On the Complexity of Lower-Order Implementations of Higher-Order Methods

Nikita Doikov* Geovani Nunes Grapiglia†

October 9, 2025

Abstract

In this work, we propose a method for minimizing non-convex functions with Lipschitz continuous pth-order derivatives, starting from $p \geq 1$. The method, however, only requires derivative information up to order $(p-1)$, since the pth-order derivatives are approximated via finite differences. To ensure oracle efficiency, instead of computing finite-difference approximations at every iteration, we reuse each approximation for m consecutive iterations before recomputing it, with $m \geq 1$ as a key parameter. As a result, we obtain an adaptive method of order $(p-1)$ that requires no more than $\mathcal{O}(\epsilon^{-\frac{p+1}{p}})$ iterations to find an ϵ-approximate stationary point of the objective function and that, for the choice $m = (p-1)n + 1$, where n is the problem dimension, takes no more than $\mathcal{O}(n^{1/p}\epsilon^{-\frac{p+1}{p}})$ oracle calls of order $(p-1)$. This improves previously known bounds for tensor methods with finite-difference approximations in terms of the problem dimension.

Keywords: Nonconvex Optimization; Higher-Order Methods; Tensor Methods; Finite Difference; Worst-Case Complexity

1 Introduction

1.1 Motivation

We consider the problem of unconstrained minimization of a smooth function $f : \mathbb{R}^n \to \mathbb{R}$, which can be non-convex. The choice of appropriate algorithms, as well as their performance guarantees, depends fundamentally on the smoothness properties of f and on the extent to which its derivative information can be computed. When for a given $x \in \mathbb{R}^n$ one can evaluate $f(x)$ as well as all derivatives $\nabla^i f(x)$ for $i = 1, \ldots, q$ for some $q \leq p$, we say that a qth-order oracle for f is available. Any algorithm that relies on such an oracle is referred to as a qth-order method.

In this context, the Gradient Method is a first-order method that, when applied to an objective function f with the Lipschitz continuous gradient, requires at most $\mathcal{O}(\epsilon^{-2})$ calls to the first-order oracle to find an ϵ-approximate stationary point [19]; that is, a point $\bar{x}$ satisfying $\|\nabla f(\bar{x})\| \leq \epsilon$.

*Machine Learning and Optimization Laboratory (MLO), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland (nikita.doikov@epfl.ch). This author was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00133.

†Department of Mathematical Engineering, ICTEAM Institute, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium (geovani.grapiglia@uclouvain.be). This author was partially supported by FRS-FNRS, Belgium (Grant CDR J.0081.23).

Turning to second-order methods, Newton's Method with Cubic Regularization [14, 22, 6], when applied to an objective function f with an Lipschitz continuous Hessian, requires at most $\mathcal{O}(\epsilon^{-3/2})$ calls to the second-order oracle to find an ϵ-approximate stationary point [22]. More generally, if f has a Lipschitz continuous pth-order derivative, the methods based on regularized pth-order Taylor models [2] require at most $\mathcal{O}(\epsilon^{-(p+1)/p})$ calls to the pth-order oracle to find an ϵ-approximate stationary point. As shown in [3], this represents an optimal worst-case complexity bound for pth-order methods applied to the class of functions with Lipschitz continuous pth-order derivatives.

In minimizing a p-times differentiable function f, one often has access only to a q-th order oracle, where $q < p$. For example, when $f(x)$ can be evaluated only through a black-box simulation or experiment, the gradient vector $\nabla f(x)$ is not readily available. In such cases, one must rely solely on function evaluations to minimize f, which motivates the use of derivative-free (zeroth-order) methods [8, 1, 16]. Another example arises when calibrating the parameters of ODE or PDE models to fit a given dataset. Although the error function may be twice differentiable, computing the Hessian via adjoint equations can be computationally prohibitive or practically challenging [24], forcing users to rely solely on a first-order oracle [15, 23, 17].

A common approach to exploiting pth-order smoothness with a lower-order oracle is to approximate the unavailable higher-order derivatives via *finite differences* of the available lower-order information. For example, by approximating gradient vectors with forward finite differences of function values, one can design zeroth-order implementations of first-order methods [11] that find ϵ-approximate stationary points using no more than $\mathcal{O}\left(n\epsilon^{-2}\right)$ calls to the zeroth-order oracle (i.e., function evaluations), where n is the problem dimension. Similarly, finite differences of gradients can be used to approximate Hessian matrices in second-order methods, leading to first-order implementations [4, 12] with worst-case oracle complexity $\mathcal{O}\left(n\epsilon^{-3/2}\right)$. Recently, in [10], we established an improved complexity bound of $\mathcal{O}\left(n^{1/2}\epsilon^{-3/2}\right)$ for a *lazy* variant of Newton's Method with Cubic Regularization [9], in which finite-difference Hessian approximations are computed only once every n iterations and kept fixed in between. Further developing the lazy technique, [7] proposed a method of order $(p-1)$ in which finite-difference approximations of the pth-order derivatives are computed only once every m iterations and updated in between via quasi-tensor rules [25], where $m \in \mathbb{N} \setminus \{0\}$ is a user-defined parameter. It was shown in [7] that this method requires at most $\mathcal{O}\Big(n \max\Big\{\epsilon_1^{-\frac{p+1}{p}}, \epsilon_2^{-\frac{p+1}{p-1}}\Big\}\Big)$ calls to the oracle of order $(p-1)$ to find an (ϵ_1, ϵ_2)-approximate second-order stationary point, i.e., an ϵ_1-approximate stationary point of f at which the smallest eigenvalue of the Hessian is at least $-\epsilon_2$.

1.2 Contributions

In this work, we propose a new method of order $(p-1)$ for minimizing p-times differentiable functions with Lipschitz continuous pth-order derivatives. To ensure oracle efficiency, we extend the lazy approach from [10] to the general case $p \geq 1$, where each finite-difference approximation of the pth-order derivative is reused for at most m consecutive iterations before a new approximation is computed. We show that, by choosing $m = (p-1)n + 1$, our method requires at most

$$\mathcal{O}\Big(n^{1/p}\epsilon^{-\frac{p+1}{p}}\Big)$$

calls to the oracle of order $(p-1)$ to find an ϵ-approximate stationary point. With respect to explicit dependence on the problem dimension n, this is, to our knowledge, the sharpest known oracle complexity bound for $(p-1)$th-order methods applied to functions with Lipschitz-continuous pth derivatives. Importantly, our method is adaptive, automatically adjusting the Lipschitz constant of the pth derivative and the parameter of the finite-difference approximation, without the need to fix them in advance.

1.3 Contents

The paper is organized as follows. Section 2 introduces the problem and presents the necessary auxiliary results. In Section 3, we describe our new method and analyze its worst-case oracle complexity. Section 4 concludes with a discussion of open problems and directions for future research.

1.4 Notation

For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we denote by $\nabla^q f(x)$ the derivative of order $q \geq 1$, which is a q-linear symmetric form. The value of this form on a set of fixed directions $h_1, \ldots, h_q \in \mathbb{R}^n$ is denoted by $\nabla^q f(x)[h_1, \ldots, h_q] \in \mathbb{R}$, which is the qth-order directional derivative of f along the given directions. When all directions are the same, $h_1 \equiv \ldots \equiv h_q \equiv h \in \mathbb{R}^n$ we use the shorthand $\nabla^q f(x)[h]^q \in \mathbb{R}$. More generally, for an arbitrary $1 \leq \ell \leq q$, we use the convenient notation $\nabla^q f(x)[h]^\ell$ for the $(q-\ell)$-form with the first ℓ directions substituted as h; by definition, it satisfies

$$\nabla^q f(x)[h]^\ell[u_1, \ldots, u_{q-\ell}] \quad \equiv \quad \nabla^q f(x)[h, \ldots h, u_1, \ldots, u_{q-\ell}] \quad \in \quad \mathbb{R}, \quad \text{for all } u_1, \ldots, u_{q-\ell} \in \mathbb{R}^n.$$

We fix the standard Euclidean inner product in our space, $\langle x, y \rangle := \sum_{i=1}^n x^{(i)} y^{(i)}$, for any $x, y \in \mathbb{R}^n$, and denote by $e_1, \ldots, e_n \in \mathbb{R}^n$ the canonical basis. Using the inner product, we treat the 1-form $\nabla f(x) \in \mathbb{R}^n$ as the gradient vector: $\langle \nabla f(x), u \rangle \equiv \nabla f(x)[u] \in \mathbb{R}$ for $x, u \in \mathbb{R}^n$. And for general $q \geq 1$ and a fixed direction $h \in \mathbb{R}^n$, we can treat $\nabla^q f(x)[h]^{q-1} \in \mathbb{R}^n$ as the vector, which satisfies $\langle \nabla^q f(x)[h]^{q-1}, u \rangle \equiv \nabla^q f(x)[h]^{q-1}[u] \in \mathbb{R}$, for all $x, h, u \in \mathbb{R}^n$.

We denote by $\| \cdot \|$ the standard Euclidean norm for vectors, $\|x\| := \langle x, x \rangle^{1/2}$, $x \in \mathbb{R}^n$. Correspondingly, for an arbitrary q-linear (not necessary symmetric) function $T[h_1, \ldots, h_q] \in \mathbb{R}$, $q \geq 1$ we use the induced operator norm:

$$\|T\| \quad := \max_{\substack{h_1, \ldots, h_q \in \mathbb{R}^n: \\ \|h_i\| \leq 1, \forall 1 \leq i \leq q}} |T[h_1, \ldots, h_q]|.$$

When T is symmetric, it holds (see, e.g., Appendix 1 in [21]) $\|T\| = \max_{h \in \mathbb{R}^n : \|h\| \leq 1} |T[h]^q|$.

