Inexact Accelerated High-Order Proximal-Point Methods

Yurii Nesterov

Noname manuscript No.
(will be inserted by the editor)

Inexact Accelerated High-Order Proximal-Point Methods

Yurii Nesterov

Received: June 19, 2020 / Accepted:

Abstract In this paper, we present a new framework of *Bi-Level Unconstrained Minimization* (BLUM) for development of accelerated methods in Convex Programming. These methods use approximations of the high-order proximal points, which are solutions of some auxiliary parametric optimization problems. For computing these points, we can use different methods, and, in particular, the lower-order schemes. This opens a possibility for the latter methods to overpass traditional limits of the Complexity Theory. As an example, we obtain a new second-order method with the convergence rate $O\left(k^{-4}\right)$, where k is the iteration counter. This rate is better than the maximal possible rate of convergence for this type of methods, as applied to functions with Lipschitz continuous Hessian. We also present new methods with the exact auxiliary search procedure, which have the rate of convergence $O\left(k^{-(3p+1)/2}\right)$, where $p \geq 1$ is the order of the proximal operator. The auxiliary problem at each iteration of these schemes is convex.

Keywords Convex Optimization · Tensor methods · Proximal-point operator · Lower complexity bounds · Optimal methods

Mathematics Subject Classification (2010) 90C25

1 Introduction

Motivation. In the last decade, in Convex Optimization we can observe a high activity in the development of the accelerated high-order methods and proving for them the lower complexity bounds. (see [1,2,5,8,11]). At this moment, for methods of any order there exists a natural problem class, for which we

Yurii Nesterov, ORCID 0000-0002-0542-8757, E-mail: Yurii.Nesterov@uclouvain.be.
Center for Operations Research and Econometrics (CORE).
Catholic University of Louvain (UCL), Louvain-la-Neuve, Belgium.

know the accelerated methods. For example, functions with Lipschitz continuous gradients, can be naturally minimized by the first-order schemes, which can demonstrate in this case an unimprovable convergence rate of the order $O\left(k^{-2}\right)$, where k is the iteration counter. For functions with Lipschits continuous Hessians, we can apply the second-order methods with the rate of convergence going up to $O\left(k^{-7/2}\right)$, etc.

This one-to-one correspondence between the type of the methods and the particular problem class allows us to speak about *optimal methods* of certain order, which have unimprovable convergence rate. However, recently in [22] there were presented new results which break down this peaceful picture. It was shown that the special *superfast* second-order methods can converge with the rate $O\left(k^{-4}\right)$, which is faster than the lower bound $O\left(k^{-7/2}\right)$ for these type of schemes. Of course, there is no contradiction with the Complexity Theory. The classical lower bound for the second-order schemes was obtained for functions with Lipschitz continuous Hessian, and in [22] we worked with the functions having Lipschitz continuous third derivative. In any case, this is the first example of successful expansion of the lower-order methods at the territory traditionally reserved for the higher-order schemes. In this paper, we are trying to analyze and explain this phenomena in some general framework.

At each iteration of the superfast methods from [22], we need to solve a serious auxiliary problem requiring additional calls of oracle (the number of these calls is bounded by the logarithm of accuracy). Therefore, in our developments we decided to employ one of the most expensive operations of Convex Optimization, the *proximal-point iteration*.

The proximal approximation of function $f(\cdot)$, defined by

$$\varphi_\lambda(x) = \min_y \left\{ f(y) + \tfrac{1}{2\lambda}\|y - x\|^2 \right\}, \quad \lambda > 0, \tag{1}$$

was introduced by Moreau [14]. Later on, Martinet [13] proposed the first proximal-point method based on this operation. The importance of this construction for computational practice was questionable up to the developments of Rockafellar [24], who used the proximal-point iteration in the dual space for justifying the Augmented Lagrangians. This dual scheme was accelerated by Güller [9], who introduced in this method some elements of the Fast Gradient Method from [15]. Some attempts were made by Teboulle and others [25,10] in studying the proximal-point iteration with non-quadratic non-Euclidean kernel. However, during decades this idea was mainly considered as a theoretical achievement which hardly can be used in the efficient optimization algorithms.

In this paper, we come back to this old idea, having in mind another type of kernel functions. Our goal is the development of accelerated methods for Unconstrained Convex Optimization. Therefore, we suggest to replace $\|y-x\|^2$ in (1) by $\|\cdot\|^{p+1}$, with $p \geq 1$. We call the corresponding proximal step the *pth-order proximal-point operation*. This terminology is justified by two facts.

Firstly, in Section 2, we show that the corresponding simple proximal-point method converges as $O\left(k^{-p}\right)$. The rate of convergence of the accelerated

version of this method is $O\left(k^{-(p+1)}\right)$. In both cases, we can use appropriate approximations of the proximal point.

Secondly, in Section 3, we show that this approximation can be computed by one step of the pth order tensor method provided that the pth derivative of the objective function is Lipschitz continuous.

Our main results are presented in Section 4. In this section, we introduce the framework of *Bi-Level Unconstrained Minimization* (BLUM), which can be used for development of new and efficient optimization methods. In this framework, we can choose independently the order of the upper-level proximal-point tensor method and the lower-level method for computing an appropriate approximation to the proximal point. It appears that this strategy opens a possibility for the lower-order methods to *overpass the limits* given by the traditional Complexity Theory. Note that a similar phenomena was already observed in the framework of non-convex optimization [7]. However, for Convex Optimization this situation is new.

For supporting this claim, we analyze efficiency of the *second-order method* in approximating the *third-order* proximal point. Using the relative smoothness condition [4,12], we develop a very efficient second-order method for computing this approximation. The global rate of convergence of our upper-level method is $O\left(k^{-4}\right)$, and the complexity of the lower-level scheme depends logarithmically on the accuracy parameters. The new second-order method can be applied to functions with Lipschitz continuous third derivative.

In the next Section 5, we introduce even more powerful operation, the proximal-point iteration *with line search*. As compared with (1) it has one more variable in the auxiliary convex minimization problem. We prove that under assumption of the exact search, the corresponding accelerated method converges as $O\left(k^{-(3p+1)/2}\right)$. Our approach has the same near-optimal complexity bound as [6]. However, its search procedures are based on convex auxiliary problems and therefore they are easier to implement.

In the last Section 6, we discuss our results and directions for future developments.

Notation and generalities. In what follows, we denote by $\mathbb{E}$ a finite-dimensional real vector space, and by $\mathbb{E}^*$ its dual space composed by linear functions on $\mathbb{E}$. For such a function $s \in \mathbb{E}^*$, we denote by $\langle s, x \rangle$ its value at $x \in \mathbb{E}$.

Let us measure distances in $\mathbb{E}$ and $\mathbb{E}^*$ in a Euclidean norm. For that, using a self-adjoint positive-definite operator $B : \mathbb{E} \to \mathbb{E}^*$ (notation $B = B^* \succ 0$), we define

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

Sometimes, it will be convenient to treat $x \in \mathbb{E}$ as a linear operator from $\mathbb{R}$ to $\mathbb{E}$, and x^* as a linear operator from $\mathbb{E}^*$ to $\mathbb{R}$. In this case, xx^* is a linear operator from $\mathbb{E}^*$ to $\mathbb{E}$, acting as follows:

$$(xx^*)g = \langle g, x \rangle x \in \mathbb{E}, \quad g \in \mathbb{E}^*.$$

For a smooth function $f : \mathbb{E} \to \mathbb{R}$ denote by $\nabla f(x)$ its gradient, and by $\nabla^2 f(x)$ its Hessian evaluated at point $x \in \mathbb{E}$. Note that

$$\nabla f(x) \in \mathbb{E}^*, \quad \nabla^2 f(x)h \ \in \ \mathbb{E}^*, \quad x, h \in \mathbb{E}.$$

Using the above norm, we can define the standard Euclidean prox-functions

$$d_{p+1}(x) = \tfrac{1}{p+1}\|x\|^{p+1}, \quad x \in \mathbb{E},$$

where $p \geq 1$ is an integer parameter. These functions have the following derivatives:

$$\begin{aligned} \nabla d_{p+1}(x) &= \|x\|^{p-1}Bx, \quad x \in \mathbb{E}, \\ \nabla^2 d_{p+1}(x) &= \|x\|^{p-1}B + (p-1)\|x\|^{p-3}Bxx^*B \ \succeq \ \|x\|^{p-1}B. \end{aligned} \tag{2}$$

Note that function $d_{p+1}(\cdot)$ is uniformly convex (see, for example, Lemma 4.2.3 in [21]):

$$d_{p+1}(y) \geq d_{p+1}(x) + \langle \nabla d_{p+1}(x), y - x\rangle + \tfrac{1}{p+1}\left(\tfrac{1}{2}\right)^{p-1}\|y - x\|^{p+1}, \quad x, y \in \mathbb{E}. \tag{3}$$

In what follows, we often work with directional derivatives. For $p \geq 1$, denote by

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

the directional derivative of function f at x along directions $h_i \in \mathbb{E}$, $i = 1, \ldots, p$. Note that $D^p f(x)[\cdot]$ is a *symmetric p-linear form*. Its *norm* is defined in a standard way:

$$\|D^p f(x)\| = \max_{h_1, \ldots, h_p} \left\{ \left| D^p f(x)[h_1, \ldots, h_p] \right| : \ \|h_i\| \leq 1, \, i = 1, \ldots, p \right\}. \tag{4}$$

If all directions $h_1, \ldots, h_p$ are the same, we apply notation

$$D^p f(x)[h]^p, \quad h \in \mathbb{E}.$$

Note that, in general, we have (see, for example, Appendix 1 in [16])

$$\|D^p f(x)\| = \max_{h} \left\{ \left| D^p f(x)[h]^p \right| : \ \|h\| \leq 1 \right\}. \tag{5}$$

In this paper, we work with functions from the problem classes $\mathcal{F}_p$, which are convex and p times continuously differentiable on $\mathbb{E}$. Denote by $M_p(f)$ the uniform upper bound for the pth derivative:

$$M_p(f) = \sup_{x \in \mathbb{E}} \|D^p f(x)\|. \tag{6}$$

One of our main results is based on the following relation between the second, third and fourth derivatives of convex function (see Lemma 3 in [19]):

$$D^3 f(\bar{x})[h] \preceq \tfrac{1}{\xi}\nabla^2 f(\bar{x}) + \tfrac{\xi}{2}M_4(f)\|h\|^2 B, \quad \bar{x}, \ h \in \mathbb{E}, \tag{7}$$

where ξ is an arbitrary positive number.

