Journal of Computational and Applied Mathematics 234 (2010) 224–232

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A superlinear space decomposition algorithm for constrained nonsmooth convex program[☆]

Yuan Lu[a,*], Li-Ping Pang[b], Fang-Fang Guo[b], Zun-Quan Xia[b]

[a] School of Sciences, Shenyang University, Shenyang 110044, China

[b] School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

ARTICLE INFO

Article history:
Received 12 June 2009
Received in revised form 10 December 2009

Keywords:
Nonsmooth optimization
Piecewise C^2
$\mathcal{VU}$ decomposition
Second-order expansion

ABSTRACT

A class of constrained nonsmooth convex optimization problems, that is, piecewise C^2 convex objectives with smooth convex inequality constraints are transformed into unconstrained nonsmooth convex programs with the help of exact penalty function. The objective functions of these unconstrained programs are particular cases of functions with primal–dual gradient structure which has connection with $\mathcal{VU}$ space decomposition. Then a $\mathcal{VU}$ space decomposition method for solving this unconstrained program is presented. This method is proved to converge with local superlinear rate under certain assumptions. An illustrative example is given to show how this method works.

1. Introduction

Nonlinear programming used to be viewed, at least for computational purposes, as the minimization of a smooth (i.e. continuously differentiable) objective function subject to finitely many equality or inequality constraints given by other smooth functions. Many applications of optimization, however, concern objective functions that are not necessarily smooth but of "max type", expressible as the pointwise maximum of certain other functions which are themselves smooth.

Consider the following constrained nonsmooth convex program:

$$\begin{cases} \min & f(x) \\ \text{s. t.} & g_j(x) \le 0,\ j \in J = \{m+1, \dots, l\}, \end{cases} \tag{1.1}$$

where f is convex and piecewise C^2, $g_j, j \in J$ are convex of class C^2. For this program, Fanwen Meng and Gongyun Zhao convert it into an unconstrained smooth convex program by using the Moreau–Yosida regulation in [1]. Then they investigate the second-order properties of the Moreau–Yosida regularization. By introducing a certain qualification, they show that the gradient of the regularized function is semismooth.

More recently, new conceptual schemes have been developed, which are based on the $\mathcal{VU}$-theory introduced in [2]; see also [3–6]. The idea is to decompose R^n into two orthogonal subspaces $\mathcal{V}$ and $\mathcal{U}$ at a point $\bar{x}$ that the nonsmoothness of f is concentrated essentially on $\mathcal{V}$, and the smoothness of f appears on the $\mathcal{U}$ subspace. More precisely, for a given $\bar{g} \in \partial f(\bar{x})$, where $\partial f(\bar{x})$ denotes the subdifferential of f at $\bar{x}$ in the sense of convex analysis. Then R^n can be decomposed into direct sum of two orthogonal subspaces, i.e., $R^n = \mathcal{U} \oplus \mathcal{V}$, where $\mathcal{V} = \text{lin}(\partial f(\bar{x}) - \bar{g})$, and $\mathcal{U} = \mathcal{V}^{\perp}$. They define the $\mathcal{U}$-Lagrangian, an approximation of the original function, and show that along certain manifolds it can be used to create a second-order expansion for a nondifferentiable function. Then this theory is applied in [2,7]. However, the objective function and the

☆ The research is supported by the National Natural Science Foundation of China under project No.10771026.

* Corresponding author.
E-mail address: luyuan626@yahoo.com.cn (Y. Lu).

0377-0427/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2009.12.018

constraint functions are convex C^2 in those papers. In this paper, we consider the constrained convex program (1.1). We convert it into an unconstrained optimization problem by using the exact penalty function. Then we show the objective function of this unconstrained optimization problem is a particular case of functions with primal–dual gradient structure, a notion related to the $\mathcal{VU}$ space decomposition. Based on the $\mathcal{U}$-Lagrangian, we investigate the second-order expansion of the objective function relative to a particular trajectory. As a result we can design a algorithm frame that makes a step in the $\mathcal{V}$ space, followed by a $\mathcal{U}$-Newton step in order to obtain superlinear convergence.