In what follows, we will use the following construction. For a given symmetric $(q-1)$-linear form E, and an arbitrary vector $e \in \mathbb{R}^n$, we denote by $T = E \otimes e$ the q-linear form, defined by

$$T[h_1, \ldots h_q] \quad \equiv \quad E[h_1, \ldots, h_{q-1}] \cdot \langle e, h_q \rangle \quad \in \quad \mathbb{R}, \qquad h_1, \ldots, h_q \in \mathbb{R}^n.$$

It holds that $\|T\| = \|E\| \cdot \|e\|$. Clearly, T is not symmetric in general. Introducing the function $g(h) = \frac{1}{q} T[h]^q \in \mathbb{R}$, $h \in \mathbb{R}^n$, we find that its gradient vector is equal to

$$\nabla g(h) \quad = \quad \tfrac{1}{q}\Big(E[h]^{q-1} e + (q-1) \langle e, h \rangle E[h]^{q-2} \Big) \quad \in \quad \mathbb{R}^n. \tag{1}$$

It is possible to make tensor T symmetric, using the following standard symmetrization operation:

$$P_{sym}(T)[h_1, \ldots, h_q] \quad := \quad \tfrac{1}{q!} \sum_{\sigma \in S_q} T[h_{\sigma(1)}, \ldots, h_{\sigma(q)}],$$

for which we have: $P_{sym}(T)[h]^q \equiv T[h]^q$, $\nabla_h P_{sym}(T)[h]^q \equiv q P_{sym}(T)[h]^{q-1}$, and $\|P_{sym}(T)\| \leq \|T\|$.

Throughout our analysis, we will use Young's inequality:

$$xy \quad \leq \quad \frac{x^\alpha}{\alpha} + \frac{y^\beta}{\beta} \qquad \text{for } x, y \geq 0 \text{ and } \frac{1}{\alpha} + \frac{1}{\beta} = 1, \; \alpha, \beta > 1. \tag{2}$$

2 Problem Formulation and Auxiliary Results

In this work we consider the smooth unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x),$$

under the following assumptions:

A1. $f : \mathbb{R}^n \to \mathbb{R}$ is a p-times differentiable function with L-Lipschitz continuous pth-order derivative ($p \geq 1$). Thus,

$$\|\nabla^p f(x) - \nabla^p f(y)\| \quad \leq \quad L\|x - y\|, \qquad \forall x, y \in \mathbb{R}^n. \tag{3}$$

A2. Function f is bounded from below: $f(x) \geq f_{low}$, for all $x \in \mathbb{R}^n$.

It follows from A1 that

$$f(y) \quad \leq \quad \Phi_{x,p}(y) + \frac{L}{(p+1)!}\|y - x\|^{p+1}, \qquad \forall x, y \in \mathbb{R}^n, \tag{4}$$

where

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

defines the pth-order Taylor polynomial of $f(\,\cdot\,)$ around x. Moreover, we also have the global bound for the gradient approximation:

$$\|\nabla f(y) - \nabla \Phi_{x,p}(y)\| \quad \leq \quad \frac{L}{p!}\|y - x\|^p, \qquad \forall x, y \in \mathbb{R}^n. \tag{5}$$

Given a point $\bar{x} \in \mathbb{R}^n$ and a regularization parameter $\sigma > 0$, we define the following augmented Taylor approximation

$$\Omega_{\bar{x},\sigma,p}(y) \quad := \quad f(\bar{x}) + \sum_{i=1}^{p-1} \frac{1}{i!}\nabla^i f(\bar{x})[y - \bar{x}]^i + \frac{1}{p!}\nabla^p f(\bar{x})[y - \bar{x}]^p + \frac{\sigma}{(p+1)!}\|y - \bar{x}\|^{p+1}. \tag{6}$$

According to (4), this is a global *upper approximation* of our objective, i.e., $f(y) \leq \Omega_{\bar{x},\sigma,p}(y)$, when the regularization parameter is large enough: $\sigma \geq L$. However, in general, this approximation might not be convex. It was shown in [20] that for convex function $f(\cdot)$, the augmented Taylor model (6) is convex for $\sigma \geq pL$, and, therefore, its minimizer can be efficiently computed.

Let us consider a p-linear tensor T that approximates pth-order derivative of the objective, $T \approx \nabla^p f(\bar{x})$, and use it in the following model:

$$M_{\bar{x},\sigma,p}(y) \quad := \quad f(\bar{x}) + \sum_{i=1}^{p-1} \frac{1}{i!}\nabla^i f(\bar{x})[y - \bar{x}]^i + \frac{1}{p!}T[y - \bar{x}]^p + \frac{\sigma}{(p+1)!}\|y - \bar{x}\|^{p+1}. \tag{7}$$

Therefore, we keep all derivatives of our function up to order $(p-1)$ in our model, while approximating the pth derivative. As we will see further, such an approximation can be done efficiently using lower derivatives of the function. Note that model (7) depends only on the symmetric part $P_{sym}(T)$ of our tensor, while we do not assume that T is symmetric. The gradient of our model is given by

$$\nabla M_{\bar{x},\sigma,p}(y) \;=\; \sum_{i=1}^{p-1} \tfrac{1}{(i-1)!}\nabla^i f(\bar{x})[y-\bar{x}]^{i-1} + \tfrac{1}{(p-1)!}P_{sym}(T)[y-\bar{x}]^{p-1} + \tfrac{\sigma}{p!}\|y-\bar{x}\|^{p-1}(y-\bar{x}),$$

and, correspondingly, the difference between $\nabla M_{\bar{x},\sigma,p}(\cdot)$ and $\nabla \Omega_{\bar{x},\sigma,p}(\cdot)$ can be bounded as

$$\begin{array}{rcl} \|\nabla M_{\bar{x},\sigma,p}(y) - \nabla \Omega_{\bar{x},\sigma,p}(y)\| & = & \tfrac{1}{(p-1)!}\|(P_{sym}(T) - \nabla^p f(\bar{x}))[y-\bar{x}]^{p-1}\| \\ \leq \;\; \tfrac{1}{(p-1)!}\|P_{sym}(T) - \nabla^p f(\bar{x})\| \cdot \|y-\bar{x}\|^{p-1} & \leq & \tfrac{1}{(p-1)!}\|T - \nabla^p f(\bar{x})\| \cdot \|y-\bar{x}\|^{p-1}. \end{array} \tag{8}$$

The following lemma is one of the main tools of our analysis that relates the length of the step $\|x^+ - \bar{x}\|$ of a method, with the gradient norm $\|\nabla f(x^+)\|$ at new point, where x^+ is an inexact minimizer to our approximate model (7).

Lemma 2.1. *Suppose that Assumption A1 holds and let x^+ be an inexact minimizer of $M_{\bar{x},\sigma,p}(\,\cdot\,)$, defined in (7), satisfying the following condition*

$$\|\nabla M_{\bar{x},\sigma,p}(x^+)\| \;\leq\; \tfrac{\sigma}{2p!}\|x^+ - \bar{x}\|^p. \tag{9}$$

If $\sigma \geq 2L$ and, for some $z \in \mathbb{R}^n$ and $\delta > 0$, we have

$$\|T - \nabla^p f(z)\| \;\leq\; \delta, \tag{10}$$

then

$$\|\nabla f(x^+)\|^{\frac{p+1}{p}} \;\leq\; 3^{1/p}\sigma^{1/p}\Big[4\sigma\|x^+ - \bar{x}\|^{p+1} + \tfrac{\delta^{p+1}}{p[(p-1)!]^{p+1}\sigma^p} + \tfrac{L^{p+1}\|z-\bar{x}\|^{p+1}}{p[(p-1)!]^{p+1}\sigma^p}\Big]. \tag{11}$$

Proof. Let us denote $r = \|x^+ - \bar{x}\|$. From the previous reasoning, triangle inequality, and the equation $\nabla \Omega_{\bar{x},\sigma,p}(x^+) = \nabla \Phi_{\bar{x},\sigma,p}(x^+) + \frac{\sigma}{p!}r^{p-1}(x^+ - x)$, it follows that

$$\begin{array}{rl} & \|\nabla f(x^+)\| \\ \leq & \|\nabla f(x^+) - \nabla \Omega_{\bar{x},\sigma,p}(x^+)\| + \|\nabla \Omega_{\bar{x},\sigma,p}(x^+) - \nabla M_{\bar{x},\sigma,p}(x^+)\| + \|\nabla M_{\bar{x},\sigma,p}(x^+)\| \\ \overset{(8),(9)}{\leq} & \|\nabla f(x^+) - \nabla \Phi_{x,p}(x^+)\| + \frac{\sigma}{p!}r^p + \frac{1}{(p-1)!}\big\|T - \nabla^p f(\bar{x}))\big\|r^{p-1} + \frac{\sigma}{2p!}r^p \\ \overset{(5)}{\leq} & \big(\frac{L}{p!} + \frac{\sigma}{p!} + \frac{\sigma}{2p!}\big)r^p + \frac{1}{(p-1)!}\|T - \nabla^p f(\bar{x})\|r^{p-1} \\ \leq & \big(\frac{L}{p!} + \frac{\sigma}{p!} + \frac{\sigma}{2p!}\big)r^p + \frac{1}{(p-1)!}\big(\|T - \nabla^p f(z)\| + \|\nabla^p f(z) - \nabla^p f(\bar{x})\|\big)r^{p-1} \\ \overset{(10),(3)}{\leq} & \frac{2\sigma}{p!}r^p + \Big(\frac{\delta}{(p-1)!} + \frac{L}{(p-1)!}\|z-\bar{x}\|\Big)r^{p-1}, \end{array} \tag{12}$$