2 Inexact high-order proximal-point steps

Consider the following optimization problem:

$$\min_{x \in \mathbb{E}} f(x), \tag{8}$$

where $f(\cdot)$ is a differentiable closed convex function. Denote by x^* one of its optimal solutions and let $f^* = f(x^*)$.

All methods presented in this paper are based on the *pth-order proximal-point operators*, defined as follows:

$$\mathrm{prox}^p_{f/H}(\bar{x}) = \arg\min_{x \in \mathbb{E}} \Big\{ f^p_{\bar{x},H}(x) \stackrel{\mathrm{def}}{=} f(x) + H d_{p+1}(x - \bar{x}) \Big\}, \tag{9}$$

where $H > 0$ and $p \geq 1$. The properties of the standard *first-order* proximal-point operator

$$\mathrm{prox}_{f/H}(\bar{x}) = \arg\min_{x \in \mathbb{E}} \Big\{ f(x) + \tfrac{H}{2}\|x - \bar{x}\|^2 \Big\}$$

are studied very well in the literature (e.g. [23]). However, we will see that the high-order proximal-point methods converge much faster. The main goal of this paper is to establish the global rate of convergence of these methods in accelerated and non-accelerated forms and complement this information by the complexity of computing an appropriate *inexact* proximal-point step (9).

Indeed, very often, the proximal-point operator (9) cannot be computed in a closed form. Instead, we have to use an approximate solution of this problem obtained by an auxiliary optimization procedure. Let us introduce the set of acceptable solutions to problem (9), that is

$$\mathcal{A}^p_H(\bar{x}, \beta) = \Big\{ x \in \mathbb{E} : \ \|\nabla f^p_{\bar{x},H}(x)\|_* \leq \beta \|\nabla f(x)\|_* \Big\}, \tag{10}$$

where $\beta \in [0, 1)$ is a tolerance parameter. Note that $\mathrm{prox}^p_{f/H}(\bar{x}) \in \mathcal{A}^p_H(\bar{x}, \beta)$. However, since $\nabla f^p_{\bar{x},H}(\bar{x}) = \nabla f(\bar{x})$, we see that $\bar{x} \notin \mathcal{A}^p_H(\bar{x}, \beta)$ unless $\bar{x} = x^*$.

Lemma 1 *Let $T \in \mathcal{A}^p_H(\bar{x}, \beta)$. Then*

$$(1 - \beta)\|\nabla f(T)\|_* \leq H\|T - \bar{x}\|^p \ \leq \ (1 + \beta)\|\nabla f(T)\|_*, \tag{11}$$

$$\langle \nabla f(T), \bar{x} - T \rangle \geq \tfrac{H}{1+\beta}\|T - \bar{x}\|^{p+1}. \tag{12}$$

Moreover, if $\beta \leq \frac{1}{p}$, then

$$\langle \nabla f(T), \bar{x} - T \rangle \geq \left[\tfrac{1-\beta}{H}\right]^{\frac{1}{p}} \|\nabla f(T)\|_*^{\frac{p+1}{p}}. \tag{13}$$

Proof Indeed, inequality (11) follows from representation

$$\nabla f^p_{\bar{x},H}(T) \stackrel{(2)}{=} \nabla f(T) + H\|T-\bar{x}\|^{p-1}B(T-\bar{x}). \tag{14}$$

Further, denote $r = \|T-\bar{x}\|$. Then, squaring both parts in inequality (10), we have

$$\|\nabla f(T)\|_*^2 + 2Hr^{p-1}\langle \nabla f(T), T-\bar{x}\rangle + H^2 r^{2p} \stackrel{(14)}{\leq} \beta^2 \|\nabla f(T)\|_*^2.$$

This inequality can be rewritten as follows:

$$\begin{aligned} \langle \nabla f(T), \bar{x}-T\rangle \;&\geq\; \kappa(r) \stackrel{\rm def}{=} \tfrac{1-\beta^2}{2Hr^{p-1}}\|\nabla f(T)\|_*^2 + \tfrac{H}{2}r^{p+1} \\ &\stackrel{(11)}{\geq} \tfrac{1-\beta^2}{2Hr^{p-1}} \cdot \tfrac{H^2 r^{2p}}{(1+\beta)^2} + \tfrac{H}{2}r^{p+1} \;=\; \tfrac{Hr^{p+1}}{1+\beta}, \end{aligned} \tag{15}$$

and this is inequality (12). Let us compute the derivative of $\kappa(\cdot)$:

$$\kappa'(\tau) = -(p-1)\tfrac{1-\beta^2}{2H\tau^p}\|\nabla f(T)\|_*^2 + (p+1)\tfrac{H}{2}\tau^p, \quad \tau > 0.$$

Note that $r \stackrel{(11)}{\geq} \hat{r} \stackrel{\rm def}{=} \left[\tfrac{1-\beta}{H}\|\nabla f(T)\|_*\right]^{\frac{1}{p}}$. Since

$$\begin{aligned} \kappa'(\hat{r}) &= -(p-1)\tfrac{(1-\beta^2)H}{2H(1-\beta)}\|\nabla f(T)\|_* + (p+1)\tfrac{H}{2}\cdot\tfrac{1-\beta}{H}\|\nabla f(T)\|_* \\ &= \|\nabla f(T)\|_*\left[\tfrac{1-\beta}{2}(p+1) - \tfrac{1+\beta}{2}(p-1)\right] \;=\; \|\nabla f(T)\|_*\,[1-\beta p] \;\geq\; 0, \end{aligned}$$

we have $\langle \nabla f(T), \bar{x}-T\rangle \geq \kappa(r) \geq \kappa(\hat{r}) + \kappa'(\hat{r})(r-\hat{r}) \geq \kappa(\hat{r})$, and this is inequality (13). □

The following corollary is a trivial consequence of convexity of $f(\cdot)$ and inequality (13).

Corollary 1 *Let $T \in \mathcal{A}^p_H(\bar{x},\beta)$ and $\beta \leq \frac{1}{p}$. Then*

$$f(\bar{x}) - f(T) \geq \left[\tfrac{1-\beta}{H}\right]^{\frac{1}{p}} \|\nabla f(T)\|_*^{\frac{p+1}{p}}. \tag{16}$$

Let us justify now the rate of convergence of the basic inexact high-order proximal-point method:

$$\boxed{x_{k+1} \in \mathcal{A}^p_H(x_k,\beta), \quad k \geq 0.} \tag{17}$$

For analyzing this scheme, we need the following Lemma A.1 from [20].

Lemma 2 *Let the sequence of positive numbers $\{\xi_k\}_{k\geq 0}$ satisfy the following condition:*

$$\xi_k - \xi_{k+1} \geq \xi_{k+1}^{1+\alpha}, \quad k \geq 0, \tag{18}$$

where $\alpha \in (0,1]$. Then for any $k \geq 0$ we have

$$\xi_k \leq \tfrac{\xi_0}{\left(1+\frac{\alpha k}{1+\alpha}\ln(1+\xi_0^\alpha)\right)^{1/\alpha}} \;\leq\; \left[\left(1+\tfrac{1}{\alpha}\right)(1+\xi_0^\alpha)\cdot\tfrac{1}{k}\right]^{\frac{1}{\alpha}}. \tag{19}$$

Denote by $D_0 = \max_{x \in \mathbb{E}} \{\|x - x^*\| : \; f(x) \leq f(x_0)\}$ the radius of the initial level set of the objective function in problem (8).