The rest of the paper is organized as follows. In Section 2, we transform constrained optimization problem (1.1) into unconstrained optimization problem. Then the $\mathcal{VU}$ space decomposition is made. Section 3 investigates a smooth trajectory, along which the second-order expansion of the objective function is given. Section 4 presents a algorithm frame and its convergence theorem. At last, we report some numerical results.

2. The $\mathcal{VU}$ space decomposition

In the convex program (1.1), f is piecewise C^2. Specifically, for all $x \in R^n$,

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

where $f_i : R^n \to R$, $i \in I$ are C^2. We refer to the function f_i, $i \in I$, as structure functions.

A classical example of f is the max-function $f(x) = \max_{i\in I} f_i$, where f_i are convex of C^2. However, this class is not restricted to max-function.

The subdifferential of f at a point $x \in R^n$ can be computed in terms of the gradients of the structure functions that are active at x. More precisely,

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

where

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

is the set of active indices at x, and

$$\Delta_s = \left\{ \alpha \in R^s \mid \alpha_i \ge 0,\ \sum_{i=1}^{s} \alpha_i = 1 \right\}.$$

Let $\bar{x} \in R^n$ be a solution of (1.1). By continuity of the structure functions, there exists a ball $B_\varepsilon(\bar{x}) \subseteq R^n$ such that

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

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

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

From now on, we consider that

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

Let $F(x, \mu)$ denote the exact penalty function of (1.1) with $g_0(x) = 0$ and $\nabla g_0(x) = 0$, where $\mu > 0$ is a penalty parameter. More precisely,

$$F(x, \mu) = f(x) + \mu G(x),$$

where

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

Call

$$J(x) = \{j \in J \cup \{0\} \mid F(x, \mu) = f(x) + \mu g_j(x)\}$$

the set of indices realizing the max at x.

The following assumption will be used in the rest of this paper.

Assumption A. The set

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

is linearly independent.

Theorem 2.1. *Suppose Assumption A holds. Then we have the following results at $\bar{x}$:*

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

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

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

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

$$\mathcal{V} = lin\{\{\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})\}_{0 \neq i \in I(\bar{x})} \cup \{\nabla g_j(\bar{x})\}_{j \in J(\bar{x})}\},$$
$$\mathcal{U} = \{d \in R^n \mid \langle d, \nabla f_i(\bar{x}) - \nabla f_0(\bar{x})\rangle = \langle d, \nabla g_j(\bar{x})\rangle = 0,\ 0 \neq i \in I(\bar{x}),\ j \in J(\bar{x})\}.$$

Proof. The subdifferential of $F(x, \mu)$ at $\bar{x}$ can be formulated in

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

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

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

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

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

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

and $\mathcal{U} = \mathcal{V}^{\perp}$ means that the second formula holds. The proof is completed. □

The class of $F(x, \mu)$ belongs to the PDG-structured family [5]. More precisely, $F(x, \mu)$ has a PDG structure at $\bar{x}$ relative to the set $B_{\varepsilon}(\bar{x})$ with primal functions

$$f_i, \quad i \in I(\bar{x}); \qquad g_j, \quad j \in J(\bar{x}),$$

and dual multiplier set

$$\Theta = \left\{\alpha \mid \alpha \in \Delta_{|I(\bar{x})|}\right\} \cup \left\{\beta \mid \beta_j \geq 0,\ j \in J(\bar{x}),\ \sum_{j \in J(\bar{x})} \beta_j \leq \mu\right\}.$$

Remark 2.1. (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

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

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 $\bar{s} \in \partial F(\bar{x}, \mu)$, we have

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

From Theorem 2.1(ii), the $\mathcal{U}$-component of a subgradient $s \in \partial F(\bar{x}, \mu)$ is the same as that of any other subgradient at $\bar{x}$, i.e., $\bar{s}_{\mathcal{U}} = \bar{U}^T s$.

