Proximal iterative Gaussian smoothing algorithm for a class of nonsmooth convex minimization problems

Sanming Liu
Zhijie Wang
Chongyang Liu
PROXIMAL ITERATIVE GAUSSIAN SMOOTHING ALGORITHM FOR A CLASS OF NONSMOOTH CONVEX MINIMIZATION PROBLEMS

SANMING LIU
Department of Mathematics and Physics, Shanghai Dianji University
1350 Ganlan Road, Shanghai, 201306, China

ZHIJIE WANG
School of Electrical Engineering, Shanghai Dianji University
1350 Ganlan Road, Shanghai, 201306, China

CHONGYANG LIU
School of Mathematics and Information Science
Shandong Institute of Business and Technology
191 Binhaizhong Road, Yantan, Shandong Province, 264005, China

Abstract. In this paper, we consider the problem of minimizing a convex objective which is the sum of three parts: a smooth part, a simple non-smooth Lipschitz part, and a simple non-smooth non-Lipschitz part. A novel optimization algorithm is proposed for solving this problem. By making use of the Gaussian smoothing function of the functions occurring in the objective, we smooth the second part to a convex and differentiable function with Lipschitz continuous gradient by using both variable and constant smoothing parameters. The resulting problem is solved via an accelerated proximal-gradient method and this allows us to recover approximately the optimal solutions to the initial optimization problem with a rate of convergence of order $O\left(\frac{\ln k}{k}\right)$ for variable smoothing and of order $O\left(\frac{1}{k}\right)$ for constant smoothing.

1. Introduction. In this paper, we are interested in solving a specific class of unconstrained convex optimization problems in finite dimensional spaces. Generally, when characterizing optimality, the convexity allows to make use of powerful results in convex analysis, separation theorems and the (Fenchel) conjugate theory here included (see [1], [9], [12]). When considering an unconstrained convex and differentiable minimization problem, there are already plenty of promising methods available (such as the steepest descent method, Newton’s method or, in an appropriate setting, fast gradient methods, see [3], [6]). However, a lot of situations occur when the objective function of the optimization problem to be solved is nondifferentiable.

The aim of this paper is to develop in finite dimensional spaces an algorithm for the problem of minimizing a convex objective which is the sum of three parts: a

2010 Mathematics Subject Classification. Primary: 90C30; Secondary: 90C25.

Key words and phrases. Gaussian smoothing, nonsmooth convex optimization, accelerated proximal-gradient methods, smoothing method.

The first author is supported by Shanghai Natural Science Foundation of China (under grant: 12ZR1411600), National Natural Science Foundation of China (under grant: 11201267.)
smooth part, a simple non-smooth Lipschitz part, and a simple non-smooth non-Lipschitz part. i.e., we consider an unconstrained nonsmooth convex optimization problem of the form

\[ \min_{x \in \mathbb{R}^n} F(x) = \varphi(x) + g(x) + h(x), \]  

(1)

where \( \varphi : \mathbb{R}^n \to \mathbb{R} \) is a convex and \( L_\varphi \)-smooth function, that is, differentiable with a Lipschitz continuous gradient: \( \| \nabla \varphi(x) - \nabla \varphi(y) \| \leq L_\varphi \| x - y \|, \forall x, y \in \mathbb{R}^n \). \( g : \mathbb{R}^n \to \mathbb{R} \) is a convex \( \rho \)-Lipschitz continuous function: \( |g(x) - g(y)| \leq \rho \| x - y \|, \forall x, y \in \mathbb{R}^n \). \( h : \mathbb{R}^n \to \mathbb{R} \cup \{ +\infty \} \) is a proper, lower semicontinuous, convex (but possibly non-continuous/non-finite) function. For example, \( h \) could be an indicator function for a convex constraint. Many optimization problems with complex structure can be written in the form (1), including robust principal component analysis (robust PCA) [4], sparse inverse covariance selection [10] [11].

We further assume that we can calculate gradients of \( \varphi \) and \( g \), \( (g_\mu \) is the Gaussian smoothing approximation of \( g \), and that \( h \) is "simple" in the sense that we can calculate its proximity operators:

\[ \text{prox}_h(x, \alpha) = \arg\min_u \left\{ \| x - u \|^2 + h(u) : u \in \mathbb{R}^n \right\}, \]