Theorem 1 *Let the sequence $\{x_k\}_{k \geq 0}$ be generated by method (17). Then for any $k \geq 0$ we have*

$$f(x_k) - f^* \leq \tfrac{1}{2} \left(\tfrac{1}{1-\beta} H D_0^{p+1} + f(x_0) - f^* \right) \cdot \left(\tfrac{2p+2}{k} \right)^p, \quad k \geq 1. \tag{20}$$

Proof Indeed, in view of Corollary 1, for any $k \geq 0$ we have

$$f(x_k) - f(x_{k+1}) \geq \left[\tfrac{1-\beta}{H} \right]^{\frac{1}{p}} \|\nabla f(x_{k+1})\|_*^{\frac{p+1}{p}} \; \geq \; \left[\tfrac{1-\beta}{H} \right]^{\frac{1}{p}} \left(\tfrac{f(x_{k+1}) - f^*}{D_0} \right)^{\frac{p+1}{p}}.$$

Denoting now $\xi_k = \frac{1-\beta}{H D_0^{p+1}} (f(x_k) - f^*)$ and $\alpha = \frac{1}{p}$, we get the condition (18) valid for all $k \geq 0$. Hence, in view of Lemma 2, we have

$$\xi_k \leq \left[\left(1 + \tfrac{1}{\alpha}\right) \left(1 + \xi_0^\alpha\right) \cdot \tfrac{1}{k} \right]^{\frac{1}{\alpha}} \; \leq \; \left(1 + \tfrac{1}{\alpha}\right)^{\frac{1}{\alpha}} 2^{\frac{1-\alpha}{\alpha}} (1 + \xi_0) \cdot \left(\tfrac{1}{k}\right)^{\frac{1}{\alpha}}.$$

And this is inequality (20). □

Note that the rate of convergence (20) of method (17) does not depend on the properties of function $f(\cdot)$. This means that the actual complexity of problem (8) for this method is reflected somehow in the complexity of finding the point $x_{k+1} \in \mathcal{A}_H^p(x_k, \beta)$. We will discuss this issue in the remaining part of this paper. To conclude this section, let us present an accelerated variant of the Inexact Proximal-Point Method.

Our presentation is very similar to the justification of Accelerated Tensor Methods in Section 2 in [22]. Therefore we omit some technical details. Denote

$$c(p) = \left[\tfrac{1-\beta}{H} \right]^{\frac{1}{p}}. \tag{21}$$

And let us choose $\beta \in [0, \frac{1}{p}]$. Then, for any $T \in \mathcal{A}_H^p(\bar{x}, \beta)$, we have

$$f(\bar{x}) - f(T) \overset{(16)}{\geq} c(p) \|\nabla f(T)\|_*^{\frac{p+1}{p}}. \tag{22}$$

Define now the sequence of scaling coefficients

$$\begin{aligned} A_k \; &= \; \left(\tfrac{1}{2} c(p)\right)^p \left(\tfrac{k}{p+1} \right)^{p+1} \overset{(21)}{=} \tfrac{2(1-\beta)}{H} \left(\tfrac{k}{2p+2} \right)^{p+1}, \\ a_{k+1} &\overset{\text{def}}{=} A_{k+1} - A_k, \quad k \geq 0. \end{aligned} \tag{23}$$

Note that $k^{p+1} \geq (k+1)^{p+1} + (p+1)(k+1)^p \cdot (-1)$. This inequality can be rewritten in the following form:

$$a_{k+1}^{\frac{p+1}{p}} \leq \tfrac{1}{2} c(p) A_{k+1}, \quad k \geq 0. \tag{24}$$

Consider the following high-order proximal method.

Inexact Accelerated pth-Order Proximal-Point Method

Initialization. Choose $x_0 \in \mathbb{E}$, $\beta \in [0, \frac{1}{p}]$, and $H > 0$. Define coefficients A_k by (23) and function $\psi_0(x) = d_{p+1}(x - x_0)$.

Iteration $k \geq 0$**.**

1. Compute $v_k = \arg\min_{x \in \mathbb{E}} \psi_k(x)$ and choose $y_k = \frac{A_k}{A_{k+1}} x_k + \frac{a_{k+1}}{A_{k+1}} v_k$.

2. Compute $T_k \in \mathcal{A}_H^p(y_k, \beta)$ and update

$$\psi_{k+1}(x) = \psi_k(x) + a_{k+1}[f(T_k) + \langle \nabla f(T_k), x - T_k \rangle].$$

3. Choose x_{k+1} with $f(x_{k+1}) \leq f(T_k)$.

(25)

Note that computation of point v_k at Step 1 can be done in a closed form.

Theorem 2 *Let sequence $\{x_k\}_{k \geq 0}$ be generated by method (25). Then, for any $k \geq 1$, we have*

$$f(x_k) - f^* \leq \frac{H}{2(p+1)(1-\beta)} \left(\frac{2p+2}{k}\right)^{p+1} \|x_0 - x^*\|^{p+1}. \tag{26}$$

Moreover, $\|v_k - x^*\|^{p+1} \leq 2^{p-1} \|x_0 - x^*\|^{p+1}$.

Proof First of all, note that by induction it is easy to see that

$$\psi_k(x) \leq A_k f(x) + d_{p+1}(x - x_0), \quad x \in \mathbb{E}. \tag{27}$$

In particular, for $\psi_k^* \stackrel{\text{def}}{=} \min_{x \in \mathbb{E}} \psi_k(x)$ and all $x \in \mathbb{E}$, we have

$$A_k f(x) + d_{p+1}(x - x_0) \stackrel{(27)}{\geq} \psi_k(x) \stackrel{(3)}{\geq} \psi_k^* + \frac{1}{p+1} \left(\frac{1}{2}\right)^{p-1} \|x - v_k\|^{p+1}. \tag{28}$$

Let us prove by induction the following relation:

$$\psi_k^* \geq A_k f(x_k), \quad k \geq 0. \tag{29}$$

For $k = 0$, we have $\psi_0^* = 0$ and $A_0 = 0$. Hence, (29) is valid. Assume it is valid for some $k \geq 0$. Then

$$\psi_{k+1}^* = \min_{x \in \mathbb{E}} \Big\{ \psi_k(x) + a_{k+1}[f(T_k) + \langle \nabla f(T_k), x - T_k \rangle] \Big\}$$

$$\stackrel{(28)}{\geq} \min_{x \in \mathbb{E}} \Big\{ \psi_k^* + \frac{1}{p+1} \left(\frac{1}{2}\right)^{p-1} \|x - v_k\|^{p+1} + a_{k+1}[f(T_k) + \langle \nabla f(T_k), x - T_k \rangle] \Big\}.$$

Note that

$$\begin{aligned}
&\psi_k^* + a_{k+1}[f(T_k) + \langle \nabla f(T_k), x - T_k\rangle] \\
&\overset{(29)}{\geq} A_k f(x_k) + a_{k+1}[f(T_k) + \langle \nabla f(T_k), x - T_k\rangle] \\
&\geq A_{k+1} f(T_k) + \langle \nabla f(T_k), a_{k+1}(x - T_k) + A_k(x_k - T_k)\rangle \\
&= A_{k+1} f(T_k) + \langle \nabla f(T_k), a_{k+1}(x - v_k) + A_{k+1}(y_k - T_k)\rangle,
\end{aligned}$$

where the last equality follows from the relation $A_k x_k = A_{k+1} y_k - a_{k+1} v_k$.

Further, for all $x \in \mathbb{E}$ we have (see, for example, Lemma 2 in [18])

$$\begin{aligned}
&\tfrac{1}{p+1}\left(\tfrac{1}{2}\right)^{p-1} \|x - v_k\|^{p+1} + a_{k+1}\langle \nabla f(T_k), x - v_k\rangle \\
&\geq -\tfrac{p}{p+1} 2^{\frac{p-1}{p}} \left(a_{k+1}\|\nabla f(T_k)\|_*\right)^{\frac{p+1}{p}}.
\end{aligned}$$

Finally, since $T_k \in \mathcal{A}_H^p(y_k, \beta)$, we get

$$\langle \nabla f(T_k), y_k - T_k\rangle \overset{(12)}{\geq} c(p)\|\nabla f(T_k)\|_*^{\frac{p+1}{p}}.$$

Putting all these inequalities together, we obtain

$$\begin{aligned}
\psi_{k+1}^* &\geq A_{k+1} f(T_k) - \tfrac{p}{p+1} 2^{\frac{p-1}{p}} \left(a_{k+1}\|\nabla f(T_k)\|_*\right)^{\frac{p+1}{p}} + A_{k+1} c(p)\|\nabla f(T_k)\|_*^{\frac{p+1}{p}} \\
&= A_{k+1} f(T_k) + \|\nabla f(T_k)\|_*^{\frac{p+1}{p}} \left(A_{k+1} c(p) - \tfrac{p}{p+1} 2^{\frac{p-1}{p}} a_{k+1}^{\frac{p+1}{p}}\right) \\
&\geq A_{k+1} f(T_k) + \|\nabla f(T_k)\|_*^{\frac{p+1}{p}} \left(A_{k+1} c(p) - 2 a_{k+1}^{\frac{p+1}{p}}\right) \\
&\overset{(24)}{\geq} A_{k+1} f(T_k) \geq A_{k+1} f(x_{k+1}).
\end{aligned}$$

It remains to note that in view of relations (27) and (29), we have

$$f(x_k) - f^* \leq \tfrac{1}{A_k} d_{p+1}(x^* - x_0) \overset{(23)}{=} \tfrac{H}{2(1-\beta)} \left(\tfrac{2p+2}{k}\right)^{p+1} \cdot \tfrac{1}{p+1}\|x^* - x_0\|^{p+1}.$$

In order to get the remaining bound for v_k, we need to apply inequalities (28) and (29) with $x = x^*$. □

We can see that method (25) is much faster than the basic method (17). Its rate of convergence is also independent on the properties of the objective function. Hence, in order to evaluate its actual performance, we need to investigate the complexity of finding a point $T \in \mathcal{A}_H^p(\bar{x}, \beta)$. This will be done in the remaining sections of the paper.

