p+\nu-2)p!}{(p+\nu)Hr_*^{p+\nu-1}} = \frac{Hp^2\Delta^2 r_*^{p+\nu-1}}{p!}.$$

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,

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

Now, using (A.26), we obtain

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

Downloaded 06/27/23 to 130.104.59.42 . Redistribution subject to CCBY license

Note that

$$\begin{aligned} \frac{(p + \nu - 1)^{\frac{1}{p+\nu-1}} p}{(p + \nu)^{\frac{p+\nu}{p+\nu-1}}} &= \frac{p(p + \nu - 1)^{\frac{1}{p+\nu-1}}}{(p + \nu)(p + \nu)^{\frac{p+\nu}{p+\nu-1}}} \geq \left(\frac{p}{p + 1}\right) \left(\frac{p - 1}{p + 1}\right) \\ &\geq \frac{1}{3}. \end{aligned}$$

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). \square

Finally, for the case in which ν is unknown, the next lemma establishes that (4.6) is also achievable when the regularization parameter is sufficiently large.

LEMMA A.7. *Let $H_{f,p}(\nu) < +\infty$ for some $\nu \in [0, 1]$, and assume that x^+ satisfies*

$$(A.27) \quad \|\nabla\Omega_{\bar{x},p,H}^{(1)}(x^+)\|_* \leq \theta \|x^+ - \bar{x}\|^p$$

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

$$(A.28) \quad \|\nabla f(x^+)\|_* \geq \delta \quad \text{and} \quad H \geq \max \left\{ C\theta(p-1)!, (CH_{f,p}(\nu))^{\frac{p}{p+\nu-1}} \left(\frac{4}{\delta}\right)^{\frac{1-\nu}{p+\nu-1}} \right\},$$

with $C \geq 4$, then

$$(A.29) \quad \langle \nabla f(x^+), \bar{x} - x^+ \rangle \geq \frac{1}{4} \left[\frac{(p-1)!}{H}\right]^{\frac{1}{p}} \|\nabla f(x^+)\|_*^{\frac{p+1}{p}}.$$

Proof. Denote $r = \|x^+ - \bar{x}\|$. Then, by (2.4), (2.6), and (A.27), we have

$$\begin{aligned} \|\nabla f(x^+) + \frac{H(p+1)}{p!} r^{p-1} B(x^+ - \bar{x})\|_* &= \|\nabla f(x^+) - \nabla\Phi_{\bar{x},p}(x^+) + \nabla\Omega_{\bar{x},p,H}^{(1)}(x^+)\|_* \\ &\leq \|\nabla f(x^+) - \nabla\Phi_{\bar{x},p}(x^+)\|_* + \|\nabla\Omega_{\bar{x},p,H}^{(1)}(x^+)\|_* \\ &\leq \frac{H_{f,p}(\nu)}{(p-1)!} r^{p+\nu-1} + \theta r^p \\ &\leq \left(\frac{H}{C(p-1)!} + \theta\right) r^p. \end{aligned}$$

Therefore,

$$\begin{aligned} \left(\frac{H}{C(p-1)!} + \theta\right)^2 r^{2p} &\geq \left\| \nabla f(x^+) + \frac{H}{p!} r^{p-1} B(x^+ - \bar{x}) \right\|_*^2 \\ &= \|\nabla f(x^+)\|_*^2 + \frac{2(p+1)}{p!} H r^{p-1} \langle f^{(1)}(x^+), x^+ - \bar{x} \rangle \frac{H^2(p+1)^2}{(p!)^2} r^{2p}, \end{aligned}$$

which gives (A.30)

$$\langle \nabla f(x^+), \bar{x} - x^+ \rangle \geq \frac{p!}{2(p+1)Hr^{p-1}} \|\nabla f(x^+)\|_*^2 + \frac{p!}{2H(p+1)} \left[\left(\frac{H(p+1)}{p!}\right)^2 - \left(\frac{H}{C(p-1)!} + \theta\right)^2 \right] r^{p+1}.$$

Since $H \geq C\theta(p-1)!$, it follows that

$$\frac{p!}{2H(p+1)} \left[\left(\frac{H(p+1)}{p!}\right)^2 - \left(\frac{H}{C(p-1)!} + \theta\right)^2 \right] r^{p+1} = \frac{Hp!}{2(p+1)} \left[\left(\frac{p+1}{p!}\right)^2 - \left(\frac{2}{C(p-1)!}\right)^2 \right]^2.$$

Because $C \geq 4$, we have

$$-\left(\frac{2}{C(p-1)!}\right)^2 \geq -\left(\frac{1}{2(p-1)!}\right)^2,$$

and so

$$\begin{aligned} \frac{p!}{2H(p+1)} \left[\left(\frac{H(p+1)}{p!}\right)^2 - \left(\frac{H}{C(p-1)!} + \theta\right)^2 \right] r^{p+1} &\geq \frac{Hp!}{2(p+1)} \left[\left(\frac{p+1}{p}\right)^2 - \left(\frac{1}{2(p-1)!}\right)^2 \right] \\ &\geq \frac{3Hp^2}{8(p+1)!}. \end{aligned}$$

