Les Cahiers du GERAD ISSN: 0711–2440

Solving Unconstrained Nonlinear Programs Using ACCPM

A. Dehghani, J.-L. Goffin,
D. Orban

G–2012–02

January 2012

Solving Unconstrained Nonlinear Programs Using ACCPM

Ahad Dehghani
Jean-Louis Goffin

GERAD & Desautels Faculty of Management
McGill University
Montréal (Québec) Canada
ahad.dehghani@mcgill.ca
jean-louis.goffin@mcgill.ca

Dominique Orban

GERAD & Department of Mathematics and Industrial Engineering
École Polytechnique de Montréal
Montréal (Québec) Canada
dominique.orban@gerad.ca

January 2012

Les Cahiers du GERAD

G–2012–02

Abstract

The analytic center cutting plane method and its proximal variant are well known techniques for solving convex programming problems. We propose two sequential convex programming variants based on convexification techniques for nonconvex unconstrained problems. The performance of the algorithm is compared with that of two other well known first-order algorithms, the steepest descent and nonlinear conjugate gradient with Armijo line search on a set of 158 problem from the CUTEr test set.

Acknowledgments: Research partially supported by NSERC Discovery Grant 299010-04 of the third author.

1 Introduction

The Analytic Center Cutting Plane Method (ACCPM) [20] for minimizing a convex function works by maintaining and refining an outer approximation to the bounded epigraph of the objective. Given an outer approximation, an approximate analytic center is computed. If this center does not determine a solution of the original problem, it determines a new (linear) cut that yields an updated outer approximation. The sequence of analytic centers converges to a (global) minimizer of the objective function [20]. Proximal ACCPM is a variant of ACCPM introduced in [21] that adds a proximal term to the logarithmic barrier to avoid zigzagging. On the other hand, when the objective function is not convex the epigraph of f is not convex therefore, the definition of the analytic center is not clear and even if we had a clear description of the analytic center, adding cuts using this center may cut off and eliminate the optimal solution.

In this paper, we introduce a generalization of proximal ACCPM for the solution of

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \;\; f(x), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and possibly nonconvex, and present convergence results. Unconstrained programs are important in their own right, and also arise as subproblems in methods for general constrained optimization such as augmented lagrangian methods, and interior-point methods. We propose two sequential convex programming approaches and show that the sequence of global minimizers of each of these convex functions converges to a (local) minimizer of f under reasonable assumptions. The two strategies are based on using the sequences of *proximal functions*

$$\varphi_k(x) := f(x) + \frac{1}{2}L\|x - x_k\|^2 \tag{2}$$

and of *potential proximal functions*

$$\psi_k(x) := -\ln(f(x_k) - f(x)) + \frac{1}{2}\hat{L}\|x - x_k\|^2 \tag{3}$$

where $L \geq 0$ and $\hat{L} \geq 0$ are sufficiently large that φ_k and ψ_k are convex. We will see in Section 2 that L can be any number larger than the Lipschitz constants of $\nabla f(x)$. The two convexifications are independent of the procedure used to solve each subproblem. In both approaches, we can apply ACCPM. In Theorems 9 and 10, we show how to advantageously reuse cuts from previous iterations in the current iteration.

The performance of our algorithm is compared to that of two other well known algorithms, the steepest descent and nonlinear conjugate gradient method with Armijo line search. These two algorithms are chosen for the comparison because they are first order methods, like ACCPM. Our implementation builds upon that of [32]. Our benchmarks compare two versions of our algorithm on 158 unconstrained problems from CUTEr test set [14] using performance profiles [6].

The rest of this paper is organized as follows. After reviewing related work in the next subsection, we discuss convexification of a nonconvex function in §2. A brief discussion of ACCPM for unconstrained convex optimization is presented in §4. In §5 and §3.2, we study ACCPM for proximal and potential functions respectively and present our convergence results. Approximating the constants L and $\hat{L}$ is discussed in §6. Numerical results are reported in §7. Finally, §8 discusses our approach and its generalization to other problem classes.

1.1 Related work

The proximal point method was first introduced by Martinet [18] for solving variational problems. Starting with Rockafellar [27, 28] it was extended to the problem of finding a zero of a maximal monotone operator and to convex optimization. A number of researchers have used the proximal point method since the late 1970's to resolve some difficulties in standard optimization problems. For example, Rockafellar [28] uses it to approximate the solutions of nonlinear equations associated with monotone operators. It is used in [3, 8] to take advantage of decomposition methods for convex minimization without assuming strong or strict convexity.

Fukushima and Mine [10] consider the minimization of $f(x) + g(x)$ where f is continuously differentiable on an open set containing the domain of g, and g is a closed proper convex function. The k-th iteration consists in minimizing the model $\nabla f(x_k)^T x + \frac{1}{2}c_k\|x - x_k\|^2 + g(x)$ where $c_k > 0$ is a parameter. This model can be interpreted as the proximal method to the function $g(x) + \nabla f(x_k)^T x$. Kaplan and Tichatschke [15] apply the proximal point method to nonsmooth and nonconvex problems. They show how proximal regularization with appropriate regularization parameters ensures convexity of the auxiliary problems.

Friedlander and Orban [25] use a proximal-like method for convex quadratic programming in order to handle free variables and rank-deficient constraint matrices in the framework of interior-point method.

Proximal ACCPM was studied in [24, 21] based on the bundle method [17], which can be loosely interpreted as Kelley's cutting plane method [16] in which proximal terms are added to the objective function. Bundle-type methods also enforce convergence of the centers to the optimal solution. The computation of the Newton direction in Proximal-ACCPM can be made more efficient than in plain ACCPM [7].

1.2 Notation

For a function $f : \mathbb{R}^n \to \mathbb{R}$, we use the notation D_f for the domain of f, x^* for a local minimizer of f, f^* for $f(x^*)$, and $\|\cdot\|$ for the Euclidian norm throughout this paper. The set of all local minimizers of a nonconvex function f is denoted $\arg\min_{x\in\mathbb{R}^n} f(x)$.

2 Convexification

A continuous function $f : \mathbb{R}^n \to \mathbb{R}$ may not be convex but in some cases there exists a constant L such that function $\varphi(x) := f(x) + L\|x\|^2$ is convex. From the convexity of $\varphi(x)$ one may get useful information about f. Given $y \in \mathbb{R}^n$, $\varphi_y(x) := f(x) + \frac{1}{2}L\|x - y\|^2$ is the *proximal function* of f at y. In the remainder of this section, K denotes a fixed convex subset of D_f.

Definition 1 *We say that f is convexifiable on K if there exists $L \geq 0$ such that $\varphi(x)$ is convex on K. The number L is a convexifier of f on K and φ is a convexification of f on K.*

Zlobec [35] studies and characterizes convexifiable functions in term of the mid-point acceleration function of f, defined as

$$\Psi(x, y) := \frac{8}{\|x - y\|^2}\left(\frac{f(x) + f(y)}{2} - f\left(\frac{x + y}{2}\right)\right), \quad \forall x, y \in K,\ x \neq y. \tag{4}$$

The following result is due to Zlobec [35].

Theorem 1 (Characterization of Convexifiable Functions) *The function f is convexifiable on K with convexifier L if and only if Ψ is bounded below on K by $-L$, i.e., $\Psi(x, y) \geq -L$ for all $x, y \in K$, $x \neq y$.*

An important class of convexifiable functions is the class of Lipschitz continuously differentiable functions.

Definition 2 *A function $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuously differentiable on a set $D \subset \mathbb{R}^n$ if there exists a constant $L > 0$ such that for any $x, y \in D$,*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|. \tag{5}$$

The constant L is called a Lipschitz constant for f.

It is clear that a Lipschitz constant is not unique. If L is a Lipschitz constant, $L' \geq L$ is also a Lipschitz constant.

Zlobec [34] shows that for Lipschitz continuously differentiable functions with Lipschitz constant L, the mid-point acceleration function (4) is bounded above by L. We state this result in the following theorem.

Theorem 2 *If $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuously differentiable on a convex set K with constant L then $|\Psi(x, y)| \leq L$.*

Using Theorems 1 and 2, we have the following result.

