

Research Article

A Decomposition Method with Redistributed Subroutine for Constrained Nonconvex Optimization

Yuan Lu,¹ Wei Wang,² Li-Ping Pang,³ and Dan Li³

¹ School of Sciences, Shenyang University, Shenyang 110044, China

² School of Mathematical, Liaoning Normal University, Dalian 116029, China

³ School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

Correspondence should be addressed to Wei Wang; wei_0713@sina.com

Received 6 September 2012; Revised 8 December 2012; Accepted 13 December 2012

Academic Editor: Jean M. Combes

Copyright © 2013 Yuan Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A class of constrained nonsmooth nonconvex optimization problems, that is, piecewise C^2 objectives with smooth inequality constraints are discussed in this paper. Based on the $\mathcal{V}\mathcal{U}$ -theory, a superlinear convergent $\mathcal{V}\mathcal{U}$ -algorithm, which uses a nonconvex redistributed proximal bundle subroutine, is designed to solve these optimization problems. An illustrative example is given to show how this convergent method works on a Second-Order Cone programming problem.

1. Introduction

Consider the following constrained nonsmooth convex program:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_j(x) \leq 0, \quad j \in J = m+1, \dots, l, \end{aligned} \quad (1)$$

where f is convex and piecewise C^2 and $g_j, j \in J$ are convex of class C^2 .

Many approaches are proposed for solving this program. For example, we have converted it into an unconstrained nonsmooth convex program via the exact penalty function in [1]. And we have showed that the objective function of this unconstrained optimization problem is a particular case of function with a primal-dual gradient structure, a notion related to the $\mathcal{V}\mathcal{U}$ -space decomposition. Based on the $\mathcal{V}\mathcal{U}$ -theory, we have designed an algorithm frame which converges with local superlinear rate.

Yet, very little systematic research has been performed on extending this convex program to a nonconvex framework. The purpose of this paper is to study the following nonconvex

program:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_j(x) \leq 0, \quad j \in J = m+1, \dots, l, \end{aligned} \quad (2)$$

where f is piecewise C^2 and $g_j, j \in J$ are of class C^2 . Based on the $\mathcal{V}\mathcal{U}$ -decomposition theory, which is first introduced in [2] for convex functions, and further studied in [3–13]. We give a $\mathcal{V}\mathcal{U}$ -algorithm using a redistributed proximal bundle subroutine to generate a sequence of approximate proximal points. When a primal-dual track exists, these points approximate the primal track points and give the algorithm's \mathcal{V} -steps. And this subroutine also approximates dual track points that are \mathcal{U} -gradients needed for the algorithm's \mathcal{U} -Newton steps. The interest in devising $\mathcal{V}\mathcal{U}$ -algorithm for (2) lies on the "smoothing" effect of \mathcal{U} -subspace and its potential to speed up the algorithm's convergence under certain conditions.

The rest of the paper is organized as follows. Section 2 breaks into two subsections. In the first part, the nonconvex program (2) is transformed into an unconstrained problem by means of the exact penalty function. Based on the Clarke subdifferential of the objective function of this unconstrained problem, we obtain the $\mathcal{V}\mathcal{U}$ -space decomposition. The second part of Section 2 is devoted to deal with the primal-dual

function and its second-order properties. Section 3 designs a conceptual Algorithm 10 and gives its convergence theorem. When a primal-dual track exists, we substitute the \mathcal{V} -step in Algorithm 10 with the redistributed proximal bundle subroutine. In the final section, this algorithm is applied to the Second-Order Cone programming problem to emphasize the theoretical findings.

2. The $\mathcal{V}\mathcal{U}$ -Decomposition Results

2.1. The $\mathcal{V}\mathcal{U}$ -Space Decomposition. In program (2), f is piecewise C^2 . Specifically, for all $x \in R^n$, f is continuous and there exists a finite collection of C^2 functions $f_i : R^n \rightarrow R$, $i \in I$ such that

$$f(x) \in \{f_i(x) \mid i \in I = \{0, \dots, m\}\}. \quad (3)$$

We refer to the function f_i , $i \in I$, as structure functions.

The Clarke subdifferential of f at a point $x \in R^n$, denoted by $\bar{\partial}f(x)$, can be computed in terms of the gradients of the structure functions that are active at x ; see [14, Lemma 1]. More precisely,

$$\bar{\partial}f(x) = \text{conv} \left\{ g \in R^n \mid g = \sum_{i \in I(x)} \alpha_i \nabla f_i(x), \alpha \in \Delta_{|I(x)|} \right\}, \quad (4)$$

where

$$I(x) = \{i \in I \mid f(x) = f_i(x)\} \quad (5)$$

is the set of active indices at x and

$$\Delta_s = \left\{ \alpha \in R^s \mid \alpha_i \geq 0, \sum_{i=1}^s \alpha_i = 1 \right\}. \quad (6)$$

Let $\bar{x} \in R^n$ be a solution of (2). By continuity of the structure functions, there exists a ball $B_\varepsilon(\bar{x}) \subseteq R^n$ such that

$$\forall x \in B_\varepsilon(\bar{x}), \quad I(x) \subseteq I(\bar{x}). \quad (7)$$

For convenience, we assume that the cardinality of $I(\bar{x})$ is $m_1 + 1$ and reorder the structure functions, so that

$$I(\bar{x}) = \{0, \dots, m_1\}. \quad (8)$$

From now on, we consider that

$$\forall x \in B_\varepsilon(\bar{x}), \quad f(x) \in \{f_i(x) \mid i \in I(\bar{x})\}. \quad (9)$$

Let $F(x, \rho)$ denote the exact penalty function of (2) with $g_0(x) = 0$ and $\nabla g_0(x) = 0$, where $\rho > 0$ is a penalty parameter. More precisely,

$$F(x, \rho) = f(x) + \rho G(x), \quad (10)$$

where

$$G(x) = \max \{g_0(x), g_{m+1}(x), \dots, g_l(x)\}. \quad (11)$$

Call

$$J(x) = \{j \in J \mid F(x, \rho) = f(x) + \rho g_j(x)\} \quad (12)$$

the set of indices realizing the max at x .

The following assumptions and definitions will be used in the rest of this paper.

Assumption 1. The set

$$\{\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})\}_{0 \neq i \in I(\bar{x})} \cup \{\nabla g_j(\bar{x})\}_{j \in J(\bar{x})} \quad (13)$$

is linearly independent.

Assumption 2. Given $x^0 \in R^N$ and $M_0 \geq 0$ there exists an open bounded set \mathcal{O} and a function H such that, $\mathcal{L}_0 := \{x \in R^n \mid f(x) \leq f(x^0) + M_0\} \subset \mathcal{O}$, H is lower- C^2 on \mathcal{O} satisfying $H \equiv f$ on \mathcal{L}_0 .