3 Approximating proximal-point operator by tensor step

Let us show that with appropriate values of parameters, the inclusion $T \in \mathcal{A}_H^p(\bar{x}, \beta)$ can be ensured by a single step of the Basic Tensor Method of degree p. Firstly, recall some simple facts from the theory of tensor methods.

For function $f(\cdot)$, let us assume that $M_{p+1}(f) < +\infty$. Define its *Taylor approximation* at point $x \in \mathbb{E}$:

$$\Omega_{x,p}(y) = f(x) + \sum_{k=1}^{p} \tfrac{1}{k!} D^k f(x)[y-x]^k.$$

Then

$$\|\nabla f(y) - \nabla \Omega_{x,p}(y)\|_* \le \tfrac{M_{p+1}(f)}{p!} \|y-x\|^p, \quad y \in \mathbb{E}. \tag{30}$$

Define now the *augmented Taylor polynomial*

$$\hat{\Omega}_{x,p,M}(y) = \Omega_{x,p}(y) + \tfrac{M}{(p+1)!} \|y-x\|^{p+1}.$$

If $M \ge M_{p+1}(f)$ then this function provides us with a uniform upper bound for the objective. Moreover, if $M \ge pM_{p+1}(f)$, then function $\hat{\Omega}_{x,p,M}(\cdot)$ is convex (see Theorem 1 in [19]). Therefore, we are able to compute the tensor step

$$T_{p,M}(x) = \arg\min_{y \in \mathbb{E}} \hat{\Omega}_{x,p,M}(y).$$

Let us allow some inexactness in computation of this point. Namely, we assume that we can compute a point T satisfying the following condition:

$$\|\nabla \hat{\Omega}_{x,p,M}(T)\|_* \le \tfrac{\gamma}{1+\gamma} \|\nabla \Omega_{x,p}(T)\|_*, \tag{31}$$

where $\gamma \in [0, \frac{\beta}{1+\beta})$ is the tolerance parameter. Thus,

$$\tfrac{\gamma}{1+\gamma} \|\nabla \Omega_{x,p}(T)\|_* \ge \|\nabla \hat{\Omega}_{x,p,M}(T)\|_* \ \ge \ \|\nabla \Omega_{x,p}(T)\|_* - \tfrac{M}{p!} \|T-x\|^p.$$

Therefore, $\|\nabla \Omega_{x,p}(T)\|_* \le (1+\gamma) \frac{M}{p!} \|T-x\|^p$. (This inequality was used as a termination criterion in [5].)

Let us prove the following simple result.

Lemma 3 *Let $M > \frac{1}{1-\gamma} M_{p+1}(f)$. Then for point T satisfying (31), we have*

$$\|\nabla f(T) + \tfrac{M}{p!} \nabla d_{p+1}(T-x)\|_* \le \tfrac{M_{p+1}(f) + \gamma M}{(1-\gamma)M - M_{p+1}(f)} \|\nabla f(T)\|_*. \tag{32}$$

Proof Denote $r = \|T-x\|$. Then

$$\begin{aligned}
\tfrac{M_{p+1}(f) r^p}{p!} &\overset{(30)}{\ge} \|\nabla f(T) - \nabla \Omega_{x,p}(T)\|_* \\
&= \|\nabla f(T) - \nabla \hat{\Omega}_{x,p,M}(T) + \tfrac{M}{p!} \nabla d_{p+1}(T-x)\|_* \\
&\overset{(31)}{\ge} \|\nabla f(T) + \tfrac{M}{p!} \nabla d_{p+1}(T-x)\|_* - \tfrac{\gamma}{1+\gamma} \|\nabla \Omega_{x,p}(T)\|_* \\
&\ge \|\nabla f(T) + \tfrac{M}{p!} \nabla d_{p+1}(T-x)\|_* - \gamma \tfrac{M r^p}{p!}.
\end{aligned}$$

Thus,

$$\frac{M_{p+1}(f)+\gamma M}{p!} r^p \;\geq\; \|\nabla f(T) + \tfrac{M}{p!}\nabla d_{p+1}(T-x)\|_*$$

$$\geq \tfrac{M}{p!}\|\nabla d_{p+1}(T-x)\|_* - \|\nabla f(T)\|_* \overset{(2)}{=} \tfrac{Mr^p}{p!} - \|\nabla f(T)\|_*.$$

Thus, $\frac{r^p}{p!} \leq \frac{1}{(1-\gamma)M - M_{p+1}(f)}\|\nabla f(T)\|_*$, and we obtain (32). □

Let us define now $M = \frac{1+\beta}{\beta(1-\gamma)-\gamma} M_{p+1}(f)$ and $H = \frac{M}{p!}$. Then the inequality (32) can be rewritten as follows:

$$\|\nabla f(T) + H\nabla d_{p+1}(T-x)\|_* \leq \beta\|\nabla f(T)\|_*.$$

In other words, these values of parameters ensure the inclusion $T \in \mathcal{A}_H^p(x,\beta)$. For such M, the convexity condition $M \geq pM_{p+1}(f)$ is satisfied with $\beta \leq \frac{1}{p-1}$.

Thus, the accelerated and non-accelerated tensor methods from [19] can be seen as particular implementations of inexact high-order proximal-point methods. Their efficiency bounds can be obtained by Theorems 1 and 2.

4 Bi-level unconstrained minimization

In solving problem (8) by inexact high-order proximal-point methods from Section 2, we have two degrees of freedom. Firstly, we need to decide on the order p of the proximal-point method. This defines the rate of convergence for the upper-level process. Note that for obtaining the rates (20) or (26), we *do not need* any assumption on the properties of the objective function.

After that, we have to choose the lower-level method for computing a point

$$T \in \mathcal{A}_H^p(x,\beta). \tag{33}$$

For analyzing efficiency of the latter method, we do need to assume something on the objective function. Thus, the overall complexity of this bi-level scheme depends on efficiency bounds of both processes. Note that the objective function in the auxiliary problem (9) has some structure (composite form, uniform convexity), which can help to increase efficiency of the lower-level method.

We call this framework *Bi-Level Unconstrained Minimization* (BLUM). Let us show that it opens new horizons in the development of very efficient optimization methods.

Indeed, as we have seen in Section 3, the condition (33) can be satisfied by one step of tensor method. This strategy does not require additional calls of the oracle. However, the high-order tensor methods need computations of the high-order derivatives and therefore quite often they are impractical. In this case, it is reasonable to solve the auxiliary problem in (9) by a cheaper method, based on the derivatives of a smaller degree than the order of the underlying proximal-point scheme.

In this section, we present an example when this strategy works very well. We are going to consider a third-order proximal-point method, which is implemented by a second-order scheme. The first confirmation that this is possible was obtained in [22], using the approximations of third derivative along two vectors by the finite differences of gradients. In the remaining part of this section, we discuss a simpler approach, based on a direct application of the relative non-degeneracy condition [4,12] to the auxiliary problem (9).

Let us consider the following unconstrained minimization problem:

$$\min_{y \in \mathbb{E}} \left\{ f_{\bar{x},H}(y) \stackrel{\text{def}}{=} f(y) + H d_4(y - \bar{x}) \right\}, \tag{34}$$

with constant $H > 0$ and central point $\bar{x} \in \mathbb{E}$. As compared with notation (9), we drop the index p since in this section we always have $p = 3$. In what follows, we assume that the fourth derivative of function $f(\cdot)$ is bounded on $\mathbb{E}$ by constant M_4.

Our main tool for solving the problem (34) is the gradient method based on *relative non-degeneracy condition*. This condition is formulated in terms of the *Bregman distance*. Recall that this is a (non-symmetric) distance function between two points x and y from $\mathbb{E}$, which is computed with respect to some convex scaling function $\rho(\cdot)$. It is defined as follows:

$$\beta_\rho(x, y) = \rho(y) - \rho(x) - \langle \nabla \rho(x), y - x \rangle, \quad x, y \in \mathbb{E}. \tag{35}$$

We say that the function $\varphi(\cdot)$ is *relatively non-degenerate* on $\mathbb{E}$ with respect to the scaling function $\rho(\cdot)$ if there exist two constants $0 < \mu \leq L$ such that

$$\mu \beta_\rho(x, y) \leq \beta_\varphi(x, y) \ \leq \ L \beta_\rho(x, y), \quad x, y \in \mathbb{E}. \tag{36}$$

The value $\varkappa = \frac{\mu}{L}$ is called the *condition number* of function $\varphi(\cdot)$ with respect to $\rho(\cdot)$. Recall that there exists a convenient sufficient condition for relations (36), this is

$$\mu \nabla^2 \rho(x) \preceq \nabla^2 \varphi(x) \ \preceq \ L \nabla^2 \rho(x), \quad x \in \mathbb{E}. \tag{37}$$

It appears that for function $f_{\bar{x},H}(\cdot)$ in the problem (34), we can point out a simple scaling function, ensuring validity of the condition (36) with a good value of $\varkappa$.