where in the last inequality we also used that $\sigma \geq 2L$. In view of (12) and using the inequality $(a+b+c)^q \leq 3^{q-1}(a^q+b^q+c^q)$ for $a,b,c\geq 0$ and $q := \frac{p+1}{p} \geq 1$, we get

$$\begin{aligned}
\|\nabla f(x^+)\|^{\frac{p+1}{p}} &\leq \left[\tfrac{2\sigma}{p!}r^p + \tfrac{\delta}{(p-1)!}r^{p-1} + \tfrac{L}{(p-1)!}\|z-\bar{x}\|r^{p-1}\right]^{\frac{p+1}{p}} \\
&\leq 3^{1/p}\left[\left(\tfrac{2}{p!}\right)^{\frac{p+1}{p}}\sigma^{\frac{p+1}{p}}r^{p+1} + \frac{\delta^{\frac{p+1}{p}}}{[(p-1)!]^{\frac{p+1}{p}}}r^{\frac{(p-1)(p+1)}{p}} + \frac{L^{\frac{p+1}{p}}\|z-\bar{x}\|^{\frac{p+1}{p}}}{[(p-1)!]^{\frac{p+1}{p}}}r^{\frac{(p-1)(p+1)}{p}}\right] \qquad (13)\\
&= 3^{1/p}\sigma^{1/p}\left[\left(\tfrac{2}{p!}\right)^{\frac{p+1}{p}}\sigma r^{p+1} + \frac{\delta^{\frac{p+1}{p}}}{[(p-1)!]^{\frac{p+1}{p}}\sigma^{1/p}}r^{\frac{(p-1)(p+1)}{p}} + \frac{L^{\frac{p+1}{p}}\|z-\bar{x}\|^{\frac{p+1}{p}}}{[(p-1)!]^{\frac{p+1}{p}}\sigma^{1/p}}r^{\frac{(p-1)(p+1)}{p}}\right].
\end{aligned}$$

When $p=1$, this inequality immediately gives (11). Now, we assume that $p\geq 2$.

Using Young's inequality (2) with $\alpha = p$ and $\beta = \frac{p}{p-1} > 1$, the second term inside the brackets in (13) can be bounded as follows

$$\begin{aligned}
\frac{\delta^{\frac{p+1}{p}}}{[(p-1)!]^{\frac{p+1}{p}}\sigma^{1/p}}r^{\frac{(p-1)(p+1)}{p}} &= \left(\frac{\delta^{\frac{p+1}{p}}}{[(p-1)!]^{\frac{p+1}{p}}\sigma}\right)\left(\sigma^{\frac{p-1}{p}}r^{\frac{(p-1)(p+1)}{p}}\right) \\
&\leq \frac{1}{p}\left(\frac{\delta^{\frac{p+1}{p}}}{[(p-1)!]^{\frac{p+1}{p}}\sigma}\right)^p + \frac{p-1}{p}\left(\sigma^{\frac{p-1}{p}}r^{\frac{(p-1)(p+1)}{p}}\right)^{\frac{p}{p-1}} \qquad (14)\\
&= \frac{\delta^{p+1}}{p[(p-1)!]^{p+1}\sigma^p} + \frac{(p-1)}{p}\sigma r^{p+1} \\
&< \frac{\delta^{p+1}}{p[(p-1)!]^{p+1}\sigma^p} + \sigma r^{p+1}.
\end{aligned}$$

Using again Young's inequality, we can also bound the third term inside the brackets in (13),

$$\begin{aligned}
\frac{L^{\frac{p+1}{p}}\|z-\bar{x}\|^{\frac{p+1}{p}}}{[(p-1)!]^{\frac{p+1}{p}}\sigma^{1/p}}r^{\frac{(p-1)(p+1)}{p}} &= \left(\frac{L^{\frac{p+1}{p}}\|z-\bar{x}\|^{\frac{p+1}{p}}}{[(p-1)!]^{\frac{p+1}{p}}\sigma}\right)\left(\sigma^{\frac{p-1}{p}}r^{\frac{(p-1)(p+1)}{p}}\right) \\
&\leq \frac{1}{p}\left(\frac{L^{\frac{p+1}{p}}\|z-\bar{x}\|^{\frac{p+1}{p}}}{[(p-1)!]^{\frac{p+1}{p}}\sigma}\right)^p + \frac{p-1}{p}\left(\sigma^{\frac{p-1}{p}}r^{\frac{(p-1)(p+1)}{p}}\right)^{\frac{p}{p-1}} \qquad (15)\\
&= \frac{L^{p+1}\|z-\bar{x}\|^{p+1}}{p[(p-1)!]^{p+1}\sigma^p} + \frac{(p-1)\sigma r^{p+1}}{p} \\
&< \frac{L^{p+1}\|z-\bar{x}\|^{p+1}}{p[(p-1)!]^{p+1}\sigma^p} + \sigma r^{p+1}.
\end{aligned}$$

Finally, combining (13), (14) and (15), and using the inequality $\left(\frac{2}{p!}\right)^{\frac{p+1}{p}} \leq 1$, $p\geq 2$, we conclude that

$$\begin{aligned}
\|\nabla f(x^+)\|^{\frac{p+1}{p}} &\leq 3^{1/p}\sigma^{1/p}\left[3\sigma r^{p+1} + \frac{\delta^{p+1}}{p[(p-1)!]^{p+1}\sigma^p} + \frac{L^{p+1}\|z-\bar{x}\|^{p+1}}{p[(p-1)!]^{p+1}\sigma^p}\right] \\
&\leq 3^{1/p}\sigma^{1/p}\left[4\sigma r^{p+1} + \frac{\delta^{p+1}}{p[(p-1)!]^{p+1}\sigma^p} + \frac{L^{p+1}\|z-\bar{x}\|^{p+1}}{p[(p-1)!]^{p+1}\sigma^p}\right].
\end{aligned}$$

That is, (11) is true for all $p\geq 1$. $\square$

The next lemma shows that minimizing an approximate upper model of the objective leads to a progress in terms of the function value.

Lemma 2.2. *Suppose that Assumption A1 holds and let x^+ be an inexact minimizer of $M_{\bar{x},\sigma,p}(\,\cdot\,)$, defined in (7), satisfying the following condition*

$$M_{\bar{x},\sigma,p}(x^+) \quad \leq \quad f(\bar{x}). \tag{16}$$

If $\sigma \geq 2L$ and, for some $z \in \mathbb{R}^n$ and $\delta > 0$, (10) holds, then

$$f(\bar{x}) - f(x^+) \quad \geq \quad \tfrac{\sigma}{4\cdot(p+1)!}\|x^+ - \bar{x}\|^{p+1} - \tfrac{(8(p+1))^p\cdot\left(\delta^{p+1}+L^{p+1}\|z-\bar{x}\|^{p+1}\right)}{\sigma^p\cdot(p+1)!}. \tag{17}$$

Proof. Let $r = \|x^+ - \bar{x}\|$. We have

$$\begin{array}{rcl} f(x^+) & \overset{(4)}{\leq} & \Omega_{\bar{x},L,p}(x^+) \;=\; \Omega_{\bar{x},\sigma,p}(x^+) + \frac{L-\sigma}{(p+1)!}r^{p+1} \\ & = & M_{\bar{x},\sigma,p}(x^+) + \frac{1}{p!}\left(\nabla^p f(\bar{x}) - T\right)[x^+ - \bar{x}]^p + \frac{L-\sigma}{(p+1)!}r^{p+1} \\ & \leq & f(\bar{x}) + \frac{1}{p!}\|\nabla^p f(\bar{x}) - T\| r^p + \frac{L-\sigma}{(p+1)!}r^{p+1} \\ & \leq & f(\bar{x}) + \frac{1}{p!}\left(\|\nabla^p f(\bar{x}) - \nabla^p f(z)\| + \|\nabla^p f(z) - T\|\right) r^p + \frac{L-\sigma}{(p+1)!}r^{p+1} \\ & \overset{(3),(10)}{\leq} & f(\bar{x}) + \frac{1}{p!}\left(L\|z-\bar{x}\| + \delta\right) r^p - \frac{\sigma}{2\cdot(p+1)!}r^{p+1}, \end{array} \tag{18}$$

where in the last inequality we used the assumption $\sigma \geq 2L$. Using Young's inequality (2) with $\alpha = p+1$ and $\beta = \frac{p+1}{p}$, and the inequality $(a+b)^q \leq 2^{q-1}(a^q + b^q)$ for $a, b \geq 0$ and $q := p+1$, we get