Definition 1 (see [15, Definition 10.29]). The function f is lower- \mathcal{E}^2 on an open set V if for each $\bar{x} \in V$ there is a neighbourhood V of \bar{x} upon which a representation $f(x) = \max_{t \in T} f_t(x)$ holds, where T is a compact set and the functions f_t are of class \mathcal{E}^2 on V such that f_t , ∇f_t , and $\nabla^2 f_t$ depend continuously not just on $x \in V$ but jointly on $(t, x) \in T \times V$.

Lemma 2 (see [19, Proposition 1]). *If Assumption 2 holds, then f is bounded below and prox-bounded.*

Definition 3 (see [16, Definition 1]). Given a lower semicontinuous function f , a point $\bar{x} \in R^n$ where $f(\bar{x})$ is finite and $\bar{\partial}f(\bar{x})$ is nonempty, and an arbitrary subgradient $g \in \bar{\partial}f(\bar{x})$, the orthogonal subspaces

$$\mathcal{V} := \text{lin}(\bar{\partial}f(\bar{x}) - g), \quad \mathcal{U} := \mathcal{V}^\perp \quad (14)$$

define the $\mathcal{V}\mathcal{U}$ -space decomposition, and $R^n = \mathcal{U} \oplus \mathcal{V}$, where \oplus is the direct sum of space decomposition.

Theorem 4. *Suppose Assumption 1 holds. Then one has the following results at \bar{x} :*

(i) *the Clarke subdifferential of $F(\bar{x}, \rho)$ has the following expression:*

$$\bar{\partial}F(\bar{x}, \rho) = \sum_{i \in I(\bar{x})} \alpha_i \nabla f_i(\bar{x}) + \sum_{j \in J(\bar{x})} \beta_j \nabla g_j(\bar{x}), \quad (15)$$

where $\alpha \in \Delta_{|I(\bar{x})|}$; $\beta_j \geq 0$, $j \in J(\bar{x})$ and $\sum_{j \in J(\bar{x})} \beta_j \leq \rho$;

(ii) *let \mathcal{V} denote the subspace generated by the Clarke subdifferential $\bar{\partial}F(\bar{x}, \rho)$. Then*

$$\begin{aligned} \mathcal{V} &= \text{lin} \left\{ \{\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})\}_{0 \neq i \in I(\bar{x})} \cup \{\nabla g_j(\bar{x})\}_{j \in J(\bar{x})} \right\}, \\ \mathcal{U} &= \left\{ j \in Jd \in R^n \mid \langle d, \nabla f_i(\bar{x}) - \nabla f_0(\bar{x}) \rangle \right. \\ &\quad \left. = \langle d, \nabla g_j(\bar{x}) \rangle = 0, \quad 0 \neq i \in I(\bar{x}), \quad j \in J(\bar{x}) \right\}. \end{aligned} \quad (16)$$

Proof. Since $f(x)$ defined in (2) belongs to the PDG-structured family and by Lemma 2.1 in [16] the Clarke subdifferential of $F(x, \rho)$ at \bar{x} can be formulated by

$$\begin{aligned}\bar{\partial}F(\bar{x}, \rho) &= \bar{\partial}f(\bar{x}) + \rho \bar{\partial}G(\bar{x}) \\ &= \bar{\partial}f(\bar{x}) + \rho \operatorname{conv} \left\{ \nabla g_j(\bar{x}) \mid j \in J(\bar{x}) \cup \{0\} \right\} \\ &= \sum_{i \in I(\bar{x})} \alpha_i \nabla f_i(\bar{x}) + \rho \sum_{j \in J(\bar{x}) \cup \{0\}} \lambda_j \nabla g_j(\bar{x}),\end{aligned}\quad (17)$$

where $\alpha \in \Delta_{|I(\bar{x})|}$; $\lambda_j \geq 0$, $j \in J(\bar{x}) \cup \{0\}$, and $\sum_{j \in J(\bar{x}) \cup \{0\}} \lambda_j = 1$. Together with $\nabla g_0(\bar{x}) = 0$, there exists

$$\begin{aligned}\bar{\partial}F(\bar{x}, \rho) &= \sum_{i \in I(\bar{x})} \alpha_i \nabla f_i(\bar{x}) + \rho \left[\lambda_0 \cdot 0 + \sum_{j \in J(\bar{x})} \lambda_j \nabla g_j(\bar{x}) \right] \\ &= \sum_{i \in I(\bar{x})} \alpha_i \nabla f_i(\bar{x}) + \sum_{j \in J(\bar{x})} \beta_j \nabla g_j(\bar{x}),\end{aligned}\quad (18)$$

where $\beta_j = \rho \lambda_j \geq 0$, $j \in J(\bar{x}) \cup \{0\}$ and $\sum_{j \in J(\bar{x})} \beta_j = \rho - \beta_0 \leq \rho$.

Letting $\alpha_0 = 1$; $\alpha_i = 0$, $0 \neq i \in I(\bar{x})$ and $\beta_0 = \rho$; $\beta_j = 0$, $j \in J(\bar{x})$, we have $\nabla f_0(\bar{x}) \in \bar{\partial}F(\bar{x}, \rho)$. Then it follows from the definition of space \mathcal{V} in Definition 3

$$\begin{aligned}\mathcal{V} &= \operatorname{lin} \left(\bar{\partial}F(\bar{x}, \rho) - \nabla f_0(\bar{x}) \right) \\ &= \operatorname{lin} \left\{ \left\{ \nabla f_i(\bar{x}) - \nabla f_0(\bar{x}) \right\}_{0 \neq i \in I(\bar{x})} \cup \left\{ \nabla g_j(\bar{x}) \right\}_{j \in J(\bar{x})} \right\},\end{aligned}\quad (19)$$

and $\mathcal{U} = \mathcal{V}^\perp$ means that the second formula holds. \square

Remark 5. (i) Since the subspaces \mathcal{U} and \mathcal{V} generate the whole space R^n , every vector can be decomposed along its \mathcal{V}/\mathcal{U} -components at \bar{x} . In particular, any $x \in R^n$ can be expressed as

$$R^n \ni x = \bar{x} + u \oplus v = \bar{x} + \bar{U}u + \bar{V}v, \quad (20)$$

where $\bar{V} = [\{\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})\}_{0 \neq i \in I(\bar{x})} \cup \{\nabla g_j(\bar{x})\}_{j \in J(\bar{x})}]$ and $\bar{U} = \bar{V}^\perp$.

(ii) For any $\bar{s} \in \bar{\partial}F(\bar{x}, \rho)$, we have

$$\bar{s} = \bar{s}_{\mathcal{U}} \oplus \bar{s}_{\mathcal{V}} = \bar{U}^T \bar{s} + \bar{V}^T \bar{s}. \quad (21)$$

From Theorem 4(ii), the \mathcal{U} -component of a subgradients $s \in \bar{\partial}F(\bar{x}, \rho)$ is the same as that of any other subgradient at \bar{x} , that is, $\bar{s}_{\mathcal{U}} = \bar{U}^T s$.

