

QUARTIC REGULARITY

Yurii Nesterov

LIDAM Discussion Paper CORE
2022 / 01

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: immaq-library@uclouvain.be

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

CORE DISCUSSION PAPER

2022/01

Quartic Regularity

Yu. Nesterov *

January 13, 2022

Abstract

In this paper, we propose new linearly convergent second-order methods for minimizing convex quartic polynomials. This framework is applied for designing optimization schemes, which can solve general convex problems satisfying a new condition of quartic regularity. It assumes positive definiteness and boundedness of the fourth derivative of the objective function. For such problems, an appropriate quartic regularization of Damped Newton Method has global linear rate of convergence. We discuss several important consequences of this result. In particular, it can be used for constructing new second-order methods in the framework of high-order proximal-point schemes [20, 21]. These methods have convergence rate $\tilde{O}(k^{-p})$, where k is the iteration counter, p is equal to 3, 4, or 5, and tilde indicates the presence of logarithmic factors in the complexity bounds for the auxiliary problems, which are solved at each iteration of the schemes.

Keywords: composite convex minimization, second-order methods, global complexity bounds, high-order proximal-point methods.

*Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL). E-mail: Yuri.Nesterov@uclouvain.be. This paper has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 788368). It was also funded by Multidisciplinary Institute in Artificial intelligence MIAI@Grenoble Alpes (ANR-19-P3IA-0003).

1 Introduction

Motivation. After appearance in 2006 of the first second-order method with global efficiency bounds [22], the question of finding the right framework for complexity analysis of the higher-order schemes became one of the most interesting research directions in Convex Optimization. Indeed, the performance estimates of Cubic Newton Method [22] and its accelerated version [16] clearly demonstrated that the existing classification of complexity of optimization problems in accordance to the *condition number* of the objective¹⁾ fits well only the first-order schemes. The higher-order methods are insensitive to this characteristics since, in some sense, the influence of the quadratic part of the objective is eliminated by a single iteration of the simplest second-order scheme. Thus, it became necessary to find new characteristics helping in ranking the performance of the new methods.

The first attempt of introducing a notion of the second-order non-degeneracy was done already in [16]. In the complexity analysis of [22], the assumption for bounding the growth of the objective was taken from the classical result of Kantorovich [14] on local quadratic convergence of Newton Method. This was the Lipschitz continuity of the Hessian of the objective function. However, in [16], the lower bound for the growth of the objective was ensured by uniform convexity of *degree three*. The justification of this combination was purely algebraic: under these assumptions it was possible to prove the global *linear* rate of converge of the second-order methods. As compared to the theory of first-order methods, this choice looks much less adequate since it is not supported by any regular source of such problems, related to the real-life applications.

Still, during the next decade, the development of the second-order methods was accomplished within this framework. For completeness of the picture, we mention interesting results on different variants of Cubic Regularization [7, 8], universal methods [13], accelerated second-order methods based on contractions [9, 11], and many others.

However, very soon it was shown that for convex functions we can efficiently implement also the higher-order optimization methods [17]. This result justifies an extensive study of these schemes [18, 10, 6, 12], including the lower complexity bounds [1, 2]. In particular, it was shown that the methods of the order three and higher have much faster rate of convergence than the best possible rate of the second-order schemes (on their “natural” problem classes).

The latter fact explains the level of the interest caused by an observation that practically all third-order methods can be implemented by a proper use of the second-order oracle, preserving at the same time their higher rate of convergence [19, 20, 21]. The first results of this type were based on the difference approximation of the product of the third-order derivative by two vectors [19], while the other papers employ the framework of the high-order proximal-point operators.

These results reveal the question on a proper attribution of optimization methods to particular problem classes. This link can be justified only by comparing the rates of convergence of the optimization schemes with the lower complexity bounds. Today it is clear that the initial naive hope that every problem class (formed by functions with uniformly

¹⁾ It is traditionally defined by the uniform lower and upper bounds for the eigenvalues of the Hessian of the objective function. These constants control the growth of the objective with respect to its linear approximation at the current test point.

bounded derivative of certain degree) can be unanimously related to the methods of certain order is too optimistic. This may be true for the first-order methods, traditionally assigned to functions with bounded second derivatives. However, the next step in this classification is already much more involved.

Indeed, if we agree that the natural problem class for the second-order methods is formed by functions with bounded third derivative, then we must accept the corresponding bound on the maximal possible global rate of convergence in functional residual, which is of the order $O(k^{-3.5})$, with k being the number of calls of oracle [2]. On the other hand, as it was shown in [17], the third derivative of convex functions is *not* an independent characteristic. It can be estimated from above by a combination of the second and the fourth derivatives. And if we assume instead the boundedness of these two derivatives, then the limits in the rate of convergence are relaxed up to $O(k^{-5})$.

From the view point of global efficiency bounds, this perspective is very interesting. One possible way of implementing this idea was already presented in the papers [20, 21]. The corresponding approach is based on the notion of the high-order proximal operators, approximated by auxiliary minimization schemes guided by the relative smoothness condition [5, 15]. In this paper, we develop a more direct approach based on the bounds for the *fourth derivative*. Our main observation is that the fourth derivative (contrary to the third one) can be *positive semidefinite*, and this property helps in further acceleration of optimization schemes.

Contents. In Section 2, we consider the problem of minimizing convex quartic polynomial in composite setting. We show that the corresponding quartic form is always convex and study its properties. In the next Section 3, we present the Damped Quartic Newton Method (DQNM), which regularizes the quadratic approximation of the objective by the fourth power of the norm, defined by the fourth derivative. We prove that the corresponding global rate of convergence is linear with the absolute reduction factor. In Section 4, we show that the regularization norm in DQNM can be replaced by a simpler one. The rate of convergence of the corresponding scheme is still linear, but it depends on the condition number of the new norm with respect to the optimal one. Note that the corresponding variant of DQNM is essentially a second-order method.

In Section 5, we present a method for solving a composite optimization problem, where the smooth part of the objective is *quartic regular*. This means that its fourth-order derivative is uniformly positive definite and bounded on the domain of the composite term. We show that such a function admits global upper and lower bounds for their growth with respect to local approximations. Using the corresponding inequalities, we justify a linear rate of convergence of some variant of DQNM.

In Section 6, we present some applications of our results. First of all, we show that DQNM can be applied to functions with positive-semidefinite fourth derivative. In terms of the number of calls of oracle, it has the rate of convergence $\tilde{O}(k^{-3})$. In the remaining part of the section we discuss applications of DQNM to the framework of proximal point methods. We show that it can be used for designing optimization methods with the rate of convergence $\tilde{O}(k^{-p})$ with $p \in \{3, 4, 5\}$. All these methods are of the order two.

Notation. Let us fix a finite-dimensional real vector space $\mathbb{E}$. We measure distances in $\mathbb{E}$ by a general norm $\|\cdot\|$. Its dual space is denoted by $\mathbb{E}^*$. It is the space of all linear

functions on $\mathbb{E}$, for which we define the norm in the standard way:

$$\|g\|^* = \max_{x \in \mathbb{E}} \{ \langle g, x \rangle : \|x\| \leq 1 \}, \quad g \in \mathbb{E}^*.$$

Using this norm, we can define an induced norm for a self-adjoint linear operator $A : \mathbb{E} \rightarrow \mathbb{E}^*$ as follows:

$$\|A\| = \max_{x \in \mathbb{E}} \{ |\langle Ax, x \rangle| : \|x\| \leq 1 \}.$$

For defining Euclidean norm, we introduce a positive-definite self-adjoint linear operator $B : \mathbb{E} \rightarrow \mathbb{E}^*$ (notation $B \succ 0$). Then for $x \in \mathbb{E}$ and $g \in \mathbb{E}^*$ we have

$$\|x\|_B = \langle Bx, x \rangle^{1/2}, \quad \|g\|_B^* = \langle g, B^{-1}g \rangle^{1/2}.$$

If no ambiguity arise, we drop index B in this notation. Note that for any self-adjoint linear operator $A : \mathbb{E} \rightarrow \mathbb{E}^*$ we have

$$\|A\| = \min \{ \lambda : \lambda B \succeq \pm A \}.$$

If in addition, A is positive semidefinite and $L \geq \|A\|$, then

$$AB^{-1}A \preceq LA. \quad (1.1)$$

For function $f(\cdot)$ with open domain $\text{dom } f \subseteq \mathbb{E}$, we denote by $\nabla f(x) \in \mathbb{E}^*$ its gradient and by $\nabla^2 f(x) : \mathbb{E} \rightarrow \mathbb{E}^*$ its Hessian at point $x \in \text{dom } f$. For the mixed directional derivative of order $p \geq 1$ along directions $h_1, \dots, h_p$ in $\mathbb{E}$, we use notation $D^p f(x)[h_1, \dots, h_p]$. Note that under the standard continuity assumptions this is a symmetric p -linear form. If all directions are the same, we use notation $D^p f(x)[h]^p$.

We use the same notation for symmetric 4-linear forms. Let $f[x_1, x_2, x_3, x_4]$ be such a form with $x_i \in \mathbb{E}$, $i = 1, \dots, 4$. If all vectors are the same, we denote the value of this form by $f[x]^4$. This is a multivariate polynomial of degree four. Let us treat it as a function $\varphi(x) = f[x]^4$. Then the vector $f[x]^3 \in \mathbb{E}^*$ is defined as follows:

$$f[x]^3 = \frac{1}{4} \nabla \varphi(x), \quad x \in \mathbb{E}.$$

Note that $\langle f[x]^3, x \rangle = f[x]^4$. Similarly,

$$f[x]^2 = \frac{1}{12} \nabla^2 \varphi(x), \quad x \in \mathbb{E}.$$

Then $f[x]^2[y]^2 = \frac{1}{12} \langle \nabla^2 \varphi(x) y, y \rangle$ and $f[x]^2[x]^2 = f[x]^4$. Finally,

$$f[x] = \frac{1}{24} D^3 \varphi(x), \quad x \in \mathbb{E},$$

and $f[x][y]^3 = \frac{1}{24} D^3 \varphi(x)[y]^3$, $f[x][x]^3 = f[x]^4$. Thus, $f = \frac{1}{24} D^4 \varphi$.