Corollary 1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz continuously differentiable function on the compact convex set K with constant L. Then L is a convexifier of f on K.*

If $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable then it is possible to write the mid-point acceleration function Ψ in term of the Hessian matrix of f.

Theorem 3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable on K. Then for any $x, y \in K$, $x \neq y$, there exist ξ and η on the line segment $[x, y]$ and line segment $[y, \frac{x+y}{2}]$, respectively, such that*

$$\Psi(x,y) = \frac{(y-x)^T}{\|y-x\|}\left(2\nabla^2 f(\xi) - \nabla^2 f(\eta)\right)\frac{y-x}{\|y-x\|}. \tag{6}$$

Proof. Since

$$\begin{aligned} \Psi(x,y) &= \frac{8}{\|x-y\|^2}\left(\frac{f(x)+f(y)}{2} - f\left(\frac{x+y}{2}\right)\right) \\ &= \frac{4}{\|x-y\|^2}\left(f(x)+f(y) - 2f\left(\frac{x+y}{2}\right)\right), \end{aligned} \tag{7}$$

using Taylor's Theorem we can find ξ on the line segment between x and y such that

$$f(x) = f(y) + \nabla f(y)^T(x-y) + \frac{1}{2}(x-y)^T\nabla^2 f(\xi)(x-y)$$

and we can find η on the line segment y and $\frac{x+y}{2}$ such that

$$\begin{aligned} f\left(\frac{x+y}{2}\right) &= f(y) + \nabla f(y)^T\left(\frac{x+y}{2} - y\right) + \frac{1}{2}\left(\frac{x+y}{2} - y\right)^T \nabla^2 f(\eta)\left(\frac{x+y}{2} - y\right) \\ &= f(y) + \frac{1}{2}\nabla f(y)^T(x-y) + \frac{1}{8}(x-y)^T\nabla^2 f(\eta)(x-y). \end{aligned}$$

Substituting these two values in (7) we get (6). □

Note that if $f(x) = \frac{1}{2}x^T Bx$ is a quadratic form then

$$\Psi(x,y) = \frac{(x-y)^T B(x-y)}{\|x-y\|^2}. \tag{8}$$

The right-hand side of (8) is the Rayleigh quotient. In this case the negative of the smallest eigenvalue of matrix B, if this eigenvalue is negative, is a convexifier of f. We can conclude that the mid-point acceleration function (6) is a generalization of Rayleigh quotient.

We note that if

$$\lim_{y\to x}\frac{y-x}{\|y-x\|} = z,$$

we have from (6) that

$$\lim_{y\to x}\Psi(x,y) = z^T\nabla^2 f(x)z,$$

which states that the limiting behavior of the mid-point acceleration function is the curvature of f at x in the direction of z.

If $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and its Hessian matrix is uniformly bounded, Theorem 3 guarantees that f is convexifiable. We state this fact in the following corollary.

Corollary 2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and suppose there exists $M > 0$ such that $\|\nabla^2 f(x)\| \leq M$ for all $x \in K$. Then $L = 3M$ is a convexifier of f on K.*

Proof. Using Theorem 3 with the fact that for $x \neq y$

$$\left\| \frac{y-x}{\|y-x\|} \right\| = 1,$$

we obtain

$$\left|\Psi(x,y)\right| \leq 3M,$$

that is $-3M$ is a lower bound of $\Psi(x,y)$ and Theorem 1 completes the proof. □

The following lemma states that if L is a convexifier of f on K then for any $y \in K$ function $\varphi_y(x) = f(x) + \frac{1}{2}L\|x-y\|^2$ is also convex on K. We use this fact frequently in our study of ACCPM for nonconvex objective functions.

Lemma 1 *$f : \mathbb{R}^n \to \mathbb{R}$ is convexifiable on K with convexifier L if and only if $\varphi_y(x) = f(x) + \frac{1}{2}L\|x-y\|^2$ is convex on K for any given $y \in \mathbb{R}^n$.*

Proof. Suppose f is convexifiable on K with convexifier L and fix $y \in \mathbb{R}^n$. Since L is a convexifier of f we have $\varphi_0(x) = \varphi(x) = f(x) + \frac{1}{2}Lx^Tx$ is convex and

$$\begin{aligned}\varphi_y(x) &= f(x) + \frac{1}{2}L(x-y)^T(x-y)\\ &= f(x) + \frac{1}{2}L(x^Tx - 2x^Ty + y^Ty)\\ &= \varphi(x) - \frac{1}{2}L(2x^Ty - y^Ty),\end{aligned}$$

the right hand side is the sum of a convex function and an affine function. Therefore, $\varphi_y(x)$ is a convex function. Now, suppose $\varphi_y(x)$ is convex on K. Since

$$\varphi(x) = \varphi_y(x) + \frac{1}{2}L(2x^Ty - y^Ty)$$

is a convex function of x for any $y \in \mathbb{R}^n$ by definition L is a convexifier of f on K. □

3 Proximal point method on the convexified and potential functions

Throughout this paper we assume, unless otherwise stated, that $f : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz continuously differentiable function with convexifier L and $K \subset \mathbb{R}^n$ is a convex and compact set that contains the set $\{\, x \in D_f \mid f(x) \leq f(x_0) \,\}$.

Our problem is to find a local minimum of the following problem

$$\min_{x \in \mathbb{R}^n} f(x). \tag{9}$$

3.1 Proximal point method on the convexified function

By Lemma 1 for any sequence $\{x_k\}$, the following proximal functions are convex

$$\varphi_k(x) = f(x) + \frac{1}{2}L\|x-x_k\|^2 \qquad k = 0, 1, 2, \cdots. \tag{10}$$

Now, we can state our algorithm as Algorithm 3.1. We show that the sequence $\{x_k\}$ generated by our algorithm converges to a local minimizer of the problem (9). In step 3 of the algorithm, we use ACCPM to find the global minimizer of the convexified function.

Algorithm 3.1 A Nonconvex ACCPM Algorithm

1. *Select* $\varepsilon > 0$, $x_0 \in K$, *and an estimate of* L. *Set* $k = 0$.
2. *If* $\|\nabla f(x_k)\| \leq \varepsilon$ *stop.*
3. *Set* x_{k+1} *as the global minimizer of* $\varphi_k(x) = f(x) + \frac{1}{2}L\|x - x_k\|^2$.
4. *Increment* k *and go to* 2.

If $\{x_k\}$ generated by Algorithm 3.1 then we have the following simple properties:

$$\varphi_k(x_k) = f(x_k), \tag{11a}$$
$$\nabla\varphi_k(x_k) = \nabla f(x_k), \tag{11b}$$
$$\nabla\varphi_k(x_{k+1}) = 0, \tag{11c}$$
$$\varphi_k(x_{k+1}) \leq \varphi_k(x) \quad \forall x \in D_{\varphi_k}, \tag{11d}$$
$$\varphi_k(x_{k+1}) < \varphi_k(x_k) \quad \text{if } x_k \text{ is not a stationary point of } f. \tag{11e}$$

The proofs are obvious but we briefly sketch the proof of (11e). If we had equality in (11e), then x_k would also be a minimizer of $\varphi_k(x)$. Therefore, $\nabla\varphi_k(x_k) = 0$ and by (11b) we get $\nabla f(x_k) = 0$ which is a contradiction.

The following lemmas state some more properties of the sequences $\{x_k\}$ and $\{f(x_k)\}$.

Lemma 2 *Suppose* $\{x_k\}$ *is generated by the Algorithm 3.1 with* $x_0 \in K$. *The sequence* $\{x_k\}$ *is a finite sequence if and only if there exits an index* k_0 *such that* $\nabla f(x_k) = 0$ *for all* $k \geq k_0$.