2.2. Primal-Dual Function and Its Second-Order Properties. In order to obtain a fast algorithm for (2), we will define an intermediate function. This function is called primal-dual function which is \mathcal{C}^2 about $u \in \mathcal{U}$.

Definition 6 (see [8, Definition 1]). We say that $(\chi(u), \gamma(u))$ is a primal-dual track leading to $(\bar{x}, 0)$, a minimizer of f and zero subgradient pair, if for all $u \in R^{\dim \mathcal{U}}$ small enough

the primal track $\chi(u) = \bar{x} + u \oplus v(u)$,

the dual track $\gamma(u) = \arg \min \{ \|g\|^2 : g \in \partial f(\chi(u)) \}$ (22)

satisfy the following:

- (i) $v : R^{\dim \mathcal{U}} \mapsto R^{\dim \mathcal{V}}$ is a \mathcal{C}^2 function satisfying $\bar{V}v(u) \in W_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}})$ for all $\bar{g} \in \operatorname{ri} \partial f(\bar{x})$,
- (ii) the Jacobian $J_\chi(u)$ is a basis matrix for $\mathcal{V}(\chi(u))^\perp$,
- (iii) the particular \mathcal{U} -Lagrangian $L_{\mathcal{U}}(u; 0)$ is a \mathcal{C}^2 -function.

When we write $v(u)$ we implicitly assume that $\dim \mathcal{U} \geq 1$. If $\dim \mathcal{U} = 0$ we define the primal-dual track to be the point $(\bar{x}, 0)$. If $\dim \mathcal{U} = n$ then $(\chi(u), \gamma(u)) = (\bar{x} + u, \nabla f(\bar{x} + u))$ for all u in a ball about $0 \in R^n$.

Theorem 7. Suppose the Assumption 1 holds. Then for all u small enough, the following hold:

- (i) the nonlinear system, with variable v and the parameter u ,

$$\begin{aligned}f_i(\bar{x} + \bar{U}u + \bar{V}v) - f_0(\bar{x} + \bar{U}u + \bar{V}v) &= 0, \quad 0 \neq i \in I(\bar{x}), \\ g_j(\bar{x} + \bar{U}u + \bar{V}v) &= 0, \quad j \in J(\bar{x}),\end{aligned}\quad (23)$$

has a unique solution $v = v(u)$ and $v : R^{\dim \mathcal{U}} \mapsto R^{\dim \mathcal{V}}$ is a \mathcal{C}^2 function;

- (ii) primal track $\chi(u) := \bar{x} + u \oplus v(u)$ is \mathcal{C}^2 , with

$$J\chi(u) = \bar{U} + \bar{V}Jv(u) \quad (24)$$

and $v(u)$ in (i) is \mathcal{C}^2 , with

$$Jv(u) = -\left(V(u)^T \bar{V}\right)^{-1} V(u)^T \bar{U}, \quad (25)$$

where

$$V(u) = \left[\left\{ \nabla f_i(x) - \nabla f_0(x) \right\}_{0 \neq i \in I(x)} \cup \left\{ \nabla g_j(x) \right\}_{j \in J(x)} \right]. \quad (26)$$

In particular, $\chi(0) = \bar{x}$, $Jv(0) = 0$, and $J\chi(0) = \bar{U}$;

- (iii) $f(\chi(u)) = fi(\chi(u))$, $i \in I(\bar{x})$ and $G(\chi(u)) = 0$.

Proof. Items (i) and (ii) follow from the assumption that f_i , g_j are \mathcal{C}^2 along the lines of [5, Theorem 5.1] and applying a Second-Order Implicit Function Theorem; see [17, Theorem 2.1]. The conclusion of (iii) can be obtained in terms of (i) and the definitions of $G(x)$ and $\chi(u)$. \square

Lemma 8 (see [7, Theorem 4.5]). Given $\bar{g} \in \bar{\partial}F(\bar{x}, \rho)$, the system with $\{\alpha_i(u)\}_{i \in I(\bar{x})}$, $\{\beta_j(u)\}_{j \in J(\bar{x}) \cup \{0\}}$

$$\bar{V}^T \left[\sum_{i \in I(\bar{x})} \alpha_i(u) \nabla f_i(\chi(u)) + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla g_j(\chi(u)) - \bar{g} \right] = 0,$$

$$\sum_{i \in I(\bar{x})} \alpha_i(u) = 1,$$

$$\sum_{j \in J(\bar{x}) \cup \{0\}} \beta_j(u) = \rho,$$
(27)

has a unique solution. In particular, $\alpha_i(0) = \bar{\alpha}_i$, $i \in I(\bar{x})$ and $\beta_j(0) = \bar{\beta}_j$, $j \in J(\bar{x}) \cup \{0\}$.

The following theorem gives the definition and properties of primal-dual function.

Theorem 9. Given $\bar{g} \in \bar{\partial}F(\bar{x}, \rho)$ and supposing Assumption 1 holds, consider the primal-dual function:

$$L_I(u, \bar{g}_{\mathcal{V}}) := F(\chi(u), \rho) - \langle \bar{g}_{\mathcal{V}}, v(u) \rangle_{\mathcal{V}}. \quad (28)$$

Then for u small enough, the following assertions are true:

- (i) L_I is a C^2 function of u ;
- (ii) the gradient of L_I is given by

$$\nabla L_I(u; \bar{g}_{\mathcal{V}}) = \bar{U}^T g(u), \quad (29)$$

where

$$g(u) = \sum_{i \in I(\bar{x})} \alpha_i(u) \nabla f_i(\chi(u)) + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla g_j(\chi(u)). \quad (30)$$

In particular, when $u = 0$, one has

$$\nabla L_I(0; \bar{g}_{\mathcal{V}}) = \bar{U}^T g(0) = \bar{U}^T \bar{g}, \quad (31)$$

where

$$g(0) = \sum_{i \in I(\bar{x})} \bar{\alpha}_i \nabla f_i(\bar{x}) + \sum_{j \in J(\bar{x})} \bar{\beta}_j \nabla g_j(\bar{x}); \quad (32)$$

- (iii) the \mathcal{U} -Hessian of F is given by

$$\nabla^2 L_I(u; \bar{g}_{\mathcal{V}}) = J\chi(u)^T M(u) J\chi(u), \quad (33)$$

where

$$M(u) = \sum_{i \in I(\bar{x})} \alpha_i(u) \nabla^2 f_i(\chi(u)) + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla^2 g_j(\chi(u)). \quad (34)$$

In particular, when $u = 0$, one has

$$\nabla^2 L_I(0; \bar{g}_{\mathcal{V}}) = \bar{U}^T M(0) \bar{U}, \quad (35)$$

where

$$M(0) = \sum_{i \in I(\bar{x})} \bar{\alpha}_i \nabla^2 f_i(\bar{x}) + \sum_{j \in J(\bar{x})} \bar{\beta}_j \nabla^2 g_j(\bar{x}). \quad (36)$$

Proof. (i) From Theorem 7(iii), we have

$$\begin{aligned} L_I(u; \bar{g}_{\mathcal{V}}) &= F(\chi(u), \rho) - \langle \bar{g}_{\mathcal{V}}, v(u) \rangle_{\mathcal{V}} \\ &= f_i(\chi(u)) - \langle \bar{g}_{\mathcal{V}}, v(u) \rangle_{\mathcal{V}}. \end{aligned} \quad (37)$$

Since f_i and $v(u)$ are C^2 , (i) holds.