3. Smooth trajectory and second-order properties

3.1. $\mathcal{U}$-Lagrangian and smooth trajectory

Given $\bar{g} \in \partial F(\bar{x}, \mu)$, the $\mathcal{U}$-Lagrangian of F can be formulated in

$$\begin{aligned}L_u(u; \bar{g}_{\mathcal{V}}) &= \inf_{v \in \mathcal{V}}\{F(\bar{x} + u \oplus v, \mu) - \langle \bar{g}_{\mathcal{V}}, v\rangle_{\mathcal{V}}\} \\ &= \inf_{v \in \mathcal{V}}\{f(\bar{x} + u \oplus v) + \mu G(\bar{x} + u \oplus v) - \langle \bar{g}_{\mathcal{V}}, v\rangle_{\mathcal{V}}\}, \qquad (3.1)\end{aligned}$$

and the minimum set

$$W(u;\bar{g}_{\mathcal{V}}) = \operatorname{Arg}\inf_{v\in\mathcal{V}}\{f(\bar{x}+u\oplus v)+\mu G(\bar{x}+u\oplus v)-\langle \bar{g}_{\mathcal{V}}, v\rangle_{\mathcal{V}}\}, \tag{3.2}$$

in terms of [2].

Theorem 3.1. *Suppose the Assumption A holds. Then for all u small enough, there exists:*

(i) *the nonlinear system with variables* (u, v)

$$\begin{cases} f_i(\bar{x}+\bar{U}u+\bar{V}v)-f_0(\bar{x}+\bar{U}u+\bar{V}v)=0, & 0\neq i\in I(\bar{x})\\ g_j(\bar{x}+\bar{U}u+\bar{V}v)=0, & j\in J(\bar{x}),\end{cases} \tag{3.3}$$

has a unique solution $v=v(u)$*;*

(ii) *trajectory* $\chi(u)=\bar{x}+u\oplus v(u)$ *is* C^2*, and*

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

where

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

with

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

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

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

Proof. (i) Differentiating the left hand side of (3.3) with respect to v gives

$$\begin{cases}[\nabla f_i(\bar{x}+\bar{U}u+\bar{V}v)-\nabla f_0(\bar{x}+\bar{U}u+\bar{V}v)]^T\bar{V}, & 0\neq i\in I(\bar{x})\\ \nabla g_j(\bar{x}+\bar{U}u+\bar{V}v)^T\bar{V}, & j\in J(\bar{x}).\end{cases}$$

This Jacobian at $(u, v)=(0,0)$ is $\bar{V}^T\bar{V}$, which is nonsingular because of Assumption A. There is also a Jacobian with respect to u, so by the implicit function theorem, there is a C^1 function $v(u)$ defined on a neighborhood of $u=0$ such that $v(0)=0$.

(ii) From (i), we have $v(u)$ is C^1. So the Jacobians $Jv(u)$ and $J\chi(u)$ exist and are continuous. Differentiating the following system with respect to u,

$$\begin{cases} f_i(\chi(u))-f_0(\chi(u))=0, & 0\neq i\in I(\bar{x})\\ g_j(\chi(u))=0, & j\in J(\bar{x}),\end{cases}$$

we obtain that

$$\begin{cases}[\nabla f_i(\chi(u))-\nabla f_0(\chi(u))]^TJ(\chi(u))=0, & 0\neq i\in I(\bar{x})\\ \nabla g_j(\chi(u))^TJ(\chi(u))=0, & j\in J(\bar{x}),\end{cases}$$

or, in matrix form, $V(u)^TJ\chi(u)=0$. Using the expression $J\chi(u)=\bar{U}+\bar{V}Jv(u)$, we have that

$$V(u)^T(\bar{U}+\bar{V}Jv(u))=0.$$

By virtue of continuity of $V(u)$, $V(u)^T\bar{V}$ is nonsingular. Hence

$$Jv(u)=-(V(u)^T\bar{V})^{-1}V(u)^T\bar{U}.$$

Furthermore, $V(u)$ is C^1 because $f_i,\ i\in I(\bar{x});\ g_j,\ j\in J(\bar{x})$ is C^2, then $Jv(u)$ is C^1. Thus $\chi(u)$ and $v(u)$ is C^2. From the definition of the $\mathcal{VU}$ spaces, we have $\mathcal{V}\perp\mathcal{U}$. Hence $\bar{V}^T\bar{U}=0$. So $Jv(0)=0$ and $J\chi(0)=\bar{U}$.