$$\begin{array}{rcl} \frac{1}{p!}\left(L\|z-\bar{x}\| + \delta\right) r^p & = & \left(\frac{4^{\frac{p}{p+1}}[(p+1)!]^{\frac{p}{p+1}}(\delta+L\|z-\bar{x}\|)}{\sigma^{\frac{p}{p+1}}\cdot p!}\right)\left(\frac{\sigma^{\frac{p}{p+1}} r^p}{4^{\frac{p}{p+1}}[(p+1)!]^{\frac{p}{p+1}}}\right) \\ & \leq & \frac{1}{p+1}\left(\frac{4^{\frac{p}{p+1}}[(p+1)!]^{\frac{p}{p+1}}(\delta+L\|z-\bar{x}\|)}{\sigma^{\frac{p}{p+1}}\cdot p!}\right)^{p+1} + \frac{p}{p+1}\left(\frac{\sigma^{\frac{p}{p+1}} r^p}{4^{\frac{p}{p+1}}[(p+1)!]^{\frac{p}{p+1}}}\right)^{\frac{p+1}{p}} \\ & = & \frac{4^p[(p+1)!]^p(\delta+L\|z-\bar{x}\|)^{p+1}}{(p+1)\cdot[p!]^{p+1}\cdot\sigma^p} + \frac{p\cdot\sigma\cdot r^{p+1}}{4(p+1)\cdot(p+1)!} \\ & \leq & \frac{(4(p+1))^p(\delta+L\|z-\bar{x}\|)^{p+1}}{\sigma^p\cdot(p+1)!} + \frac{\sigma}{4\cdot(p+1)!}r^{p+1} \\ & \leq & \frac{(8(p+1))^p\cdot\left(\delta^{p+1}+L^{p+1}\|z-\bar{x}\|^{p+1}\right)}{\sigma^p\cdot(p+1)!} + \frac{\sigma}{4\cdot(p+1)!}r^{p+1}. \end{array} \tag{19}$$

Now, combining (18) and (19), it follows that

$$f(x^+) \quad \leq \quad f(\bar{x}) + \tfrac{(8(p+1))^p\cdot\left(\delta^{p+1}+L^{p+1}\|z-\bar{x}\|^{p+1}\right)}{\sigma^p\cdot(p+1)!} - \tfrac{\sigma}{4\cdot(p+1)!}r^{p+1},$$

which implies that (17) is true. □

Now, we can combine the two previous lemmas to obtain progress in the function value in terms of the gradient norm.

Lemma 2.3. *Suppose that Assumption A1 holds and let x^+ be an inexact minimizer of $M_{\bar{x},\sigma,p}(\cdot)$ defined in (7), satisfying conditions (9) and (16). If $\sigma \geq 2L$ and, for some $z \in \mathbb{R}^n$ and $\delta > 0$, (10) holds, then*

$$f(\bar{x}) - f(x^+) \quad \geq \quad \frac{\|\nabla f(x^+)\|^{\frac{p+1}{p}}}{2^5 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!} + \frac{\sigma}{8 \cdot (p+1)!}\|x^+ - \bar{x}\|^{p+1} - \frac{c_p \cdot \left(\delta^{p+1} + L^{p+1}\|z - \bar{x}\|^{p+1}\right)}{\sigma^p \cdot (p+1)!}, \tag{20}$$

where $c_p := \frac{1}{2^5 \cdot p[(p-1)!]^{p+1}} + (8(p+1))^p \leq 2 \cdot (8(p+1))^p$.

Proof. Let $r = \|x^+ - \bar{x}\|$. In view of Lemma 2.1, we have

$$\frac{\sigma}{8 \cdot (p+1)!} r^{p+1} \quad \overset{(11)}{\geq} \quad \frac{\|\nabla f(x^+)\|^{\frac{p+1}{p}}}{2^5 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!} - \frac{\delta^{p+1} + L^{p+1}\|z - \bar{x}\|^{p+1}}{2^5 \cdot p[(p-1)!]^{p+1} \cdot \sigma^p \cdot (p+1)!}. \tag{21}$$

Then, combining (21) with inequality (17) in Lemma 2.2, we conclude that

$$\begin{aligned} f(\bar{x}) - f(x^+) &\overset{(17)}{\geq} \frac{\sigma}{8 \cdot (p+1)!} r^{p+1} + \frac{\sigma}{8 \cdot (p+1)!} r^{p+1} - \frac{(8(p+1))^p \cdot \left(\delta^{p+1} + L^{p+1}\|z - \bar{x}\|^{p+1}\right)}{\sigma^p \cdot (p+1)!} \\ &\overset{(21)}{\geq} \frac{\|\nabla f(x^+)\|^{\frac{p+1}{p}}}{2^5 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!} + \frac{\sigma}{8 \cdot (p+1)!} r^{p+1} - \frac{\delta^{p+1} + L^{p+1}\|z - \bar{x}\|^{p+1}}{\sigma^p \cdot (p+1)!} \cdot \left[\frac{1}{2^5 \cdot p[(p-1)!]^{p+1}} + (8(p+1))^p\right], \end{aligned}$$

that is (20) is true. □

Up to this moment, we have considered an arbitrary tensor $T \approx \nabla^p f(\bar{x})$ satisfying the δ-approximation guarantee (10), for some $\delta > 0$. In this work, we are interested in using for T the finite-difference approximation provided by the $(p-1)$th-order derivatives. To this end, we use the following lemma.

Lemma 2.4. *Suppose that Assumption A1 holds. Given $z \in \mathbb{R}^n$ and $h > 0$, let T be the p-linear form defined by*

$$T \quad = \quad \sum_{i=1}^{n} \left(\frac{\nabla^{p-1} f(z + h e_i) - \nabla^{p-1} f(z)}{h} \right) \otimes e_i, \tag{22}$$

where e_i is the ith vector of the canonical basis of $\mathbb{R}^n$. Then

$$\|T - \nabla^p f(z)\| \quad \leq \quad \frac{L\sqrt{n}}{2} h. \tag{23}$$

Proof. Let us fix arbitrary directions $u_1, \ldots, u_{n-1} \in \mathbb{R}^n$ s.t. $\|u_j\| \leq 1$ for all j. Then, for an arbitrary $1 \leq i \leq n$, we have

$$\begin{aligned} &|(T - \nabla^p f(z))[u_1, \ldots, u_{n-1}, e_i]| \\ &= \frac{1}{h}\left|\left(\nabla^{p-1} f(z + h e_i) - \nabla^{p-1} f(z) - h \nabla^p f(z)[e_i]\right)[u_1, \ldots, u_{n-1}]\right| \\ &= \left|\int_0^1 (\nabla^p f(z + \tau h e_i) - \nabla^p f(z))[u_1, \ldots, u_{n-1}, e_i] d\tau\right| \overset{(3)}{\leq} L \int_0^1 \tau h d\tau = \frac{L}{2} h, \end{aligned} \tag{24}$$

where we used that $\nabla^p f$ is a symmetric form, and the standard Newton-Leibniz formula. Therefore,

$$\begin{array}{rcl} \|T - \nabla^p f(z)\| & = & \max\limits_{\substack{u_1,\dots,u_{n-1},x\in\mathbb{R}^n \\ \forall j\,\|u_j\|\leq 1, \|x\|\leq 1}} |(T - \nabla^p f(z))[u_1,\dots,u_{n-1},x]| \\ & = & \max\limits_{\substack{u_1,\dots,u_{n-1},x\in\mathbb{R}^n \\ \forall j\,\|u_j\|\leq 1, \|x\|\leq 1}} |(T - \nabla^p f(z))[u_1,\dots,u_{n-1},\sum\limits_{i=1}^n x^{(i)}e_i]| \\ & \leq & \max\limits_{\substack{u_1,\dots,u_{n-1},x\in\mathbb{R}^n \\ \forall j\,\|u_j\|\leq 1, \|x\|\leq 1}} \sum\limits_{i=1}^n |x^{(i)}| \cdot |(T - \nabla^p f(z))[u_1,\dots,u_{n-1},e_i]| \\ & \overset{(24)}{\leq} & \max\limits_{x\in\mathbb{R}^n :\, \sum\limits_{i=1}^n [x^{(i)}]^2\leq 1} \sum\limits_{i=1}^n |x^{(i)}| \cdot \frac{L}{2}h \;\; = \;\; \frac{L\sqrt{n}}{2}h, \end{array}$$

where the last equation follows from the Cauchy-Schwartz inequality. □

Remark 2.5. *Note that the tensor T defined by* (22) *is not symmetric. Since our model* (7) *depends only on the symmetric part of T, one can replace T by $P_{sym}(T)$ (see Section 1.4), which possesses the same approximation guarantee:*

$$\|P_{sym}(T) - \nabla^p f(z)\| \;\; \leq \;\; \|T - \nabla^p f(z)\| \;\; \overset{(23)}{\leq} \;\; \frac{L\sqrt{n}}{2}h.$$

All our results remain valid regardless of whether T is symmetric or not. However, the use of symmetrization affects the implementation of an inner solver for our subproblem (see expression (1) for computing the gradient of pth-order term of our model).

From (23), we see that the key parameter that controls the finite-difference approximation error is the discretization step $h > 0$. If h is chosen sufficiently small, we can ensure the same progress as in the method using the exact pth derivative.