(ii) In view of the chain rule, differentiating the following system with respect to u :

$$\begin{aligned} L_I(u; \bar{g}_{\mathcal{V}}) &= f_i(\chi(u)) - \langle \bar{g}_{\mathcal{V}}, v(u) \rangle_{\mathcal{V}}, \\ g_j(\chi(u)) &= 0, \quad j \in J(\bar{x}), \end{aligned} \quad (38)$$

we have

$$\begin{aligned} \nabla L_I(u; \bar{g}_{\mathcal{V}}) &= J\chi(u)^T \nabla f_i(\chi(u)) - Jv(u)^T \bar{V}^T \bar{g}, \\ J\chi(u)^T \nabla g_j(\chi(u)) &= 0, \quad j \in J(\bar{x}). \end{aligned} \quad (39)$$

Multiplying each equation by the appropriate $\alpha_i(u)$ and $\beta_j(u)$, respectively, summing the results, and using the fact that $\sum_{i \in I(\bar{x})} \alpha_i(u) = 1$ yields

$$\nabla L_I(u; \bar{g}_{\mathcal{V}}) = J\chi(u)^T g(u) - Jv(u)^T \bar{V}^T \bar{g}, \quad (40)$$

where

$$g(u) = \sum_{i \in I(\bar{x})} \alpha_i(u) \nabla f_i(\chi(u)) + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla g_j(\chi(u)). \quad (41)$$

Using the transpose of the expression of $J\chi(u)$, we get

$$\nabla L_I(u; \bar{g}_{\mathcal{V}}) = \bar{U}^T g(u) + Jv(u)^T \bar{V}^T (g(u) - \bar{g}), \quad (42)$$

which together with (6.11) in [5] yields the desired result.

In particular, if $u = 0$, then $v(0) = 0$ and $\chi(0) = \bar{x}$. It follows from Remark 5(ii) that

$$\nabla L_I(0; \bar{g}_{\mathcal{V}}) = \bar{U}^T g(0) = \bar{U}^T \bar{g}, \quad (43)$$

where

$$g(0) = \sum_{i \in I(\bar{x})} \bar{\alpha}_i \nabla f_i(\bar{x}) + \sum_{j \in J(\bar{x})} \bar{\beta}_j \nabla g_j(\bar{x}). \quad (44)$$

- (iii) Differentiating (ii) with respect to u , we obtain

$$\begin{aligned} \nabla^2 L_I(u; \bar{g}_{\mathcal{V}}) &= \bar{U}^T M(u) J\chi(u) \\ &+ \bar{U}^T \left[\sum_{i \in I(\bar{x})} \alpha_i(u) \nabla f_i(\chi(u)) J\alpha_i(u) \right. \\ &\quad \left. + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla g_j(\chi(u)) J\beta_j(u) \right], \end{aligned} \quad (45)$$

where

$$M(u) = \sum_{i \in I(\bar{x})} \alpha_i(u) \nabla^2 f_i(\chi(u)) + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla^2 g_j(\chi(u)). \quad (46)$$

According to the proof of Theorem 6.3 in [5], we get

$$\begin{aligned} & \sum_{i \in I(\bar{x})} \alpha_i(u) \nabla f_i(\chi(u)) J\alpha_i(u) + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla g_j(\chi(u)) J\beta_j(u) \\ &= -V(u) \left(\bar{V}^T V(u) \right)^{-1} \bar{V}^T M(u) J\chi(u). \end{aligned} \quad (47)$$

Then

$$\begin{aligned} \nabla^2 L_I(u; \bar{g}_{\mathcal{V}}) &= \bar{U}^T M(u) J\chi(u) \\ &\quad - \bar{U}^T V(u) \left(\bar{V}^T V(u) \right)^{-1} \bar{V}^T M(u) J\chi(u) \\ &= \bar{U}^T M(u) J\chi(u) + Jv(u)^T \bar{V}^T M(u) J\chi(u) \\ &= \left[\bar{U}^T + Jv(u)^T \bar{V}^T \right] M(u) J\chi(u) \\ &= J\chi(u)^T M(u) J\chi(u), \end{aligned} \quad (48)$$

when $u = 0$,

$$\nabla^2 L_I(0; \bar{g}_{\mathcal{V}}) = \bar{U}^T M(0) \bar{U},$$

$$\text{where } M(0) = \sum_{i \in I(\bar{x})} \bar{\alpha}_i(u) \nabla^2 f_i(\bar{x}) + \sum_{j \in J(\bar{x})} \bar{\beta}_j(u) \nabla^2 g_j(\bar{x}). \quad (49) \quad \square$$

3. Algorithm and Convergence Analysis

Supposing $0 \in \bar{\partial}F(\bar{x}, \rho)$, we give an algorithm frame which can solve (2). This algorithm makes a step in the \mathcal{V} -subspace, followed by a \mathcal{U} -Newton step in order to obtain superlinear convergence rate.

Algorithm 10 (algorithm frame).

Step 0 (Initialization). Given $\varepsilon > 0$, choose a starting point $x^{(0)}$ close to \bar{x} enough, and a Clarke subgradient $\tilde{g}^{(0)} \in \bar{\partial}F(x^{(0)}, \rho)$, set $k = 0$.

Step 1. Stop if

$$\|\tilde{g}^{(k)}\| \leq \varepsilon. \quad (50)$$

Step 2. Find the active index set $I(\bar{x})$ and $J(\bar{x})$.

Step 3. Construct $\mathcal{V}\mathcal{U}$ -decomposition at \bar{x} , that is, $R^n = \mathcal{V} \oplus \mathcal{U}$. Compute

$$\nabla^2 L_I(0; 0) = \bar{U}^T M(0) \bar{U}, \quad (51)$$

where

$$M(0) = \sum_{i \in I(\bar{x})} \bar{\alpha}_i \nabla^2 f_i(\bar{x}) + \sum_{j \in J(\bar{x})} \bar{\beta}_j \nabla^2 g_j(\bar{x}). \quad (52)$$

Step 4. Perform \mathcal{V} -step. Compute $\delta_{\mathcal{V}}^{(k)}$ which denotes $v(u)$ in (23) and set $\bar{x}^{(k)} = x^{(k)} + 0 \oplus \delta_{\mathcal{V}}^{(k)}$.