where \( x \in \mathbb{R}^n, \alpha \in \mathbb{R}_{++} \). By replacing the functions \( g \) through its Gaussian smoothing function, we approximate (1) by the problem of minimizing a convex objective which is the sum of two parts: a smooth part with Lipschitz continuous gradient and a simple non-smooth non-Lipschitz part. For solving the resulting smoothed problem, we use an accelerated gradient descent approach on \( \varphi \) plus the smoothed version of \( g \), as in Nesterov [7]. We also use the ideas of partial linearization to handle the component \( h \): instead of attempting to linearize it as in gradient descent, we include it as is in each iteration, as in Duchi and Singer [5] and Beck and Teboule [2], i.e., we propose an extension of the accelerated first order method of Beck and Teboule (cf. [2]) for convex optimization problems involving variable smoothing parameters which are updated in each iteration. This scheme yields for the minimization of the objective of the initial problem a rate of convergence of order \( O\left(\frac{\ln k}{k}\right) \), while, in the particular case when the smoothing parameters are constant, the order of the rate of convergence becomes \( O\left(\frac{1}{k}\right) \). Nonetheless, using variable smoothing parameters has an important advantage, although the theoretical rate of convergence is not as good as when these are constant. In the first case the approach generates a sequence of iterates \( \{x_k\} \) such that \( \{\varphi(x_k) + g(x_k) + h(x_k)\} \) converges to the optimal objective value of (1). In the case of constant smoothing variables the approach provides a sequence of iterates which solves the problem (1) with a priori given accuracy, however, the sequence \( \{\varphi(x_k) + g(x_k) + h(x_k)\} \) may not converge to the optimal objective value of the problem to be solved.

The structure of this paper is as follows. In Section 2 we give some preliminaries and smoothing of nonsmooth functions. Section 3 is mainly devoted to the description of the iterative methods for solving (1) and of their convergence properties for both variable and constant smoothing. In Section 4, some conclusions are given.

2. Preliminaries and Smoothing of Nonsmooth Functions. In order to handle the non-smooth component \( g \) in (1), we approximate it using a smooth function. Such a smoothing plays a central role in our method, and we devote this section to carefully presenting it. For a function \( f : \mathbb{R}^n \to \mathbb{R} \), let us form its Gaussian
smoothing approximation
\[ f_\mu(x) = \frac{1}{\kappa} \int_{\mathbb{R}^n} f(x + \mu u) e^{-\frac{1}{2} \|u\|^2} du, \]
where \( \kappa = \int_{\mathbb{R}^n} e^{-\frac{1}{2} \|u\|^2} du = (2\pi)^{\frac{n}{2}} \). In this definition, \( \mu > 0 \) plays a role of smoothing parameter. The function \( f_\mu \) is a smoothing approximation to \( f \), as summarized in the following lemma.

**Lemma 2.1** \cite{8} Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a convex, \( \rho \)-Lipschitz function, then
(i). \( f_\mu(x) \) is convex;
(ii). \( f_\mu(x) \geq f(x) \);
(iii). \( f_\mu(x) \) is \( 2\rho \sqrt{n} \)-smooth.

**Lemma 2.2** \cite{8} Let \( M_p = \frac{1}{\kappa} \int_{\mathbb{R}^n} \|u\|^p e^{-\frac{1}{2} \|u\|^2} du = \frac{1}{\kappa} \int_{\mathbb{R}^n} \|u\|^p e^{-\frac{1}{2} \|u\|^2} du = (2\pi)^{\frac{n}{2}} \). Then for \( p \in [0, 2] \), we have
\[ M_p \leq n^{\frac{p}{2}}. \]
If \( p \geq 2 \), then we have two-side bounds
\[ n^{\frac{p}{2}} \leq M_p \leq (p + n)^{\frac{p}{2}}. \]

**Lemma 2.3** \cite{8} Let \( f : \mathbb{R}^n \to \mathbb{R} \), be \( \rho \)-Lipschitz function, then for \( \mu > 0 \),
\[ f(x) - \mu \rho \sqrt{n} \leq f_\mu(x) \leq f(x) + \mu \rho \sqrt{n}. \]

**Lemma 2.4** Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a convex, \( \rho \) Lipschitz function, then for \( \mu > 0 \),
\[ f(x) \leq f_\mu(x) \leq f(x) + \mu \rho \sqrt{n}. \]

**Proof.** It follows from Lemma 2.1(ii) and Lemma 2.3. \qed

**Lemma 2.5** Let \( f : \mathbb{R}^n \to \mathbb{R} \), be \( \rho \)-Lipschitz function, then for \( \mu \geq \mu' \), we have
\[ |f_\mu(x) - f_{\mu'}(x)| \leq \rho(\mu - \mu') \sqrt{n}. \]