Finally, we need one result, which is usually proved under more restrictive assumptions.

Lemma 1 *Let function $f(\cdot)$ be convex and twice continuously differentiable at its open domain. For two points $x, y \in \text{dom } f$, denote $M_{x,y} = \max_{0 \leq \alpha \leq 1} \|\nabla^2 f(\alpha x + (1 - \alpha)y)\|$, where the norm is Euclidean. Then*

$$\|\nabla f(x) - \nabla f(y)\|_B^* \leq M_{x,y} \langle \nabla f(x) - \nabla f(y), x - y \rangle. \quad (1.2)$$

Proof:

Denote $G = \int_0^1 \nabla^2 f(x + \tau(y - x)) d\tau$. Then $G(y - x) = \nabla f(y) - \nabla f(x)$ and $\|G\| \leq M_{x,y}$. Therefore, inequality (1.2) follows from (1.1). $\square$

2 Positive semidefinite quartic forms

In the first part of this paper, we are interested in numerical methods for finding an approximate solution to the following minimization problem:

$$F^* = \min_{x \in \text{dom } \psi} \left[F(x) = f(x) + \psi(x) \right], \quad (2.1)$$

where $\psi(\cdot)$ is a *simple* closed convex function and $f(\cdot)$ is a convex *quartic polynomial*. Since, $f(\cdot)$ is a polynomial of degree four, its Taylor expansion with respect to any point in $\mathbb{E}$ is exact. Namely, for any $x, y \in \mathbb{E}$ we have

$$\begin{aligned} f(y) &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\quad + \frac{1}{6} D^3 f(x)[y - x]^3 + f_4[y - x]^4. \end{aligned} \quad (2.2)$$

where $f_4[\cdot, \cdot, \cdot, \cdot]$ is a symmetric four-linear form in $\mathbb{E}$. Thus, $f_4 = \frac{1}{24} D^4 f$.

Convexity of function $f(\cdot)$ has several important consequences.

Theorem 1 *For any x and y in $\mathbb{E}$ we have*

$$0 \leq f_4[x]^2[y]^2. \quad (2.3)$$

This property implies the following relations:

$$2f_4[x, y] \leq f_4[x]^2 + f_4[y]^2, \quad (2.4)$$

$$(f_4[x]^2[y]^2)^2 \leq f_4[x]^4 \cdot f_4[y]^4, \quad (2.5)$$

$$(f_4[x]^3[y])^2 \leq f_4[x]^4 \cdot f_4[x]^2[y]^2. \quad (2.6)$$

Proof:

Indeed, taking in (2.2) $x = 0$, we have $\nabla^2 f(y) = \nabla^2 f(0) + D^3 f(0)[y] + \frac{1}{2} D^4 f(0)[y]^2$. Therefore,

$$0 \leq \langle \nabla^2 f(y)h, h \rangle = \langle \nabla^2 f(0)h, h \rangle + D^3 f(0)[y][h]^2 + 12f_4[y]^2[h]^2.$$

Replacing now y by τy and taking the limit as $\tau \rightarrow \infty$, we get inequality (2.3). It implies that $f_4[x - y]^2 \succeq 0$. And this is exactly the relation (2.4).

Further, in view of (2.3), for any $\tau \in \mathbb{R}$ we have $f_4[x - \tau y]^2[x + \tau y]^2 \geq 0$. If we open the brackets, then we get

$$0 \leq f_4[x]^4 + \tau^4 f_4[y]^4 - 2\tau^2 f_4[x]^2[y]^2.$$

If $f_4[y]^4 = 0$, then $f_4[x]^2[y]^2 = 0$ and (2.5) holds. If not, then by minimizing the right-hand side of the latter inequality in τ , we get the relation (2.5). Finally, minimizing the right-hand side of inequality

$$0 \leq f_4[x]^2[x - \tau y]^2$$

in τ , we get inequality (2.6). □

Define $d(x) = f_4[x]^4$. Note that

$$d(tx) = |t|^4 d(x), \quad x \in \mathbb{E}, t \in \mathbb{R}. \quad (2.7)$$

Let us look at some properties of this function. First of all, note that all derivatives of this function can be written in terms of the form f_4 as follows:

$$\begin{aligned} \langle \nabla d(x), h \rangle &= 4f_4[x]^3[h], \quad \langle \nabla^2 d(x)h, h \rangle = 12f_4[x]^2[h]^2, \\ D^3 d(x)[h]^3 &= 24f_4[x][h]^3, \quad D^4 d(x)[h]^4 = 24f_4[h]^4. \end{aligned} \quad (2.8)$$

where x and h are arbitrary vectors in $\mathbb{E}$. Thus, by inequality (2.3), we can see that function $d(\cdot)$ is convex.

Let us assume now that the form f_4 is positive definite:

$$f_4[x]^4 > 0 \quad \forall x \in \mathbb{E} \setminus \{0\}. \quad (2.9)$$

Then, we can define in $\mathbb{E}$ the following norm:

$$\|x\|_f = [d(x)]^{1/4}, \quad x \in \mathbb{E}. \quad (2.10)$$

This norm has several interesting properties. First of all, it justifies the uniform convexity of function $d(\cdot)$ itself.

Lemma 2 *Function $d(\cdot)$ is uniformly convex of degree four with constant $\sigma_d = \frac{4}{3}$:*

$$d(y) \geq d(x) + \langle d(x), y - x \rangle + \frac{1}{4}\sigma_d \|y - x\|_f^4, \quad x, y \in \mathbb{E}. \quad (2.11)$$

Moreover, we have the following lower quadratic bound for the growth of function $d(\cdot)$:

$$d(y) \geq d(x) + \langle d(x), y - x \rangle + \frac{1}{6} \langle \nabla^2 d(x)(y - x), y - x \rangle, \quad x, y \in \mathbb{E}. \quad (2.12)$$

Proof:

Indeed, for any x and y from $\mathbb{E}$, we have

$$d(y) - d(x) - \langle d(x), y - x \rangle = \frac{1}{2} \langle \nabla^2 d(x)(y - x), y - x \rangle$$

$$+ \frac{1}{6} D^3 d(x)[y - x]^3 + \frac{1}{24} D^4 d(x)[y - x]^4$$

$$\stackrel{(2.8)}{=} 6f_4[x]^2[y - x]^2 + 4f_4[x][y - x]^3 + f_4[y - x]^4$$

$$\stackrel{(2.6)}{\geq} 6f_4[x]^2[y - x]^2 - 4(f_4[x]^4 \cdot f_4[x]^2[y - x]^2)^{1/2} + f_4[y - x]^4.$$

Minimizing the right-hand side of this inequality in $f_4[x]^2[y - x]^2$, we get

$$d(y) - d(x) - \langle d(x), y - x \rangle \geq (1 - \frac{2}{3})f_4[y - x]^4 = \frac{1}{3}\|y - x\|_f^4.$$

Finally, minimizing the same inequality in $f_4[x]^4$, we get

$$d(y) - d(x) - \langle d(x), y - x \rangle \geq (6 - 4)f_4[x]^2[y - x]^2,$$

and this is (2.12). □

Consider now the function $Q_f(x) = \frac{1}{2}\|x\|_f^2$.

Lemma 3 *Function $Q_f(\cdot)$ is strongly convex and has Lipschitz continuous gradients with respect to the norm $\|\cdot\|_f$:*

$$\frac{1}{12\|x\|_f^2} \langle \nabla^2 d(x)h, h \rangle \leq \langle \nabla^2 Q_f(x)h, h \rangle \leq 3\|h\|_f^2, \quad x, h \in \mathbb{E}. \quad (2.13)$$

Proof:

Since $Q_f(x) = \frac{1}{2}d^{1/2}(x)$, for any x and h from $\mathbb{E}$, we have

$$\begin{aligned} \langle \nabla Q_f(x), h \rangle &= d^{-1/2}(x) f_4(x) [x]^3 [h], \\ \langle \nabla^2 Q_f(x)h, h \rangle &= 3d^{-1/2}(x) f_4(x) [x]^2 [h]^2 - 2d^{-3/2}(x) (f_4[x]^3 [h])^2 \\ &\stackrel{(2.6)}{\geq} d^{-1/2}(x) f_4(x) [x]^2 [h]^2. \end{aligned}$$

On the other hand,

$$\langle \nabla^2 Q_f(x)h, h \rangle \leq 3d^{-1/2}(x) f_4(x) [x]^2 [h]^2 \stackrel{(2.5)}{\leq} 3\|h\|_f^2. \quad \square$$

3 Minimizing convex polynomials of degree four

In this section, we are going to check which properties of function $d(\cdot)$ are inherited by function $f(\cdot)$ and how we can use them for constructing minimization schemes. The next statement is a variant of Theorem 1 in [17].