The conclusion of (iii) can be obtained in terms of (i) and the definitions of $G(x)$ and $\chi(u)$. □

3.2. Second-order expansion

Theorem 3.2. *Given* $\bar{g}\in\partial F(\bar{x},\mu)$*, we have*

$$L_u(u;\bar{g}_{\mathcal{V}})=f_i(\chi(u))-\langle\bar{g}_{\mathcal{V}}, v(u)\rangle_{\mathcal{V}},\quad i\in I(\bar{x}).$$

Proof. According to (3.1) and Theorem 3.1, we get

$$\begin{aligned}L_u(u;\bar{g}_{\mathcal{V}}) &= f(\chi(u))+\mu G(\chi(u))-\langle\bar{g}_{\mathcal{V}}, v(u)\rangle_{\mathcal{V}}\\ &= f_i(\chi(u))-\langle\bar{g}_{\mathcal{V}}, v(u)\rangle_{\mathcal{V}},\quad i\in I(\bar{x}).\end{aligned}$$ □

Lemma 3.1 ([8]). *Given $\bar{g} \in \partial F(\bar{x}, \mu)$, the system with $\{\alpha_i(u)\}_{i\in I(\bar{x})}$, $\{\beta_j(u)\}_{j\in J(\bar{x})\cup\{0\}}$,*

$$\begin{cases} \bar{V}^T\left[\sum_{i\in I(\bar{x})}\alpha_i(u)\nabla f_i(\chi(u)) + \sum_{j\in J(\bar{x})}\beta_j(u)\nabla g_j(\chi(u)) - \bar{g}\right] = 0 \\ \sum_{i\in I(\bar{x})}\alpha_i(u) = 1 \\ \sum_{j\in J(\bar{x})\cup\{0\}}\beta_j(u) = \mu, \end{cases}$$

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

Theorem 3.3. *Given $\bar{g} \in \partial F(\bar{x}, \mu)$ and suppose Assumption A holds. Then for u small enough, the following assertions are true:*

(i) *the gradient of L_u is given by*

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

where

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

In particular, when $u = 0$, we have

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

where

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

(ii) *the Hessian of L_u is given by*

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

where

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

In particular, when $u = 0$, we have

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

where

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

Proof. (i) Using the chain rule to differentiate the following system with respect to u,

$$\begin{cases} L_u(u; \bar{g}_{\mathcal{V}}) = f_i(\chi(u)) - \langle \bar{g}_{\mathcal{V}}, v(u)\rangle_{\mathcal{V}}, & i \in I(\bar{x}) \\ g_j(\chi(u)) = 0, & j \in J(\bar{x}), \end{cases}$$

we obtain

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

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

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

where

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

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

$$\nabla L_u(u; \bar{g}) = \bar{U}^T g(u) - Jv(u)^T\bar{V}^T(g(u) - \bar{g}),$$

which together with (6.11) in [6] yields the desired result.

If $u = 0$, $v(0) = 0$ and $\chi(0) = \bar{x}$. By Remark 2.1(ii), we have

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

where

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

(ii) Differentiating (i) respect to u, we obtain

$$\nabla^2 L_u(u; \bar{g}_{\mathcal{V}}) = \bar{U}^T M(u) J\chi(u) + \bar{U}^T \left[\sum_{i \in I(\bar{x})} \alpha_i(u) \nabla f_i(\chi(u)) J\alpha_j(u) + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla g_j(\chi(u)) J\beta_j(u) \right],$$

where

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

It follows from the proof of Theorem 6.3 in [6] that

$$\sum_{i \in I(\bar{x})} \alpha_i(u) \nabla f_i(\chi(u)) J\alpha_j(u) + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla g_j(\chi(u)) J\beta_j(u) = -V(u)(\bar{V}V(u))^{-1} \bar{V}^T M(u) J\chi(u).$$