**Proof.**
\[
|f_\mu(x) - f_{\mu'}(x)| \\
= \frac{1}{\kappa} \int_{\mathbb{R}^n} |f(x + \mu u) - f(x + \mu' u)| e^{-\frac{1}{2} \|u\|^2} du \\
\leq \frac{1}{\kappa} \int_{\mathbb{R}^n} |f(x + \mu u) - f(x + \mu' u)| e^{-\frac{1}{2} \|u\|^2} du \\
\leq \frac{1}{\kappa} \int_{\mathbb{R}^n} \rho(\mu - \mu') \|u\| e^{-\frac{1}{2} \|u\|^2} du \\
= \frac{\rho(\mu - \mu')}{\kappa} \int_{\mathbb{R}^n} \|u\| e^{-\frac{1}{2} \|u\|^2} du \\
= \rho(\mu - \mu') M_1 \\
\leq \rho(\mu - \mu') \sqrt{n}. \]
\qed
Lemma 2.6 [1] The convex function $f$ is $L$-smooth if and only if
\[ f(x) \leq f(y) + \langle x - y, \nabla f(y) \rangle + \frac{L}{2} \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n. \]

Lemma 2.7 [6] For a $\beta$-strongly convex function $\psi$, and $\hat{v} = \arg\min \psi$,
\[ \psi(w) - \psi(\hat{v}) \geq \frac{\beta}{2} \|w - \hat{v}\|^2, \quad \forall w \in \mathbb{R}^n. \]

3. The variable smoothing and the constant smoothing algorithms. The variable smoothing algorithm (A1) which we present at the beginning of this section can be seen as an extension of the accelerated first order method of Beck and Teboule (cf. [2]) by using variable smoothing parameters, which we update in each iteration.

Our proposed method, Algorithm (A1), is given as follows.

Algorithm (A1):
Parameters $\{\mu_k > 0 : k \in N\}$
Initialization: $x_1 = y_1 \in \text{dom} h$, $\theta_1 = 1$, $L_1 = L_\varphi + \frac{2\rho_\varphi}{\mu_1}$ for $k = 1, 2, \ldots,$

\[
L_{k+1} = L_\varphi + \frac{2\rho_\varphi}{\mu_{k+1}},
\]

\[
\theta_{k+1} = \frac{2}{1 + \sqrt{1 + \frac{4\mu_{k+1}}{\rho_\varphi}L_k}},
\]

\[
x_{k+1} = \text{prox}_h \left( y_k - \frac{1}{L_k} \nabla \varphi(y_k) - \frac{1}{L_k} \nabla g_{\mu_k}(y_k), \frac{1}{L_k} \right),
\]

\[
y_{k+1} = x_{k+1} + \theta_{k+1} \left( \frac{1}{\theta_k} - 1 \right) (x_{k+1} - x_k).
\]

As is standard in "accelerated" first order methods (e.g. Beck and Teboule [2]), we keep track of two sequences of iterates, $x_k$ and $y_k$. At each iteration we perform a proximal gradient update $x_{k+1} = \text{prox}_h \left( y_k - \frac{1}{L_k} \nabla \varphi(y_k) - \frac{1}{L_k} \nabla g_{\mu_k}(y_k), \frac{1}{L_k} \right)$, where $\mu_k$ is some sequence of parameters. We then set $y_{k+1}$ to be a linear combination of $x_{k+1}$ and $x_k$. Note also that the recursive formula for the sequence $\mu_k$ satisfies

\[
\frac{1}{\theta_{k+1}^2} - \frac{1}{\theta_k^2} = \frac{L_{k+1}}{L_k} \frac{1}{\theta_k^2}.
\]

We establish the following guarantee on the values of the objective function.

Theorem 3.1. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a convex and $L_\varphi$-smooth function, $g : \mathbb{R}^n \to \mathbb{R}$ be a convex and $\rho$-Lipschitz continuous function, $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous convex function and $x^*$ be an optimal solution to (1). If the sequence $\mu_k$ is nonincreasing, then the iterates of Algorithm (A1) satisfy

\[
F(x_{k+1}) - F(x^*) \leq \frac{\theta_k^2 L_k}{2} \|x^* - x_1\|^2 + \theta_k^2 L_k \sum_{i=1}^{k} \rho \sqrt{n} (\theta_i \mu_i + \mu_i - \mu_{i+1}) + \rho \sqrt{n} \mu_{k+1}.
\]

Proof. Denote

\[
F_{\mu_k}(x) = \varphi(x) + g_{\mu_k}(x) + h(x),
\]

\[
L_1 = L_\varphi + \frac{2\rho \sqrt{n}}{\mu_1},
\]

\[
L_{k+1} = L_\varphi + \frac{2\rho \sqrt{n}}{\mu_{k+1}}.
\]
Fix an arbitrary $k \in \{1, 2, \ldots\}$, since $\varphi + g_{\mu_k}$ is $L_k$-smooth by Lemma 2.1, applying Lemma 2.6 yields

$$\varphi(x_{k+1}) + g_{\mu_k}(x_{k+1})$$

$$\leq \varphi(y_k) + g_{\mu_k}(y_k) + (x_{k+1} - y_k, \nabla \varphi(y_k) + \nabla g_{\mu_k}(y_k)) + \frac{L_k}{2} \|x_{k+1} - y_k\|^2. \quad (3)$$