Lemma 2.6. *Let $\epsilon > 0$ be fixed. Suppose that Assumption A1 holds, and let x^+ be an inexact minimizer of the model $M_{\bar{x},\sigma,p}(\cdot)$ defined in (7), where the tensor T is constructed by finite differences as in Lemma 2.4, with stepsize*

$$h \;\; \leq \;\; \frac{4}{\sigma\sqrt{n}}\left[\frac{\sigma^p\cdot\epsilon^{\frac{p+1}{p}}}{(8(p+1))^p\cdot 2^7\cdot 3^{1/p}\sigma^{1/p}}\right]^{\frac{1}{p+1}}. \tag{25}$$

Assume further that x^+ satisfies conditions (9) and (16), and that $\|\nabla f(x^+)\| \geq \epsilon$. If $\sigma \geq 2L$, then

$$f(\bar{x}) - f(x^+) \;\; \geq \;\; \frac{\epsilon^{\frac{p+1}{p}}}{2^6\cdot 3^{1/p}\cdot\sigma^{1/p}\cdot(p+1)!} + \frac{\sigma}{8\cdot(p+1)!}\|x^+ - \bar{x}\|^{p+1} - \frac{c_p L^{p+1}\|z-\bar{x}\|^{p+1}}{\sigma^p\cdot(p+1)!}. \tag{26}$$

Proof. Since $\sigma \geq 2L$, it follows from (25) that

$$h \;\; \leq \;\; \frac{2}{L\sqrt{n}}\left[\frac{\sigma^p\cdot\epsilon^{\frac{p+1}{p}}}{(8(p+1))^p\cdot 2^7\cdot 3^{1/p}\sigma^{1/p}}\right]^{\frac{1}{p+1}}.$$

Thus, as T is constructed by finite differences with stepsize h, Lemma 2.4 implies that

$$\|T - \nabla^p f(z)\| \leq \frac{L\sqrt{n}}{2}h \leq \delta, \tag{27}$$

where

$$\delta := \left[\frac{\sigma^p \cdot \epsilon^{\frac{p+1}{p}}}{(8(p+1))^p \cdot 2^7 \cdot 3^{1/p}\sigma^{1/p}}\right]^{\frac{1}{p+1}}. \tag{28}$$

Then, by (27), Lemma 2.3 and the assumption $\|\nabla f(x^+)\| > \epsilon$, we have

$$\begin{aligned} f(\bar{x}) - f(x^+) \geq & \frac{\epsilon^{\frac{p+1}{p}}}{2^5 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!} - \frac{c_p \cdot \delta^{p+1}}{\sigma^p \cdot (p+1)!} \\ & + \frac{\sigma}{8 \cdot (p+1)!}\|x^+ - \bar{x}\|^{p+1} - \frac{c_p \cdot L^{p+1}\|z - \bar{x}\|^{p+1}}{\sigma^p \cdot (p+1)!}. \end{aligned} \tag{29}$$

Note that by (28) and the bound $c_p \leq 2 \cdot (8(p+1))^p$, we have

$$\frac{c_p \cdot \delta^{p+1}}{\sigma^p \cdot (p+1)!} = \frac{c_p}{2(8(p+1))^p} \cdot \frac{\epsilon^{\frac{p+1}{p}}}{2^6 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!} \leq \frac{\epsilon^{\frac{p+1}{p}}}{2^6 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!}. \tag{30}$$

Then, combining (29) and (30) we conclude that (26) is true. □

We are now ready to prove the main result of this section, which serves as a building block of our algorithm. We start with a fixed initialization $x_0 \in \mathbb{R}^n$ and perform $m \geq 1$ steps of the method using a fixed finite-difference approximation tensor T, computed once at $z := x_0$, and a fixed regularization parameter $\sigma > 0$. We show that for an appropriate choice of σ, we guarantee strict progress for the iterates of the method.

Theorem 2.7. *Let $\epsilon > 0$ be fixed. Suppose that Assumption A1 holds. Given $z \in \mathbb{R}^n$, $\sigma > 0$, and $m \in \mathbb{N} \setminus \{0\}$, let $\{x_t\}_{t=0}^m$ be a sequence of points defined as follows*

$$\begin{cases} x_0 & = z, \\ x_{t+1} & \in \left\{y \in \mathbb{R}^n : M_{x_t,\sigma,p}(y) \leq f(x_t) \text{ and } \|\nabla M_{x_t,\sigma,p}(y)\| \leq \frac{\sigma}{2p!}\|y - x_t\|^p\right\}, \quad t = 0, \ldots, m-1, \end{cases} \tag{31}$$

where, for every $t \in \{0, \ldots, m-1\}$, the model is given by

$$M_{x_t,\sigma,p}(y) \equiv f(x_t) + \sum_{i=1}^{p-1} \frac{1}{i!}\nabla^i f(x_t)[y - x_t]^i + \frac{1}{p!}T[y - x_t]^p + \frac{\sigma}{(p+1)!}\|y - x_t\|^{p+1}, \tag{32}$$

with T being a fixed tensor defined by finite differences, as in Lemma 2.4, with stepsize h satisfying (25). If

$$\sigma \geq 11(p+1)Lm, \tag{33}$$

and

$$\|\nabla f(x_{i+1})\| \geq \epsilon \quad \text{for } i = 0, \ldots, t, \tag{34}$$

for some $t \in \{0, \ldots, m-1\}$, then

$$f(x_0) - f(x_{t+1}) \geq \frac{\epsilon^{\frac{p+1}{p}}}{2^6 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!}(t+1). \tag{35}$$

Proof. Assume that $t \geq 1$ and consider $i \in \{0, \ldots, t\}$. In view of (31)–(34) and the construction of the tensor T, Lemma 2.6 applies with $x^+ = x_{i+1}$ and $\bar{x} = x_i$. Hence, we obtain

$$f(x_i) - f(x_{i+1}) \geq \frac{\epsilon^{\frac{p+1}{p}}}{2^6 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!} + \frac{\sigma}{8 \cdot (p+1)!} \|x_{i+1} - x_i\|^{p+1} - \frac{c_p \cdot L^{p+1} \|z - x_i\|^{p+1}}{\sigma^p \cdot (p+1)!}. \tag{36}$$

Let us denote $r_i = \|x_{i+1} - x_i\|$. Summing inequalities (36) over $i = 0, \ldots, t$ and using $z = x_0$, we obtain

$$\begin{aligned} f(x_0) - f(x_{t+1}) \geq\ & \frac{\epsilon^{\frac{p+1}{p}}}{2^6 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!}(t+1) + \frac{\sigma}{8 \cdot (p+1)!} \sum_{i=0}^{t} r_i^{p+1} \\ & - \frac{c_p \cdot L^{p+1}}{\sigma^p \cdot (p+1)!} \sum_{i=0}^{t} \|x_i - x_0\|^{p+1}. \end{aligned} \tag{37}$$

Note that

$$\begin{aligned} \sum_{i=0}^{t} \|x_i - x_0\|^{p+1} &= \sum_{i=1}^{t} \|x_i - x_0\|^{p+1} = \sum_{i=1}^{t} \left\| \sum_{j=0}^{i-1} x_{j+1} - x_j \right\|^{p+1} \\ &\leq \sum_{i=1}^{t} \left(\sum_{j=1}^{i-1} \|x_{j+1} - x_j\| \right)^{p+1} = \sum_{i=1}^{t} \left(\sum_{j=0}^{i-1} r_j \right)^{p+1}. \end{aligned} \tag{38}$$

In addition, by the Hölder inequality, we also have

$$\sum_{j=0}^{i-1} r_j \leq \left(\sum_{j=0}^{i-1} 1^{\frac{p+1}{p}} \right)^{\frac{p}{p+1}} \left(\sum_{j=0}^{i-1} r_j^{p+1} \right)^{\frac{1}{p+1}} \leq i^{\frac{p}{p+1}} \left(\sum_{j=0}^{i-1} r_j^{p+1} \right)^{\frac{1}{p+1}}$$

and so, since $i \leq t$, it follows that

$$\left(\sum_{j=0}^{i-1} r_j \right)^{p+1} \leq i^p \left(\sum_{j=0}^{i-1} r_j^{p+1} \right) \leq t^p \left(\sum_{j=0}^{t} r_j^{p+1} \right). \tag{39}$$

Thus, combining (38) and (39) we get

$$\sum_{i=0}^{t} \|x_i - x_0\|^{p+1} \leq t^{p+1} \left(\sum_{i=0}^{t} r_i^{p+1} \right). \tag{40}$$

It follows from (37) and (40) that

$$f(x_0) - f(x_{t+1}) \geq \frac{\epsilon^{\frac{p+1}{p}}}{2^6 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!}(t+1) + \left[\frac{\sigma}{8 \cdot (p+1)!} - \frac{c_p \cdot L^{p+1} \cdot t^{p+1}}{\sigma^p \cdot (p+1)!} \right] \sum_{i=0}^{t} r_i^{p+1}. \tag{41}$$

By (33) we have

$$\sigma \geq 11(p+1)Lm \geq (8c_p)^{\frac{1}{p+1}} Lt,$$

which implies that

$$\left[\frac{\sigma}{8 \cdot (p+1)!} - \frac{c_p \cdot L^{p+1} \cdot t^{p+1}}{\sigma^p \cdot (p+1)!} \right] \geq 0. \tag{42}$$

Thus, by combining (41) and (42), we conclude that (35) holds for $t \geq 1$. Next, observe that when $t = 0$, inequality (36) also holds for $i = t = 0$, and the last term on its right-hand side vanishes, since in this case $x_i = x_0 = z$. Consequently, (35) also holds for $t = 0$, which completes the proof. □

3 A Lazy Method of Order $(p-1)$

In view of Theorem 2.7, let us define the algorithm

$$(z^+,\alpha) = \texttt{LazyTensorSteps}\,(z,T,\sigma,m,\epsilon)$$

that attempts to perform m inexact pth-order steps, starting from z, using the same tensor T and regularization parameter σ, and recomputing the derivatives of f up to order $p-1$ at each step. The algorithm stops earlier whenever an ϵ-approximate stationary point of f is found, or when the functional decrease with respect to $f(z)$ does not satisfy condition (35). If the m steps are performed, or the algorithm stops due to the violation of (35), the output z^+ is the point with smallest function value among those generated by the algorithm. Otherwise, if an ϵ-approximate stationary point is found, that point is returned as z^+. The output α specifies the reason why the algorithm stopped, and thus characterizes the type of the output.