Lemma 4 *For any point x and direction $h \in \mathbb{E}$, we have*

$$D^3 f(x)[h] \leq \frac{1}{\tau} \nabla^2 f(x) + 12\tau f_4[h]^2 \quad (3.1)$$

Proof:

Since $f(\cdot)$ is a convex quartic polynomial, for any $\tau > 0$ we have

$$0 \leq \nabla^2 f(x - \tau h) = \nabla^2 f(x) - \tau D^3 f(x)[h] + \frac{\tau^2}{2} D^4 f(x)[h]^2.$$

Dividing this inequality by τ , we get

$$D^3 f(x)[h] \leq \frac{1}{\tau} \nabla^2 f(x) + \frac{\tau}{2} D^4 f(x)[h]^2,$$

and this is the relation (3.1). $\square$

Corollary 1 *For all $x, y \in \mathbb{E}$ and $\tau > 0$, we have*

$$\begin{aligned} &f(y) - f(x) - \langle f(x), y - x \rangle - \frac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\leq \frac{1}{6\tau} \langle \nabla^2 f(x)(y - x), y - dx \rangle + (1 + 2\tau) \|y - x\|_f^4, \end{aligned} \quad (3.2)$$

$$\begin{aligned} &f(y) - f(x) - \langle f(x), y - x \rangle - \frac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\geq -\frac{1}{6\tau} \langle \nabla^2 f(x)(y - x), y - dx \rangle + (1 - 2\tau) \|y - x\|_f^4. \end{aligned} \quad (3.3)$$

Proof:

In view of relation (3.1), we have $|D^3 f(x)[h]^3| \leq \frac{1}{\tau} \langle \nabla^2 f(x)h, h \rangle + 12\tau \|h\|_f^4$. Therefore,

$$\begin{aligned} f(y) - f(x) - \langle f(x), y - x \rangle - \frac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ \stackrel{(2.2)}{=} \frac{1}{6} D^3 f(x)[y - x]^3 + \|y - x\|_f^4 \leq \frac{1}{6\tau} \langle \nabla^2 f(x)h, h \rangle + (1 + 2\tau) \|h\|_f^4. \end{aligned}$$

The second inequality can be justified in the same way. $\square$

Using inequalities (3.2) and (3.3), we can write down the lower and upper approximations for the smooth part of the objective function:

$$\begin{aligned} \bar{f}_{x,\tau}(y) &= f(x) + \langle f(x), y - x \rangle + \frac{3\tau-1}{6\tau} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\quad + (1 - 2\tau) \|y - x\|_f^4 \leq f(y), \\ \hat{f}_{x,\tau}(y) &= f(x) + \langle f(x), y - x \rangle + \frac{1+3\tau}{6\tau} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\quad + (1 + 2\tau) \|y - x\|_f^4 \geq f(y), \quad x, y \in \mathbb{E}. \end{aligned} \tag{3.4}$$

Note that we are interested in $\tau \in [\frac{1}{3}, \frac{1}{2}]$.

Let us assume that the norm $\|\cdot\|_f$ is simple enough for minimizing the following auxiliary function,

$$\hat{f}_{x,\tau}(y) + \psi(y), \quad y \in \text{dom } \psi,$$

in a closed form. Thus, we can use it for justifying methods for solving the problem (2.1).

Let us define the following parametric function:

$$\begin{aligned} \bar{f}_{x,\tau}(\alpha, y) &= f(x) + \langle f(x), y - x \rangle + \frac{3\tau-1}{6\tau\alpha} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\quad + \frac{1-2\tau}{\alpha^3} \|y - x\|_f^4, \quad y \in \mathbb{E}, \alpha > 0. \end{aligned} \tag{3.5}$$

Note that $\bar{f}_{x,\tau}(1, y) = \bar{f}_{x,\tau}(y) \leq f(y)$ for all $y \in \mathbb{E}$. On the other hand, function $\bar{f}_{x,\tau}(\cdot, \cdot)$ is jointly convex in its arguments. Therefore, the univariate function

$$F_{x,\tau}^*(\alpha) = \min_{y \in \text{dom } \psi} \left[\bar{f}_{x,\tau}(\alpha, y) + \psi(y) \right], \quad \alpha > 0,$$

is convex. Let us prove the following result.

Theorem 2 *Let $\tau \in (\frac{1}{3}, \frac{1}{2})$ and the parameter $\alpha > 0$ satisfies the following condition:*

$$\alpha \leq \min \left\{ \frac{3\tau-1}{3\tau+1}, \left[\frac{1-2\tau}{1+2\tau} \right]^{1/3} \right\}. \tag{3.6}$$

Then

$$\min_{y \in \text{dom } \psi} \left[\hat{f}_{x,\tau}(y) + \psi(y) \right] \leq \alpha F^* + (1 - \alpha) F(x). \tag{3.7}$$

Proof:

Note that $F_{x,\tau}^*(1) \leq F^*$. On the other hand, by continuity, we can define $F_{x,\tau}^*(0) = F(x)$. At the same time, condition (3.6) ensures the following relations:

$$\frac{3\tau-1}{6\tau\alpha} \geq \frac{3\tau+1}{6\tau}, \quad \frac{1-2\tau}{\alpha^3} \geq 1 + 2\tau.$$

This means that $\bar{f}_{x,\tau}(\alpha, y) \geq \hat{f}_{x,\tau}(y)$ for all $y \in \mathbb{E}$. Hence,

$$\begin{aligned} \min_{y \in \text{dom } \psi} \left[\hat{f}_{x,\tau}(y) + \psi(y) \right] &\leq F_{x,\tau}^*(\alpha) \leq \alpha F_{x,\tau}^*(1) + (1-\alpha)F_{x,\tau}^*(0) \\ &\leq \alpha F^* + (1-\alpha)F(x). \quad \square \end{aligned}$$

Inequality (3.7) is important for justification of the following optimization scheme. Denote by τ_* the unique positive root of the following equation:

$$\frac{3\tau-1}{3\tau+1} = \left[\frac{1-2\tau}{1+2\tau} \right]^{1/3}. \quad (3.8)$$

Damped Quartic Newton Method	
Initialization. Choose $x_0 \in \text{dom } \psi$. kth iteration ($k \geq 0$). Iterate $x_{k+1} = \arg \min_{y \in \text{dom } \psi} \left[\hat{f}_{x_k, \tau_*}(y) + \psi(y) \right]$.	(3.9)

Theorem 3 For all $k \geq 0$, we have

$$F(x_k) - F^* \leq (1 - \alpha_*)^k (F(x_0) - F^*), \quad (3.10)$$

where $\alpha_* = \frac{3\tau_*-1}{3\tau_*+1} > 0.193$ with $\tau_* = \frac{1}{6}\sqrt{3 + \sqrt{33}}$.

Proof:

Note that for any $x, y \in \mathbb{E}$ we have $\bar{f}_{x,\tau_*}(\alpha_*, y) = \hat{f}_{x,\tau_*}(y)$. Therefore,

$$F(x_{k+1}) \stackrel{(3.4)}{\leq} \hat{f}_{x_k, \tau_*}(x_{k+1}) + \psi(x_{k+1}) \stackrel{(3.7)}{\leq} \alpha_* F^* + (1 - \alpha_*)F(x_k).$$

It remains to note that τ_* is the root of the equation

$$(3\tau - 1)^3(1 + 2\tau) \stackrel{(3.8)}{=} (3\tau + 1)^3(1 - 2\tau).$$

After simplification, it is reduced to $54\tau^4 - 9\tau^2 = 1$, which gives us $\tau_* = \frac{1}{6}\sqrt{3 + \sqrt{33}}$. $\square$

Method (3.9) is implementable if the structure of the norm $\|\cdot\|_f$ is quite simple. Let us end up this section with an important source of such problems.

Example 1 For a convex function, consider its Taylor polynomials of degree three, augmented by a fourth power of Euclidean norm:

$$\begin{aligned} f_{\bar{x},H}(y) &= f(\bar{x}) + \langle \nabla f(\bar{x}), y - \bar{x} \rangle + \frac{1}{2} \langle \nabla^2 f(\bar{x})(y - \bar{x}), y - \bar{x} \rangle \\ &\quad + \frac{1}{6} D^3 f(\bar{x})[y - \bar{x}]^3 + \frac{H}{24} \|y - \bar{x}\|^4, \end{aligned}$$

where $\|\cdot\|$ is a Euclidean norm. The possibility to minimize this polynomial is essential for the third-order schemes. It can be proved [17] that for $H \geq 3L_3$, where L_3 is the Lipschitz constant for third derivative of function $f(\cdot)$, the quartic polynomial $f_{\bar{x},H}(\cdot)$ is convex. Note that it fits our framework since $D^4 f_{\bar{x},H}(x)[h]^4 = H\|h\|^4$. Thus, in this case, the auxiliary problem in method (3.9) can be solved by using the standard tools of Linear Algebra.

4 Simplified approximations

Let us introduce in the space $\mathbb{E}$ some norm $\|\cdot\|$, which is simpler than the norm $\|\cdot\|_f$. Since the space $\mathbb{E}$ is finite-dimensional, assumption (2.9) implies existence of two constants $0 < \mu \leq L$ such that

$$\mu\|x\|^4 \leq f_4[x]^4 \leq L\|x\|^4, \quad x \in \mathbb{E}. \quad (4.1)$$

In the particular case when $\|\cdot\| = \|\cdot\|_f$, we have $\mu = L = 1$. Denote by $q = \frac{\mu}{L}$ the condition number of the norm $\|\cdot\|$ with respect to the optimal norm $\|\cdot\|_f$.

Now, in view of (3.4), we can define the following simplified lower and upper bounds for the smooth part of the objective function:

$$\begin{aligned} \bar{\phi}_{x,\tau}(y) &\stackrel{\text{def}}{=} f(x) + \langle f(x), y - x \rangle + \frac{3\tau-1}{6\tau} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\quad + (1 - 2\tau)\mu\|y - x\|^4 \leq f(y), \\ \hat{\phi}_{x,\tau}(y) &\stackrel{\text{def}}{=} f(x) + \langle f(x), y - x \rangle + \frac{1+3\tau}{6\tau} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\quad + (1 + 2\tau)L\|y - x\|^4 \geq f(y), \quad x, y \in \mathbb{E}. \end{aligned} \quad (4.2)$$

As in (3.5), define the following parametric functions

$$\begin{aligned} \bar{\phi}_{x,\tau}(\alpha, y) &\stackrel{\text{def}}{=} f(x) + \langle f(x), y - x \rangle + \frac{3\tau-1}{6\tau\alpha} \cdot \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\quad + (1 - 2\tau) \cdot \frac{\mu}{\alpha^3} \|y - x\|^4, \end{aligned} \quad (4.3)$$

$$\Phi_{x,\tau}^*(\alpha) = \min_{y \in \text{dom } \psi} \left[\bar{\phi}_{x,\tau}(\alpha, y) + \psi(y) \right], \quad \alpha > 0.$$

Consider the following optimization method.