Adding $h(x_{k+1})$ on two sides of (3), we have

$$F^{\mu_k}(x_{k+1})$$

$$\leq \varphi(y_k) + g_{\mu_k}(y_k) + (x_{k+1} - y_k, \nabla \varphi(y_k) + \nabla g_{\mu_k}(y_k)) + \frac{L_k}{2} \|x_{k+1} - y_k\|^2 + h(x_{k+1}). \quad (4)$$

We now use the strong convexity of the function

$$\langle \nabla \varphi(y_k) + \nabla g_{\mu_k}(y_k), x \rangle + \frac{L_k}{2} \|x - y_k\|^2 + h(x),$$

whose minimizer equals $x_{k+1}$. Using Lemma 2.7 with

$$w = (1 - \theta_k)x_k + \theta_k x^*$$

and

$$\psi(x) = \langle \nabla \varphi(y_k) + \nabla g_{\mu_k}(y_k), x \rangle + \frac{L_k}{2} \|x - y_k\|^2 + h(x),$$

we get

$$\langle \nabla \varphi(y_k) + \nabla g_{\mu_k}(y_k), (1 - \theta_k)x_k + \theta_k x^* \rangle + \frac{L_k}{2} \|(1 - \theta_k)x_k + \theta_k x^* - y_k\|^2$$

$$+ h((1 - \theta_k)x_k + \theta_k x^*)$$

$$\geq \langle \nabla \varphi(y_k) + \nabla g_{\mu_k}(y_k), x_{k+1} \rangle + \frac{L_k}{2} \|x_{k+1} - y_k\|^2 + h(x_{k+1})$$

$$+ \frac{L_k}{2} \|(1 - \theta_k)x_k + \theta_k x^* - x_{k+1}\|^2,$$

i.e., we obtain

$$\langle \nabla \varphi(y_k) + \nabla g_{\mu_k}(y_k), x_{k+1} - y_k \rangle + \frac{L_k}{2} \|x_{k+1} - y_k\|^2 + h(x_{k+1})$$

$$\leq \langle \nabla \varphi(y_k) + \nabla g_{\mu_k}(y_k), (1 - \theta_k)x_k + \theta_k x^* - y_k \rangle + \frac{L_k}{2} \|(1 - \theta_k)x_k + \theta_k x^* - y_k\|^2$$

$$+ h((1 - \theta_k)x_k + \theta_k x^*) - \frac{L_k}{2} \|(1 - \theta_k)x_k + \theta_k x^* - x_{k+1}\|^2. \quad (5)$$

By (4) and (5), we get

$$F^{\mu_k}(x_{k+1}) \leq \varphi(y_k) + g_{\mu_k}(y_k) + h((1 - \theta_k)x_k + \theta_k x^*)$$

$$+ \langle \nabla \varphi(y_k) + \nabla g_{\mu_k}(y_k), (1 - \theta_k)x_k + \theta_k x^* - y_k \rangle$$

$$+ \frac{L_k}{2} \|(1 - \theta_k)x_k + \theta_k x^* - y_k\|^2 - \frac{L_k}{2} \|(1 - \theta_k)x_k + \theta_k x^* - x_{k+1}\|^2. \quad (6)$$

We let $z_k := \frac{1}{\theta_k}y_k + (1 - \frac{1}{\theta_k})x_k$ for every $k = 1, 2, \ldots$, and note that $z_{k+1} = x_k + \frac{1}{\theta_k}(x_{k+1} - x_k)$ by the algorithmic construction of $y_{k+1}$. According to $z_k = \frac{1}{\theta_k}y_k + (1 - \frac{1}{\theta_k})x_k$ and $z_{k+1} = x_k + \frac{1}{\theta_k}(x_{k+1} - x_k)$, we obtain

$$(1 - \theta_k)x_k + \theta_k x^* - y_k = \theta_k(x^* - z_k) \quad (7)$$
and

\[(1 - \theta_k)x_k + \theta_k x^* - x_{k+1} = \theta_k(x^* - z_{k+1}).\]  

Using the convexity of the function \(h(x)\), we get

\[h((1 - \theta_k)x_k + \theta_k x^*) \leq (1 - \theta_k)h(x_k) + \theta_k h(x^*).\]