Theorem 3 *Let $H \geq M_4(f)$. Then, the scaling function*

$$\rho_{\bar{x},H}(x) = \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(x - \bar{x}), x - \bar{x} \rangle + H d_4(x - \bar{x}), \tag{38}$$

and function $f_{\bar{x},H}(\cdot)$ satisfy the condition (37) on $\mathbb{E}$ with constants

$$\mu = 1 - \tfrac{1}{\xi}, \quad L \ = \ 1 + \tfrac{1}{\xi}, \quad \varkappa = \tfrac{\xi - 1}{\xi + 1}, \tag{39}$$

where $\xi \geq 1$ is the unique solution of the following quadratic equation:

$$\xi(1 + \xi) = \tfrac{2H}{M_4(f)}. \tag{40}$$

Proof For the sake of notation, denote $M_4 = M_4(f)$ and assume that $\bar{x} = 0 \in \mathbb{E}$. Then, for any $x \in \mathbb{E}$, we have

$$\begin{aligned}
\nabla^2 f(x) &= \nabla^2 f(0) + D^3 f(0)[x] + \int_0^1 (1-\tau) D^4 f(\tau x)[x]^2 d\tau \\
&\overset{(4)}{\preceq} \nabla^2 f(0) + D^3 f(0)[x] + \tfrac{1}{2} M_4 \|x\|^2 B \\
&\overset{(7)}{\preceq} \left(1 + \tfrac{1}{\xi}\right) \nabla^2 f(0) + \tfrac{1}{2} M_4 \|x\|^2 (1+\xi) B \\
&\overset{(2)}{\preceq} \left(1 + \tfrac{1}{\xi}\right) \nabla^2 f(0) + \tfrac{1}{2} M_4 (1+\xi) \nabla^2 d_4(x).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\nabla^2 f_{0,H}(x) &\overset{(40)}{\preceq} \left(1 + \tfrac{1}{\xi}\right) \nabla^2 f(0) + \left[\tfrac{\xi(1+\xi)}{2} M_4 + \tfrac{1}{2} M_4 (1+\xi)\right] \nabla^2 d_4(x) \\
&= \left(1 + \tfrac{1}{\xi}\right) \nabla^2 \rho_{0,\xi}(x).
\end{aligned}$$

Similarly, using again (4), we have

$$\begin{aligned}
\nabla^2 f(x) &\overset{(7)}{\succeq} \left(1 - \tfrac{1}{\xi}\right) \nabla^2 f(0) - \tfrac{1}{2} M_4 \|x\|^2 (1+\xi) B \\
&\overset{(2)}{\succeq} \left(1 - \tfrac{1}{\xi}\right) \nabla^2 f(0) - \tfrac{1}{2} M_4 (1+\xi) \nabla^2 d_4(x).
\end{aligned}$$

Hence,

$$\begin{aligned}
\nabla^2 f_{0,H}(x) &\overset{(40)}{\succeq} \left(1 - \tfrac{1}{\xi}\right) \nabla^2 f(0) + \left[\tfrac{\xi(1+\xi)}{2} M_4 - \tfrac{1}{2} M_4 (1+\xi)\right] \nabla^2 d_4(x) \\
&= \left(1 - \tfrac{1}{\xi}\right) \nabla^2 \rho_{0,\xi}(x). \qquad \square
\end{aligned}$$

From now on, we fix the following values for our parameters:

$$\xi = 2, \quad H = 3M_4(f), \quad \mu = \tfrac{1}{2}, \quad L = \tfrac{3}{2}, \quad \varkappa = \tfrac{1}{3}. \tag{41}$$

Note that these values satisfy relations (39) and (40). Consequently, we can use a simpler notation for the corresponding scaling function:

$$\rho_{\bar{x}}(y) \overset{\text{def}}{=} \tfrac{1}{2} \langle \nabla^2 f(\bar{x})(y - \bar{x}), y - \bar{x} \rangle + 3M_4(f) d_4(y - \bar{x}). \tag{42}$$

Let us present now an optimization method for solving efficiently the problem (34). For our goals, the most appropriate variant of this method can be

found in [20].

Choose $\bar{x} \in \mathbb{E}$ **and** $H = 3M_4(f)$**. Set** $x_0 = \bar{x}$**.**

For $k \geq 0$**, iterate:**

$$x_{k+1} = \arg \min_{x \in \mathrm{dom}\, \mathbb{E}} \Bigl\{ \langle \nabla f_{\bar{x},H}(x_k), x - x_k \rangle + L \beta_{\rho_{\bar{x}}}(x_k, x) \Bigr\}. \tag{43}$$

Note that this is a *first-order* method for solving the problem (34) provided that the Hessian $\nabla^2 f(x_k)$ is represented in an appropriate basis (this can be done before the iterations start). It forms a sequence of points $\{x_k\}_{k \geq 0}$ with monotonically decreasing values of the objective function.

Applying now Lemma 3 in [20], we come to the following result.

Lemma 4 *Let sequence $\{x_k\}_{k \geq 0}$ be generated by method (43). Then, for any $k \geq 1$ and any $x \in \mathrm{dom}\, \psi$ we have*

$$\beta_{\rho_{\bar{x}}}(x_k, x) \leq (1 - \varkappa)^k \, \beta_{\rho_{\bar{x}}}(x_0, x) + \tfrac{1}{L} [f_{\bar{x},H}(x) - f_{\bar{x},H}(x_k)]. \tag{44}$$

Let us show how this method can be used on the lower level of the proximal-point method (25) with $p = 3$. Our optimization problem (8) is characterized by the following parameters:

$$\begin{aligned} M_4(f) \; < \; +\infty, \quad R_0 \; = \; \|x_0 - x^*\|, \quad M_2(f) \; < \; +\infty, \\ D_0 \; = \; \max_{x \in \mathbb{E}} \{ \|x - x^*\| : \; f(x) \leq f(x_0) \} \; < \; +\infty. \end{aligned} \tag{45}$$

For our analysis, parameters $M_4(f)$ and R_0 are critical. The remaining parameters $M_2(f)$ and D_0 appear in the efficiency bounds only inside the logarithms. Using the constant $M_2(f)$, we can bound the variation of the objective function as follows:

$$f(y) - f^* \leq \tfrac{1}{2} M_2(f) \|y - x^*\|^2, \quad y \in \mathbb{E}. \tag{46}$$

Let us write down the full version of the combination of method (25) with (43). We choose $\beta = \frac{1}{p} = \frac{1}{3}$ and other parameters by (41).

Define the following sequences:

$$A_k = \tfrac{4}{9 M_4(f)} \left(\tfrac{k}{8} \right)^4, \quad a_{k+1} \; = \; A_{k+1} - A_k, \quad k \geq 0. \tag{47}$$

Inexact Accelerated 3rd-Order Proximal-Point Method

Initialization. Choose $x_0 \in \mathbb{E}$. Define function $\psi_0(x) = d_4(x - x_0)$.

Iteration $k \geq 0$.

1. Compute $v_k = \arg\min_{x \in \mathbb{E}} \psi_k(x)$ and choose $y_k = \frac{A_k}{A_{k+1}} x_k + \frac{a_{k+1}}{A_{k+1}} v_k$.

2. Compute $x_{k,i_k^*} \in \mathcal{A}^3_{3M_4(f)}\left(y_k, \frac{1}{3}\right)$ by the following procedure.

- Define functions $\varphi_k(x) = f(x) + 3M_4(f) d_4(x - y_k)$

  and $\rho_k(x) = \frac{1}{2}\langle \nabla^2 f(y_k)(x - y_k), x - y_k\rangle + 3M_4(f) d_4(x - y_k)$.

- Set $x_{k,0} = y_k$. For $i \geq 0$, iterate

  $$x_{k,i+1} = \arg\min_{x \in \mathbb{E}} \left\{ \langle \nabla \varphi_k(x_{k,i}), x - x_{k,i}\rangle + \frac{3}{2}\beta_{\rho_k}(x_{k,i}, x) \right\}$$

  up to the first iteration i_k^* with $\|\nabla \varphi_k(x_{k,i_k^*})\|_* \leq \frac{1}{3}\|\nabla f(x_{k,i_k^*})\|_*$.

3. Update $\psi_{k+1}(x) = \psi_k(x) + a_{k+1}[f(x_{k,i_k^*}) + \langle \nabla f(x_{k,i_k^*}), x - x_{k,i_k^*}\rangle]$.

4. Define $x_{k+1} = \arg\min_x \left\{ f(x) : \; x \in \{x_k, x_{k,i_k^*}\} \right\}$.

(48)

The major difference of this method from the earlier tensor methods [19,22] consists in the necessity to call oracle of the objective function at each iteration of the internal loop.

Clearly, this is a second-order method, which implements the inexact third-order proximal-point method (25). Let us assume for a moment, that at each upper-level iteration of this scheme, the numbers i_k^* are well defined. Then by Theorem 2, we get the following rate of convergence:

$$f(x_k) - f^* \leq 9M_4(f)\left(\tfrac{4}{k}\right)^4 R_0^4, \quad k \geq 1. \tag{49}$$

Thus, it remains to find an upper bound for the numbers i_k^*. For that, we need to get an upper bound for the size of points $x_{k,i}$. We will do this under the following assumption:

$$f(x_{k,i}) - f^* \geq \epsilon, \quad 0 \leq i \leq i_k^*, \; k \geq 0, \tag{50}$$

where $\epsilon > 0$ is the desired accuracy of the solution to problem (8). Note that we need this assumption only for estimating the number of steps, which are necessary to violate it.