Relaxed Quartic Newton Method
Initialization. Choose $x_0 \in \text{dom } \psi$ and parameter $\tau \in (\frac{1}{3}, \frac{1}{2})$. kth iteration ($k \geq 0$). Iterate $x_{k+1} = \arg \min_{y \in \text{dom } \psi} [\hat{\phi}_{x_k, \tau}(y) + \psi(y)]$.

(4.4)

As compared with method (3.9), we use here a conservative approximation of the smooth part of the objective function: $\hat{\phi}_{x, \tau}(y) \stackrel{(4.1)}{\geq} \hat{f}_{x, \tau}(y)$ for all $x, y \in \mathbb{E}$.

Theorem 4 *Let the parameter τ of method (4.4) satisfy the following condition:*

$$\alpha \leq \min \left\{ \frac{3\tau-1}{3\tau+1}, \left[q \cdot \frac{1-2\tau}{1+2\tau} \right]^{1/3} \right\} \quad (4.5)$$

for some $\alpha > 0$. Then, for all $k \geq 0$, we have

$$F(x_k) - F^* \leq (1 - \alpha)^k (F(x_0) - F^*). \quad (4.6)$$

Proof:

Indeed, in view of condition (4.5), we have

$$\frac{3\tau+1}{6\tau} \leq \frac{3\tau-1}{6\tau\alpha}, \quad (1+2\tau)L \leq (1-2\tau) \cdot \frac{\mu}{\alpha^3}.$$

Hence,

$$\begin{aligned} F(x_{k+1}) &\stackrel{(4.2)}{\leq} \hat{\phi}_{x_k, \tau}(x_{k+1}) + \psi(x_{k+1}) \stackrel{(4.4)}{=} \min_{y \in \text{dom } \psi} [\hat{\phi}_{x_k, \tau}(y) + \psi(y)] \\ &\stackrel{(4.5)}{\leq} \min_{y \in \text{dom } \psi} [\bar{\phi}_{x_k, \tau}(\alpha, y) + \psi(y)] \stackrel{(4.3)}{=} \Phi_{x_k, \tau}^*(\alpha). \end{aligned}$$

On the other hand, since function $\Phi_{x_k, \tau}^*(\cdot)$ is convex, we have

$$\Phi_{x_k, \tau}^*(\alpha) \leq \alpha \Phi_{x_k, \tau}^*(1) + (1 - \alpha) \Phi_{x_k, \tau}^*(0) \stackrel{(4.2)}{\leq} \alpha F^* + (1 - \alpha) F(x_k).$$

Thus, we obtain the bound (4.6). □

Since we are interested in the maximal possible value of α , let us try to understand which value of $\tau \in (\frac{1}{3}, \frac{1}{2})$ makes the right-hand side of inequality (4.5) big. Note that

$$\frac{3\tau-1}{3\tau+1} \geq \frac{2}{5}(3\tau-1), \quad \tau \in (\frac{1}{3}, \frac{1}{2}).$$

Let us look now at the function $\xi(\tau) = \omega^{1/3}(\tau)$ with $\omega(\tau) = \frac{1-2\tau}{1+2\tau}$. Since

$$\begin{aligned}\xi'(\tau) &= \frac{1}{3}\omega^{-2/3}(\tau)\omega'(\tau) & \xi''(\tau) &= -\frac{2}{9}\omega^{-5/3}(\tau)(\omega'(\tau))^2 + \frac{1}{3}\omega^{-2/3}\omega''(\tau), \\ \omega'(\tau) &= -\frac{4}{(1+2\tau)^2}, & \omega''(\tau) &= \frac{16}{(1+2\tau)^3},\end{aligned}$$

we conclude that $\xi''(\tau) = \frac{16\omega^{-5/3}(\tau)}{9(1+2\tau)^4}(-2 + 3(1-2\tau)) \leq 0$ for $\tau \in (\frac{1}{3}, \frac{1}{2})$. Hence, for this interval, we have

$$\xi(\tau) \geq \xi(\frac{1}{3}) \cdot 3(1-2\tau) = 3(1-2\tau) \cdot 5^{-1/3}.$$

Thus, the condition (4.5) is satisfied for

$$\alpha = \min \left\{ \frac{2}{5}(3\tau - 1), 3(1-2\tau) \cdot \left(\frac{q}{5}\right)^{1/3} \right\}. \quad (4.7)$$

Therefore, the best choice of τ is as follows:

$$\tau_{\#} = \frac{1}{2} - \frac{1}{6(1+5\kappa)}, \quad \kappa = \left(\frac{q}{5}\right)^{1/3}. \quad (4.8)$$

In this case, $\alpha_{\#} = \frac{2}{5}(3\tau_{\#} - 1) = \frac{\kappa}{1+5\kappa}$. Thus, we have proved the following theorem.

Theorem 5 *Let the sequence of points $\{x_k\}_{k \geq 0}$ be generated by the method (4.4) with parameter $\tau = \tau_{\#}$. Then, for all $k \geq 0$, we have*

$$F(x_k) - F^* \leq (1 - \alpha_{\#})^k (F(x_0) - F^*). \quad (4.9)$$

In order to generate an ϵ -solution of problem (2.1) in function value, method (4.4) needs at most

$$\left(5 + \left(\frac{5}{q}\right)^{1/3}\right) \ln \frac{F(x_0) - F^*}{\epsilon} \quad (4.10)$$

iterations. Note that for $q = 1$, we have $\alpha_{\#} = \frac{1}{5+5^{1/3}} > 0.149$, which is quite close to α_* . Thus, the general rule (4.8) is not much worse than the optimal specific rule (3.8).

For practical applications, it is reasonable to use the Euclidean norm $\|\cdot\|$. In this case, the complexity of one iteration of method (4.4) is the same as that of the usual Cubic Newton Method [22]

5 Quartic regularity for convex functions

Let us check to what extent the results of the previous sections can be propagated onto the class of general convex functions. Consider the convex unconstrained minimization problem

$$F^* = \min_{x \in \text{dom } \psi} \left[F(x) = f(x) + \psi(x) \right], \quad (5.1)$$

where $\psi(\cdot)$ is a simple closed convex function, and the smooth part $f(\cdot)$ of the objective satisfies an additional assumption of *quartic regularity*.

Definition 1 Function $f(\cdot)$ with convex and open domain $\text{dom } f \subseteq \mathbb{E}$ is called quartic regular (Q -regular) on convex set S , if it is four times continuously differentiable on $\text{dom } f$, and there exist two parameters $L \geq \mu \geq 0$ such that

$$D^4 f(x)[h]^4 \geq \mu \|h\|^4, \quad \forall h \in \mathbb{E}, x \in S, \quad (5.2)$$

$$D^4 f(x)[h]^4 \leq L \|h\|^4, \quad \forall h \in \mathbb{E}, x \in S_{1/2}, \quad (5.3)$$

where $S_{1/2} = \{z = x - \frac{1}{2}(y - x) : x, y \in S\} \subseteq \text{dom } f$.

This definition can be applied to general norms.²⁾

Note that quartic regularity itself does not imply convexity. Thus, for this section, our assumption is as follows.

Assumption 1 Smooth part of the objective function in problem (5.1) is quartic regular on the set $S \supseteq \text{dom } \psi$.

Let us mention some trivial properties of Q -regular functions.

- If functions $f_i(\cdot)$ are Q -regular on sets S_i with parameters (μ_i, L_i) , $i = 1, 2$, then their weighted sum $f(x) = \alpha f_1(x) + \beta f_2(x)$, $\alpha, \beta \geq 0$, is Q -regular on $S_1 \cap S_2$ with parameters

$$(\alpha\mu_1 + \beta\mu_2, \alpha L_1 + \beta L_2).$$

- If function $\phi(\cdot)$ is Q -regular on S_1 with parameters (μ, L) and $A : \mathbb{E} \rightarrow \mathbb{E}_1$ is a non-degenerate linear operator, then function $f(x) = \phi(Ax)$ is Q -regular on $S = \{x : Ax \in S_1\}$ with parameters $(\mu \sigma_{\min}^4(A), L \sigma_{\max}^4(A))$.
- In particular, if function $f(\cdot)$ is Q -regular on S with parameters (μ, L) , then function $\varphi_\tau(x) = f(\tau x)$, with scaling factor $\tau \neq 0$, is Q -regular on $\tau^{-1}S$ with parameters $(\mu\tau^4, L\tau^4)$.
- Function $d(x) = \frac{1}{24}\|x\|^4$, defined by a Euclidean norm, is Q -regular on $\mathbb{E}$ with parameters $\mu = L = 1$.

The abilities of optimization methods in solving the problem (5.1) are supported by the following inequalities.

Theorem 6 Let function $f(\cdot)$ be Q -regular on S with parameters μ and L . Then for all $x, y \in S$ and $\gamma \in [\frac{1}{3}, \frac{1}{2}]$, we have

$$\begin{aligned} f(y) &\geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{3\gamma-1}{6\gamma} \langle \nabla^2 f(x)(y-x), y-x \rangle \\ &\quad + \frac{\mu}{24} \|y-x\|^4 \left[(1-2\gamma) - \frac{16}{125} \left(\frac{L}{\mu} - 1 \right) \left(\frac{3\gamma-1}{\gamma} \right)^3 \right]. \end{aligned} \quad (5.4)$$

Moreover, for all $\gamma > 0$, we have

$$\begin{aligned} f(y) &\leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{3\gamma+1}{6\gamma} \langle \nabla^2 f(x)(y-x), y-x \rangle \\ &\quad + \frac{2\gamma+1}{24} L \|y-x\|^4. \end{aligned} \quad (5.5)$$

²⁾ As compared to (4.1), it is more convenient now to work directly with the fourth derivative.