By (6), (7), (8) and (9), we get

\[
F^{\mu_k}(x_{k+1}) \leq \varphi(y_k) + g_{\mu_k}(y_k) + (1 - \theta_k)h(x_k) + \theta_k h(x^*)
+ (\nabla \varphi(y_k) + \nabla g_{\mu_k}(y_k), (1 - \theta_k)x_k + \theta_k x^* - y_k)
+ \frac{\theta_k^2 L_k}{2} \| x^* - z_k \|^2 - \frac{\theta_k^2 L_k}{2} \| x^* - z_{k+1} \|^2
\leq (1 - \theta_k)F^{\mu_k}(x_k) + \theta_k F^{\mu_k}(x^*)
+ \frac{\theta_k^2 L_k}{2} \| x^* - z_k \|^2 - \frac{\theta_k^2 L_k}{2} \| x^* - z_{k+1} \|^2,
\]

where we have used the convexity of the functions \(\varphi, g_{\mu_k}, h\). Using Lemma 2.3 we have

\[
F^{\mu_k}(x_{k+1}) - F(x^*) \leq \frac{\theta_k^2 L_k}{2} (\| x^* - z_k \|^2 - \| x^* - z_{k+1} \|^2)
+ (1 - \theta_k)[F^{\mu_k}(x_k) - F(x^*)] + \theta_k \mu_k \rho \sqrt{n}.
\]

Since \(\mu_{k+1} \leq \mu_k\), we obtain by Lemma 2.5

\[
F^{\mu_{k+1}}(x_{k+1}) \leq F^{\mu_k}(x_{k+1}) + \rho (\mu_k - \mu_{k+1}) \sqrt{n}.
\]

Denoting \(D_k = F^{\mu_k}(x_k) - F(x^*)\), we have by (11)

\[
D_{k+1} = F^{\mu_{k+1}}(x_{k+1}) - F(x^*)
\leq F^{\mu_k}(x_{k+1}) - F(x^*) + \rho (\mu_k - \mu_{k+1}) \sqrt{n}
\leq \frac{\theta_k^2 L_k}{2} (\| x^* - z_k \|^2 - \| x^* - z_{k+1} \|^2) + \rho (\mu_k - \mu_{k+1}) \sqrt{n}
+ (1 - \theta_k)[F^{\mu_k}(x_k) - F(x^*)] + \theta_k \mu_k \rho \sqrt{n}
\]

\[= \frac{\theta_k^2 L_k}{2} (\| x^* - z_k \|^2 - \| x^* - z_{k+1} \|^2) + (1 - \theta_k)D_k
+ \theta_k \mu_k \rho \sqrt{n} + \rho (\mu_k - \mu_{k+1}) \sqrt{n},\]

i.e.,

\[
D_{k+1} \leq \frac{\theta_k^2 L_k}{2} (\| x^* - z_k \|^2 - \| x^* - z_{k+1} \|^2) + (1 - \theta_k)D_k
+ \theta_k \mu_k \rho \sqrt{n} + \rho (\mu_k - \mu_{k+1}) \sqrt{n}.
\]

Furthermore, we get

\[
\frac{D_{k+1}}{\theta_k^2 L_k} \leq \frac{1}{2} (\| x^* - z_k \|^2 - \| x^* - z_{k+1} \|^2) + \frac{1 - \theta_k}{\theta_k^2 L_k} D_k
+ \frac{\theta_k \mu_k \rho \sqrt{n} + \rho (\mu_k - \mu_{k+1}) \sqrt{n}}{\frac{\theta_k^2 L_k}{2}}.
\]

Using the definition of \(\theta_k\) in (2), and summing we obtain

\[
\frac{D_{k+1}}{\theta_k^2 L_k} \leq \frac{1}{2} (\| x^* - z_1 \|^2 - \| x^* - z_{k+1} \|^2)
+ \sum_{i=1}^{k} \frac{\rho \sqrt{n}(\theta_i \mu_i + \mu_i - \mu_{i+1})}{\theta_i^2 L_i},
\]
i.e.,
\[
D_{k+1} \leq \frac{\theta_k^2 L_k}{2} (\|x^* - z_1\|^2 - \|x^* - z_{k+1}\|^2) + \theta_k^2 L_k \sum_{i=1}^{k} \frac{\rho \sqrt{n}(\theta_i \mu_i + \mu_i - \mu_{i+1})}{\theta_i^2 L_i}
\]
\[
\leq \frac{\theta_k^2 L_k}{2} \|x^* - x_1\|^2 + \theta_k^2 L_k \sum_{i=1}^{k} \frac{\rho \sqrt{n}(\theta_i \mu_i + \mu_i - \mu_{i+1})}{\theta_i^2 L_i}.
\]
(13)
We apply Lemma 2.3 to obtain
\[
F(x_{k+1}) \leq F(x_k) + \rho \sqrt{n} \mu_{k+1}.
\]
Therefore, we have by (13)
\[
F(x_{k+1}) - F(x^*) \leq F(x_k) - F(x^*) + \rho \sqrt{n} \mu_{k+1}
\]
\[
\leq \frac{\theta_k^2 L_k}{2} \|x^* - x_1\|^2 + \theta_k^2 L_k \sum_{i=1}^{k} \frac{\rho \sqrt{n}(\theta_i \mu_i + \mu_i - \mu_{i+1})}{\theta_i^2 L_i} + \rho \sqrt{n} \mu_{k+1},
\]
which concludes the proof.