Assume that at some iteration $k \geq 0$ the points x_k and v_k are well defined. Since $f(x_k) \leq f(x_0)$, in view of Theorem 2 we have

$$\|y_k - x^*\| \;\leq\; \max\{\|x_k - x^*\|, \|v_k - x^*\|\} \;\leq\; \max\left\{D_0, \sqrt{2}R_0\right\}.$$

At the same time, since $\varphi_k(x_{k,i}) \leq \varphi_k(x_{i,0}) = f(y_k)$, we get

$$\tfrac{3}{4}M_4(f)\|x_{k,i} - y_k\|^4 \leq f(y_k) - f(x_{k,i}) \overset{(46)}{\leq} \tfrac{1}{2}M_2(f)D_0^2.$$

Therefore, $\|x_{k,i} - y_k\| \leq D_1 \overset{\text{def}}{=} \left[\frac{2M_2(f)}{3M_4(f)}D_0^2\right]^{\frac{1}{4}}$ and

$$\|x_{k,i} - x^*\| \leq D_2 \overset{\text{def}}{=} D_1 + \max\left\{D_0, \sqrt{2}R_0\right\}.$$

Hence,

$$\|\nabla f(x_{k,i})\|_* \geq \tfrac{f(x_{k,i}) - f^*}{D_2} \overset{(50)}{\geq} \tfrac{\epsilon}{D_2}. \tag{51}$$

However, in view of Lemma 4, $\varphi_k(x_{k,i}) \to \min\limits_{x \in \mathbb{E}} \varphi_k(x)$ as $i \to \infty$. This implies

$$\|\nabla \varphi_k(x_{k,i})\|_* \to 0,$$

ensuring that the auxiliary minimization process at iteration k is finite and x_{k+1} and v_{k+1} are well defined. Let us estimate its length.

In view of Lemma 3.2 in [22], for all $u \in \mathbb{E}$ with $\|u\| \leq D$, we have

$$\beta_{d_4}(u, v) \leq \tfrac{5}{2}D^2\|v - u\|^2 + \tfrac{1}{2}\|v - u\|^4, \quad v \in \mathbb{E}.$$

Therefore,

$$\begin{aligned}
\beta_{\rho_k}(x, y) \;&\leq\; \tfrac{1}{2}M_2(f)\|y - x\|^2 + 3M_4(f)\beta_{d_4}(x - y_k, y - y_k) \\
&\leq\; \tfrac{1}{2}M_2(f)\|y - x\|^2 + 3M_4(f)\left[\tfrac{5}{2}D_1^2\|y - x\|^2 + \tfrac{1}{2}\|y - x\|^4\right] \\
&=\; \tfrac{1}{2}\left[M_2(f) + 15M_4(f)D_1^2\right]\|y - x\|^2 + \tfrac{3}{2}M_4(f)\|y - x\|^4 \\
&\overset{\text{def}}{=} \theta(\|y - x\|).
\end{aligned}$$

for all x with $\|x - y_k\| \leq D_1$ and $y \in \mathbb{E}$. At the same time, in view of Lemma 3.3 in [22], the last bound and Theorem 3 imply

$$L\theta_*(\tfrac{1}{L}\|\nabla \varphi_k(x_{k,i})\|_*) \leq \varphi_k(x_{k,i}) - \varphi_k(x_k^*).$$

Hence,

$$\begin{aligned}
L\theta_*(\tfrac{1}{L}\|\nabla \varphi_k(x_{k,i})\|_*) \leq \varphi_k(x_{k,i}) - \varphi_k(x_k^*) &\overset{(44)}{\leq} L\left(\tfrac{2}{3}\right)^i \beta_{\rho_k}(y_k, x_k^*) \\
&\leq L\left(\tfrac{2}{3}\right)^i \left[\tfrac{1}{2}M_2(f)\|x_k^* - y_k\|^2 + \tfrac{3}{4}M_4\|x_k^* - y_k\|^4\right]
\end{aligned} \tag{52}$$

where $x_k^* = \arg\min_{x \in \mathbb{E}} \varphi_k(x)$ and $\theta_*(\lambda) = \max_\tau [\lambda\tau - \theta(\tau)]$. Since $\|x_k^* - y_k\| \le D_1$, we can get an upper bound for i_k^* from the following inequality:

$$\left(\tfrac{2}{3}\right)^i \left[\tfrac{1}{2} M_2(f) D_1^2 + \tfrac{3}{4} M_4 D_1^4\right] \overset{(51)}{\le} \theta_* \left(\tfrac{\epsilon}{L D_2}\right).$$

Using Lemma 7 in [22], we can estimate for function $\theta(\tau) = \frac{a}{2}\tau^2 + \frac{b}{4}\tau^4$ its dual function as follows:

$$\theta_*(\lambda) \ge \tfrac{\lambda^2}{2[a + b^{1/3}\lambda^{2/3}]}.$$

In our case, $a = M_2(f) + 15 M_4(f) D_1^2$ and $b = 6 M_4(f)$. Therefore,

$$\theta_*(\lambda) \ge \tfrac{\lambda^2}{2[M_2(f) + 15 M_4(f) D_1^2] + [6 M_4(f)]^{1/3}\lambda^{2/3}}.$$

Thus, we can see that all values i_k^* are bounded by $O(\ln \frac{1}{\epsilon})$. A similar reasoning shows that the length of the last iteration, stopped at the moment when the condition (50) be violated, is also bounded by $O(\ln \frac{1}{\epsilon})$. Hence, we have proved the following theorem.

Theorem 4 *The second-order method (48) finds an ϵ-solution of problem (8) in*

$$4 \left(\tfrac{9 M_4(f)}{\epsilon}\right)^{\frac{1}{4}} R_0$$

iterations. At each iteration, it calls the second-order oracle once and the first-order oracle $O\left(\ln \frac{1}{\epsilon}\right)$ times at most.

Let us discuss now the implementation details of method (48). At each inner iteration of this scheme, it is necessary to solve an auxiliary optimization problem for finding the point $x_{k,i+1} \in \mathbb{R}^n$. For doing this efficiently, it is reasonable to start with computation of the tri-diagonal factorization of matrix $\nabla^2 f(y_k)$:

$$\nabla^2 f(y_k) = U_k \Lambda_k U_k^T,$$

where $U_k \in \mathbb{R}^{n \times n}$ is an orthogonal matrix and $\Lambda_k \in \mathbb{R}^{n \times n}$ is a symmetric tri-diagonal matrix. Then we can change variables:

$$x = y_k + U_k w, \quad w \in \mathbb{R}^n,$$

and minimize the function $\hat{\varphi}_k(w) = \varphi_k(y_k + U_k w)$. The advantage of this formulation is that in the new variables the scaling function $\rho_k(\cdot)$ becomes very simple:

$$\rho_k(x) = \hat{\rho}_k(w) \;=\; \tfrac{1}{2}\langle \Lambda_k w, w\rangle + \tfrac{3}{4} M_4(f) \|w\|_{(2)}^4,$$

where $\|\cdot\|_{(2)}$ is the standard Euclidean norm in $\mathbb{R}^n$. Thus, the computation of the new point $w_{k,i+1} = U_k^T (x_{k,i+1} - y_k)$ can be done in a linear time. Therefore, the total complexity of each iteration of the inner method will be quadratic in n (plus one computation of the first-order oracle).

Note that in the method (48) we have a possibility of computing the lower bounds for the optimal value of the objective function, provided that we have an upper bound for the distance to the minimum:

$$\|x_0 - x^*\| \le R.$$

Then, for $k \ge 1$, we can compute the value

$$\ell_k^* = \frac{1}{A_k} \min_x \left\{ \sum_{j=0}^{k-1} a_{j+1} [f(x_{j,i_j^*}) + \langle \nabla f(x_{j,i_j^*}), x - x_{j,i_j^*} \rangle] : \ \|x - x_0\| \le R \right\} \le f(x^*),$$

and use it in the termination criterion. Note that with this value the inequality (49) is valid if we replace f^* by ℓ_k^* and R_0 by R. If the bound R is not known, we can update the initial guess dynamically using the observed distance between x_k and x_0.

5 High-order proximal-point methods with line search

In this section, we consider new methods for solving the problem (8), which are based on pth-order proximal-point operator *with line search* ($p \ge 1$). It is defined as follows:

$$\mathrm{prox}_{f/H}^p(\bar{x}, u) = \arg\min_{\substack{x \in \mathbb{E}, \\ \tau \in \mathbb{R}}} \left\{ f(x) + H d_{p+1}(x - \bar{x} - \tau u) \right\} \ \in \ \mathbb{E} \times \mathbb{R}, \tag{53}$$

where the point $\bar{x}$ and direction u belong to $\mathbb{E}$ and the proximal coefficient H is positive. Note that the value of this operator is a solution of a convex optimization problem. As compared with operation (9), we increased the dimension of the search variable by one. Hence, it should not create a significant additional complexity. In this paper, we will analyze only the exact computation in (53).

Let us mention the main properties of operator (53).