Step 5. Perform \mathcal{U} -step. Compute $\delta_{\mathcal{U}}^{(k)}$ from the system

$$\bar{U}^T M(0) \bar{U} \delta_{\mathcal{U}} + \bar{U}^T \tilde{g}^{(k)} = 0, \quad (53)$$

where

$$\begin{aligned} & \sum_{i \in I(\bar{x})} \alpha_i(u) \nabla f_i(\bar{x}^{(k)}) + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla g_j(\bar{x}^{(k)}) \\ &= \tilde{g}^{(k)} \in \bar{\partial}F(\bar{x}^{(k)}, \rho), \end{aligned} \quad (54)$$

is such that $\bar{V}^T \tilde{g}^{(k)} = 0$. Compute $x^{(k+1)} = \bar{x}^{(k)} + \delta_{\mathcal{U}}^{(k)} \oplus 0 = x^{(k)} + \delta_{\mathcal{U}}^{(k)} \oplus \delta_{\mathcal{V}}^{(k)}$.

Step 6 (update). Set $k = k + 1$, and return to Step 1.

Theorem 11. *Suppose the starting point $x^{(0)}$ close to \bar{x} enough and $0 \in \text{ri} \bar{\partial}F(\bar{x}, \rho)$, $\nabla^2 L_I(0; 0) \succ 0$. Then the iteration points $\{x^{(k)}\}_{k=1}^{\infty}$ generated by the algorithm converge and satisfy*

$$\|x^{(k+1)} - \bar{x}\| = o(\|x^{(k)} - \bar{x}\|). \quad (55)$$

Proof. Let $u^{(k)} = (x^{(k)} - \bar{x})_{\mathcal{U}}$, $v^{(k)} = (x^{(k)} - \bar{x})_{\mathcal{V}} + \delta_{\mathcal{V}}^{(k)}$. It follows from Theorem 7(i) that

$$\begin{aligned} \|(x^{(k+1)} - \bar{x})_{\mathcal{V}}\| &= \|(\bar{x}^{(k)} - \bar{x})_{\mathcal{V}}\| \\ &= o(\|x^{(k)} - \bar{x}\|) = o(\|x^{(k)} - \bar{x}\|). \end{aligned} \quad (56)$$

Since $\nabla^2 L_I(0; 0)$ exists and $\nabla L_I(0; 0) = 0$, we have from the definition of \mathcal{U} -Hessian matrix that

$$\begin{aligned} \nabla L_I(u^{(k)}; 0) &= \bar{U}^T \tilde{g}^{(k)} \\ &= 0 + \nabla^2 L_I(0; 0) u^{(k)} + o(\|u^{(k)}\|_{\mathcal{U}}). \end{aligned} \quad (57)$$

By virtue of (53), we have $\nabla^2 L_I(0; 0)(u^{(k)} + \delta_{\mathcal{U}}^{(k)}) = o(\|u^{(k)}\|_{\mathcal{U}})$. It follows from the hypothesis $\nabla^2 L_I(0; 0) \succ 0$ that $\nabla^2 L_I(0; 0)$ is invertible and hence $\|u^{(k)} + \delta_{\mathcal{U}}^{(k)}\| = o(\|u^{(k)}\|_{\mathcal{U}})$. In consequence, one has

$$\begin{aligned} (x^{(k+1)} - \bar{x})_{\mathcal{U}} &= (x^{(k+1)} - \bar{x}^{(k)})_{\mathcal{U}} \\ &\quad + (\bar{x}^{(k)} - x^{(k)})_{\mathcal{U}} + (x^{(k)} - \bar{x})_{\mathcal{U}} \\ &= u^{(k)} + \delta_{\mathcal{U}}^{(k)} = o(\|u^{(k)}\|_{\mathcal{U}}) \\ &= o(\|x^{(k)} - \bar{x}\|). \end{aligned} \quad (58)$$

The proof is completed by combining (56) and (58). \square

Since Algorithm 10 relies on knowing the subspaces \mathcal{U} and \mathcal{V} and converges only locally, it needs significant modification for implemental. Our $\mathcal{V}\mathcal{U}$ -algorithm defined below finds \mathcal{V} -step by approximating equivalent proximal points.

Given a positive scalar parameter λ , the proximal point function depending on f is defined by

$$p_\lambda(x) := \arg \min_{p \in \mathbb{R}^n} \left\{ f(p) + \frac{\lambda}{2} \|p - x\|^2 \right\} \quad \text{for } x \in \mathbb{R}^n. \quad (59)$$

If Assumption 2 holds, then the proximal point p_λ is single-valued; see [18, Theorem 1].

Corresponding to the primal track, the dual track is defined by

$$\gamma(u) = \arg \min \left\{ \|g\|^2 \mid g \in \bar{\partial}F(\chi(u), \rho) \right\}. \quad (60)$$

For its properties, one can refer to [16].

The next theorem shows that \mathcal{V} -steps in Algorithm 10 can be replaced by proximal steps, at least in the locality of a minimizer, if Assumptions 1 and 2 hold.

Theorem 12. *Suppose that Assumptions 1 and 2 hold, and that $0 \in \text{ri } \bar{\partial}G(\bar{x})$. Then for all $\lambda > 0$ sufficiently large and for any sequence $x_k \rightarrow \bar{x}$, one has*

$$p_\lambda(x_k) = \chi(u_k), \quad (61)$$

$$\gamma_k = \arg \min \left\{ \|g\|^2 \mid g \in \bar{\partial}F(p_\lambda(x_k), \rho) \right\},$$

for all large k , where $u_k := (p_\lambda(x_k) - \bar{x})_{\mathcal{U}}$.

Proof. Since g_j are C^2 , $j = m+1, \dots, l$ and g_j are lower- C^2 . Functions defined by sums, maximums are lower- C^2 [15, Example 10.35]; therefore F is lower- C^2 . From Lemma 2 and [15, Proposition 13.33], we have F is prox-bounded and -regular. Applying the definition of F and the fact f, g_j are C^2 , respectively, we have that F is subdifferentially regular. So f is a function with pdg structure satisfying strong transversality and prox-regular at \bar{x} , and $0 \in \text{ri } \bar{\partial}G(\bar{x})$, $0 \in \text{ri } \bar{\partial}F(\bar{x}, \rho)$, by [16, Theorem 5.3] we get the result. \square

In order to define a nonconvex $\mathcal{V}\mathcal{U}$ -algorithm for (2) problem, we will use a nonconvex bundle method to approximate proximal points. Many practically nonconvex bundle algorithms are modifications of some convex forerunner, with a fixed model function. Basically, such fixes consist in redefining linearization errors to enforce nonnegativity. However, a redistributed proximal bundle method for nonconvex optimization [19] based on [18] is a different picture. This work proposes an approach based on generating cutting-planes models, not of the objective function as most bundle methods do, but of a local convexification of the objective function. They deal with the augmented functions at x :

$$F_{\eta_n}^x(\cdot, \rho) := F(\cdot, \rho) + \frac{1}{2} \eta_n \|\cdot - x\|^2, \quad (62)$$

where η_n denotes convexification parameter; in the following μ_n is model prox-parameter and λ_n strands for the prox-parameter, which satisfies $\lambda_n = \eta_n + \mu_n$.