Proof. Suppose that the sequence $\{x_k\}$ is a finite sequence then there exist an index k_0 such that $x_k = x_{k_0}$ for all $k \geq k_0$. In this case, x_{k+1} must be a stationary point of f because x_{k+1} is the minimizer of $\varphi_k(x)$ and $x_{k+1} = x_k$ we get from (11d) and (10)

$$\begin{aligned} 0 = \nabla\varphi_k(x_{k+1}) &= \nabla f(x_{k+1}) + L(x_{k+1} - x_k) \\ &= \nabla f(x_{k+1}) \\ &= \nabla f(x_k). \end{aligned} \tag{12}$$

Now, suppose $\nabla f(x_k) = 0$ for all $k \geq k_0$. Replacing k by k_0 in equation (12) and the fact that $L \neq 0$ we get, $x_{k_0} = x_{k_0+1}$. The same reasoning Shows that $x_{k_0+n} = x_{k_0+n+1}$. Therefore, by induction on n we have $x_{k_0+n} = x_{k_0}$ for all integer $n \geq 1$. □

Lemma 3 *Suppose* $\{x_k\}$ *is generated by the Algorithm 3.1 with* $x_0 \in K$ *and for all* k, $\nabla f(x_k) \neq 0$, *then the sequence* $\{x_k\}$ *is an infinite sequence for which* $\{f(x_k)\}$ *converges (to some* $\bar{f}$*) and*

$$\|x_{k+1} - x_k\| \to 0 \quad as \quad k \to \infty.$$

Proof. Using (11e), we have

$$f(x_{k+1}) \leq f(x_{k+1}) + \frac{1}{2}L\|x_{k+1} - x_k\|^2 = \varphi_k(x_{k+1}) < \varphi_k(x_k) = f(x_k). \tag{13}$$

We conclude that $f(x_k)$ is a decreasing sequence and since f is bounded on the compact set K the sequence $\{f(x_k)\}$ must converge to some $\bar{f}$. From (13) it is clear that $\|x_{k+1} - x_k\| \to 0$ as $k \to \infty$. □

Lemma 4 states that the direction $(x_{k+1} - x_k)$ is a descent direction for f at x_k for all k.

Lemma 4 *Suppose $\{x_k\}$ is generated by the Algorithm 3.1 and for all k, $\nabla f(x_k) \neq 0$, then $(x_{k+1} - x_k)$ is a descent direction for f at x_k.*

Proof. By convexity of $\varphi_k(x)$ we have

$$\varphi_k(x_{k+1}) \geq \varphi_k(x_k) + \nabla\varphi_k(x_k)^T(x_{k+1} - x_k).$$

Since x_{k+1} is the minimizer of $\varphi_k(x)$, using (11e) we get

$$\nabla\varphi_k(x_k)^T(x_{k+1} - x_k) \leq \varphi_k(x_{k+1}) - \varphi_k(x_k) < 0. \tag{14}$$

Using (11b) we have $\nabla\varphi_k(x_k) = \nabla f(x_k)$, substituting this equality in (14) the desired result is obtained. □

Lemma 5 states that the sequence $\{x_k\}$ has a convergent subsequence converging to a local minimizer of f.

Lemma 5 *If $\{x_k\}$ is generated by Algorithm 3.1 with $x_0 \in K$, then $\lim_{k\to\infty} \nabla f(x_k) = 0$ and it has a limit point $\hat{x}$ for which $\nabla f(\hat{x}) = 0$.*

Proof. Since x_{k+1} is the minimizer of $\varphi_k(x)$ we must have $\nabla\varphi_k(x_{k+1}) = 0$. By definition of $\varphi_k(x)$ we get

$$\nabla f(x_{k+1}) + L(x_{k+1} - x_k) = 0$$

consequently

$$\big\|\nabla f(x_{k+1})\big\| = L\|x_{k+1} - x_k\|.$$

Lemma 3 gives the first result. By (13), $\{x_k\} \subset K$, so it must be bounded and hence, have a limit point $\hat{x}$, using continuity of gradient the second result is obtained. □

Remark 1 *For any c such that $f^* < c \leq f(x_0)$ where $x_0 \in K$, define*

$$\mathcal{L}_c = \{\, x \in D_f \mid f(x) \leq c \,\},$$

it is easy to see that

$$\mathcal{L}_c \subset \bigcup_{i\in I} \Omega_i^c, \tag{15}$$

where Ω_i^c $(i \in I)$ is a convex set and I is a finite index set. The reason is that $\mathcal{L}_c \subset \mathcal{L}_{f(x_0)} \subset K$ is a compact set and therefore it can be covered by finitely many convex balls.

The following theorem shows under which conditions x_k converges to a local minimizer of f.

Theorem 4 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable and convexifiable function with convexifier L on $\mathbb{R}^n$. Suppose for some $x_0 \in K$ define $c = f(x_0)$, $\bar{c}$, and $d > 0$ that $\bar{c} \in [f^*, c)$ and some finite index set J,*

$$\|\nabla f(x)\| > d \qquad \forall x \in \bigcup_{i\in I\cap J} \big(\Omega_i^c \setminus \Omega_i^{\bar{c}}\big) \tag{16}$$

and that f is convex on $\Omega_j^{\bar{c}}$ $(j \in J)$. Then $\{x_k\}$ generated by Algorithm 3.1 converges to a local minimizer of (9).

Proof. Suppose that $x_0 \in \Omega_j^c \setminus \Omega_j^{\bar{c}}$ and Algorithm 3.1 generates an infinite sequence $\{x_k\}$. That is $\nabla f(x_k) \neq 0$ for all k. Since $I \cap J$ is a finite index set, we must have an index $i \in I \cap J$ and an infinite subsequence $\{x_k\}_{k\in\mathcal{K}}$ such that $x_k \in \Omega_i^c$ for all $k \in \mathcal{K}$. Now, we show that the sequence $\{x_k\}_{k\in\mathcal{K}}$ finally reaches to the region $\Omega_i^{\bar{c}}$. Suppose by contradiction that for all $k \in \mathcal{K}$ we have $x_k \in \Omega_i^c \setminus \Omega_i^{\bar{c}}$. By definition of x_{k+1}, we have $\nabla\varphi_k(x_{k+1}) = 0$ or equivalently $\nabla f(x_{k+1}) + L(x_{k+1} - x_k) = 0$. Using assumption (16) we get

$$\|x_{k+1} - x_k\| > \frac{d}{L} \qquad \forall k \in \mathcal{K}.$$

This is a contradiction with Lemma 3. Therefor, our sequence finally reaches to the region $\Omega_i^{\bar{c}}$. Using Lemma 5 and the fact that f is a convex function on convex set $\Omega_i^{\bar{c}}$ the proof of the theorem is completed. □

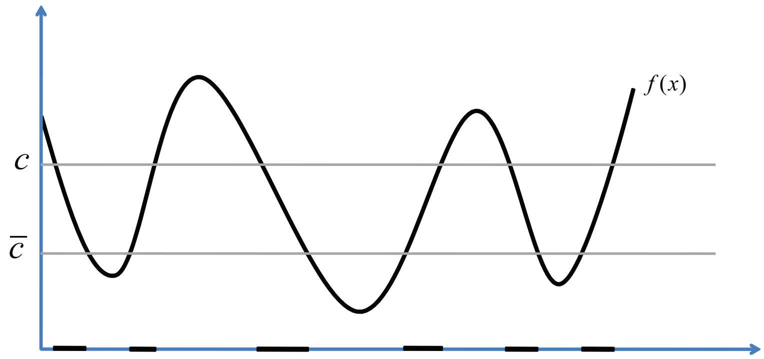

Figure 1: Representation of the set $\cup_{i\in I\cap J}(\Omega_i^c \setminus \Omega_i^{\bar{c}})$

The convergence of the proximal point method to a global minimizer of nonconvex and nondifferentiable objective functions is introduced in [15].

Example 1 *Consider the following function*

$$f(x_1, x_2) = x_1^2 + x_2^2 + 4x_1^2x_2^2.$$

This function is not convex because $\nabla^2 f(-1,1)$ is not positive semi definite. Since $(0,0)$ is a stationary point and $\nabla^2 f(x)$ is positive definite on $B(0,\frac{1}{2}) = \{x \in \mathbb{R}^2 \mid \|x\| < \frac{1}{2}\}$, we conclude that $(0,0)$ is a strict local minimum of f. It is easy to see that $\|\nabla f(x)\| > 1$ for any x outside ball $B(0,\frac{1}{2})$. Therefore, $c = f(1,0) = 1$ and $\bar{c} = \frac{1}{2}$ satisfy the condition of Theorem 4.