Algorithm 1: $(z^+,\alpha) = \texttt{LazyTensorSteps}(x,T,\sigma,m,\epsilon)$

Step 0. Set $x_0 := x$, $\tilde{x}_0 := x$ and $t := 0$.

Step 1. If $t = m$ then stop and **return** ($\tilde{x}_t$, `success`).

Step 2. Compute x_{t+1} as an approximate solution to the subproblem

$$\min_{y\in\mathbb{R}^n} M_{x_t,\sigma,p}(y) \quad\equiv\quad f(x_t) + \sum_{i=1}^{p-1}\tfrac{1}{i!}\nabla^i f(x_t)[y-x_t]^i + \tfrac{1}{p!}T[y-x_t]^p + \tfrac{\sigma}{(p+1)!}\|y-x_t\|^{p+1},$$

such that

$$M_{x_t,\sigma,p}(x_{t+1}) \quad\leq\quad f(x_t) \quad\text{and}\quad \|\nabla M_{x_t,\sigma,p}(x_{t+1})\| \quad\leq\quad \tfrac{\sigma}{2p!}\|x_{t+1}-x_t\|^p.$$

Define $\tilde{x}_{t+1} = \arg\min\{f(y)\,:\,y\in\{\tilde{x}_t,x_{t+1}\}\}$.

Step 3. If $\|\nabla f(x_{t+1})\|\leq\epsilon$ then stop and **return** (x_{t+1}, `solution`).

Step 4. If

$$f(x_0) - f(\tilde{x}_{t+1}) \quad\geq\quad \frac{\epsilon^{\frac{p+1}{p}}}{2^6\cdot 3^{1/p}\cdot\sigma^{1/p}\cdot(p+1)!}(t+1). \tag{43}$$

holds then set $t := t+1$ and go to Step 1. Otherwise, stop and **return** ($\tilde{x}_{t+1}$, `halt`).

Algorithm 1 outputs both the point z^+ and the status indicator

$$\alpha \in \{\texttt{success}, \texttt{solution}, \texttt{halt}\},$$

which specifies the termination condition. The value `success` indicates that all prescribed m steps were completed; `solution` signals that an ϵ-approximate stationary point of f was identified; and `halt` means that progress in reducing the objective function was insufficient. As a direct consequence of Theorem 2.7, we have the following result.

Lemma 3.1. *Suppose that Assumption A1 holds. Given $z\in\mathbb{R}^n$, $\epsilon>0$, $\sigma>0$ and $m\in\mathbb{N}\setminus\{0\}$, let (z^+,α) be the corresponding output of Algorithm 1 with*

$$T \quad=\quad \sum_{i=1}^{n}\left(\frac{\nabla^{p-1}f(z+he_i)-\nabla^{p-1}f(z)}{h}\right)\otimes e_i, \tag{44}$$

for some $h > 0$. *If*

$$\sigma \quad \geq \quad 11(p+1)Lm \quad \textit{and} \quad h \quad \leq \quad \frac{4}{\sigma\sqrt{n}}\left[\frac{\sigma^p \cdot \epsilon^{\frac{p+1}{p}}}{(8(p+1))^p \cdot 2^7 \cdot 3^{1/p}\sigma^{1/p}}\right]^{\frac{1}{p+1}}. \tag{45}$$

then either $\alpha = \texttt{solution}$ *(and so* $\|\nabla f(z^+)\| \leq \epsilon$*) or* $\alpha = \texttt{success}$ *and so*

$$f(z) - f(z^+) \quad \geq \quad \frac{\epsilon^{\frac{p+1}{p}}}{2^6 \cdot 3^{1/p} \cdot \sigma^{1/p} \cdot (p+1)!}m. \tag{46}$$

Proof. Suppose that $\alpha \neq \texttt{solution}$. Then $z^+ = \tilde{x}_{t_*+1}$ for some $t_* < m$ and

$$\|\nabla f(x_{i+1})\| \quad \geq \quad \epsilon \quad \text{for } i = 0, \ldots, t_*. \tag{47}$$

In view of (44), (45), (47), and Theorem 2.7, it follows that

$$f(x_0) - f(x_{t_*+1}) \quad \geq \quad \frac{\epsilon^{\frac{p+1}{p}}}{2^6 \cdot 3^{1/p} \cdot \sigma^{1/p}(p+1)!}(t_* + 1). \tag{48}$$

Since $f(\tilde{x}_{t_*+1}) \leq f(x_{t_*+1})$, this means that condition (43) was satisfied for $t = t_*$. Therefore, the only way Algorithm 1 could have returned $z^+ = \tilde{x}_{t_*+1}$ is if $t_* + 1 = m$. Consequently, $\alpha = \texttt{success}$, and by (48) we conclude that (46) holds. □

Building upon Algorithm 1, we can design an adaptive lazy scheme that emulates a pth-order method while using only a lower-order oracle of order $p-1$. At the kth iteration of the lazy method, we have the current iterate z_k and an estimate L_k of the Lipschitz constant L. Defining

$$\sigma_k \quad = \quad 11(p+1)L_k m \quad \text{and} \quad h_k \quad = \quad \frac{4}{\sigma_k\sqrt{n}}\left[\frac{\sigma_k^p \cdot \epsilon^{\frac{p+1}{p}}}{(8(p+1))^p \cdot 2^7 \cdot 3^{1/p} \cdot \sigma_k^{1/p}}\right]^{\frac{1}{p+1}},$$

we then compute a pth-order tensor T_k as in Lemma 3.1 (with $z = z_k$ and $h = h_k$). The next iterate z_{k+1} is computed by calling Algorithm 1:

$$(z_{k+1}, \alpha_k) \quad = \quad \texttt{LazyTensorSteps}(z_k, T_k, \sigma_k, m, \epsilon).$$

If $\alpha_k = \texttt{solution}$, this means that $\|\nabla f(z_{k+1})\| \leq \epsilon$, and the algorithm terminates. If $\alpha_k = \texttt{success}$, then it follows from Lemma 3.1 that

$$f(z_k) - f(z_{k+1}) \quad \geq \quad \frac{\epsilon^{\frac{p+1}{p}}}{2^6 \cdot 3^{1/p} \cdot \sigma_k^{1/p} \cdot (p+1)!}m.$$

In this case, we say that the iteration was *successful* and, to allow a larger step in the next iteration, we set $L_{k+1} = L_k/2$. Conversely, if $\alpha_k = \texttt{halt}$, we say that the kth iteration was *unsuccessful* and set $L_{k+1} = 2L_k$. In what follows we provide a detailed description of this algorithm.

Note that the number of consecutive unsuccessful iterations is finite: by doubling L_k, we will eventually have $\sigma_k = 11(p+1)L_k m \geq 11(p+1)Lm$, and by Lemma 3.1 this ensures that $\alpha_k \in \{\texttt{solution}, \texttt{success}\}$. This fact allows us to establish the following upper bound for the sequence of estimates L_k.

Algorithm 2: Lower-Order Implementation of a pth-Order Method

Step 0. Given $z_0 \in \mathbb{R}^n$, $L_0 > 0$, $\epsilon > 0$, and $m \in \mathbb{N} \setminus \{0\}$, set $k := 0$.

Step 1. If $\|\nabla f(z_k)\| \leq \epsilon$, stop.

Step 2. Using

$$\sigma_k \quad = \quad 11(p+1)L_k m \quad \text{and} \quad h_k \quad = \quad \frac{4}{\sigma_k\sqrt{n}}\left[\frac{\sigma_k^p \cdot \epsilon^{\frac{p+1}{p}}}{(8(p+1))^p \cdot 2^7 \cdot 3^{1/p}\sigma_k^{1/p}}\right]^{\frac{1}{p+1}}, \tag{49}$$

compute the finite difference tensor:

$$T_k \quad = \quad \sum_{i=1}^{n}\left(\frac{\nabla^{p-1}f(z_k+h_k e_i)-\nabla^{p-1}f(z_k)}{h_k}\right)\otimes e_i. \tag{50}$$

Step 3. Attempt to perform m lazy tensor steps using the same tensor T_k:

$$(z_{k+1},\alpha_k) = \texttt{LazyTensorSteps}(z_k, T_k, \sigma_k, m, \epsilon). \tag{51}$$

Step 4. Update the estimate of the Lipschitz constant:

$$L_{k+1} = \begin{cases} 2L_k & \text{if } \alpha_k = \texttt{halt}, \\ L_k & \text{if } \alpha_k = \texttt{solution}, \\ L_k/2 & \text{if } \alpha_k = \texttt{success}. \end{cases} \tag{52}$$

Step 5. Set $k := k+1$ and go back to Step 1.

Lemma 3.2. *Suppose that Assumption A1 holds and let $\{z_k\}_{k=0}^{K}$ be generated by Algorithm 2. Then*

$$L_k \quad \leq \quad L_{\max} \quad \equiv \quad \max\{L_0, 2L\} \tag{53}$$

for all $k \in \{0,\ldots,K\}$.