Then

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

when $u = 0$,

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

where

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

Theorem 3.4. *Suppose Assumption A holds and $\bar{g} \in \partial F(\bar{x}, \mu)$. Then for u small enough, there holds the second-order expansion of f along the trajectory $\chi(u) = \bar{x} + u \oplus v(u)$,*

$$f(\chi(u)) = f(\bar{x}) + \langle \bar{g}, u \oplus v(u) \rangle + \frac{1}{2} u^T \nabla^2 L_u(0; \bar{g}_{\mathcal{V}}) u + o(\|u\|^2_{\mathcal{U}}).$$

Proof. From the definition of L_u and the fact $G(\chi(u)) = 0$, we have

$$\begin{aligned} L_u(u; \bar{g}_{\mathcal{V}}) &= f(\chi(u)) + \mu G(\chi(u)) - \langle \bar{g}_{\mathcal{V}}, v(u) \rangle_{\mathcal{V}} \\ &= f(\chi(u)) - \langle \bar{g}_{\mathcal{V}}, v(u) \rangle_{\mathcal{V}}. \end{aligned}$$

Since $L_u \in C^2$, we get

$$\begin{aligned} L_u(u; \bar{g}_{\mathcal{V}}) &= L_u(0; \bar{g}_{\mathcal{V}}) + \langle \nabla L_u(0; \bar{g}_{\mathcal{V}}), u \rangle_{\mathcal{U}} + \frac{1}{2} u^T \nabla^2 L_u(0; \bar{g}_{\mathcal{V}}) u + o(\|u\|^2_{\mathcal{U}}) \\ &= f(\bar{x}) + \langle \bar{g}_{\mathcal{U}}, u \rangle_{\mathcal{U}} + \frac{1}{2} u^T \nabla^2 L_u(0; \bar{g}_{\mathcal{V}}) u + o(\|u\|^2_{\mathcal{U}}). \end{aligned}$$

Therefore,

$$\begin{aligned} f(\chi(u)) &= f(\bar{x}) + \langle \bar{g}_{\mathcal{U}}, u \rangle_{\mathcal{U}} + \langle \bar{g}_{\mathcal{V}}, v(u) \rangle_{\mathcal{V}} + \frac{1}{2} u^T \nabla^2 L_u(0; \bar{g}_{\mathcal{V}}) u + o(\|u\|^2_{\mathcal{U}}) \\ &= f(\bar{x}) + \langle \bar{g}, u \oplus v(u) \rangle + \frac{1}{2} u^T \nabla^2 L_u(0; \bar{g}_{\mathcal{V}}) u + o(\|u\|^2_{\mathcal{U}}). \quad \square \end{aligned}$$

4. Algorithm and convergence

Suppose $0 \in \partial F(\bar{x}, \mu)$, we give a algorithm frame which can solve (1.1). This algorithm makes a step in the $\mathcal{V}$ subspace, followed by a $\mathcal{U}$-Newton step in order to obtain superlinear convergence.

Algorithm 4.1. Step 0 Initialization. Given $\varepsilon > 0$. Choose a starting point $x^{(0)}$ close to $\bar{x}$ enough, and a subgradient $\tilde{g}^{(0)} \in \partial F(x^{(0)}, \mu)$, set $k = 0$.

Step 1 Stop if

$$\|\tilde{g}^{(k)}\| \le \varepsilon. \tag{4.1}$$

Step 2 Find the active index set.

Step 3 Construct $\mathcal{VU}$ decomposition at $\bar{x}$, i.e., $R^n = \mathcal{U} \oplus \mathcal{V}$. Compute

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

where

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

Step 4 Perform $\mathcal{V}$-Step. Compute

$$\delta_{\mathcal{V}}^{(k)} \in \operatorname{Arg\,min}\{F(x^{(k)} + 0 \oplus \delta_{\mathcal{V}}) : \delta_{\mathcal{V}} \in \mathcal{V}\}.$$

Set $\tilde{x}^{(k)} = x^{(k)} + 0 \oplus \delta_{\mathcal{V}}^{(k)}$.