Lemma 5 *Let $(T, \tau) = prox_{f/H}^p(\bar{x}, u)$. Denote $y = \bar{x} + \tau u$ and $r = \|T - y\|$. Then*

$$\nabla f(T) + H r^{p-1} B(T - y) = 0, \tag{54}$$

$$\|\nabla f(T)\|_* = H r^p, \quad \langle \nabla f(T), y - T \rangle \ = \ H r^{p+1}, \tag{55}$$

$$\langle B(T - y), u \rangle = 0, \quad \langle \nabla f(T), u \rangle \ = \ 0. \tag{56}$$

Moreover, $f(\bar{x}) - f(T) \ge \frac{H}{p+1} r^{p+1}$.

Proof Equation (54) and the first equation in (56) are the first-order optimality conditions for the objective function in the problem (53). The first equation in (55) follows from (54), and we get the second one by multiplying (54) by $y - T$. Second equation in (56) follows from the first one in view of (54).

Finally, for proving the remaining inequality, we choose in the optimization problem in (53) $x = \bar{x}$ and $\tau = 0$. □

Clearly, the smaller H is, the better is the result of (53). However, the small values of H make this computation more difficult. Thus, a reasonable choice of H must be dictated by the problem class and the auxiliary methods, which will be used for solving (53) approximately. We keep the detailed analysis of different possibilities for the future research.

Consider the following optimization scheme.

pth-Order Proximal-Point Method With Line Search

Initialization. Choose $x_0 \in \mathbb{E}$, $H > 0$, and $\psi_0(x) = \frac{1}{2}\|x - x_0\|^2$.

Iteration $k \geq 0$**.**

1. Compute $v_k = \arg\min_{x \in \mathbb{E}} \psi_k(x)$.

2. Compute $(x_{k+1}, \tau_k) = \mathrm{prox}^p_{f/H}(x_k, v_k - x_k)$.

3. Define $y_k = x_k + \tau_k(v_k - x_k)$ and $r_k = \|x_{k+1} - y_k\|$.

4. Define a_{k+1} by equation $\frac{a^2_{k+1}}{A_{k+1}} = \frac{1}{H r_k^{p-1}}$ with $A_{k+1} = A_k + a_{k+1}$.

5. Set $\psi_{k+1}(x) = \psi_k(x) + a_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1}\rangle]$.

(57)

Let $B_0 = 0$ and denote $B_k = \frac{H}{2} \sum\limits_{i=1}^{k} A_i r_{i-1}^{p+1}$ for $k \geq 1$. Let us prove the following result.

Lemma 6 *Let the sequence $\{x_k\}_{k \geq 0}$ be generated by method (57). Then for all $k \geq 0$ and $x \in \mathbb{E}$ we have*

$$A_k f(x_k) + B_k \leq \psi_k^* = \min_{x \in \mathbb{E}} \psi_k(x). \tag{58}$$

Proof Let us prove this relation by induction. For $k = 0$, we have $A_0 = 0$, $B_0 = 0$, and $\psi_0(x) = \frac{1}{2}\|x - x_0\|$. Thus, in this case inequality (58) is valid.

Let us assume that it is valid for some $k \geq 0$. Then

$$\begin{aligned}
\psi_{k+1}^* &= \min_{x \in \mathbb{E}} \Bigl\{ \psi_k(x) + a_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1} \rangle] \Bigr\} \\
&\geq \min_{x \in \mathbb{E}} \Bigl\{ \psi_k^* + \tfrac{1}{2}\|x - v_k\|^2 + a_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1} \rangle] \Bigr\} \\
&\geq \min_{x \in \mathbb{E}} \Bigl\{ A_k f(x_k) + B_k + \tfrac{1}{2}\|x - v_k\|^2 \\
&\qquad + a_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1} \rangle] \Bigr\} \\
&\geq \min_{x \in \mathbb{E}} \Bigl\{ A_{k+1} f(x_{k+1}) + B_k + \tfrac{1}{2}\|x - v_k\|^2 \\
&\qquad + \langle \nabla f(x_{k+1}), a_{k+1}(x - x_{k+1}) + A_k(x_k - x_{k+1}) \rangle \Bigr\} \\
&\overset{(56)}{=} \min_{x \in \mathbb{E}} \Bigl\{ A_{k+1} f(x_{k+1}) + B_k + \tfrac{1}{2}\|x - v_k\|^2 \\
&\qquad + \langle \nabla f(x_{k+1}), a_{k+1}(x - v_k) + A_{k+1}(y_k - x_{k+1}) \rangle \Bigr\} \\
&= A_{k+1}\, f(x_{k+1}) + B_k - \tfrac{1}{2} a_{k+1}^2 \|\nabla f(x_{k+1})\|_*^2 \\
&\qquad + A_{k+1} \langle \nabla f(x_{k+1}), y_k - x_{k+1} \rangle \\
&\overset{(55)}{=} A_{k+1}\, f(x_{k+1}) + B_k - \tfrac{1}{2} a_{k+1}^2 H^2 r_k^{2p} + A_{k+1} H r_k^{p+1} \\
&= A_{k+1}\, f(x_{k+1}) + B_k + \tfrac{1}{2} A_{k+1} H r_k^{p+1} \;=\; A_{k+1} f(x_{k+1}) + B_{k+1}. \quad \square
\end{aligned}$$

Let us prove now the main result of this section. In the proof, we closely follow the arguments justifying Lemma 4.3.5 in [21].

Theorem 5 *For any $k \geq 1$, we have*

$$f(x_k) - f^* \leq \frac{2^p H R_0^{p+1}}{\bigl(1 + \frac{2(k-1)}{p+1}\bigr)^{\frac{3p+1}{2}}}. \tag{59}$$

Proof Note that

$$\sqrt{A_{k+1}} - \sqrt{A_k} = \frac{a_{k+1}}{\sqrt{A_{k+1}} + \sqrt{A_k}} \;=\; \frac{\sqrt{A_{k+1}}}{(\sqrt{A_{k+1}} + \sqrt{A_k}) H^{1/2} r_k^{\frac{p-1}{2}}} \;\geq\; \frac{1}{2\sqrt{H r_k^{p-1}}}.$$

Thus, denoting $\xi_k = 2\sqrt{H r_k^{p-1}}$, we get $A_k \geq \Bigl(\sum\limits_{i=0}^{k-1} \frac{1}{\xi_i} \Bigr)^2$. For $p = 1$ this proves that $A_k \geq \frac{k^2}{4H}$. So, let us assume that $p > 1$.

On the other hand, in view of Lemma 6, we have

$$A_k f(x_k) + B_k \le \psi_k^* \;\le\; f(x^*) + \tfrac{1}{2}R_0^2,$$

where $R_0 = \|x_0 - x^*\|$. Therefore,

$$\tfrac{1}{2}R_0^2 \ge \tfrac{H}{2}\sum_{i=0}^{k-1} A_{i+1} r_i^{p+1} \;=\; \tfrac{H}{2}\sum_{i=0}^{k-1} A_{i+1}\left[\tfrac{\xi_i^2}{4H}\right]^{\frac{p+1}{p-1}}.$$

In other words, we have the following bound:

$$\sum_{i=0}^{k-1} A_{i+1}\xi_i^{\frac{2(p+1)}{p-1}} \le D \;\stackrel{\text{def}}{=}\; (2^{p+1}H)^{\frac{2}{p-1}}R_0^2. \tag{60}$$

We need to minimize now the sum $\sum\limits_{i=0}^{k-1}\frac{1}{\xi_i}$ subject to this bound. Since the bound is active, we can introduce for it a Lagrange multiplier $\lambda > 0$ and find the optimal ξ_i from the equation

$$\tfrac{\lambda}{\xi_i^2} = A_{i+1}\xi_i^{\frac{p+3}{p-1}}, \quad i = 0, \dots, k-1.$$

Thus, $\xi_i = \left[\frac{\lambda}{A_{i+1}}\right]^{\frac{p-1}{3p+1}}$. Substituting these values in the constraint (60), we get equation for optimal λ:

$$D = \sum_{i=0}^{k-1} A_{i+1}\left[\tfrac{\lambda}{A_{i+1}}\right]^{\frac{2(p+1)}{3p+1}} \;=\; \lambda^{\frac{2(p+1)}{3p+1}}\sum_{i=0}^{k-1} A_{i+1}^{\frac{p-1}{3p+1}}.$$

Therefore,

$$\sum_{i=0}^{k-1}\tfrac{1}{\xi_i} \ge \left(\tfrac{1}{\lambda}\right)^{\frac{p-1}{3p+1}}\sum_{i=0}^{k-1} A_{i+1}^{\frac{p-1}{3p+1}} \;=\; \left(\tfrac{1}{D}\right)^{\frac{p-1}{2(p+1)}}\left(\sum_{i=0}^{k-1} A_{i+1}^{\frac{p-1}{3p+1}}\right)^{\frac{3p+1}{2(p+1)}}.$$

This means that we have proved the following inequality:

$$A_k \ge \left(\tfrac{1}{D}\right)^{\frac{p-1}{p+1}}\left(\sum_{i=0}^{k-1} A_{i+1}^{\frac{p-1}{3p+1}}\right)^{\frac{3p+1}{p+1}}. \tag{61}$$