The following theorem shows that all nonconstant Lipschitz continuously differentiable function satisfy the conditions of Theorem 4.

Theorem 5 *If $f : \mathbb{R}^n \to \mathbb{R}$ be a nonconstant Lipschitz continuously differentiable function with Lipschitz constant L then the assumptions of the Theorem 4 hold.*

Proof. Suppose that the conditions of the Theorem 4 don't hold. That is, for any $c, \bar{c}, d > 0$, and f^* that $c \in [f^*, \bar{c})$ there exists $\hat{x} \in \Omega_i^c \setminus \Omega_i^{\bar{c}}$ such that $\|\nabla f(\hat{x})\| \le d$. Fix $c, \bar{c}$, and f^* we can find a sequence $\{\hat{x}_k\} \subset \Omega_i^c \setminus \Omega_i^{\bar{c}}$ such that $\|\nabla f(\hat{x}_k)\| \le \frac{1}{k}$. Since $\Omega_i^c \setminus \Omega_i^{\bar{c}}$ is a bounded set and f is continuously differentiable, without loss of generality we can assume $\{\hat{x}_k\}$ converges to $\bar{x} \in \overline{(\Omega_i^c \setminus \Omega_i^{\bar{c}})}$ and $\nabla f(\bar{x}) = 0$.

Now, consider x and y for which $f(x) < f(y)$ and let $c = f(x)$ and $\bar{c} = f(y)$ then by using the above discussion, there exists an $\bar{x}$ such that $f(x) \le f(\bar{x}) \le f(y)$ and $\nabla f(\bar{x}) = 0$. Therefore, we can find a sequence $\{\bar{x}_k\}$ such that $f(\bar{x}_k) \to f(x)$ and $\nabla f(\bar{x}_k) = 0$ for $k = 1, 2, \cdots$. We now, prove that f must be a constant function or equivalently $\nabla f(x) = 0$. let define a sequence of constant functions as follow

$$g_k, g : \mathbb{R}^n \to \mathbb{R} \text{ by } g_k(u) = f(\bar{x}_k) \;\; k = 1, 2, \cdots \text{ and } g(u) = f(x)$$

It is obvious that $g_k(u) \to g(u)$ as $k \to \infty$ and $\nabla g_k(u) = 0$ for all $k = 1, 2, \cdots$. Since

$$\sup_{u\in\mathbb{R}^n} \|\nabla g_k(u)\| = 0$$

we conclude that $\nabla g_k \to 0$ uniformly. Therefore, by Theorem 7.17 [29] we must have $\lim_{k\to\infty} \nabla g_k(u) = \nabla g(u)$. That is $\nabla f(x) = 0$, this is contradict that f is not a constant function. □

When f is convex, the proximal point method that satisfies conditions (17) converges to the local (global) minimizer of f [28]. In our case, according to Theorem 4 the sequence $\{x_k\}$ finally reaches to the convex region $\Omega_i^{\bar{c}}$ and f is convex on this region. Therefore, to use the convergence result of [28] we need conditions (17), although in the proof of Theorem 6 we just need this fact that $\varepsilon_k \to 0$.

Theorem 6 *Suppose the assumptions of Theorem 4 hold and $\{\tilde{x}_k\}$ be a sequence such that*

$$\|\tilde{x}_{k+1} - x_{k+1}\| \le \varepsilon_k \ , \qquad \sum_{k=0}^{\infty} \varepsilon_k < \infty, \tag{17}$$

with $\varepsilon_k > 0$. If we define $\tilde{\varphi}_k(x) = f(x) + \frac{1}{2}L\|x - \tilde{x}_k\|^2$, then $\tilde{x}_k \to x^*$ and $\nabla\tilde{\varphi}_k(\tilde{x}_{k+1}) \to 0$ as $k \to \infty$.

Proof. By Theorem 4 we have $x_{k+1} \to x^*$ and using conditions (17), we get $\varepsilon_k \to 0$ and consequently $\tilde{x}_{k+1} \to x^*$. We also have

$$\begin{aligned} \|\nabla\tilde{\varphi}_k(\tilde{x}_{k+1})\| &= \|\nabla f(\tilde{x}_{k+1}) + L(\tilde{x}_{k+1} - \tilde{x}_k)\| \\ &\le \|\nabla f(\tilde{x}_{k+1})\| + L\|\tilde{x}_{k+1} - \tilde{x}_k\|. \end{aligned} \tag{18}$$

Since $\tilde{x}_{k+1} \to x^*$ and $\nabla f(x)$ is continuous, the right side of (18) must converge to zero. □

3.2 Proximal point method on the potential function

In this section, we extend the results of the previous section to the potential function of the epigraph of $f(x)$, that is we consider the function

$$\psi_k(x) = -\ln(f(x_k) - f(x)) + \frac{1}{2}\hat{L}_k\|x - x_k\|^2 \tag{19}$$

where $\hat{L}_k$ is large enough so that $\psi_k(x)$ is strictly convex. The gradient and the Hessian of the function $\psi_k(x)$ are as follows:

$$\nabla\psi_k(x) = \frac{\nabla f(x)}{f(x_k) - f(x)} + \hat{L}_k(x - x_k)$$

and

$$\nabla^2\psi_k(x) = \frac{\nabla^2 f(x)}{f(x_k) - f(x)} + \frac{\nabla f(x)\nabla f(x)^T}{[f(x_k) - f(x)]^2} + \hat{L}_k I. \tag{20}$$

From (20), it is not difficult to see that if $\hat{L}_k$ satisfies the following inequality

$$\hat{L}_k \ge -\min_{x\in\mathbb{R}^n} \min_i \lambda_i \left(\frac{\nabla^2 f(x)}{f(x_k) - f(x)} + \frac{\nabla f(x)\nabla f(x)^T}{[f(x_k) - f(x)]^2} \right)$$

then $\psi_k(x)$ is convex. The merit of using potential proximal approach is that log barrier $-\ln(f(x_k) - f(x))$ in $\psi_k(x)$ prevents x_{k+1} from being too close to the current minimizer x_k. Like the proximal approach, the term $\hat{L}_k\|x - x_k\|^2$ also prevents choosing new point far from current x_k.

We will show that under some conditions the following procedure generates a sequence converging to a local minimizer of the unconstrained optimization problem (9).

Our procedure starts with $x_0 \in K$ and iterate as follows:

$$x_{k+1} \in \operatorname*{argmin}_{x\in D_{\psi_k}} \psi_k(x), \tag{21}$$

where $D_{\psi_k} = \{x \in \mathbb{R}^n \mid f(x) < f(x_k)\}$. The minimizer x_{k+1} is the global minimizer because ψ_k is strictly convex. We assume also that the level set

$$\Omega_0 = \{\, x \in \mathbb{R}^n \mid f(x) \le f(x_0) \,\} \subset K.$$

If we define $\Omega_k = \{\, x \in \mathbb{R}^n \mid f(x) \le f(x_k) \,\}$, then it is easy to see that for any $k = 0, 1, \ldots$ the inclusion $\Omega_0 \supseteq \Omega_1 \supseteq \ldots$ holds.

Replacing $\varphi_k(x)$ by $\psi_k(x)$ we can design an algorithm analogous to Algorithm 3.1 base on the proximal potential function (19).

Since $x_{k+1} \in D_{\psi_k} = \text{int}\Omega_k$, we must have $f(x_{k+1}) < f(x_k)$. If f is bounded from below we conclude that the sequence $\{f(x_k)\}$ converges to some limit point $\bar{f}$. Since $x_k \in \Omega_k \subset \Omega_0$ and Ω_0 is bounded, $\{x_{k+1} - x_k\}$ is a bounded sequence. Therefore, it has a convergent subsequence. Without loss of generality we can assume that $\{x_{k+1} - x_k\}$ is a convergent sequence. Therefore, we have just proved the following lemma which is analogous to Lemma 3.

Lemma 6 *If $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuously differentiable and bounded from below, then there exist $\bar{f} \in \mathbb{R}$ and $\hat{x} \in \mathbb{R}^n$ such that*

$$f(x_k) \to \bar{f} \quad \textit{and} \quad \|x_{k+1} - x_k\| \to \hat{x} \quad \textit{as} \quad k \to \infty.$$

The following lemma shows that the result of Lemma 4 also holds for our potential function.