Bundle subroutine accumulates information from past points x_i in the form

$$\bigcup_{i \in \mathcal{B}} (e_i, d_i, \Delta_i, g^i), \quad (63)$$

where \mathcal{B} is some index set containing an index i such that $x_i = x$, $e_i = F(x, \rho) - (F(x_i, \rho) + \langle g^i, x - x_i \rangle)$, $d_i = (1/2) \|x_i - x\|^2$, $\Delta_i = x_i - x$, and $g^i \in \partial F(x_i, \rho)$. This information is used at each iteration to define a V -model underestimating $F_{\eta_n}^x$ via the cutting-plane function.

$$\varphi_n(y) = F(x, \rho) + \max_{i \in \mathcal{B}} \left\{ - (e_i + \eta_n d_i) + \langle (g^i + \eta_n \Delta_i), y - x \rangle \right\}. \quad (64)$$

To approximate a proximal point we solve a first quadratic programming subproblem $\chi_{\text{-QP}}$, which has the following form and properties.

The problem $\chi_{\text{-QP}}$

$$\begin{aligned} \min \quad & r + \frac{1}{2} \mu_n \|p - x\|^2 \\ \text{s.t.} \quad & r \geq F(x, \rho) - (e_i + \eta_n d_i) \\ & + \langle (g^i + \eta_n \Delta_i), p - x \rangle, \quad i \in \mathcal{B} \end{aligned} \quad (65)$$

has a dual

$$\begin{aligned} \min \quad & \frac{1}{2\mu_n} \left\| \sum_{i \in \mathcal{B}} \alpha_i (g^i + \eta_n \Delta_i) \right\|^2 + \sum_{i \in \mathcal{B}} \alpha_i (e_i + \eta_n d_i) \\ \text{s.t.} \quad & \alpha_i \geq 0, \quad i \in \mathcal{B}, \quad \sum_{i \in \mathcal{B}} \alpha_i = 1. \end{aligned} \quad (66)$$

Their respective solutions, denoted by (\hat{r}, \hat{p}) and $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_{|\mathcal{B}|})$, satisfy

$$\hat{r} = \varphi_n(\hat{p}), \quad \hat{p} = x - \frac{1}{\mu_n} \hat{g}, \quad \text{where } \hat{g} := \sum_{i \in \mathcal{B}} \hat{\alpha}_i (g^i + \eta_n \Delta_i). \quad (67)$$

In addition, $\hat{\alpha}_i = 0$ for all $i \in \mathcal{B}$ such that

$$\begin{aligned} \hat{r} &> F(x, \rho) - (e_i + \eta_n d_i) + \langle (g^i + \eta_n \Delta_i), \hat{p} - x \rangle, \\ \varphi_n(\hat{p}) &= F(x, \rho) - \sum_{i \in \mathcal{B}} \hat{\alpha}_i (e_i + \eta_n d_i) \\ &+ \left\langle \sum_{i \in \mathcal{B}} \hat{\alpha}_i (g^i + \eta_n \Delta_i), \hat{p} - x \right\rangle \\ &= F(x, \rho) - \sum_{i \in \mathcal{B}} \hat{\alpha}_i (e_i + \eta_n d_i) - \frac{1}{\mu_n} \|\hat{g}\|^2. \end{aligned} \quad (68)$$

For convenience, in the sequel we denote the output of these calculations by

$$(\widehat{p}, \widehat{r}) = \chi_{\text{-QP}}(\mu_n, \eta_n, \{(e_i + \eta_n d_i, g^i + \eta_n \Delta_i)\}_{i \in \mathcal{B}}). \quad (69)$$

The vector \widehat{p} is an estimate of a proximal point and, hence, approximates a primal track point when the latter exists. To proceed further we define new data, corresponding to a new index i_+ , by letting $x_{i_+} := \widehat{p}$ and computing

$$\widehat{\varepsilon} := F(\widehat{p}, \rho) + \frac{\eta_n}{2} \|\widehat{p} - x\|^2 - \varphi_n(\widehat{p}). \quad (70)$$

An approximate dual track point, denoted by \widehat{s} , is constructed by solving a second quadratic problem, which depends on a new index set:

$$\begin{aligned} \widehat{\mathcal{B}} := \{ & i \in \mathcal{B} : \widehat{r} = F(x, \rho) - (e_i + \eta_n d_i) \\ & + \langle (g^i + \eta_n \Delta_i), \widehat{p} - x \rangle\} \cup \{i_+\}. \end{aligned} \quad (71)$$

The second quadratic programming problem, denoted by $\gamma\text{-QP}$,

$$\begin{aligned} \min \quad & r + \frac{1}{2} \|p - x\|^2 \\ \text{s.t.} \quad & r \geq \langle (g^i + \eta_n \Delta_i), p - x \rangle, \quad i \in \widehat{\mathcal{B}}, \end{aligned} \quad (72)$$

has a dual problem similar to (66),

$$\begin{aligned} \min \quad & \frac{1}{2} \left\| \sum_{i \in \widehat{\mathcal{B}}} \alpha_i (g^i + \eta_n \Delta_i) \right\|^2 \\ \text{s.t.} \quad & \alpha_i \geq 0, \quad i \in \widehat{\mathcal{B}}, \quad \sum_{i \in \widehat{\mathcal{B}}} \alpha_i = 1. \end{aligned} \quad (73)$$

Similar to (67), the respective solutions, denoted by (\bar{r}, \bar{p}) and $\bar{\alpha}$, satisfy

$$\bar{p} - x = -\widehat{s} \quad \text{where} \quad \widehat{s} := \sum_{i \in \widehat{\mathcal{B}}} \bar{\alpha}_i (g^i + \eta_n \Delta_i). \quad (74)$$

Let an active index set be defined by

$$\widehat{\mathcal{B}}_{\text{act}} := \{i \in \widehat{\mathcal{B}} : \bar{r} = (g^i + \eta_n \Delta_i)^T (\bar{p} - x)\}. \quad (75)$$

Then, from (74), $\bar{r} = -(g^j + \eta_n \Delta_j)^T \widehat{s}$, $j \in \widehat{\mathcal{B}}_{\text{act}}$, so

$$\left[(g^j + \eta_n \Delta_j) - (g^l + \eta_n \Delta_l) \right]^T \widehat{s} = 0, \quad (76)$$

for all such i and for a fixed $l \in \widehat{\mathcal{B}}_{\text{act}}$. Define a full column rank matrix \widehat{V} by choosing the largest number of indices i satisfying (76) such that the corresponding vectors $(g^i + \eta_n \Delta_i) - (g^l + \eta_n \Delta_l)$ are linearly independent and by letting these vectors be the columns of \widehat{V} . Then let \widehat{U} be a matrix whose columns form an orthonormal basis for the null-space of \widehat{V}^T . And let $\widehat{U} = I$ if \widehat{V} is vacuous.