Proof:

By Taylor formula, we have

$$\begin{aligned}\Delta(x, y) &\stackrel{\text{def}}{=} f(y) - f(x) - \langle \nabla f(x), y - x \rangle - \frac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &= \frac{1}{6} D^3 f(x)[y - x]^3 + \frac{1}{3!} \int_0^1 (1 - \tau)^3 D^4 f(x + \tau(y - x))[y - x]^4 d\tau.\end{aligned}$$

Since $f(\cdot)$ is convex,

$$\begin{aligned}0 \leq \nabla^2 f(x + \gamma(y - x)) &= \nabla^2 f(x) + \gamma D^3 f(x)(y - x) \\ &\quad + \int_0^\gamma (\gamma - \tau) D^4 f(x + \tau(y - x))[y - x]^2 d\tau.\end{aligned}$$

Hence,

$$\begin{aligned}\Delta(x, y) &\geq -\frac{1}{6\gamma} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\quad + \int_0^\gamma \left[\frac{(1-\tau)^3}{3!} - \frac{1}{6} \left(1 - \frac{\tau}{\gamma} \right) \right] D^4 f(x + \tau(y - x))[y - x]^4 d\tau \quad (5.6) \\ &\quad + \frac{1}{3!} \int_\gamma^1 (1 - \tau)^3 D^4 f(x + \tau(y - x))[y - x]^4 d\tau.\end{aligned}$$

Denote $\theta(\tau) = (1 - \tau)^3 + \frac{\tau}{\gamma} - 1 = \tau \left(\frac{1}{\gamma} - 3 \right) + 3\tau^2 - \tau^3$. This cubic polynomial has three real roots, one at $\tau = 0$ and two others at the points

$$\tau_{0,1}(\gamma) = \frac{3}{2} \pm \sqrt{\frac{1}{\gamma} - \frac{3}{4}}.$$

Note that by conditions of the theorem, we have

$$0 \leq \tau_0(\gamma) \leq \gamma.$$

Therefore, at the interval $\tau \in [0, 1]$, we have the following relations:

$$\theta(\tau) \leq 0, \quad \tau \in [0, \tau_0(\gamma)], \quad \theta(\tau) \geq 0, \quad \tau_0(\tau) \leq \tau \leq 1 \quad (\leq \tau_1(\gamma)).$$

Denoting the first integral in (5.6) by I_1 , the second one by I_2 , and $r = \|y - x\|$, we have

$$\begin{aligned}6I_1 &= \int_0^{\tau_0(\gamma)} \theta(\tau) D^4 f(x + \tau(y - x))[y - x]^4 d\tau + \int_{\tau_0(\gamma)}^\gamma \theta(\tau) D^4 f(x + \tau(y - x))[y - x]^4 d\tau \\ &\stackrel{(5.3)}{\geq} Lr^4 \int_0^{\tau_0(\gamma)} \theta(\tau) d\tau + \mu r^4 \int_{\tau_0(\gamma)}^\gamma \theta(\tau) d\tau \\ &= (L - \mu)r^4 \int_0^{\tau_0(\gamma)} \theta(\tau) d\tau + \mu r^4 \int_0^\gamma \theta(\tau) d\tau.\end{aligned}$$

On the other hand, $6I_2 \stackrel{(5.2)}{\geq} \mu r^4 \int_{\gamma}^1 (1-\tau)^3 d\tau$. Note that

$$\int_0^{\gamma} \theta(\tau) d\tau + \int_{\gamma}^1 (1-\tau)^3 d\tau = \int_0^1 (1-\tau)^3 d\tau + \int_0^{\gamma} \left(\frac{\tau}{\gamma} - 1\right) d\tau = \frac{1-2\gamma}{4},$$

and we conclude that

$$6(I_1 + I_2) \geq (L - \mu) r^4 \int_0^{\tau_0(\gamma)} \theta(\tau) d\tau + \mu r^4 \frac{1-2\gamma}{4}.$$

It remains to estimate the integral in the latter inequality from below.

Denoting for short $\tau_0 = \tau_0(\gamma)$, we have

$$\begin{aligned} I_0 &\stackrel{\text{def}}{=} \int_0^{\tau_0} \theta(\tau) d\tau = \int_0^{\tau_0} \left[(1-\tau)^3 + \frac{\tau}{\gamma} - 1 \right] d\tau \\ &= -\frac{1}{4}(1-\tau)^4 + \frac{\tau^2}{2\gamma} - \tau \Big|_0^{\tau_0} = -\frac{1}{4}(1-\tau_0)^4 + \frac{\tau_0^2}{2\gamma} - \tau_0 + \frac{1}{4} \end{aligned}$$

Since $\theta(\tau_0) = 0$, we conclude that

$$I_0 = -\frac{1}{4}(1-\tau_0) \left(1 - \frac{\tau_0}{\gamma} \right) + \frac{\tau_0^2}{2\gamma} - \tau_0 + \frac{1}{4} = -\frac{3\gamma-1}{4\gamma} \tau_0 + \frac{\tau_0^2}{4\gamma}.$$

Denote $\xi = \frac{3\gamma-1}{\gamma}$ and $\lambda = \frac{4}{9}\xi$. Then $0 \leq \xi \leq 1$, $0 \leq \lambda \leq \frac{4}{9}$, and

$$\tau_0 = \frac{3}{2} - \sqrt{\frac{1}{\gamma} - \frac{3}{4}} = \frac{3}{2} - \sqrt{\frac{9}{4} - \xi} = \frac{3}{2}(1 - \sqrt{1-\lambda}) \leq \frac{3}{2} \cdot \frac{3}{5}\lambda = \frac{2}{5}\xi.$$

Note that in our new notation we have

$$3\tau_0 - \tau_0^2 = 3 - \frac{1}{\gamma} = \xi. \quad (5.7)$$

Therefore,

$$\begin{aligned} 4I_0 &= -\xi\tau_0 + \frac{1}{\gamma}\tau_0^2 \stackrel{(5.7)}{=} -(3\tau_0 - \tau_0^2)\tau_0 + \frac{1}{\gamma}\tau_0^2 = \tau_0^3 - \xi\tau_0^2 \\ &\stackrel{(5.7)}{=} \tau_0^3 - (3\tau_0 - \tau_0^2)\tau_0^2 = -2\tau_0^3 + \tau_0^4 \geq -2\tau_0^3 \geq -\frac{16}{125}\xi^3. \end{aligned}$$

Thus, we get the following lower bound:

$$\begin{aligned} f(y) - f(x) - \langle \nabla f(x), y - x \rangle &\geq \frac{\xi}{6} \langle \nabla^2 f(x)(y - x), y - x \rangle \\ &\quad + \frac{\mu r^4}{6} \left[\frac{1-2\gamma}{4} - \frac{4}{125} \left(\frac{L}{\mu} - 1 \right) \xi^3 \right]. \end{aligned}$$

In order to prove the upper bound, note that

$$\begin{aligned} 0 \leq \nabla^2 f(x - \gamma(y - x)) &= \nabla^2 f(x) - \gamma D^3 f(x)(y - x) \\ &\quad + \int_0^{\gamma} (\gamma - \tau) D^4 f(x - \tau(y - x)) [y - x]^2 d\tau. \end{aligned}$$

Hence,

$$\begin{aligned}
\Delta(x, y) &\leq \frac{1}{6\gamma} \langle \nabla^2 f(x)(y-x), y-x \rangle \\
&\quad + \frac{1}{6\gamma} \int_0^\gamma (\gamma - \tau) D^4 f(x - \tau(y-x)) [y-x]^4 d\tau \\
&\quad + \frac{1}{3!} \int_0^1 (1 - \tau)^3 D^4 f(x + \tau(y-x)) [y-x]^4 d\tau \\
&\stackrel{(5.3)}{\leq} \frac{1}{6\gamma} \langle \nabla^2 f(x)(y-x), y-x \rangle + \frac{L}{6} \|y-x\|^4 \left(\frac{\gamma}{2} + \frac{1}{4} \right). \quad \square
\end{aligned} \tag{5.8}$$

Now we can estimate performance of method (4.4) as applied to problem (5.1). Define

$$\begin{aligned}
\hat{\xi}_{x,\gamma}(y) &\stackrel{\text{def}}{=} f(x) + \langle f(x), y-x \rangle + \frac{1+3\gamma}{6\gamma} \langle \nabla^2 f(x)(y-x), y-x \rangle \\
&\quad + \frac{1+2\gamma}{24} L \|y-x\|^4 \stackrel{(5.5)}{\geq} f(y), \quad x, y \in S,
\end{aligned} \tag{5.9}$$

where $f(\cdot)$ is a Q-regular function. Consider the following optimization method.