The convergence of Algorithm (A1) is stated by the following theorem.

**Theorem 3.2.** Let \( \phi : \mathbb{R}^n \to \mathbb{R} \) be a convex and \( L_\phi \)-smooth function, \( g : \mathbb{R}^n \to \mathbb{R} \)
be a convex and \( \rho \)-Lipschitz continuous function, \( h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) be a proper lower semicontinuous convex function and \( x^* \) be an optimal solution to (1). Then, when choosing
\[
\mu_k = \frac{1}{\alpha_k}, \quad \forall k \geq 1,
\]
where \( \frac{L_\phi}{2\rho \sqrt{n}} > \alpha > 0 \), Algorithm (A1) generates a sequence \( \{x_k\} \) satisfying
\[
F(x_{k+1}) - F(x^*) \leq 2 L_\phi \left( \frac{2\rho \sqrt{n}\alpha k}{(k+1)^2} \|x^* - x_1\|^2 + \frac{\rho \sqrt{n}}{\alpha(k+1)} \right)
\]
\[
+ \frac{8\rho \sqrt{n}(L_\phi + (2\rho \sqrt{n})\alpha k)}{\alpha(k+1)^2} \left( \frac{3}{4\rho \sqrt{n}\alpha} \ln \frac{L_\phi + (2\rho \sqrt{n})\alpha k}{L_\phi + (2\rho \sqrt{n})\alpha} + \frac{2}{L_\phi + (2\rho \sqrt{n})\alpha} \right).
\]

**Proof.** Firstly we prove that for \( k = 1, 2, \ldots, \)
\[
\frac{1}{k+1} \leq \theta_k \leq \frac{2}{k+1}.
\]
We prove the result by induction. By assumption, it is true for \( k = 1 \). We assume it is true for \( k = i, \) i.e.,
\[
\frac{1}{i+1} \leq \theta_i \leq \frac{2}{i+1}.
\]
(15)
In the following, we prove that it is true for \( k = i + 1 \). By (15) and \( L_k = L_\phi + (2\rho \sqrt{n})\alpha k \), we obtain that \( \frac{1}{2} \geq \frac{i+1}{2} \) and \( \frac{L_{i+1}}{L_i} > 1 \). Therefore,
\[
\theta_{i+1} = \frac{2}{1 + \left(1 + \frac{4L_{i+1}}{\theta_i^2 L_i}\right)^{\frac{1}{2}}} \leq \frac{2}{1 + \left(1 + (i+1)^2 \frac{L_{i+1}}{L_i}\right)^{\frac{1}{2}}},
\]
\[
\begin{align*}
&\leq \frac{2}{1 + \sqrt{1 + (i + 1)^2}} \\
&\leq \frac{2}{i + 2}.
\end{align*}
\]

By (15), we get \( \frac{1}{\theta_i} < i + 1 \). Therefore,
\[
\begin{align*}
\theta_{i+1} - \frac{1}{i + 2} &= \frac{2}{1 + \sqrt{1 + 4(i + 1)^2 \frac{L_{i+1}}{L_i}}} - \frac{1}{i + 2} \\
&> \frac{2}{1 + \sqrt{1 + 4(i + 1)^2 \frac{L_{i+1}}{L_i}}} - \frac{1}{i + 2} \\
&= \frac{2i + 4 - 1 - \sqrt{1 + 4(i + 1)^2 \frac{L_{i+1}}{L_i} + (i + 1) \alpha (2\rho \sqrt{n})}}{1 + \sqrt{1 + 4(i + 1)^2 \frac{L_{i+1}}{L_i} + (i + 1) \alpha (2\rho \sqrt{n})}} (i + 2).
\end{align*}
\]

Since
\[
\begin{align*}
2i + 4 - 1 - \sqrt{1 + 4(i + 1)^2 \frac{L_{i+1}}{L_i} + (i + 1) \alpha (2\rho \sqrt{n})} &= (2i + 3)^2 - \left[ 1 + 4(i + 1)^2 \frac{L_{i+1} + (i + 1) \alpha (2\rho \sqrt{n})}{L_i + \alpha (2\rho \sqrt{n})} \right] \\
&= \left( \frac{4i + 4}{L_{i+1} - \alpha (2\rho \sqrt{n})} \right) \left( L_{i+1} + \alpha (2\rho \sqrt{n}) \right),
\end{align*}
\]

hence choosing \( \alpha < \frac{L_{i+1}}{2\rho \sqrt{n}} \), we have
\[
\begin{align*}
\frac{(4i + 4)(L_{i+1} - \alpha (2\rho \sqrt{n}))}{(2i + 3 + \sqrt{1 + 4(i + 1)^2 \frac{L_{i+1}}{L_i} + (i + 1) \alpha (2\rho \sqrt{n})}) (L_{i+1} + \alpha (2\rho \sqrt{n}))} > 0.
\end{align*}
\]