For convenience, in the sequel we denote the output from these calculation by

$$(\widehat{s}, \widehat{U}) = \gamma\text{-QP}(\{g^i + \eta_n \Delta_i\}_{i \in \widehat{\mathcal{B}}}). \quad (77)$$

The bundle subprocedure is terminated and \widehat{p} is declared to be approximation of $p_\lambda(x)$ if

$$\widehat{\varepsilon} \leq \|\widehat{s}\|^2. \quad (78)$$

Otherwise, \mathcal{B} above is replaced by $\widehat{\mathcal{B}}$ and new iterate data are computed by solving the updated two quadratic programming problems above.

Now we consider a heuristic algorithm depending on the $\mathcal{V}\mathcal{U}$ -theory and the primal-dual track point approximations above.

Algorithm 13 (nonconvex $\mathcal{V}\mathcal{U}$ -Algorithm for (2)).

Step 0. Select initial starting point p_0 and positive parameter $M_0, \lambda_0, \xi, \rho$, a convexification growth parameter $\Gamma > 1$. Compute the oracle values $F(p_0, \rho)$ and $g^0 \in \partial F(p_0, \rho)$, and the additional bundle information $(e_0, d_0, \Delta_0) := (0, 0, 0)$, with $(\mu_0, \eta_0) = (\lambda_0, 0)$. Also, let U_0 be a matrix with orthonormal n -dimensional columns estimating an optimal \mathcal{U} -basis. Set $s_0 = g^0$ and $k := 0$.

Step 1. Stop if $\|s_k\|^2 \leq \xi$.

Step 2. Choose an $n_k \times n_k$ positive definite matrix H_k , where n_k is the number of columns of U_k .

Step 3. Compute an \mathcal{U} -Newton step by solving the linear system

$$H_k \delta_k = -U_k^T s_k. \quad (79)$$

Set $x'_{k+1} := p_k + U_k \delta_k$.

Step 4. Initialize \mathcal{B} and run the bundle subprocedure with $x = x'_{k+1}$. Compute recursively,

$$(\widehat{p}, \widehat{r}) = \chi_{\text{-QP}}(x, \{(e_i + \eta_n d_i, g^i + \eta_n \Delta_i)\}_{i \in \mathcal{B}}),$$

$$\widehat{\varepsilon} = F(\widehat{p}, \rho) + \frac{\eta_n}{2} \|\widehat{p} - x\|^2 - \varphi_n(\widehat{p}), \quad \widehat{\mathcal{B}} \text{ given by (71),}$$

$$(\widehat{s}, \widehat{U}) = \gamma\text{-QP}(\{g^i + \eta_n \Delta_i\}_{i \in \widehat{\mathcal{B}}}) \quad (80)$$

until satisfaction of (78). Then set $(\varepsilon'_{k+1}, p'_{k+1}, s'_{k+1}, U'_{k+1}) := (\widehat{\varepsilon}, \widehat{p}, \widehat{s}, \widehat{U})$.

Step 5. If

$$F(p'_{k+1}, \rho) - F(p_k, \rho) \leq -\|s'_{k+1}\|^2, \quad (81)$$

then set

$$\begin{aligned} & (x_{k+1}, \varepsilon_{k+1}, p_{k+1}, s_{k+1}, U_{k+1}) \\ & := (x'_{k+1}, \varepsilon'_{k+1}, p'_{k+1}, s'_{k+1}, U'_{k+1}). \end{aligned} \quad (82)$$

TABLE I: Numerical results.

n	$\#f/g$	$\ x^k - \bar{x}\ $	$\ x^0 - \bar{x}\ $
40	5	1.0084×10^{-12}	3.3503×10^3
100	10	5.0028×10^{-12}	5.8699×10^3
200	15	6.7345×10^{-12}	8.1666×10^3
500	20	1.2895×10^{-11}	1.3440×10^4
1000	25	1.6550×10^{-11}	1.8329×10^4

And apply rule

$$\begin{aligned} \eta_{n+1} &:= \eta_n \quad \text{if } \eta_{n+1}^{\min} \leq \eta_n, \\ \eta_{n+1} &:= \Gamma \eta_{n+1}^{\min}, \quad \lambda_n := \mu_n + \eta_{n+1} \quad \text{if } \eta_{n+1}^{\min} > \eta_n, \end{aligned} \quad (83)$$

where

$$\eta_{n+1}^{\min} := \max_{\substack{i \in \mathcal{B} \\ d_i > 0}} -\frac{e_i}{d_i}. \quad (84)$$

Otherwise, execute a line search on the line determined by p_k and p'_{k+1} to find x_{k+1} thereon satisfying $f(x_{k+1}) \leq f(p_k)$; reinitialize \mathcal{B} and restart the bundle subroutine with $x = x_{k+1}$, and set $\eta_0 := \eta$, $\mu_0 := \Gamma \mu_n$, $\lambda_0 := \eta_0 + \mu_0$, $(e_0, d_0, \Delta_0) := (0, 0, 0)$ to find new values for $(\hat{\varepsilon}, \hat{p}, \hat{s}, \hat{U})$; then set $(\varepsilon_{k+1}, p_{k+1}, s_{k+1}, U_{k+1}) = (\hat{\varepsilon}, \hat{p}, \hat{s}, \hat{U})$.

Step 6. Replace k by $k + 1$ and go to Step 1.

4. An Illustration Numerical Example

Now we report numerical result to illustrate Algorithm 13. Our numerical experiment is carried out in Matlab 7.8.0 running on a PC Intel Core 2 Duo CPU 2.93 GHz and 2.00 GB memory.

We consider the following Second-Order Cone programming problem (SOCP):

$$\begin{aligned} \min \quad & \sum_{j=1}^p \frac{1}{2} x^{jT} D_j x^j \\ \text{s.t.} \quad & x^j \in \mathcal{K}^{n_j}, \quad j = 1, \dots, p, \end{aligned} \quad (85)$$

where $D_j \in R^{n_j} \times R^{n_j}$ is $n_j \times n_j$ symmetric infinite matrix and $x^j = (x_0^j, \bar{x}^j)$ with $\bar{x}^j = (x_1^j, \dots, x_{n_j-1}^j)$.