Quartic Regularization of Newton Method for Q-Regular Functions	
Initialization. Choose $x_0 \in \text{dom } \psi$ and parameter $\gamma > 0$. kth iteration ($k \geq 0$). Iterate $x_{k+1} = \arg \min_{y \in \text{dom } \psi} \left[\hat{\xi}_{x_k, \gamma}(y) + \psi(y) \right]$.	(5.10)

Theorem 7 Let function $f(\cdot)$ be Q-regular on the sets $S \supseteq \text{dom } \psi$ with parameters $0 < \mu \leq L$. And let the parameter $\gamma \in (\frac{1}{3}, \frac{1}{2})$ of method (5.10) satisfy the following condition:

$$\kappa(\gamma) \stackrel{\text{def}}{=} (1 - 2\gamma) - \frac{16}{125} \left(\frac{L}{\mu} - 1 \right) \left(\frac{3\gamma-1}{\gamma} \right)^3 \geq 0, \tag{5.11}$$

$$\alpha \leq \min \left\{ \frac{3\gamma-1}{3\gamma+1}, \left[\frac{\mu}{L(1+2\gamma)} \kappa(\gamma) \right]^{1/3} \right\}$$

for some $\alpha > 0$. Then, for all $k \geq 0$, we have

$$F(x_k) - F^* \leq (1 - \alpha)^k (F(x_0) - F^*). \tag{5.12}$$

Proof:

As in the proof of Theorem 4, define the following parametric functions:

$$\begin{aligned}
\hat{\xi}_{x,\gamma}(\alpha, y) &= f(x) + \langle \nabla f(x), y-x \rangle + \frac{3\gamma-1}{6\alpha\gamma} \langle \nabla^2 f(x)(y-x), y-x \rangle \\
&\quad + \frac{\mu}{24\alpha^3} \|y-x\|^4 \kappa(\gamma), \\
\Xi_{x,\gamma}^*(\alpha) &= \min_{y \in \text{dom } \psi} \left[\hat{\xi}_{x,\gamma}(\alpha, y) + \psi(y) \right], \quad \alpha > 0.
\end{aligned}$$

Note that $\Xi_{x,\gamma}^*(0) = F(x)$ and $\hat{\xi}_{x,\gamma}(1, y) \stackrel{(5.4)}{\leq} f(y)$. Thus, $\Xi_{x,\gamma}^*(1) \leq F^*$. On the other hand, in view of conditions (5.11), we have

$$\frac{3\gamma-1}{6\alpha\gamma} \geq \frac{3\gamma+1}{6\gamma}, \quad \frac{\mu}{24\alpha^3}\kappa(\gamma) \geq \frac{1+2\gamma}{24}L.$$

Therefore, as in the proof of Theorem 4, we conclude that

$$F(x_{k+1}) \leq \Xi_{x,\gamma}^*(\alpha) \leq \alpha F^* + (1-\alpha)F(x_k). \quad \square$$

Note that method (5.10) can be endowed with a termination criterion, which supports its rate of convergence (5.12). Indeed, in this method, in parallel with the minimizing sequence $\{x_k\}_{k \geq 0}$, we can update the following lower bounds for the optimal value of problem (5.1):

$$\xi_0^* = \Xi_{x_0,\gamma}^*(1), \quad \xi_{k+1}^* = \max \left\{ \xi_k^*, \Xi_{x_k,\gamma}^*(1) \right\}, \quad k \geq 0. \quad (5.13)$$

Clearly, under conditions of Theorem 7, for all $k \geq 0$, we have

$$(a) : \xi_k^* \leq \xi_{k+1}^* \leq F^*, \quad (b) : \xi_{k+1}^* \geq \Xi_{x_k,\gamma}^*(1). \quad (5.14)$$

Theorem 8 *Let the conditions of Theorem 7 be satisfied. Then, for all $k \geq 0$, we have*

$$F(x_k) - F^* \leq F(x_k) - \xi_k^* \leq (1-\alpha)^k (F(x_0) - \xi_0^*). \quad (5.15)$$

Proof:

Indeed, by the same reasons as in the proof of Theorem 4, we can see that

$$\begin{aligned} F(x_{k+1}) - \xi_{k+1}^* &\leq \alpha \Xi_{x_k,\gamma}^*(1) + (1-\alpha)F(x_k) - \xi_{k+1}^* \\ &\stackrel{(5.14b)}{\leq} (1-\alpha)(F(x_k) - \xi_{k+1}^*) \stackrel{(5.14a)}{\leq} (1-\alpha)(F(x_k) - \xi_k^*). \quad \square \end{aligned}$$

Thus, if the inequality

$$F(x_k) - \xi_k^* \leq \epsilon \quad (5.16)$$

is satisfied, then the point x_k is an ϵ -solution of problem (5.1).

Finally, let us provide method (5.10) with a reasonable value of parameter γ . In accordance to the relation (5.11), it could be found from the balance equation

$$\left(\frac{3\gamma-1}{3\gamma+1} \right)^3 = \frac{\mu\kappa(\gamma)}{L(1+2\gamma)}.$$

However, it is quite complicated. Moreover, we need to emphasize the asymptotic dependence of the coefficient α of (5.11) in the condition number q . Therefore, we are going to use an approximate solution of the above equation.

Theorem 9 *Let $q = \frac{\mu}{L} \leq 1$. For coefficient*

$$\gamma = \gamma_* \stackrel{\text{def}}{=} \frac{1}{3(1-\frac{3}{11}q^{1/3})} \leq \frac{11}{24}, \quad (5.17)$$

we have

$$\frac{3\gamma-1}{3\gamma+1} \leq \left[\frac{q\kappa(\gamma)}{1+2\gamma} \right]^{1/3}. \quad (5.18)$$

Proof:

Let $\gamma = \gamma_* \stackrel{\text{def}}{=} \frac{1}{3(1-\tau_*q^{1/3})}$ with $\tau_* = \frac{3}{11}$. Denote $\lambda = \frac{3\gamma-1}{3\gamma} = \tau_*q^{1/3}$. We need to prove that

$$\begin{aligned} (1+2\gamma) \left(\frac{3\gamma-1}{3\gamma+1} \right)^3 &\leq q \left((1-2\gamma) - \frac{16}{125} (q^{-1}-1) \left(\frac{3\gamma-1}{\gamma} \right)^3 \right) \\ &= q \left((1-2\gamma) + 16 \left(\frac{3}{5} \right)^3 \lambda^3 \right) - 16 \left(\frac{3}{5} \right)^3 \lambda^3. \end{aligned}$$

Since $\gamma \leq \frac{1}{2}$, we have $\left(\frac{3\gamma-1}{3\gamma+1} \right)^3 = \left(\frac{3\gamma}{3\gamma+1} \right)^3 \lambda^3 \leq \left(\frac{3}{5} \right)^3 \lambda^3$. Thus, it is enough to prove that

$$\left[(1+2\gamma) \left(\frac{3}{5} \right)^3 + 16 \left(\frac{3}{5} \right)^3 \right] \lambda^3 \leq q(1-2\gamma+16\lambda^3).$$

However, $\lambda^3 = O(q)$. Hence, we can relax our goal up to the following target inequality:

$$18 \left(\frac{3}{5} \right)^3 \lambda^3 \leq \frac{q}{12} \left(\stackrel{(5.17)}{\leq} q(1-2\gamma) \right). \quad (5.19)$$

By definition of λ , this is the inequality $18 \left(\frac{9}{55} \right)^3 \leq \frac{1}{12}$, which is valid since $18 \cdot 12 = 6^3$.

Thus, inequality (5.19) is valid and we prove the relation (5.18). $\square$

Now, using $\gamma = \gamma_*$ in inequality (5.4), and denoting $\tau_* = \frac{3}{11}$, we get the following lower bound:

$$\begin{aligned} &f(y) - f(x) - \langle \nabla f(x), y - x \rangle \\ &\geq \frac{3q^{1/3}}{22} \langle \nabla^2 f(x)(y-x), y-x \rangle + \frac{L(1+2\gamma_*)}{24} \|y-x\|^4 \left(\frac{3\gamma_*-1}{3\gamma_*+1} \right)^3 \\ &\geq \frac{3q^{1/3}}{22} \langle \nabla^2 f(x)(y-x), y-x \rangle + \frac{5\mu}{3 \cdot 24} \|y-x\|^4 \left(\frac{\tau_*}{2-\tau_*q^{1/3}} \right)^3 \\ &\geq \frac{3q^{1/3}}{22} \langle \nabla^2 f(x)(y-x), y-x \rangle + \frac{5\mu}{3 \cdot 24} \left(\frac{3}{22} \right)^3 \|y-x\|^4. \end{aligned} \quad (5.20)$$

Relation (5.18) can be also used for justifying the rate of convergence of method (5.10)]. Indeed, if we choose in (5.10) $\gamma = \gamma_*$, then in accordance to Theorems 7 and 9, the rate of convergence (5.12) is described by

$$\alpha = \alpha_* \stackrel{\text{def}}{=} \frac{3\gamma_*-1}{3\gamma_*+1}.$$

Note that

$$\alpha_* \stackrel{(5.17)}{=} \frac{3q^{1/3}}{22-3q^{1/3}}.$$

Thus, in order to get ϵ -solution of problem (2.1), the pure second-order method (5.10) with $\gamma = \gamma_*$ needs

$$\frac{22}{3q^{1/3}} \ln \frac{F(x_0)-F^*}{\epsilon} \quad (5.21)$$

iterations at most. Note that this bound is approximately in four times worse than the bound (4.10) for efficiency of a specialized method (4.4) as applied to the quartic polynomials.

6 Applications in Convex Optimization

Let us show now that the notion of Q-regularity can be used for constructing new and efficient optimization schemes. In this section, we are working with Euclidean norm $\|x\| = \langle Bx, x \rangle^{1/2}$ defined by a self-adjoint linear operator $B \succ 0$.

Consider first the problem (5.1) with bounded fourth derivative of function $f(\cdot)$:

$$0 \leq D^4 f(x)[h]^4 \leq L \|h\|^4, \quad \forall x \in S, h \in \mathbb{E}. \quad (6.1)$$

Since $\mu = 0$, such a problem cannot be solved directly by method (5.10). However, we can treat it by an appropriate regularization technique.