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

$$\bar{U}^T M(0) \bar{U} \delta_{\mathcal{U}} + \bar{U}^T \tilde{g}^{(k)} = 0, \tag{4.2}$$

where

$$\sum_{i \in I(\bar{x})} \alpha_i(u) \nabla f_i(\tilde{x}^{(k)}) + \sum_{j \in J(\bar{x})} \beta_j(u) \nabla g_j(\tilde{x}^{(k)}) = \tilde{g}^{(k)} \in \partial F(\tilde{x}^{(k)}, \mu)$$

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

Step 6 Update. Set $k = k + 1$, and return to Step 1.

Theorem 4.1. *Suppose* $0 \in ri\, \partial F(\bar{x}, \mu)$, $\nabla^2 L_u(0; 0) \succ 0$. *Then the iteration points* $\{x^{(k)}\}_{k=1}^{\infty}$ *generated by the algorithm converge and satisfy*

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

Proof. Let $u^{(k)} = (x^{(k)} - \bar{x})_{\mathcal{U}}$, $v^{(k)} = (x^{(k)} - \bar{x})_{\mathcal{V}} + \delta_{\mathcal{V}}^{(k)}$. Then $\bar{x} + u^{(k)} \oplus v^{(k)} = x^{(k)} + 0 \oplus \delta_{\mathcal{V}}$, and

$$\begin{aligned} \delta_{\mathcal{V}}^{(k)} &\in \operatorname{Arg\,min}\{F(x^{(k)} + 0 \oplus \delta_{\mathcal{V}}) : \delta_{\mathcal{V}} \in \mathcal{V}\} \\ &= \operatorname{Arg\,min}\{F(\bar{x} + u^{(k)} \oplus v^{(k)}) : \delta_{\mathcal{V}} \in \mathcal{V}\}. \end{aligned}$$

Hence, $v^{(k)} \in W(u^{(k)}; 0)$. It follows from Corollary 3.5 in [2] that

$$\|(x^{(k+1)} - \bar{x})_{\mathcal{V}}\| = \|(\tilde{x}^{(k)} - \bar{x})_{\mathcal{V}}\| = o\|(x^{(k)} - \bar{x})_{\mathcal{U}}\| = o\|(x^{(k)} - \bar{x})\|. \tag{4.3}$$

Since $\nabla^2 L_u(0; 0)$ exists and $\nabla L_u(0; 0) = 0$, we have

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

By virtue of (4.2), we have $\nabla^2 L_u(0; 0)(u^{(k)} + \delta_{\mathcal{U}}^{(k)}) = o(\|u^{(k)}\|_{\mathcal{U}})$. It follows from the hypothesis $\nabla^2 L_u(0; 0) \succ 0$ that $\nabla^2 L_u(0; 0)$ is invertible and hence $\|u^{(k)} + \delta_{\mathcal{U}}^{(k)}\| = o(\|u^{(k)}\|_{\mathcal{U}})$. In consequence, one has

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

Then

$$\|(x^{(k+1)} - \bar{x})_{\mathcal{U}}\| = o(\|u^{(k)}\|_{\mathcal{U}}) = o(\|x^{(k)} - \bar{x}\|). \tag{4.4}$$

The proof is completed by combining (4.3) and (4.4). □

Table 1
Problem data.

Problem	dim	dim-$\mathcal{V}$	dim-$\mathcal{U}$	$\bar{x}$	$f\bar{x}$	$x^{(0)}$
NS2	2	1	1	(0.2 0.8)	0.8	(10, −20)
NS3-U3	3	0	3	(0.288462 −2.788462 10)	1.217825	(100, 34, −90)
NS3-U2	3	1	2	(−0.858059 −1.641941 10)	−1.641941	(100, 33, −90)
NS3-U1	3	2	1	(−1.75 −0.75 10)	2.5625	(100, 33, −100)
NS3-U0	3	3	0	(−1.75 −0.75 10)	3.0625	(101, 33, −100)

5. An illustrative numerical example

Now we report numerical result to illustrate Algorithm 4.1 for solving (1.1). Our numerical experiment is carried out in Matlab 7.1 running on a PC Intel Pentium IV of 1.70 GHz CPU and 256 MB memory.