Lemma 7 *Let $\{x_k\}$ be generated by the Algorithm 3.1, then $\{x_{k+1} - x_k\}$ is a sequence of descent directions of f at x_{k+1} for all $k = 0, 1, 2, \cdots$.*

Proof. Since $x_{k+1} \in \text{argmin}\psi_k(x)$, we must have $\nabla\psi_k(x_{k+1}) = 0$. That is

$$\frac{\nabla f(x_{k+1})}{f(x_k) - f(x_{k+1})} + \hat{L}(x_{k+1} - x_k) = 0,$$

multiplying both side by the vector $x_{k+1} - x_k$ we have

$$\frac{\nabla f(x_{k+1})^T(x_{k+1} - x_k)}{f(x_k) - f(x_{k+1})} = -\hat{L}\|x_{k+1} - x_k\|^2 \leq 0.$$

Since $f(x_{k+1}) < f(x_k)$ and $x_{k+1} \neq x_k$ we get

$$\nabla f(x_{k+1})^T(x_k - x_{k+1}) > 0,$$

which is the desired result. □

The following lemma shows that the sequence $\{x_k\}$ has a convergent subsequence converging to a stationary point of f.

Lemma 8 *If $\{x_k\}$ is generated by the Algorithm 3.1, then $\lim_{k\to\infty} \nabla f(x_k) = 0$ and it has a limit point $\hat{x}$ for which $\nabla f(\hat{x}) = 0$.*

Proof. Since $\nabla\psi_k(x_{k+1}) = 0$ we have

$$\frac{\nabla f(x_{k+1})}{f(x_k) - f(x_{k+1})} + \hat{L}(x_{k+1} - x_k) = 0$$

or equivalently

$$\nabla f(x_{k+1}) = \hat{L}(x_k - x_{k+1})\big[f(x_k) - f(x_{k+1})\big]. \tag{22}$$

Since the right hand side of the (22) tends to zero as $k \to +\infty$ the first part of the proof is obtained. By continuity of $\nabla f(x)$ and the fact that K is compact and $\{x_k\} \subset K$ the existence of $\hat{x}$ is obvious. □

Keeping Remark 1 in mind the following theorem shows sufficient conditions for the sequence $\{x_k\}$, generated by the Algorithm 3.1, to converge to a local minimizer of f.

Theorem 7 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Suppose for some c, $\bar{c}$, and $d > 0$ that $\bar{c} \in [f^*, c)$ and some finite index set J,*

$$\|\nabla f(x)\| > d \qquad \forall x \in \bigcup_{i\in I\cap J} \left(\Omega_i^c \setminus \Omega_i^{\bar{c}}\right) \tag{23}$$

and that f is convex on $\Omega_j^{\bar{c}}$ $(j \in J)$. Then the sequence $\{x_k\}$ generated by the Algorithm 3.1 converges to a local minimizer of (9).

Proof. The proof basically is the same of the proof of Theorem 4. Suppose that $x_0 \in \Omega_j^c \setminus \Omega_j^{\bar{c}}$ and our algorithm generates an infinite sequence $\{x_k\}$. That is $\nabla f(x_k) \neq 0$ for all k. Since $I \cap J$ is a finite index set, we must have an index $i \in I \cap J$ and an infinite subsequence $\{x_k\}_{k\in\mathcal{K}}$ such that $x_k \in \Omega_i^c$ for all $k \in \mathcal{K}$. Now, we show that the sequence $\{x_k\}_{k\in\mathcal{K}}$ finally reaches to the region $\Omega_i^{\bar{c}}$. Suppose by contradiction that for all $k \in \mathcal{K}$ we have $x_k \in \Omega_i^c \setminus \Omega_i^{\bar{c}}$. By definition of x_{k+1}, we have $\nabla\psi_k(x_{k+1}) = 0$ or equivalently

$$\frac{\nabla f(x_{k+1})}{f(x_k) - f(x_{k+1})} + \hat{L}(x_{k+1} - x_k) = 0,$$

from this we get

$$\begin{aligned}\|x_{k+1} - x_k\| &= \frac{\|\nabla f(x_{k+1})\|}{\hat{L}\big(f(x_k) - f(x_{k+1})\big)} \\ &> \frac{d}{\hat{L}\big(f(x_k) - f(x_{k+1})\big)}.\end{aligned}$$

Now, if $k \to \infty$ we must have $\|x_{k+1} - x_k\| \to +\infty$. This is a contradiction with Lemma 6. Therefore, $\{x_k\}$ finally reaches to the convex region $\Omega_i^{\bar{c}}$. To finish the proof of the theorem, it is enough to note that f is a convex function on the convex set $\Omega_i^{\bar{c}}$ and use Lemma 8. □

The following theorem is the same as Theorem 5 and we omit its proof.

Theorem 8 *If $f : \mathbb{R}^n \to \mathbb{R}$ be a nonconstant Lipschitz continuously differentiable function with Lipschitz constant L, then the assumptions of Theorem 7 hold.*

4 Background on ACCPM

ACCPM is widely used in many areas of optimization both in theory and applications including, integer programming [9], variational inequalities [5, 4], semidefinite programming [26], conic optimization [2], and stochastic programming [1]. ACCPM is an efficient method to compute a center of a polyhedron so called *analytic center*. In this section, we see how analytic center can be used in order to find a minimizer of a convex function.

For any convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$ and any $x, y \in \mathbb{R}^n$, we have

$$\varphi(x) \geq \varphi(y) + \nabla\varphi(y)^T(x - y). \tag{24}$$

Therefore, if $\nabla\varphi(y)^T(x-y) > 0$, then $\varphi(x) > \varphi(y)$ and x cannot be a minimizer of $\varphi(\cdot)$ over $\mathbb{R}^n$. We conclude that for any $y \in \mathbb{R}^n$, the inequality

$$\nabla\varphi(y)^T(x - y) \leq 0$$

is satisfied by any minimizer of φ. This inequality is called a *valid cut* or an *optimality cut* at query point y. In this section, we briefly review how these optimality cuts can be used to find the optimal solution of the unconstrained minimizer of the convex function.

Let the optimality cuts $\nabla\varphi(y^q)^T(x - y^q) \leq 0 \quad (q = 1, 2, \cdots m)$ be generated and

$$\bar{\theta} = \min_q \varphi(y^q).$$

The optimization problems $\min_x \varphi(x)$ and

$$\begin{gathered}\min_{x,t} t \\ \varphi(x) \leq t\end{gathered}$$

are equivalent and the optimal solution of the second problem is contained in the following set called *localization set*

$$\mathcal{L}_{\bar{\theta}} = \{(x,t) \in \mathbb{R}^{n+1} \mid \nabla\varphi(y^q)^T(x - y^q) \leq 0,\ q = 1, 2, \cdots m, \quad t \leq \bar{\theta}\}. \tag{25}$$

We note that since φ is convex, $\mathcal{L}_{\bar{\theta}}$ is an outer approximation of the epigraph of φ bounded by $t \leq \bar{\theta}$. Therefore, the localization set is a bounded polyhedron.

The basic steps of the cutting plane method can be stated as follows:

1. select a new query point (y^{m+1}, t^{m+1}) in the localization set,
2. add the optimality cut corresponding to y^{m+1} to the localization set,
3. test for termination by measuring $\|\nabla\varphi(y^{m+1})\|$.

There are several ways to choose the new query point in the localization set. For example, the center of gravity of the defined as

$$\mathbf{Cg}(\mathcal{L}_{\bar{\theta}}) = \frac{\int_{\mathcal{L}_{\bar{\theta}}} z dz}{\int_{\mathcal{L}_{\bar{\theta}}} dz}$$

and center of max-volume ellipsoid inscribing the localization set [33]. In ACCPM, the new query point is chosen as the analytic center of the bounded polyhedron $\mathcal{L}_{\bar{\theta}}$. The localization set (25) in matrix form can be written as $A^T y \leq c$, where

$$A = \begin{bmatrix} \nabla\varphi(y^1) & \nabla\varphi(y^2) & \cdots & \nabla\varphi(y^m) & e_n \end{bmatrix} \in \mathbb{R}^{n\times(m+1)},$$

$y = (x, t)^T, e_n = (0, 0, \cdots, 0, 1)^T \in \mathbb{R}^n, c_i = \nabla\varphi(y^i)^T x^i,\ i = 1, 2, \cdots, m$ and $c_{m+1} = \bar{\theta}$.