Indeed, let $\epsilon > 0$ be our target accuracy in function value for an approximate solution of problem (5.1). Suppose that we know an upper bound for the distance to the optimal solution:

$$\|x_0 - x^*\| \leq R. \quad (6.2)$$

Then we can define the following regularized version of the objective function:

$$\begin{aligned} F_{x_0, H}(x) &\stackrel{\text{def}}{=} F(x) + \frac{H}{24} \|x - x_0\|^4, \quad x \in \text{dom } \psi, \\ H &= \frac{12\epsilon}{R^4}. \end{aligned} \quad (6.3)$$

Note that this function is Q-regular with the following parameters:

$$\mu_\epsilon = H, \quad L_\epsilon = L + H, \quad q_\epsilon^{-1} = 1 + \frac{LR^4}{12\epsilon}. \quad (6.4)$$

Hence, since $F_{x_0, H}(x_0) = F(x_0)$ and $F_{x_0, H}^* \geq F^*$, inequality (5.21) provides us with the following bound on the number of iterations of method (5.10) needed for finding an $(\epsilon/2)$ -solution of the regularized problem:

$$\frac{22}{3} \left[1 + \frac{LR^4}{12\epsilon} \right]^{1/3} \ln \frac{2(F(x_0) - F^*)}{\epsilon}. \quad (6.5)$$

Let $\bar{x}$ be such a point, and x^* be an optimal solution of problem (5.1). Then

$$\begin{aligned} F(\bar{x}) &\leq F_{x_0, H}(\bar{x}) \leq F_{x_0, H}^* + \frac{\epsilon}{2} \leq F_{x_0, H}(x^*) + \frac{\epsilon}{2} \\ &= F(x^*) + \frac{\epsilon}{2R^4} \|x^* - x_0\|^4 + \frac{\epsilon}{2} \stackrel{(6.2)}{\leq} F^* + \epsilon. \end{aligned}$$

Thus, the point $\bar{x}$ is an ϵ -solution of our initial problem (5.1).

However, the notion of Q-regularity finds much more important applications in the framework of high-order proximal-point methods [20, 3]. In this approach, we solve the general problems of composite minimization in the form

$$F^* = \min_{x \in \text{dom } \psi} \left[F(x) = f(x) + \psi(x) \right], \quad (6.6)$$

where $\psi(\cdot)$ is a simple closed convex function, and the smooth part $f(\cdot)$ of the objective function is convex and has bounded fourth derivative:

$$|D^4 f(x)[h]^4| \leq M_4 \|h\|^4, \quad \forall x \in (\text{dom } \psi)_{1/2}, h \in \mathbb{E}. \quad (6.7)$$

The problem (6.6) can be solved by different third-order proximal-point methods [20, 21], provided that we are able to compute in a reasonable time an approximate solution to the following auxiliary optimization problems:

$$\min_{x \in \text{dom } \psi} \left[F_{\bar{x}, H}(x) = F(x) + \frac{H}{24} \|x - \bar{x}\|^4 \right] \quad (6.8)$$

with different prox-centers $\bar{x} \in \text{dom } \psi$. Note that the objective function of this problem can be represented as follows:

$$\begin{aligned} F_{\bar{x}, H}(x) &= f_{\bar{x}, H}(x) + \psi(x), \\ f_{\bar{x}, H}(x) &= f(x) + \frac{H}{24} \|x - \bar{x}\|^4. \end{aligned}$$

Hence, for H big enough, the smooth part of the objective function in (6.8) is Q -regular.

Lemma 5 *Let convex function $f(\cdot)$ satisfy condition (6.7). Then for $H \geq M_4$, function $f_{\bar{x}, H}(\cdot)$ is Q -regular with parameters $\mu = H - M_4$ and $L = H + M_4$.*

Proof:

Indeed, since the norm $\|\cdot\|$ is Euclidean, for function $d(x) = \frac{1}{24} \|x\|^4$ we have

$$D^4 d(x)[h]^4 = \|h\|^4, \quad \forall x, h \in \mathbb{E}.$$

It remains to use inequality (6.7). □

Thus, for H big enough, problem (6.8) can be efficiently solved by the Quartic Newton Method (5.10). This method has linear rate of convergence in function value. Let us find an upper bound for the number of iterations, which is sufficient for finding an appropriate approximate solution of problem (6.8).

We need two auxiliary results. Let $\Phi(x) = \varphi(x) + \psi(x)$, where $\psi(\cdot)$ is a simple closed convex function, and function $\varphi(\cdot)$ is twice continuously differentiable on $\text{dom } \psi$. Consider the following problem of composite minimization:

$$\Phi_* = \min_{x \in \text{dom } \psi} \Phi(x).$$

Lemma 6 *Let us fix a point $\bar{x} \in \text{dom } \psi$ with $\Phi(\bar{x}) > \Phi_*$. Define the point*

$$T = T(\bar{x}) \stackrel{\text{def}}{=} \arg \min_{x \in \text{dom } \psi} \left[\langle \nabla \varphi(\bar{x}), x \rangle + \frac{M}{2} \|x - \bar{x}\|^2 + \psi(x) \right], \quad (6.9)$$

where $M \geq \hat{M} = \sup_{x \in \mathcal{F}} \|\nabla^2 \varphi(x)\|$ and $\mathcal{F} = \{x \in \text{dom } \psi : \Phi(x) \leq \Phi(\bar{x})\}$. Then

$$\begin{aligned} \Phi'(T) &\stackrel{\text{def}}{=} \nabla \varphi(T) - \nabla \varphi(\bar{x}) - MB(T - \bar{x}) \in \partial \Phi(T), \\ (\|\Phi'(T)\|^*)^2 &\leq 2M[\Phi(\bar{x}) - \Phi(T)]. \end{aligned} \quad (6.10)$$

Proof:

Consider the point $\hat{T} = \arg \min_{x \in \mathcal{F}} \left[\langle \nabla \varphi(\bar{x}), x \rangle + \frac{M}{2} \|x - \bar{x}\|^2 + \psi(x) \right]$. The first-order optimality condition for this problem is as follows:

$$\langle \nabla \varphi(\bar{x}) + MB(\hat{T} - \bar{x}), x - \hat{T} \rangle + \psi(x) \geq \psi(\hat{T}), \quad \forall x \in \mathcal{F}. \quad (6.11)$$

If $\hat{T} = \bar{x}$, then for $x_* \in \text{Arg} \min_{\text{dom} \psi} \Phi(x)$, we have

$$\Phi(x_*) = \varphi(x_*) + \psi(x_*) \geq \varphi(\bar{x}) + \langle \nabla \varphi(\bar{x}), x_* - \bar{x} \rangle + \psi(x_*) \stackrel{(6.11)}{\geq} \Phi(\bar{x}),$$

and this contradicts to the assumptions of the theorem.

Thus, $\hat{T} \neq \bar{x}$. Applying (6.11) to $x = \bar{x}$ and since both $\bar{x}$ and $\hat{T}$ belong to $\mathcal{F}$, we have

$$\begin{aligned} \Phi(\hat{T}) &= \varphi(\hat{T}) + \psi(\hat{T}) \\ &\leq \varphi(\bar{x}) + \langle \nabla \varphi(\bar{x}), \hat{T} - \bar{x} \rangle + \frac{M}{2} \|\hat{T} - \bar{x}\|^2 + \psi(\hat{T}) \\ &\stackrel{(6.11)}{\leq} \Phi(\bar{x}) - \frac{M}{2} \|\hat{T} - \bar{x}\|^2 < \Phi(\bar{x}). \end{aligned}$$

Hence, the functional boundary of the level set $\mathcal{F}$ is not active at $\hat{T}$ and therefore $\hat{T} = T$ as defined by (6.9).

Note that the first-order optimality condition for problem (6.9) is as follows:

$$\langle \nabla \varphi(\bar{x}) + MB(T - \bar{x}), x - T \rangle + \psi(x) \geq \psi(T), \quad \forall x \in \text{dom} \psi. \quad (6.12)$$

This means that $\psi'(T) \stackrel{\text{def}}{=} -\nabla \varphi(\bar{x}) - MB(T - \bar{x}) \in \partial \psi(T)$ and the inclusion

$$\Phi'(T) = \nabla \varphi(T) + \psi'(T) \in \partial \varphi(T)$$

is valid. On the other hand, since both $\bar{x}$ and T belong to $\mathcal{F}$, we conclude that

$$\begin{aligned} (\|\Phi'(T)\|^*)^2 &= (\|\varphi(T) - \nabla \varphi(\bar{x}) - MB(T - \bar{x})\|^*)^2 \\ &= (\|\varphi(T) - \nabla \varphi(\bar{x})\|^*)^2 - 2M \langle \varphi(T) - \nabla \varphi(\bar{x}), T - \bar{x} \rangle + M^2 \|T - \bar{x}\|^2 \\ &\stackrel{(1.2)}{\leq} M^2 \|T - \bar{x}\|^2. \end{aligned}$$

This gives us the bound (6.10). □

We are going to apply Lemma 6 in the situation when function $\varphi(\cdot)$ is Q-regular. Hence, we need to use an explicit bound for its second derivative.

Lemma 7 *Let function $\varphi(\cdot)$ be Q-regular on $\text{dom} \psi$ with constants $L \geq \mu > 0$. Then, for any $x \in \mathcal{F} = \{x \in \text{dom} \psi : \Phi(x) \leq \Phi(\bar{x})\}$ we have*

$$\|\nabla^2 \varphi(x)\| \leq \hat{M} \stackrel{\text{def}}{=} 4\|\nabla^2 \varphi(\bar{x})\| + \frac{2}{3}LD^2, \quad (6.13)$$

where $D \stackrel{\text{def}}{=} 2 \left[\frac{72}{\mu} (\Phi(\bar{x}) - \Phi_*) \right]^{1/4}$.