We consider the following constrained problem:

$$\begin{cases} \min & f(x) \\ \text{s. t.} & Ax - b \le 0, \end{cases}$$

where A is $l \times n$ finite matrix and b is an $l \times 1$ vector. The objective functions of all test examples are of the form

$$f = \max_{j \in J} f_j,$$

where J is finite and each f_j is C^2 on R^n.

For our runs we used the following examples:

- NS2: the objective function is given in [9], defined for $x \in R^n$ by

$$F2d(x) := \max\left\{\frac{1}{2}(x_1^2 + x_2^2) - x_2, x_2\right\} \quad \text{and} \quad A = \begin{pmatrix} -1 & -1 \\ 1 & -1.5 \end{pmatrix}, \quad b = (-1, -1)^T.$$

- NS3-Uv: the objective functions are given in [9], four functions of three variables, where $v = 3, 2, 1, 0$ denotes the corresponding dimension of the $\mathcal{U}$ subspace. Given $e := (0, 1, 1)^T$ and four parameter vectors $\beta^v \in R^4$, for $x \in R^3$

$$F3d - Uv(x) := \max\left\{\frac{1}{2}(x_1^2 + x_2^2 + 0.1x_3^2) - e^T x - \beta_1^v, x_1^2 - 3x_1 - \beta_2^v, x_2 - \beta_3^v, x_2 - \beta_4^v\right\},$$

  where $\beta^3 := (-5.5, 10, 11, 20)$, $\beta^2 := (-5, 10, 0, 10)$, $\beta^1 := (0, 10, 0, 0)$ and $\beta^0 := (0.5, -2, 0, 0)$. In these examples, we set $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$ and $b = (10, -2.5, -2.5)^T$ for all $v = 3, 2, 1, 0$.

In the implementation, we use a bundle technique to generate $\mathcal{V}$-step. Each subgradient g we use a gradient of $f_{j_i}(x) = f(x)$. As for the $M(0)$ at x, one can refer to [9].

The parameters have values $\mu = 100$ and $\varepsilon = 1.0 \times 10^{-6}$. Optimality is declared when stopping criterion (4.1) is satisfied.

In Table 1, we show some relevant data for the problems described above including the dimension of the problem (dim), the dimension of $\mathcal{V}$ space (dim-$\mathcal{V}$) and $\mathcal{U}$ space (dim-$\mathcal{U}$), the optimal solutions $\bar{x}$ and optimal function values $f\bar{x}$ and the starting points $x^{(0)}$. The optimal solutions and optimal function values are calculated.

We show the exhibitions for the proposed algorithm compared with bundle algorithm in Table 2, where $\mathcal{VU}$ indicates solving the related program by $\mathcal{VU}$ decomposition Algorithm 4.1 proposed in this paper while *Bundle* indicates solving the program by Bundle subroutine. #f/g denotes the number of function and subgradient evaluations, and x is the calculated solution. The solutions x in bold, such as (**−22.21 19.71 −19.71**), means these solutions are not feasible. $x - \bar{x}$ is the difference of x and the optimal solution $\bar{x}$ and $fx - f\bar{x}$ is the difference between the function value at x and the optimal function value $f\bar{x}$.

It can be seen from Table 2 that the proposed $\mathcal{VU}$ decomposition algorithm costs much less function and subgradient evaluations than the Bundle algorithm. What is more, we obtained more accurate solutions by the proposed algorithm. The exhibition in NS2 problem is subtle: the function and subgradient evaluation number of the proposed $\mathcal{VU}$ decomposition algorithm is more. The reason is that in this case, $\mathcal{U}$-step is not executed when the iteration points approach the optimal solution.

This favorable results demonstrate that it is worthwhile to continue development of the space decomposition method for the constrained program.

6. Conclusions

In this paper, we use $\mathcal{VU}$ space decomposition theory to deal with constrained nonsmooth convex programs. With the help of exact penalty function, we transform the constrained nonsmooth convex programs into an unconstrained nonsmooth

Table 2
Numerical results of $\mathcal{VU}$ decomposition algorithm.