The analytic center of the bounded polyhedron $A^T y \leq c$ exists and is unique if there exists a y such that $A^T y < c$ [13]. This center is defined as the unique solution of the following optimization problem

$$\min_{y,s} -\sum_{i=1}^{m} \log s_i \ \text{ subject to } \ A^T y + s = c. \tag{26}$$

Using the KKT conditions we get the system

$$\begin{aligned} -S^{-1}e + \lambda &= 0, \\ A\lambda &= 0, \\ A^T y + s &= c, \\ s &> 0, \end{aligned}$$

where $S = \mathrm{diag}(s_1, s_2, \cdots, s_m)$. Part of the challenge of computing the analytic center is that we are not given an initial point $s = c - A^T y > 0$. Goffin and Mokhtarian [12] suggested to use an infeasible Newton method. The infeasible Newton method can be started from any y and $s > 0$. For instance, we can start with any y and choose s as

$$s_i = \begin{cases} c_i - a_i^T y, & \text{if } c_i - a_i^T y > 0; \\ 1, & \text{otherwise,} \end{cases} \qquad i = 1, 2, \cdots, m.$$

The Newton step at a point (y, s, λ) is defined by the system of linear equations

$$\begin{bmatrix} 0 & 0 & A \\ 0 & S^{-2} & I \\ A^T & I & 0 \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta s \\ \Delta\lambda \end{bmatrix} = - \begin{bmatrix} A\lambda \\ -S^{-1}e + \lambda \\ A^T y + s - c \end{bmatrix},$$

where $S^{-2} = \mathrm{diag}(s_i^{-2})$ is the Hessian of the Lagrangian of the problem (26). When A has full row rank, the coefficient matrix is nonsingular and the Newton step is obtained by solving this system and using the expressions

$$\Delta y = -(AS^{-2}A^T)^{-1}(AS^{-1}e - AS^{-2}r_p), \tag{27a}$$
$$\Delta s = -A^T\Delta y - r_p, \tag{27b}$$
$$\Delta\lambda = -S^{-2}\Delta s - S^{-1}e - \lambda \tag{27c}$$

where $r_p = A^Ty + s - c$. We can compute Δy from (27a) by finding the Cholesky factorization of $AS^{-2}A^T$ and performs backward and forward substitution. Other alternatives are possible. We could equivalently compute Δy by solving the linear least-squares problem

$$\Delta y \in \operatorname{argmin}_z \left\|S^{-1}A^Tz - S^{-1}r_p + Sg\right\|_2,$$

and then compute Δs and $\Delta\lambda$ form (27b) and (27c). For the convergence results of infeasible Newton method and the algorithm, we refer to [12].

5 ACCPM for proximal and potential function

The following theorem shows how optimality cuts at iteration k are related to the optimality cuts that were previously generated.

Theorem 9 *Let y^q, $(q = 1, 2, \cdots, m)$ be query points at iteration $k-1$ to generate the optimality cuts*

$$\nabla\varphi_k(y^q)^T(x - y^q) \le 0 \qquad q = 1, 2, \cdots, m. \tag{28}$$

Then in the k-th iteration of the Algorithm 3.1 all of the following cuts are valid

$$\left[\nabla\varphi_{k-1}(y^q) + L(x_{k-1} - x_k)\right]^T(x - y^q) \le 0 \qquad q = 1, 2, \cdots, m\,.$$

Proof. At iteration k, we need to solve problem

$$\min_{x\in\mathbb{R}^n} \varphi_k(x). \tag{29}$$

From (24), all cuts (28) are valid. Using the definition of $\varphi_k(x)$ we get

$$\begin{aligned}\nabla\varphi_k(y^q)^T(x - y^q) &= \left[\nabla f(y^q) + L(y^q - x_k)\right]^T(x - y^q)\\ &= \left[\nabla f(y^q) + L(y^q - x_{k-1}) + L(x_{k-1} - x_k)\right]^T(x - y^q)\\ &= \left[\nabla\varphi_{k-1}(y^q) + L(x_{k-1} - x_k)\right]^T(x - y^q).\end{aligned}$$

Therefore, the following cuts are valid

$$\left[\nabla\varphi_{k-1}(y^q) + L(x_{k-1} - x_k)\right]^T(x - y^q) \le 0 \qquad (q = 1, 2, \cdots, m).$$

and the proof is complete. □

The following theorem is analogous of Theorem 9 for the potential function.

Theorem 10 *Let y^q, $(q = 1, 2, \cdots, m)$ be query points at iteration $k-1$ to generate the optimality cuts*

$$\nabla\psi_{k-1}(y^q)^T(x - y^q) \le 0 \qquad (q = 1, 2, \cdots, m).$$

Then the following cuts are valid at iteration k

$$\left[\nabla\psi_{k-1}(y^q) + \hat{L}(\beta - 1)(y^q - x_{k-1}) + \hat{L}\beta(x_{k-1} - x_k)\right]^T(x - y^q) \le 0 \tag{30}$$

for all $q = 1, 2, \cdots, m$ for which $f(y^q) < f(x_k)$, and where we define

$$\beta = \frac{f(x_k) - f(y^q)}{f(x_{k-1}) - f(y^q)}.$$

Proof. Let $Q = \{\, q = 1, 2, \cdots, m \mid f(y^q) < f(x_k) \,\}$. Since $\psi_k(x)$ is convex, the following cuts are valid at the k-th iteration

$$\nabla \psi_k(y^q)^T (x - y^q) \le 0 \qquad (q \in Q).$$

Using the definition of ψ_k, we get, for all $q \in Q$,

$$\nabla \psi_k(y^q)^T (x - y^q) = \left[\frac{\nabla f(y^q)}{f(x_k) - f(y^q)} + \hat{L}(y^q - x_k) \right]^T (x - y^q) \le 0. \tag{31}$$

Since $\beta > 0$ for $q \in Q$, multiplying inequality (31) by β we get the following valid cuts

$$\left[\frac{\nabla f(y^q)}{f(x_{k-1}) - f(y^q)} + \hat{L}\beta(y^q - x_k) \right]^T (x - y^q) \le 0 \qquad (q \in Q).$$

These inequalities are equivalent to

$$\left[\frac{\nabla f(y^q)}{f(x_{k-1}) - f(y^q)} \pm \hat{L}(y^q - x_{k-1}) + \hat{L}\beta(y^q \pm x_{k-1} - x_k) \right]^T (x - x) \le 0 \qquad (q \in Q).$$

After some simple algebraic manipulations, we can see that these cuts are exactly the same as (30). □

6 Calculation of Lipschitz constant

Finding the Lipschitz constant of a function is itself a global optimization problem. In fact, if $f : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz function with Lipschitz constant L. Then

$$L = \max_{\substack{t \in \mathbb{R} \\ d \in \mathbb{R}^n}} \left| \frac{d}{dt} f(x + td) \right|.$$

Existing methods dealing with the Lipschitz constant estimation problem in the literature fall into two categories. First, the analytical form of the objective function and its derivatives are known explicitly. Second, this analytic form is unknown and only the function value can be evaluated. These two categories are known as white box and black box functions, respectively. For the white box problem, Shubert [30] gives a univariate example of Lipschitz constant estimation using the upper bound of the derivative. Mladineo [19] discusses the two dimensional case and choses the upper bound of $\sqrt{(\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2}$ as the estimate.

On the other hand, for the black box problem, one has to find an upper bound on the magnitude of the gradient of the function using only function evaluations. Strongin [31] proposes a method for univariate functions. After k evaluations, the ordered evaluation points $x_1 < x_2 < ... < x_k$ and corresponding function values $f(x_1), f(x_2), \cdots, f(x_k)$ are available and an under-estimation of the Lipschitz constant is given by $\hat{L} = \max \frac{|f(x_i) - f(x_{i-1})|}{x_i - x_{i-1}}$. Strongin's estimate is then obtained by multiplying $\hat{L}$ by a factor $\rho > 1$. There is no guarantee, however, that the estimate $\rho\hat{L}$ is greater than or equal to the true Lipschitz constant. A stochastic method for estimating the Lipschitz constant a univariable function is presented in [11] based on the cumulative distribution function of the random variable $\frac{|f(X) - f(Y)|}{|X - Y|}$.