By (16), (17) and (18), we get
\[
\frac{1}{i + 2} \leq \theta_{i+1} \leq \frac{2}{i + 2}.
\]

By induction, we have that for \( k = 1, 2, \ldots \),
\[
\begin{align*}
&\frac{1}{k + 1} \leq \theta_k \leq \frac{2}{k + 1}.
\end{align*}
\]

Because \( i + 1 > \frac{1}{\theta_i} > \frac{i + 1}{2} \), we obtain
\[
\begin{align*}
\sum_{i=1}^{k} \mu_i \theta_i L_i &\leq \frac{1}{\alpha} \sum_{i=1}^{k} \left( 1 + \frac{1}{i} \right) \frac{1}{L_i + (2\rho \sqrt{n}) \alpha i} \\
&= \frac{1}{\alpha} \left[ \frac{2}{L_{i+1} + (2\rho \sqrt{n}) \alpha} + \sum_{i=2}^{k} \left( 1 + \frac{1}{i} \right) \frac{1}{L_i + (2\rho \sqrt{n}) \alpha} \right] \\
&\leq \frac{1}{\alpha} \left[ \frac{2}{L_{i+1} + (2\rho \sqrt{n}) \alpha} + \frac{3}{2} \sum_{i=2}^{k} \frac{1}{L_i + (2\rho \sqrt{n}) \alpha} \right].
\end{align*}
\]
Theorem 1 when

Proof. In order to prove this statement, one has only to reproduce the result of

and

The convergence of the Algorithm (A2) is proved by the following theorem.

Algorithm (A2):
Parameters \( \mu > 0 \)
Initialize: \( x_1 = y_1 \in \text{dom} h, \theta_1 = 1, L(\mu) = L_\varphi + \frac{2\rho\sqrt{n}}{\mu} \) for \( k = 1, 2, \ldots \),
\[ \theta_{k+1} = \frac{2}{1 + \sqrt{1 + \frac{4\rho\sqrt{n}}{\mu} \frac{1}{L(\mu)}}} \]
\[ x_{k+1} = \text{prox}_h \left( y_k - \frac{1}{L(\mu)} \nabla \varphi (y_k) - \frac{1}{L(\mu)} \nabla g_\mu (y_k), \frac{1}{L(\mu)} \right) \]
\[ y_{k+1} = x_{k+1} + \theta_{k+1} \left( \frac{1}{L(\mu)} - 1 \right) (x_{k+1} - x_k) \]
The convergence of the Algorithm (A2) is proved by the following theorem.

Theorem 3.3. Let \( \varphi : \mathbb{R}^n \to \mathbb{R} \) be a convex and \( L_\varphi \)-smooth function, \( g : \mathbb{R}^n \to \mathbb{R} \) be a convex and \( \rho \)-Lipschitz continuous function, \( h : \mathbb{R}^n \to \mathbb{R} \cup \{ +\infty \} \) be a proper lower semicontinuous convex function and \( \mathbf{x}^* \) be an optimal solution to (1). Let \( 1 > \varepsilon > 0 \), and set \( \mu = \frac{\varepsilon}{5\rho\sqrt{n}} \) in Algorithm 2, then Algorithm (A2) generates a sequence \( \{x_k\} \) which provides an \( \varepsilon \)-optimal solution to (1) with a rate of convergence for the objective of order \( O\left( \frac{1}{k} \right) \).

Proof. In order to prove this statement, one has only to reproduce the result of

\[ \mu_k = \mu, \quad L_k = L(\mu) = L_\varphi + \frac{2\rho\sqrt{n}}{\mu}, \quad \forall k \geq 1. \]
This inequality reads in this particular situation

\[ F(x_{k+1}) - F(x^*) \leq \frac{\theta_k^2 L(\mu)}{2} \|x^* - x_1\|^2 + \rho \sqrt{n} \mu \theta_k \sum_{i=1}^{k} \frac{1}{\theta_i} + \rho \sqrt{n} \mu. \]