This (SOCP) can be formulated in the form:

$$\begin{aligned} \min \quad & \sum_{j=1}^p \frac{1}{2} x^{jT} D_j x^j \\ \text{s.t.} \quad & x_0^j \geq \|\bar{x}^j\|, \quad j = 1, \dots, p, \end{aligned} \quad (86)$$

equivalently,

$$\begin{aligned} \min \quad & \sum_{j=1}^p \frac{1}{2} x^{jT} D_j x^j \\ \text{s.t.} \quad & \sum_{i=1}^{n_j-1} (x_i^j)^2 - (x_0^j)^2 \leq 0, \quad j = 1, \dots, p \\ & -x_0^j \leq 0, \quad j = 1, \dots, p, \end{aligned} \quad (87)$$

Let

$$\tilde{x}^j = \begin{cases} \sum_{i=1}^{n_j-1} (x_i^j)^2 - (x_0^j)^2 \leq 0, & j = 1, \dots, p \\ -x_0^{j-p} \leq 0, & j = p+1, \dots, 2p. \end{cases} \quad (88)$$

Then (SOCP) problem is equivalent to the nonlinear programming problem:

$$\begin{aligned} \min \quad & \sum_{j=1}^p \frac{1}{2} x^{jT} D_j x^j \\ \text{s.t.} \quad & \tilde{x}^j \leq 0, \quad j = 1, \dots, 2p. \end{aligned} \quad (89)$$

Let $D = \text{diag}(D_1, \dots, D_p)$, $x = (x^{1T}, \dots, x^{pT})^T$, then the exact penalty function of this nonlinear programming problem is

$$\min F(x, \rho) = \frac{1}{2} x^T D x + \rho \max \{ \tilde{x}^0, \tilde{x}^1, \dots, \tilde{x}^{2p} \}, \quad (90)$$

with $\tilde{x}^0 = 0$.

In the implementation, the initial starting point is chosen arbitrarily, and the parameters have values $M_0 = 10$, $\lambda_0 = 10$, $\xi = 1.0 \times 10^{-5}$ and $\Gamma = 2$. Optimality is declared when stopping criterion is satisfied.

Numerical results are summarized in Table 1 in which n denotes the number of variables, $\#f/g$ denotes the number of function and one subgradient evaluation.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China under Projects nos. 11226230, 11171138 and 11171049, 11226238 and General Project of the Education Department of Liaoning Province no. L2012427.

References

- [1] Y. Lu, L.-P. Pang, F.-F. Guo, and Z.-Q. Xia, "A superlinear space decomposition algorithm for constrained nonsmooth convex program," *Journal of Computational and Applied Mathematics*, vol. 234, no. 1, pp. 224–232, 2010.
- [2] C. Lemaréchal, F. Oustry, and C. Sagastizábal, "The U -Lagrangian of a convex function," *Transactions of the American Mathematical Society*, vol. 352, no. 2, pp. 711–729, 2000.
- [3] R. Mifflin and C. Sagastizábal, "VU-decomposition derivatives for convex max-functions," in *Ill-Posed Variational Problems and Regularization Techniques*, R. Tichatschke and M. A. Théra, Eds., vol. 477 of *Lecture Notes in Economics and Mathematical Systems*, pp. 167–186, Springer, Berlin, Germany, 1999.
- [4] C. Lemaréchal and C. Sagastizábal, "More than first-order developments of convex functions: primal-dual relations," *Journal of Convex Analysis*, vol. 3, no. 2, pp. 255–268, 1996.
- [5] R. Mifflin and C. Sagastizábal, "On VU-theory for functions with primal-dual gradient structure," *SIAM Journal on Optimization*, vol. 11, no. 2, pp. 547–571, 2000.
- [6] R. Mifflin and C. Sagastizábal, "Functions with primal-dual gradient structure and U -Hessians," in *Nonlinear Optimization and Related Topics*, G. Pillo and F. Giannessi, Eds., vol. 36 of *Applied Optimization*, pp. 219–233, Kluwer Academic Publishers, 2000.
- [7] R. Mifflin and C. Sagastizábal, "Primal-dual gradient structured functions: second-order results; links to epi-derivatives and partly smooth functions," *SIAM Journal on Optimization*, vol. 13, no. 4, pp. 1174–1194, 2003.
- [8] R. Mifflin and C. Sagastizábal, "A VU-algorithm for convex minimization," *Mathematical Programming B*, vol. 104, no. 2-3, pp. 583–608, 2005.
- [9] F. Shan, L.-P. Pang, L.-M. Zhu, and Z.-Q. Xia, "A UV-decomposed method for solving an MPEC problem," *Applied Mathematics and Mechanics*, vol. 29, no. 4, pp. 535–540, 2008.
- [10] Y. Lu, L.-P. Pang, J. Shen, and X.-J. Liang, "A decomposition algorithm for convex nondifferentiable minimization with errors," *Journal of Applied Mathematics*, vol. 2012, Article ID 215160, 15 pages, 2012.
- [11] A. Daniilidis, C. Sagastizábal, and M. Solodov, "Identifying structure of nonsmooth convex functions by the bundle technique," *SIAM Journal on Optimization*, vol. 20, no. 2, pp. 820–840, 2009.
- [12] W. L. Hare, "A proximal method for identifying active manifolds," *Computational Optimization and Applications*, vol. 43, no. 2, pp. 295–306, 2009.
- [13] W. L. Hare, "Functions and sets of smooth substructure: relationships and examples," *Computational Optimization and Applications*, vol. 33, no. 2-3, pp. 249–270, 2006.
- [14] R. Mifflin, L. Qi, and D. Sun, "Properties of the Moreau-Yosida regularization of a piecewise C^2 convex function," *Mathematical Programming A*, vol. 84, no. 2, pp. 269–281, 1999.
- [15] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, vol. 317 of *Fundamental Principles of Mathematical Sciences*, Springer, Berlin, Germany, 1998.
- [16] R. Mifflin and C. Sagastizábal, "VU-smoothness and proximal point results for some nonconvex functions," *Optimization Methods & Software*, vol. 19, no. 5, pp. 463–478, 2004.
- [17] S. Lang, *Real and Functional Analysis*, Springer, New York, NY, USA, 3rd edition, 1993.
- [18] W. Hare and C. Sagastizábal, "Computing proximal points of nonconvex functions," *Mathematical Programming B*, vol. 116, no. 1-2, pp. 221–258, 2009.
- [19] W. L. Hare and C. Sagastizábal, "A redistributed proximal bundle method for nonconvex optimization," *SIAM Journal on Optimization*, vol. 20, no. 5, pp. 2442–2473, 2010.