Nesterov and Polyak [22, 23] use the Lipschitz constant of the objective Hessian to establish a better global complexity bound than that achieved by the steepest descent method in unconstrained optimization. Rather than regular second-order approximation of the objective function, they use the following model to be minimized at iteration k. This model is again a second-order approximation of the objective function, but in degree three. More specifically they use

$$f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)\nabla^2 f(x_k)(x - x_k) + \frac{L}{6}\|(x - x_k)\|^3. \tag{32}$$

they term their approach *cubic regularization of Newton method*.

We use the mid-point acceleration function (4) to estimate Lipschitz constant L in our implementation. Since ACCPM is applied to $\varphi_k(x)$ or $\psi_k(x)$ to compute x_{k+1}, this algorithm generates a sequence of analytic centers converging to x_{k+1}. Let this sequence be $\{x_{kn}\}_{n=0}^{\infty}$, that is $x_{kn} \to x_{k+1}$ as $n \to \infty$. Using acceleration function (4) and $\{x_{kn}\}_{n=0}^{\infty}$ the following estimation L is obtained:

$$L = \max |\Psi(x_{kn}, x_{km})| \qquad x_{kn} \neq x_{km}. \tag{33}$$

According to Lemma 1, it is enough to calculate L after x_1 is calculated. When x_k is close to a local minimum of f, L can be set to (or close to) zero because of the convexity of f around a local minimum. Since by Lemma 3 $(x_{k+1} - x_k) \to 0$, we can use $\|x_{k+1} - x_k\|$ and update the estimated Lipschitz constant L using

$$\frac{\|x_{k+1} - x_k\|}{1 + \|x_{k+1} - x_k\|} L. \tag{34}$$

We can also use $\|\nabla f(x_k)\|$ and update L using

$$\frac{\|\nabla f(x_k)\|}{1 + \|\nabla f(x_k)\|} L. \tag{35}$$

In both (34) and (35), when we are close to a local minimizer the numerators are close to zero and we reduce the effect of L. In our implementation, we update L using (34).

7 Numerical results

In the implementation of our algorithm, we use the implementation of proximal ACCPM given in [32]. We use MATLAB to code and run our algorithm on a Intel Core dual CPU processor $T8300$ @ 2.4 GHZ, with 3 GB of RAM. We tested our algorithm on problems from the CUTEr collection [14]. In our sequential convex programming approach, we need to find a global minimizer of a convex function at each iteration as a subproblem. We don't find the exact minimizer of the subproblem at each iteration, an approximate solution of the minimizer is calculated instead. We run our algorithm on 158 test problems of this set with two versions of our proposed algorithm, ACCPM_AdapTol and ACCPM_FixTol. More specifically, in ACCPM_AdapTol we use an adaptive tolerance and depends on the iteration count k indirectly. When ACCPM_AdapTol satisfies the following tolerance will stop and return x_{k+1} as the global minimizer of the subproblem at iteration k ..

$$\begin{aligned} \text{AdapTol} = & \quad \min\left(10^{-4}, \sqrt{\|\nabla\varphi_(x_k)\|}\right) \|\nabla\varphi_(x_k)\| \\ = & \quad \min\left(10^{-4}, \sqrt{\|\nabla f(x_k)\|}\right) \|\nabla f(x_k)\|. \end{aligned} \tag{36}$$

In ACCPM_FixTol algorithm, the tolerance is kept fixed for all iterations to 10^{-4}. The stopping criteria for termination of the algorithms in both cases is set to

$$\begin{aligned} \text{StopTol} = & \quad 10^{-6} + 10^{-4}\|\nabla\varphi_(x_k)\| \\ = & \quad 10^{-6} + 10^{-4}\|\nabla f(x_k))\|, \end{aligned} \tag{37}$$

where the second equalities in (36) and (37) are valid due to (11b). The two versions of our algorithm are compared with two well known algorithms, steepest descent and nonlinear conjugate gradient method with Armijo line search.

Each iteration of the proximal ACCPM is expensive. On the other hand, each iteration of the steepest descent and conjugate gradient method are cheap. Therefore, one iteration of ACCPM cannot be compared directly with one iteration of conjugate gradient or steepest descent. In order to compensate for this, we let the steepest descent and conjugate gradients algorithm iterations reach to 1000 and we run our algorithm just in 100 iterations and then compare the performance.

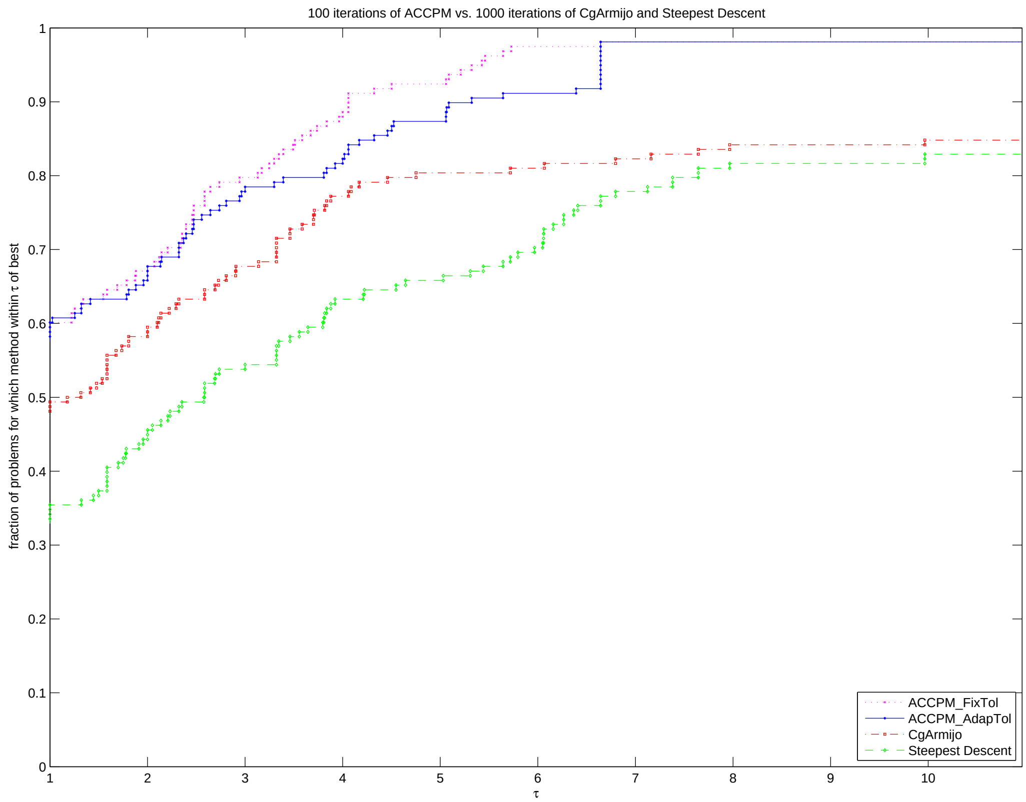

Figure 2: 100 iterations of ACCPM vs. 1000 iterations of steepest descent and conjugate gradient algorithms

The interpretation of the Figure 2 is interesting. It indicates that for any real number $\tau \geq 1$, algorithm ACCPM_AdapTol and ACCPM_FixTol solve more problems within a factor of τ of the best algorithm. More specifically, if $\tau = 1$ or $\tau = 10$ Figure ref1001000 shows that ACCPM_FixTol acts as the best solver on approximately 60% of the problems and for 98% of the problems, the number of iteration that this solver needs is not more than 10 times of the number of iteration of the best solver.

8 Conclusion

The main contribution of this work is to introduce a generalization of proximal ACCPM for nonconvex objective function and accompanying convergent results. We propose two sequences of convex functions and show that the global minimizers of these sequences converge to a local minimizer of the original unconstrained nonconvex objective function f under reasonable assumptions.