Proof. Let us show it by induction over k. By the definition of $L_{\max}$, it follows that (53) is true for $k=0$. Suppose $K \geq 1$ and that (53) holds for some $k \in \{0,\ldots,K-1\}$.

Case I: $\alpha_k \in \{\texttt{solution}, \texttt{success}\}$.

In this case, by (52) and the induction assumption we have

$$L_{k+1} \quad \leq \quad L_k \quad \leq \quad L_{\max},$$

that is, (53) holds for $k+1$.

Case II: $\alpha_k = \texttt{halt}$.

In this case, we must have

$$L_k \quad < \quad L, \tag{54}$$

since otherwise we would obtain $\sigma_k = 11(p+1)L_k m \geq 11(p+1)Lm$. By Lemma 3.1, this would imply $\alpha_k \in \{\texttt{solution}, \texttt{success}\}$, contradicting the hypothesis of the present case. Hence, in view of (52), we deduce

$$L_{k+1} \quad = \quad 2L_k \quad \overset{(54)}{<} \quad 2L \quad \leq \quad L_{\max},$$

which shows that (53) also holds for $k+1$. □

Given $\{z_k\}_{k=0}^{K}$ generated by Algorithm 2 with $K \geq 1$, let

$$\mathcal{S}_K = \{k \in \{0,\ldots,K-1\} \,:\, \alpha_k = \texttt{success}\}\,, \quad (55)$$
$$\mathcal{U}_K = \{k \in \{0,\ldots,K-1\} \,:\, \alpha_k = \texttt{halt}\}\,. \quad (56)$$

Lemma 3.3. *Suppose that Assumptions A1-A2 hold and let $\{z_k\}_{k=0}^{K}$ be generated by Algorithm 2 with $K \geq 2$. Then*

$$|\mathcal{S}_{K-1}| \leq \frac{2^6\cdot(3\cdot 11(p+1)\cdot L_{\max})^{1/p}(p+1)!(f(z_0)-f_{low})}{m^{(p-1)/p}}\cdot\epsilon^{-\frac{p+1}{p}}. \quad (57)$$

Proof. Since z_K has been generated, it follows that

$$\|\nabla f(z_{k+1})\| \geq \epsilon, \quad k \in \{0,\ldots,K-2\}\,.$$

By Lemma 3.1, we then have

$$\alpha_k \in \{\texttt{success},\texttt{halt}\}\,, \quad k \in \{0,\ldots,K-2\}\,.$$

Consequently, from (51) together with Steps 1 and 4 of Algorithm 1, we have

$$f(z_{k+1}) \leq f(z_k), \quad k \in \{0,\ldots,K-2\}\,. \quad (58)$$

Finally, combining A1, (58), (55) and Lemma 3.2 we deduce

$$\begin{aligned}
f(z_0)-f_{low} &\geq f(z_0)-f(z_{T-1}) = \sum_{k=0}^{K-2} f(z_k)-f(z_{k+1}) \\
&\geq \sum_{k\in\mathcal{S}_{K-1}} f(z_k)-f(z_{k+1}) \\
&\geq \sum_{k\in\mathcal{S}_{K-1}} \frac{\epsilon^{\frac{p+1}{p}}}{2^6\cdot 3^{1/p}\cdot\sigma_k^{1/p}\cdot(p+1)!}m \\
&= \sum_{k\in\mathcal{S}_{K-1}} \frac{\epsilon^{\frac{p+1}{p}}}{2^6\cdot 3^{1/p}\cdot(11(p+1)L_k m)^{1/p}\cdot(p+1)!}m \\
&\geq \sum_{k\in\mathcal{S}_{K-1}} \frac{\epsilon^{\frac{p+1}{p}}}{2^6\cdot(3\cdot 11(p+1)L_{\max}m)^{1/p}\cdot(p+1)!}m \\
&= \frac{\epsilon^{\frac{p+1}{p}}}{2^6\cdot(3\cdot 11(p+1)L_{\max})^{1/p}\cdot(p+1)!}m^{(p-1)/p}|\mathcal{S}_{K-1}|,
\end{aligned}$$

and therefore,

$$|\mathcal{S}_{K-1}| \leq \frac{2^6\cdot(3\cdot 11(p+1)\cdot L_{\max})^{1/p}(p+1)!(f(z_0)-f_{low})}{m^{(p-1)/p}}\epsilon^{-\frac{p+1}{p}}.$$

□

Lemma 3.4. *Suppose that Assumptions A1-A2 hold and let* $\{z_k\}_{k=0}^{K}$ *be generated by Algorithm 2 with* $K \geq 2$. *Then*

$$|\mathcal{U}_{K-1}| \quad \leq \quad \frac{2^6 \cdot (3 \cdot 11(p+1) \cdot L_{\max})^{1/p}(p+1)!(f(z_0)-f_{low})}{m^{(p-1)/p}} \epsilon^{-\frac{p+1}{p}} + \log_2\left(\frac{L_{\max}}{L_0}\right). \tag{59}$$

Proof. As in the proof of Lemma 3.3, the fact that z_K has been generated implies

$$\alpha_k \in \{\texttt{success}, \texttt{halt}\}, \quad k \in \{0, \ldots, K-2\}.$$

Therefore, by (52) and Lemma 3.2, we obtain

$$L_0 \left(\tfrac{1}{2}\right)^{|\mathcal{S}_{K-1}|} \cdot 2^{|\mathcal{U}_{K-1}|} \quad = \quad L_{K-1} \quad \leq \quad L_{\max}.$$

Dividing both sides by L_0 and then taking the logarithm, it follows that

$$-|\mathcal{S}_{K-1}| + |\mathcal{U}_{K-1}| \quad \leq \quad \log_2\left(\frac{L_{\max}}{L_0}\right),$$

and hence

$$|\mathcal{U}_{K-1}| \quad \leq \quad |\mathcal{S}_{K-1}| + \log_2\left(\frac{L_{\max}}{L_0}\right).$$

Thus, by Lemma 3.3, relation (59) follows. □

Now we can establish an iteration-complexity bound of $\mathcal{O}\left(L_{\max}^{1/p}(f(z_0) - f_{low})\epsilon^{-\frac{p+1}{p}}\right)$ for Algorithm 2.

Theorem 3.5. *Let* $\epsilon > 0$ *be fixed. Suppose that Assumptions A1–A2 hold, and let* $K(\varepsilon) \in \mathbb{N} \cup \{+\infty\}$ *denote the termination time of Algorithm 2, i.e., the smallest index such that* $\|\nabla f(x_{K(\varepsilon)})\| \leq \varepsilon$. *Then*

$$K(\epsilon) \quad \leq \quad 1 + \frac{2^7 \cdot (3 \cdot 11(p+1) \cdot L_{\max})^{1/p}(p+1)!(f(z_0)-f_{low})}{m^{(p-1)/p}} \epsilon^{-\frac{p+1}{p}} + \log_2\left(\frac{L_{\max}}{L_0}\right). \tag{60}$$

Proof. If $K(\varepsilon) \leq 1$, then (60) follows immediately. Hence, suppose that $K(\varepsilon) \geq 2$. In this case, Lemmas 3.3 and 3.4 imply that

$$\begin{aligned} K(\varepsilon) - 1 \quad &= \quad |\mathcal{S}_{K(\varepsilon)-1}| + |\mathcal{U}_{K(\varepsilon)-1}| \\ &\leq \quad \frac{2^7\,(3 \cdot 11(p+1)\,L_{\max})^{1/p}(p+1)!\,(f(z_0) - f_{\mathrm{low}})}{m^{(p-1)/p}}\,\varepsilon^{-\frac{p+1}{p}} + \log_2\left(\frac{L_{\max}}{L_0}\right), \end{aligned}$$

and therefore (60) holds. □

For $x \in \mathbb{R}^n$, a single call to the oracle of order $(p-1)$ corresponds to the computation of any nonempty subset of

$$\{f(x), \nabla f(x), \ldots, \nabla^{p-1} f(x)\}.$$

At iteration t, Algorithm 1 computes

$$\{f(x_t), \nabla f(x_t), \ldots, \nabla^{p-1} f(x_t)\} \quad \text{and} \quad \{f(x_{t+1}), \nabla f(x_{t+1})\},$$

which amounts to two oracle calls. Since Algorithm 1 performs at most m iterations, each run requires at most $2m$ oracle calls. In turn, at iteration k, Algorithm 2 executes Algorithm 1, computes $\nabla f(z_k)$

at Step 1, and computes $\{\nabla^{p-1} f(z_k + h_k e_i)\}_{i=1}^n$ at Step 2. Therefore, each iteration of Algorithm 2 requires at most $2m + (1+n) = \mathcal{O}\left(m+n\right)$ oracle calls. Thus, by Theorem 3.5, Algorithm 2 requires at most

$$\mathcal{O}\left(\left[\tfrac{m+n}{m^{(p-1)/p}}\right] L_{\max}^{1/p}(f(z_0) - f_{low})\epsilon^{-\frac{p+1}{p}}\right) \tag{61}$$

calls to the oracle of order $(p-1)$ to find an ε-approximate stationary point of f. Recall that $m \in \mathbb{N}\setminus\{0\}$ is a user-defined parameter specifying the maximum number of iterations of Algorithm 2 (i.e., the number of lazy tensor steps). A natural question then arises: *how should one choose m?* One approach is to select m so as to minimize the corresponding factor in the oracle complexity bound (61). Let

$$\varphi(m) \;\; = \;\; \tfrac{m+n}{m^{(p-1)/p}}.$$

For $p=1$, the optimal value is $m=1$. Consider $p \geq 2$. Note that the minimum of the unimodal function $\varphi(\,\cdot\,)$ is achieved when

$$0 \;\; = \;\; \varphi'(m) \;\; = \;\; m^{-\frac{2p-1}{p}}\left[m - \tfrac{(p-1)}{p}(m+n)\right]$$

which holds if, and only if, $m = (p-1)n$. Therefore, with such a choice of m we obtain the following oracle complexity bound for Algorithm 2, combining both cases $p=1$ and $p \geq 2$.