		$\mathcal{VU}$	Bundle
NS2	$\#f/g$	246	33
	x	(0.199 0.801)	(0.200 0.800)
	$x-\bar{x}$	(−5.728e−4 5.728e−4)	(4.774e−15 −4.330e−15)
	$fx-f\bar{x}$	5.73E−04	1.13E−12
NS3-U3	$\#f/g$	55	807
	x	(−1.720 −0.750 10.000)	(−25.949 20.432 −20.804)
	$x-\bar{x}$	(−1.296e−6 1.296e−6 1.251e−12)	(−2.420e+1 2.118e+1 −3.080e+1)
	$fx-f\bar{x}$	1.68E−12	7.38E+02
NS3-U2	$\#f/g$	60	824
	x	(−1.750 −0.750 10.000)	**(−22.21 19.71 −19.71)**
	$x-\bar{x}$	(−1.834e−5 1.834e−5 −1.776e−15)	(−2.046e+1 2.046e+1 −2.971e+1)
	$fx-f\bar{x}$	3.36E−10	5.47E+02
NS3-U1	$\#f/g$	89	918
	x	(−0.858 −1.642 10.000)	(−14.280 11.780 −11.780)
	$x-\bar{x}$	(0.000e+0 0.000e+0 0.000e+0)	(−1.342e+1 1.342e+1 −2.178e+1)
	$fx-f\bar{x}$	0.00E+0	2.38E+2
NS3-U0	$\#f/g$	80	1411
	x	(0.288 −2.788 10.000)	**(−20.54 16.75 −19.89)**
	$x-\bar{x}$	(0.000e+0 0.000e+0 0.000e+0)	(−2.083e+1 1.954e+1 −2.990e+1)
	$fx-f\bar{x}$	0.000e+0	6.698e+2

convex programs. Then the $\mathcal{VU}$-theory is applied and a space decomposition algorithm is obtained. This method can operate well in practice for the programs of the form (1.1) and is proved to convergent with local superlinear rate under certain assumptions. We compare the proposed Algorithm 4.1 with the Bundle algorithm and find that the proposed algorithm generated more accurate solutions and cost less function and subgradient evaluations. In addition, we find that the function and subgradient evaluation number of Algorithm 4.1 is more in NS2 problem. The reason is that, in this case the $\mathcal{U}$-step is not executed when the iteration points approach the optimal solution. And this will be a subject of future work.

References

[1] Fanwen Meng, Gongyun Zhao, On second-order properties of the Moreau–Yosida regulation for constrained nonsmooth convex programs, Numerical Functional Analysis and Optimization 25 (2007) 515–530.
[2] C. Lemaréchal, F. Oustry, C. Sagastizábal, The $\mathcal{U}$-Lagrangian of a convex function, Transactions of the American Mathematical Society 352 (2000) 711–729.
[3] R. Mifflin, C. Sagastizábal, $\mathcal{VU}$-decomposition derivatives for convex max-functions, in: Tichatschke, Théra (Eds.), Ill-posed Variational Problems and Regularization Techniques, in: Lecture Notes in Economics and Mathematical Systems, vol. 477, 1999, pp. 167–186.
[4] R. Mifflin, C. Sagastizábal, Proximal points are on the fast track, Journal of Convex Analysis 9 (2) (2002) 563–579.
[5] R. Mifflin, C. Sagastizábal, $\mathcal{VU}$-smoothness and proximal point results for some nonconvex functions, Optimization Methods and Software 19 (5) (2004) 463–478.
[6] R. Mifflin, C. Sagastizábal, On $\mathcal{VU}$-theory for functions with primal-dual gradient structure, SIAM Journal on Optimization 11 (2) (2000) 547–571.
[7] Yuan Lu, Wei Wang, The $\mathcal{VU}$-decomposition to the proper convex function. http://www.springerlink.com/content/t750736057402lw6/.
[8] R. Mifflin, C. Sagastizábal, Primal-dual gradient structured functions: Second-order results; links to epi-derivatives and partly smooth functions, SIAM Journal on Optimization 13 (2003) 1174–1197.
[9] R. Mifflin, C. Sagastizábal, A $\mathcal{VU}$-algorihtm for convex minimization, Mathematical Programming, Series B 104 (2005) 583–608.