Since \( \frac{1}{\theta_{k+1}} = \frac{1}{\theta_k} + \frac{1}{\theta_{k+1}} \) for any \( k \geq 1 \), one can inductively prove that \( \frac{1}{\theta_{k+1}} = \sum_{i=1}^{k+1} \frac{1}{\theta_i} \), which, together with the fact that \( \frac{1}{\theta_k} \geq \frac{k+2}{2} \) for any \( k \geq 1 \) in (15), yields

\[ F(x_{k+1}) - F(x^*) \leq \frac{2L(\mu)}{(k+1)^2} \|x^* - x_1\|^2 + 2\rho \sqrt{n} \mu, \quad \forall k \geq 1. \]

In order to obtain \( \varepsilon \)-optimality for the objective of the problem (1), where \( 1 > \varepsilon > 0 \) is a given level of accuracy, we choose \( \mu = \frac{\varepsilon}{\sqrt{\rho \sqrt{n}}} \) and, thus, we have only to force the first term in the right-hand side of the above estimate to be less than or equal to \( \frac{\varepsilon}{3} \). Taking also into account that in this situation \( L(\mu) = L_\varphi + \frac{6 \rho^2 n}{\varepsilon} \), it holds

\[ \frac{\varepsilon}{3} \geq \frac{2L(\mu)}{(k+1)^2} \|x^* - x_1\|^2 = \frac{2(L_\varphi + \frac{6 \rho^2 n}{\varepsilon})}{(k+1)^2} \|x^* - x_1\|^2, \]

i.e.,

\[ \frac{\varepsilon}{3} \geq \frac{2(\varepsilon L_\varphi + 6 \rho^2 n)}{\varepsilon (k+1)^2} \|x^* - x_1\|^2. \]

(19) is equivalent to the following

\[ \frac{\varepsilon^2}{9} \geq \frac{2(\varepsilon L_\varphi + 6 \rho^2 n)}{(k+1)^2} \|x^* - x_1\|^2. \]

Since \( 1 > \varepsilon > 0 \), we have

\[ \frac{2(\frac{\varepsilon}{3} L_\varphi + 2 \rho^2 n)}{(k+1)^2} \|x^* - x_1\|^2 < \frac{2(\frac{1}{3} L_\varphi + 2 \rho^2 n)}{(k+1)^2} \|x^* - x_1\|^2. \]

When

\[ \frac{\varepsilon^2}{9} \geq \frac{2(\frac{1}{3} L_\varphi + 2 \rho^2 n)}{(k+1)^2} \|x^* - x_1\|^2 \]

holds, we obtain

\[ \frac{\varepsilon^2}{9} \geq \frac{2(\frac{\varepsilon}{3} L_\varphi + 2 \rho^2 n)}{(k+1)^2} \|x^* - x_1\|^2. \]

It is obvious that

\[ \frac{\varepsilon^2}{9} \geq \frac{2(\frac{\varepsilon}{3} L_\varphi + 2 \rho^2 n)}{(k+1)^2} \|x^* - x_1\|^2 \]

is equivalent to the following

\[ \frac{\varepsilon}{3} \geq \frac{\sqrt{2(\frac{\varepsilon}{3} L_\varphi + 2 \rho^2 n)}}{(k+1)} \|x^* - x_1\|, \]

which shows that an \( \varepsilon \)-optimal solution to (1) can be provided with a rate of convergence for the objective of \( O(\frac{1}{k}) \). which concludes the proof.
4. Conclusions. Motivated by the recent applications, we consider the problem of minimizing a convex objective which is the sum of three parts: a smooth part, a simple non-smooth Lipschitz part, and a simple non-smooth non-Lipschitz part. By making use of the Gaussian smoothing function of the functions occurring in the objective, we smooth the second part to a convex and differentiable function with Lipschitz continuous gradient by using both variable and constant smoothing parameters. The resulting problem is solved via Algorithm (A1) and Algorithm (A2). We showed that the rate of convergence of Algorithm (A1) may not be as good as the one proved for the Algorithm (A2) with constant smoothing parameters depending on a fixed level of accuracy $\varepsilon (1 > \varepsilon > 0)$. However, the main advantage of the variable smoothing methods is given by the fact that the sequence of objective values $\{\varphi(x_k) + g(x_k) + h(x_k)\}$ converges to the optimal objective value of (1), whereas, when generated by algorithm (A2), despite of the fact that it approximates the optimal objective value with a better convergence rate, this sequence may not converge to this optimal objective value.

Acknowledgments This work was supported by Shanghai Natural Science Foundation of China, Grant No. 12ZR1411600, and by National Natural Science Foundation of China, Grant No. 11201267. We would like to thank Professor Shu-Zhong Zhang, and part of this work was performed during a research visit by the first author to the University of Minnesota. We would like to thank the reviewer for his valuable suggestions and the editor for his helpful assistance.

REFERENCES


Received December 2014; 1st revision February 2015; final revision March 2015.

E-mail address: liusanmingxyz@163.com
E-mail address: wzjsdstu@163.com
E-mail address: chongyangliu@yahoo.cn