Proof:

Since $\mu > 0$, function $\Phi(\cdot)$ is uniformly convex. Let us fix $\gamma = \frac{1}{3}$. By inequality (5.4), we get

$$\frac{\mu}{72} \|\bar{x} - x_*\|^4 \leq \Phi(\bar{x}) - \Phi(x_*) \stackrel{\text{def}}{=} \Delta,$$

where $x_* = \arg \min_{x \in \text{dom } \psi} \Phi(x)$. Thus, $\|\bar{x} - x_*\| \leq R \stackrel{\text{def}}{=} \left[\frac{72}{\mu} \Delta \right]^{1/4}$. Note that the same bound on the distance to the optimum is valid for any point from the level set $\mathcal{F}$. Hence, for any $y \in \mathcal{F}$ we have $\|y - \bar{x}\| \leq D = 2R$.

In view of (5.3), for any $y \in \text{dom } \psi$ we have

$$0 \preceq \nabla^2 \varphi(\bar{x} - \gamma(y - \bar{x})) \preceq \nabla^2 \varphi(\bar{x}) - \gamma D^3 \varphi(\bar{x})[y - \bar{x}] + \frac{\gamma^2}{2} L \|y - \bar{x}\|^2 B.$$

Therefore, for any $y \in \mathcal{F}$, we get the following bound:

$$\begin{aligned} \nabla^2 \varphi(y) &\preceq \nabla^2 \varphi(\bar{x}) + D^3 \varphi(\bar{x})[y - \bar{x}] + \frac{1}{2} L \|y - \bar{x}\|^2 B \\ &\preceq \frac{1+\gamma}{\gamma} \nabla^2 \varphi(\bar{x}) + \frac{1+\gamma}{2} L \|y - \bar{x}\|^2 B. \end{aligned}$$

Thus, $\hat{M} = \frac{1+\gamma}{\gamma} \|\nabla^2 \varphi(\bar{x})\| + \frac{1+\gamma}{2} L D^2$ is an upper bound for the norms of Hessians of function $\varphi(\cdot)$ on the set $\mathcal{F}$. $\square$

Let us show now how we can use all this machinery in the framework of high-order proximal-point methods of degree three. Recall that an approximation $\hat{x} \in \text{dom } \psi$ of the exact minimum of problem (6.8) is acceptable if it satisfies the following condition:

$$\exists \hat{g} \in \partial \psi(\hat{x}) : \quad \|\nabla F_{\bar{x}, H}(\hat{x}) + \hat{g}\|^* \leq \beta \|\nabla f(\hat{x}) + \hat{g}\|^*, \quad (6.14)$$

where $\beta \in (0, 1)$ is the tolerance parameter (see [3]; compare with [20]). Note that at the exact solution of problem (6.8) we can enforce the left-hand side of this condition to be zero.

Let us estimate complexity of finding such a point by QRNM (5.10). The whole process consists of two stages, dependent on positive parameteres δ and M .

- **Stage 1.** Suppose that we know the constant M_4 in (6.7). Then, taking $H = 2M_4$, we can ensure Q-regularity of function $f_{\bar{x}, h}(\cdot)$ with parameters

$$\mu = M_4, \quad L = 3M_4, \quad q = \frac{1}{3}$$

(see Lemma 5). Hence, choosing in (5.10) $\gamma \stackrel{(5.17)}{=} \gamma_*$, we get the minimizing sequence convergent with the following rate:

$$F_{\bar{x}, H}(x_k) - \xi_k^* \stackrel{(5.15)}{\leq} (1 - \alpha_*)^k (F(\bar{x}) - F^*), \quad (6.15)$$

where ξ_k^* are computable lower bounds for the optimal value of (6.8) updated in accordance to (5.13). This stage is terminated when $F_{\bar{x}, H}(x_k) - \xi_k^* \leq \delta$.

- **Stage 2.** Compute point $\hat{x} = T(x_k)$ using the iteration (6.9) with $\varphi(\cdot) = f_{\bar{x}, H}(\cdot)$ and an appropriate constant M .

Let us point out the values of parameters δ and M ensuring validity of the condition (6.14). In view of Lemma 7, we can take

$$M = 4\|\nabla^2 f(\bar{x})\| + 2D^2, \quad D = 2 \left[\frac{72}{M_4} (F(\bar{x}) - F^*) \right]^{1/4}.$$

Thus, we come to the following conclusion.

Lemma 8 *Let problem (6.6) satisfy the following assumptions:*

$$\begin{aligned} (a) : \quad & \|\nabla^2 f(x)\| \leq M_2, \quad x \in \text{dom } \psi, \\ (b) : \quad & F(x) - F(y) \leq \Delta_F, \quad x, y \in \text{dom } \psi. \end{aligned} \tag{6.16}$$

Then we can take

$$M = 4M_2 + 48 \left[\frac{2\Delta_F}{M_4} \right]^{1/2}. \tag{6.17}$$

Suppose now that $F(\hat{x}) - F^* \geq \epsilon$. Then in the condition (6.14) we have

$$\|\nabla f(\hat{x}) + \hat{g}\|^* \geq (F(\hat{x}) - F^*)/D_\psi \geq \epsilon/D_\psi,$$

where $D_\psi < +\infty$ is the diameter of $\text{dom } \psi$. Hence, in view of inequality (6.10), in order to satisfy condition (6.14), it is sufficient to ensure

$$2M(F_{\bar{x},H}(x_k) - \xi_k^*) \leq \left(\frac{\beta\epsilon}{D_\psi} \right)^2.$$

Taking into account the rate of convergence (6.15), we get the following upper bound for the number of iterations of method (5.10) for solving the auxiliary problem (6.8):

$$\frac{22}{3^{2/3}} \ln \frac{2M\Delta_F D_\psi^2}{\beta^2 \epsilon^2} \tag{6.18}$$

Having this machinery at hand, we can design different second-order methods for solving the problem (6.6) with the rate of convergence $O(k^{-p})$ with $p = 3, 4, 5$ (see [3, 4]).

References

- [1] N. Agarwal, E. Hazan. Lower Bounds for Higher-Order Convex Optimization. *Proceedings of the 31st Conference On Learning Theory*, PMLR 75:774-792 (2018).
- [2] Arjevani, O. Shamir, and R. Shiff. Oracle Complexity of Second-Order Methods for Smooth Convex Optimization. *Mathematical Programming*, **178**(6), 327 - 360 (2019).
- [3] M. Ahookhosh and Yu. Nesterov. High-order methods beyond the classical complexity bounds, I: inexact high-order proximal-point methods. *arXiv*: 2107.05958v1 (2021).
- [4] M. Ahookhosh and Yu. Nesterov. High-order methods beyond the classical complexity bounds, II: inexact high-order proximal-point methods with segment search. *arXiv*: 2109.12303v1 (2021).

- [5] H.H. Bauschke, J. Bolte, M. Teboulle.. A descent lemma beyond Lipschitz gradient continuity: first order methods revisited and applications. *Math. Oper. Res.* **42**, 330–348 (2016).
- [6] S. Bubeck, Q. Jiang, Y. T. Lee, Y. Li, and A. Sidford. Near-optimal method for highly smooth convex optimization. *COLT*, 492-507 (2019).
- [7] C. Cartis, N. Gould, and Ph. Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Mathematical Programming*, **127**(2), 245-295 (2011).
- [8] C. Cartis, N. Gould, and Ph. Toint. On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming. *SIOPT*, **21**(4), 1721-1739 (2011).
- [9] N. Doikov and, Yu. Nesterov. Contracting proximal methods for smooth convex optimization. *SIAM Journal on Optimization*, **30**(4): 3146–3169 (2020).
- [10] N. Doikov and Yu. Nesterov. Inexact tensor methods with dynamic accuracies. In *International Conference on Machine Learning*, 2577–2586. PMLR (2020).
- [11] N. Doikov and Yu. Nesterov. Affine-invariant contracting-point methods for convex optimization. *Mathematical Programming*, DOI 10.1007/s10107-021-01761-9 (2022).
- [12] A. Gasnikov, E. Gorbunov, D. Kovalev, A. Mohhamed, and E. Chernousova. The global rate of convergence for optimal tensor methods in smooth convex optimization. arXiv: 1809.00382, (2018).
- [13] G.N. Grapiglia and Yu. Nesterov. Regularized Newton methods for minimizing functions with Hölder continuous Hessians. *SIOPT*, **27**(1), 478-506 (2017).
- [14] L.V. Kantorovich. Functional analysis and applied mathematics. *Uspehi Matem. Nauk*, **3**(1), 89–185 (1948), (in Russian). Translated as N.B.S. Report 1509, Washington D.C. (1952)
- [15] H. Lu, R. Freund, and Yu. Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIOPT*, **28**(1), 333-354 (2018).
- [16] Yu. Nesterov. Accelerating the cubic regularization of Newton’s method on convex problems. *Mathematical Programming*, **112**(1) 159-181 (2008)
- [17] Yu. Nesterov. Implementable Tensor Methods in Unconstrained Convex Optimization. *Mathematical Programming*, **186**: 157–183 (2021)
- [18] Yu.Nesterov. Inexact basic tensor methods for some classes of convex optimization problems. *Optimization Methods and Software* DOI 10.1080/10556788.2020.1854252 (2020)
- [19] Yu. Nesterov. Superfast second-order methods for Unconstrained Convex Optimization. *JOTA*, **191**, 1-30 (2021)
- [20] Yu. Nesterov. Inexact accelerated high-order proximal-point methods. *Mathematical Programming*, DOI 10.1007/s10107-021-01727-x (2021).
- [21] Yu. Nesterov. Inexact High-Order Proximal-Point Methods with Auxiliary Search Procedure. *SIOPT*, **31**(4), 2807-2828 (2021). DOI 10.1137/20M134705X.
- [22] Yu. Nesterov, B. Polyak. “Cubic regularization of Newton’s method and its global performance”. *Mathematical Programming*, **108**(1), 177-205 (2006).