Denote $C_k = \left(\sum\limits_{i=1}^{k} A_i^{\frac{p-1}{3p+1}}\right)^{\frac{2}{p+1}}$ and $\theta = \left(\frac{1}{D}\right)^{\frac{p-1}{p+1}}$. Then inequality (61) can be written as follows:

$$C_{k+1}^{\frac{p+1}{2}} - C_k^{\frac{p+1}{2}} = A_{k+1}^{\frac{p-1}{3p+1}} \;\ge\; \theta^{\frac{p-1}{3p+1}}\left(\sum_{i=1}^{k+1} A_i^{\frac{p-1}{3p+1}}\right)^{\frac{p-1}{p+1}} \;=\; \theta^{\frac{p-1}{3p+1}}C_{k+1}^{\frac{p-1}{2}}.$$

Denoting $\gamma = \theta^{\frac{p-1}{3p+1}}$, we see that $C_1 \ge \gamma$. Moreover, since $\tau^{\frac{p+1}{2}}$ with $\tau \ge 0$ is a convex function, we have $\tau_+^{\frac{p+1}{2}} - \tau^{\frac{p+1}{2}} \le \frac{p+1}{2}\tau_+^{\frac{p-1}{2}}(\tau_+ - \tau)$. Therefore,

$$\gamma C_{k+1}^{\frac{p-1}{2}} \le C_{k+1}^{\frac{p+1}{2}} - C_k^{\frac{p+1}{2}} \;\le\; \tfrac{p+1}{2}C_{k+1}^{\frac{p-1}{2}}(C_{k+1} - C_k).$$

Consequently, $C_{k+1} \geq C_k + \frac{2\gamma}{p+1}$ and we conclude that $C_k \geq \gamma + \frac{2\gamma(k-1)}{p+1}$, $k \geq 1$. Substituting this bound in (61), we get

$$A_k \geq \theta C_k^{\frac{3p+1}{2}} \geq \theta \left(\gamma + \frac{2\gamma(k-1)}{p+1}\right)^{\frac{3p+1}{2}} = \left(\frac{1}{D}\right)^{\frac{p-1}{2}} \left(1 + \frac{2(k-1)}{p+1}\right)^{\frac{3p+1}{2}}.$$

It remains to note that in view of inequality (58), we have

$$f(x_k) - f^* \leq \frac{1}{2A_k} R_0^2 \leq \frac{D^{\frac{p-1}{2}} R_0^2}{2\left(1+\frac{2(k-1)}{p+1}\right)^{\frac{3p+1}{2}}} = \frac{2^p H R_0^{p+1}}{\left(1+\frac{2(k-1)}{p+1}\right)^{\frac{3p+1}{2}}}. \qquad \square$$

6 Conclusion

In this paper we present a new framework BLUM, where the development of the accelerated minimization scheme consists of two steps. Firstly, we choose the order of the proximal-point iteration. At this moment, we are not restricted by any properties of the objective function except its differentiability (which can be dropped) and convexity.

These properties play a crucial role at the second step, where we decide on the type of the scheme we use for approximating the proximal point. The overall complexity of the method can be computed then as the product of the number of steps of the upper-level process and the estimate for the number of steps in the lower level process.

In this way, we managed to justify a second-order scheme, which uses only the second-order information for an approximate computation of the third-order proximal point. It is interesting that the overall complexity bound of our method is essentially the same as the bound for the number of iterations of the accelerated third-order methods [3,19]. At the same time, these bounds overpass the limits for the maximal efficiency established for the second-order methods by functions with bounded third derivative (see [1,2] and Section 4.3.1 in [21]).

It is interesting to understand why this improvement was possible. One of the reasons is that in this paper we are working with a subclass of functions with Lipschitz continuous third derivative. Indeed, in view of Lemma 4 in [22],

$$M_3(f) \leq \sqrt{2M_2(f)M_4(f)}.$$

Therefore, the lower bounds of [1,2] are not valid anymore.

Another question is if it is possible to improve the rate of convergence of the lower-order methods in the framework of BLUM. We are not ready to give now a comprehensive answer to this question. However, let us look at the worst-case functions, which justify the lower complexity bound for the tensor methods. In accordance to [19], for pth-order methods they have the form

$$f_p(x) = |x^{(1)}|^{p+1} + \sum_{i=1}^{n-1} |x^{(i+1)} - x^{(i)}|^{p+1}, \quad x \in \mathbb{R}^n.$$

This function justifies the maximal rate of convergence $O(k^{-2})$ for the first order methods (e.g. [21]).

Note that, for $p = 1$ this is a quadratic function. Hence, all its derivatives $D^s f_1(\cdot)$ of the order s with $s \geq 3$ are equal to zero. Therefore, we cannot expect for the first-order methods any improvements from any assumptions on the boundedness of these derivatives.

The situation with the second-order methods is different. Their worst-case function $f_2(\cdot)$ has *discontinuous* third derivative. Hence, the corresponding lower bound $O(k^{-7/2})$ may be not valid if we assume the existence and boundedness of the forth derivative. And the results of Section 4 show that this is indeed the case. Our second order method (48) has the rate of convergence $O(k^{-4})$, and the results of Section 5 gives us a hope that there exist the second-order methods with the rate of convergence $O(k^{-5})$ (this is the maximal rate of convergence for the third-order methods).

Another consequence of the above observation is that we cannot speak anymore about the *pth-order optimal methods*. Instead, we should switch to speaking about the optimal methods for different *problem classes*. Indeed, we have seen that for the same problem class we can have methods of different order, which have the same rate of convergence. In this situation, it is natural to agree that the lower-order method is better. This means, that our complexity scale, instead being one-dimensional, should be two-dimensional at least. But of course all these questions need further investigations.

We hope that our results create interesting directions of research related to further increase of the efficiency of the lower-order methods as applied to the problem classes which were traditionally out of their scope.

Declarations

Acknowledgments. The author would like to thank two anonymous referees for the careful reading of the text and valuable suggestions.

Funding. 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 supported by Multidisciplinary Institute in Artificial intelligence MIAI@Grenoble Alpes (ANR-19-P3IA-0003).

Conflict of interests. None.

Availability of data and material. None.

Code availability. Not applicable.

References

1. N. Agarwal, E. Hazan. Lower Bounds for Higher-Order Convex Optimization. arXiv: 1710.10329v1 [math.OC] (2017).

2. Arjevani, O. Shamir, and R. Shiff. Oracle Complexity of Second-Order Methods for Smooth Convex Optimization. arXiv:1705.07260 [math.OC] (2017).
3. M. Baes. Estimate sequence methods: extensions and approximations. *Optimization Online* (2009).
4. 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).
5. E. G. Birgin, J. L. Gardenghi, J. M. Martinez, S. A. Santos, and Ph. L. Toint. Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models. *Mathematical Programming*, **163**, 359-368 (2017).
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. Y. Carmon, J. C. Duchi, O. Hinder, and A. Sidford. Accelerated methods for nonconvex optimization. *SIAM Journal on Optimization*, **28**(2), 1751-1772 (2018).
8. 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).
9. O. Güller. New proximal point algorithms for convex minimization. *SIAM Journal on Control and Optimization*, **14**(5), 877-898 (1992)
10. A.N. Iusem, B.F. Svaiter, and M. Teboulle. Entropy-like proximal methods in Convex Programming. *Math. oper. res.* **19**(4) 790-814 (1994).
11. B. Jiang, H. Wang, and S. Zang. An optimal high-order tensor method fo convex optimization. arXiv: 1812.06557, (2018).
12. H. Lu, R. Freund, and Yu. Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIOPT*, **28**(1), 333-354 (2018).
13. B. Martinet. Perturbation des methods d'Optimization. Application., R.A.I.R.O. *Numer. Anal.* (1978).
14. J.J. Moreau. Proximité et Dualité dans un Espace Hilbertien. *Bull. Soc. Math. France* **93** 273-299 (1965).
15. Yu. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $O(\frac{1}{k^2})$. *Doklady AN SSSR* (translated as Soviet Math. Docl.), **269**(3), 543-547 (1983).
16. Yu. Nesterov, A. Nemirovskii. *Interior point polynomial methods in convex programming: Theory and Applications,*, SIAM, Philadelphia, 1994.
17. Yu. Nesterov, B. Polyak. "Cubic regularization of Newton's method and its global performance". *Mathematical Programming*, **108**(1), 177-205 (2006).
18. Yu. Nesterov. Accelerating the cubic regularization of Newton's method on convex problems. *Mathematical Programming*, **112**(1) 159-181 (2008)
19. Yu. Nesterov. Implementable Tensot Methods in Unconstrained Convex Optimization. *Mathematical Programming*, DOI 10.1007/s10107-019-01449-1 (2019)
20. Yu. Nesterov. Inexact Basic Tensor Methods. *CORE DP* #2019/23, (2019),
21. Yu. Nesterov. Lectures on Convex Optimization. *Springer* (2018).
22. Yu. Nesterov. Superfast second-order methods for Unconstraine Convex Optimization. CORE DP, #2020/7, (2020)
23. N. Parikh and S. Boyd. Proximal Algorithms. *Foundations and Trends in Optimization*, **1**(3), 123–231 (2013).
24. R.T. Rockafellar. Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming. *Math. Oper. Res.* **1**, 97-116 (1976).
25. M. Teboulle. Entropic proximal mapping with applications to Nonlinear Programming. *Math. Oper. Res.* **17**, 670-690 (1992)