Therefore,

$$(A.31) \quad \langle \nabla f(x^+), \bar{x} - x^+ \rangle \geq \frac{p!}{2H(p+1)r^{p-1}} \|\nabla f(x^+)\|^2 + \frac{3Hp^2}{8(p+1)!} r^{p+1}.$$

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

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

Let us find the optimal r_* as a solution to the first-order optimality condition for function h :

$$\frac{(p-1)g^2 p!}{2(p+1)Hr_*^p} = \frac{3Hp^2(p+1)r_*^p}{8(p+1)!} = \frac{3Hp^2 r_*^p}{8p!}.$$

Solving this equation for r_* , we obtain

$$r_*^p = \frac{g(p-1)!}{H\Delta} \sqrt{\frac{8(p-1)}{3(p+1)}}.$$

Consequently,

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

Therefore, (A.29) holds. \square

Acknowledgment. The authors are very grateful to the two anonymous referees, whose comments helped to improve the first version of this paper.

REFERENCES

- [1] Y. ARJEVANI, O. SHAMIR, AND R. SHIFF, *Oracle complexity of second-order methods for smooth convex optimization*, Math. Program., 178 (2019), pp. 327–360.
- [2] M. BAES, *Estimate sequence methods: Extensions and approximations*, Optimization Online, 2009.

- [3] S. BANACH, *Über homogene Polynome in (L_2)* , *Studia Math.*, 7 (1938), pp. 36–44.
- [4] E.G. BIRGIN, J.L. GARDENGI, J.M. MARTÍNEZ, S.A. SANTOS, AND PH.L. TOINT, *Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models*, *Math. Program.*, 163 (2017), pp. 359–368.
- [5] A. BOUARICHA, *Tensor methods for large, sparse unconstrained optimization*, *SIAM J. Optim.*, 7 (1997), pp. 732–756, <https://doi.org/10.1137/S1052623494267723>.
- [6] C. CARTIS, N.I. GOULD, AND P.L. TOINT, *Universal regularized methods: Varying the power, the smoothness and the accuracy*, *SIAM J. Optim.*, 29 (2019), pp. 595–615, <https://doi.org/10.1137/16M1106316>.
- [7] X. CHEN, PH.L. TOINT, AND H. WANG, *Complexity of partially separable convexly constrained optimization with non-Lipschitzian singularities*, *SIAM J. Optim.*, 29 (2019), pp. 874–903, <https://doi.org/10.1137/18M1166511>.
- [8] G.N. GRAPIGLIA AND YU. NESTEROV, *Regularized Newton methods for minimizing functions with Hölder continuous Hessians*, *SIAM J. Optim.*, 27 (2017), pp. 478–506, <https://doi.org/10.1137/16M1087801>.
- [9] G.N. GRAPIGLIA AND Y. NESTEROV, *Accelerated regularized Newton methods for minimizing composite convex functions*, *SIAM J. Optim.*, 29 (2019), pp. 77–99, <https://doi.org/10.1137/17M1142077>.
- [10] G.N. GRAPIGLIA AND YU. NESTEROV, *On inexact solution of auxiliary problems in tensor methods for convex optimization*, *Optim. Methods Softw.*, (2020), <https://doi.org/10.1080/10556788.2020.1731749>.
- [11] J.M. MARTÍNEZ, *On high-order model regularization for constrained optimization*, *SIAM J. Optim.*, 27 (2017), pp. 2447–2458, <https://doi.org/10.1137/17M1115472>.
- [12] A. NEMIROVSKII AND D. YUDIN, *Problem Complexity and Method Efficiency in Optimization*, John Wiley & Sons, New York, 1983.
- [13] YU. NESTEROV, *Accelerating the cubic regularization of Newton’s method on convex problems*, *Math. Program.*, 112 (2008), pp. 159–181.
- [14] YU. NESTEROV, *Implementable tensor methods in unconstrained convex optimization*, *Math. Program.*, (2019), <https://doi.org/10.1007/s10107-019-01449-1>.
- [15] YU. NESTEROV, *Lectures on Convex Optimization*, 2nd ed., Springer, Cham, 2018.
- [16] YU. NESTEROV AND B.T. POLYAK, *Cubic regularization of Newton method and its global performance*, *Math. Program.*, 108 (2006), pp. 177–205.
- [17] R.B. SCHNABEL AND T.-T. CHOW, *Tensor methods for unconstrained optimization using second derivatives*, *SIAM J. Optim.*, 1 (1991), pp. 293–315, <https://doi.org/10.1137/0801020>.

Downloaded 06/27/23 to 130.104.59.42 . Redistribution subject to CCBY license