Corollary 3.6. *Suppose that A1-A2 holds. Then Algorithm 2 with* $\boxed{m = (p-1)n+1}$ *requires at most*

$$\mathcal{O}\left(n^{1/p} L_{\max}^{1/p}(f(z_0) - f_{low})\epsilon^{-\frac{p+1}{p}}\right) \tag{62}$$

calls to the oracle of order $(p-1)$ to find an ϵ-approximate stationary point of $f(\,\cdot\,)$.

In particular, our analysis shows that for $p=1$ (the zeroth-order approximation of the first-order method), the best strategy is $m=1$, which means not using stale derivatives and updating the information at every iteration.

For $p \geq 2$, we show that the best strategy is to choose $m \propto n$ (i.e., updating the pth-order tensor approximation once every $(p-1)n$ iterations). It is also interesting that, as the order of the method p increases, we obtain an improved dependence on the problem dimension in the oracle complexity (62).

4 Discussion

In summary, for objectives with Lipschitz continuous pth-order derivatives, it is known [3] that the optimal oracle complexity of pth-order tensor methods for finding an ϵ-approximate stationary point is $\mathcal{O}\left(\epsilon^{-\frac{p+1}{p}}\right)$, a bound attained by regularized pth-order methods (e.g., [2]). Our contribution is a method of order $(p-1)$ which achieves complexity $\mathcal{O}\left(n^{1/p}\epsilon^{-\frac{p+1}{p}}\right)$, thereby extending the result of [10] beyond the case $p=2$. To complete the picture, complexity bounds of order $\mathcal{O}(n\epsilon^{-2})$ and $\mathcal{O}(n^{3/2}\epsilon^{-3/2})$ have been established in [11] and [10], respectively, for zeroth-order methods applied to the classes of functions with Lipschitz continuous gradient and Lipschitz continuous Hessian. These results are summarized in Table 1, where the present work fills the entries highlighted in gray.

Method/Class	∇f Lip.	$\nabla^2 f$ Lip.	$\nabla^3 f$ Lip.	...	$\nabla^{p-1} f$ Lip.	$\nabla^p f$ Lip.
zeroth-order	$\mathcal{O}\left(n\epsilon^{-2}\right)$	$\mathcal{O}\left(n^{\frac{3}{2}}\epsilon^{-\frac{3}{2}}\right)$	?		?	?
1st-order	$\mathcal{O}\left(\epsilon^{-2}\right)$	$\mathcal{O}\left(n^{\frac{1}{2}}\epsilon^{-\frac{3}{2}}\right)$	?		?	?
2nd-order	–	$\mathcal{O}\left(\epsilon^{-\frac{3}{2}}\right)$	$\mathcal{O}\left(n^{\frac{1}{3}}\epsilon^{-\frac{4}{3}}\right)$		?	?
3rd-order	–	–	$\mathcal{O}\left(\epsilon^{-\frac{4}{3}}\right)$	$\ddots$	?	?
$\vdots$	–	–	–	$\ddots$	$\mathcal{O}\left(n^{\frac{1}{p-1}}\epsilon^{-\frac{p}{p-1}}\right)$	?
$(p-1)$**th-order**	–	–	–		$\mathcal{O}\left(\epsilon^{-\frac{p}{p-1}}\right)$	$\mathcal{O}\left(n^{\frac{1}{p}}\epsilon^{-\frac{p+1}{p}}\right)$
p**th-order**	–	–	–		–	$\mathcal{O}\left(\epsilon^{-\frac{p+1}{p}}\right)$

Table 1: Summary of the best-known complexity bounds for finding ϵ-approximate stationary points using qth-order methods applied to functions with Lipschitz continuous pth-order derivatives ($p \geq q$).

In view of Table 1, to complete our understanding of the worst-case complexity of qth-order method applied to functions with Lipschitz continuous pth-order derivatives ($p \geq q$), the case

$$\boxed{q < p-1, \quad \text{for} \quad p \geq 3}$$

still needs to be investigated. Even for the known bounds, outside the main diagonal ($q = p$) it is unclear whether they can be improved, as the question of their tightness ramains open. From a practical perspective, it would be interesting to design a less lazy variant of Algorithm 1 that incorporates quasi-tensor updates [25], as in [7], while still preserving the improved complexity bound with respect to the problem dimension n. Additionally, exploring universal lower-order implementations of higher-order methods for convex and nonconvex functions with Hölder-continuous derivatives would also be of interest [13, 18, 5]. We plan to address these questions in future work.

Acknowledgements

The authors are very grateful to Coralia Cartis and Sadok Jerad for the interesting and motivating discussion, which resulted in an improved version of this paper.

References

[1] Charles Audet and Warren Hare. *Derivative-Free and Blackbox Optimization.* Springer Series in Operations Research and Financial Engineering. Springer, 2017.

[2] Ernesto G. Birgin, Jo ao L. Gardenghi, José M. Martínez, Sandra A. Santos, and Philippe L. Toint. Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models. *Mathematical Programming*, 163(1-2):359–368, May 2017.

[3] Yair Carmon, John C. Duchi, Oliver Hinder, and Aaron Sidford. Lower bounds for finding stationary points i. *Mathematical Programming*, 184:71–120, 2020.

[4] Coralia Cartis, Nicholas I. M. Gould, and Philippe L. Toint. On the oracle complexity of first-order and derivative-free algorithms for smooth nonconvex minimization. *SIAM Journal on Optimization*, 22(1):66–86, 2012.

[5] Coralia Cartis, Nicholas I. M. Gould, and Philippe L. Toint. Universal regularized methods – varying the power, the smoothness, and the accuracy. *SIAM Journal on Optimization*, 29(2):595–615, 2019.

[6] Coralia Cartis, Nicholas IM Gould, and Philippe L Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Mathematical Programming*, 127(2):245–295, 2011.

[7] Coralia Cartis and Sadok Jerad. On global rates for regularization methods based on secant derivative approximations. arXiv preprint arXiv:2509.07580, 2025.

[8] Andrew R. Conn, Katya Scheinberg, and Luis N. Vicente. *Introduction to Derivative-Free Optimization*, volume 8 of *MPS-SIAM Series on Optimization*. SIAM, 2009.

[9] Nikita Doikov, El Mahdi Chayti, and Martin Jaggi. Second-order optimization with lazy Hessians. In *International Conference on Machine Learning*. PMLR, 2023.

[10] Nikita Doikov and Geovani Nunes Grapiglia. First and zeroth-order implementations of the regularized Newton method with lazy approximated Hessians. *Journal of Scientific Computing*, 103(1):32, 2025.

[11] Geovani Nunes Grapiglia. Quadratic regularization methods with finite-difference gradient approximations. *Computational Optimization and Applications*, 85(3):683–703, 2023.

[12] Geovani Nunes Grapiglia, Max LN Gonçalves, and GN Silva. A cubic regularization of Newton's method with finite-difference Hessian approximations. *Numerical Algorithms*, 90(2):607–630, 2022.

[13] Geovani Nunes Grapiglia and Yurii Nesterov. Tensor methods for finding approximate stationary points of convex functions. *Optimization Methods and Software*, 37(2):605–638, 2022.

[14] Andreas Griewank. The modification of Newton's method for unconstrained optimization by bounding cubic terms. Technical report, Technical report NA/12, 1981.

[15] Eva-Maria Kapfer, Paul Stapor, and Jan Hasenauer. Challenges in the calibration of large-scale ordinary differential equation models. *IFAC-PapersOnLine*, 52(26):58–64, 2019.

[16] Jeffrey Larson, Matt Menickelly, and Stefan M. Wild. Derivative-free optimization methods. *Acta Numerica*, 28:287–404, 2019.

[17] Qiancheng Liu, Stephen Beller, Wenjie Lei, Daniel Peter, and Jeroen Tromp. Pre-conditioned BFGS-based uncertainty quantification in elastic full-waveform inversion. *Geophysical Journal International*, 228(2):796–815, 2022.

[18] José M. Martínez. On high-order model regularization for constrained optimization. *SIAM Journal on Optimization*, 27(4):2447–2458, 2017.

[19] Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[20] Yurii Nesterov. Implementable tensor methods in unconstrained convex optimization. *Mathematical Programming*, 186(1):157–183, 2021.

[21] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[22] Yurii Nesterov and Boris T. Polyak. Cubic regularization of Newton method and its global performance. *Mathematical Programming*, 108(1):177–205, 2006.

[23] R.-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, 2006.

[24] Paul Stapor, Fabian Fröhlich, and Jan Hasenauer. Optimization and profile calculation of ode models using second-order adjoint sensitivity analysis. *Bioinformatics*, 34(13):i151–i159, 2018.

[25] Karl Welzel and Raphael A. Hauser. Approximating higher-order derivative tensors using secant updates. *SIAM Journal on Optimization*, 34(1):893–917, 2024.