The results of the Theorems 9 and 10 provide guidelines for how to reuse old cuts instead of restarting the optimization from scratch. In a sophisticated implementation of the proposed scheme which is in progress, we want to add this capability to current implementation of ACCPM.

In this work, we study ACCPM for the proximal function (10) and potential function (19). An extension of these functions could be the following function

$$\hat{\psi}_k(x) \quad = \quad -\ln\left(f(x_k) - f(x) - \frac{1}{2}\alpha\|x - x_k\|^2\right) + \frac{1}{2}\beta\|x - x_k\|^2. \tag{38}$$

This function is a combination of φ and ψ. When α is chosen such that $f(x_k) - f(x) + \frac{1}{2}\alpha\|x - x_k\|^2$ becomes convex then β can be any nonnegative number. Otherwise, we need to chose large β in (38) in order to $\hat{\psi}_k(x)$ becomes convex. In (38), two parameter α and β control the convexity of $\hat{\psi}_k(x)$. β could be chosen small when α is large. The Hessian of $\hat{\psi}_k(x)$ is

$$\begin{aligned}\nabla^2\hat{\psi}_k(x) &= \frac{\nabla^2 f(x) + \alpha I}{f(x_k) - f(x) - \frac{1}{2}\alpha\|x - x_k\|^2} + \frac{\nabla f(x)\nabla f(x)^T}{\left(f(x_k) - f(x) - \frac{1}{2}\alpha\|x - x_k\|^2\right)^2} \\ &+ \frac{\alpha\nabla f(x)(x - x_k)^T}{\left(f(x_k) - f(x) - \frac{1}{2}\alpha\|x - x_k\|^2\right)^2} + \frac{\alpha(x - x_k)\nabla f(x)^T}{\left(f(x_k) - f(x) - \frac{1}{2}\alpha\|x - x_k\|^2\right)^2} \\ &+ \frac{\alpha^2(x - x_k)(x - x_k)^T}{\left(f(x_k) - f(x) - \frac{1}{2}\alpha\|x - x_k\|^2\right)^2} + \beta I.\end{aligned}$$

From $\nabla^2\hat{\psi}_k(x)$ we understand that it is possible to make $\hat{\psi}_k(x)$ convex by controlling β and α. We also note that in the third term of Hessian the rank one matrix $\nabla f(x)(x-x_k)^T$ has the eigenvalue $\nabla f(x)^T(x-x_k)$ with corresponding eigenvector $\nabla f(x)$. Particularly, if we set $\alpha = 0$ and $\beta = \hat{L}$ in (38) then $\hat{\psi}_k(x)$ is exactly $\psi_k(x)$. Therefore, our study of potential function (19) is an special case of (38).

References

[1] O. Bahn, O. Du Merle, J.-L. Goffin, and J.-Ph. Vial. A cutting plane method from analytic centers for stochastic programming. *Mathematical Programming*, 69:45–73, 1995.

[2] V. L. Basescu and J. E. Mitchell. An analytic center cutting plane approach for conic programming. *Mathematics of Operations Research*, 33(3):529–551, 2008.

[3] G. Chen and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Mathematical Programming*, 64:81–101, 1994.

[4] M. Denault and J.-L. Goffin. Solving variational inequalities with a quadratic cut method: a primal-dual, jacobian-free approach. *Computers & OR*, 31(5):721–743, 2004.

[5] M. Denault and J.-L. Goffin. The analytic-center cutting-plane method for variational inequalities: A quadratic-cut approach. *INFORMS Journal on Computing*, 17(2):192–206, 2005.

[6] E. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91:201–213, 2002.

[7] O. du Merle and J.-P. Vial. Proximal accpm, a cutting plane method for column generation and lagrangian relaxation: application to the p-median problem. Technical report, Logilab, University of Geneva, 40 Bd du Pont dArve, CH-1211, Geneva, Switzerland, 2002.

[8] J. Eckstein and D. P. Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[9] S. Elhedhli and J.-L. Goffin. The integration of an interior-point cutting plane method within a branch-and-price algorithm. *Mathematical Programming*, 100(2):267–294, 2004.

[10] M. Fukushima and H. Mine. A generalized proximal point algorithm for certain nonconvex minimization problems. *International Journal of Systems Science*, 12(1):989–1000, 1981.

[11] B. P. Zhang G. R. Wood. Estimation of the lipschitz constant of a function. *Journal of Global Optimization*, 8(1):91–103, 1996.

[12] J.-L. Goffin and F. S. Mokhtarian. Using the primal dual infeasible newton method in the analytic center method for problems defined by deep cutting planes. *Journal of Optimization Theory and Applications*, 101:35–58, 1999.

[13] J.-L. Goffin and J.-P. Vial. Multiple cuts in the analytic center cutting plane method. *SIAM Journal on Optimization*, 11(1):266–288, 2000.

[14] N. I. M. Gould, D. Orban, and P. L. Toint. CUTEr and SifDec: A constrained and unconstrained testing environment, revisited. *ACM Transactions on Mathematical Software*, 29(4):373–394, 2003.

[15] A. Kaplan and R. Tichatschke. Proximal point methods and nonconvex optimization. *Journal of Global Optimization*, 13(4):389–406, 1998.

[16] J. E. Kelley. The cutting plane method for solving convex programs. *Journal of the SIAM*, 8:703–712, 1960.

[17] C. Lemaréchal and C. Sagastizábal. Variable metric bundle methods: From conceptual to implementable forms. *Mathematical Programming*, 76:393–410, 1997.

[18] B. Martinet. Brève communication. Régularisation d'inéquations variationnelles par approximations successives, 1970.

[19] F. H. Mladineo. An algorithm for finding the global maximum of a multimodal, multivariate function. *Mathematical Programming*, 34:188–200, 1986.

[20] F. S. Mokhtarian and J.-L. Goffin. A nonlinear analytic center cutting plane method for a class of convex programming problems. *SIAM Journal on Optimization*, 8:1108–1131, 1998.

[21] Y. Nesterov and J.-P.Vial. Augmented self-concordant barriers and nonlinear optimization problems with finite complexity. *Mathematical Programming*, 99(1):149–174, 2004.

[22] Y. E. Nesterov. Accelerating the cubic regularization of Newton's method on convex problems. *Mathematical Programming*, 112(1):159–181, 2008.

[23] Y. E. Nesterov and B. T. Polyak. Cubic regularizatin of Newton method and its global performance. *Mathematical Programming*, 108(1):177–205, 2006.

[24] Y. E. Nesterov and J.-P. Vial. Homogeneous analytic center cutting plane methods for convex problems and variational inequalities. *SIAM Journal on Optimization*, 9(3):707–728, 1999.

[25] M. P. Friedlander D. Orban. A primal-dual regularized interior-point method for convex quadratic programs. *Cahier du GERAD*, G-2010-47, Les Cahiers du GERAD HEC Montral 3000, chemin de la Cte-Sainte-Catherine Montral (Qubec) Canada, 2010.

[26] M. R. Oskoorouchi and J.-L. Goffin. An interior point cutting plane method for the convex feasibility problem with second-order cone inequalities. *Mathematics of Operations Research*, 30(1):127–149, 2005.

[27] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.

[28] R.T. Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.

[29] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, NY, USA, 3rd edition, 1976.

[30] B. O. Shubert. A sequential method seeking the global maximum of a function. *SIAM Journal on Numerical Analysis*, 9(3):379–388, 1972.

[31] R.G. Strongin. On the convergence of an algorithm for finding a global extremum. *Engineering Cybernetics*, 11:549–555, 1973.

[32] F. Babonneau, C. Beltran, A. Haurie, C. Tadonki and J.-P. Vial. *Proximal-ACCPM: A Versatile Oracle Based Optimisation Method*, volume 9 of *Advances in Computational Management Science*. Springer Berlin Heidelberg, 2007.

[33] Y. Ye. *Interior Point Algorithms: Theory and Analysis*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, 1997.

[34] S. Zlobec. On the liu-floudas convexification of smooth programs. *Journal of Global Optimization*, 32:401–407, 2005.

[35] S. Zlobec. Characterization of convexifiable functions. *Optimization*, 55(3):251–261, 2006.