Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2012, Article ID 215160, 15 pages
doi:10.1155/2012/215160

Research Article

A Decomposition Algorithm for Convex Nondifferentiable Minimization with Errors

Yuan Lu,[1] Li-Ping Pang,[2] Jie Shen,[3] and Xi-Jun Liang[2]

[1] *School of Sciences, Shenyang University, Shenyang 110044, China*
[2] *School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China*
[3] *School of Mathematics, Liaoning Normal University, Dalian 116029, China*

Correspondence should be addressed to Li-Ping Pang, lppang@dlut.edu.cn

Received 29 July 2011; Accepted 25 October 2011

Academic Editor: Chein-Shan Liu

A decomposition algorithm based on proximal bundle-type method with inexact data is presented for minimizing an unconstrained nonsmooth convex function f. At each iteration, only the approximate evaluation of f and its approximate subgradients are required which make the algorithm easier to implement. It is shown that every cluster of the sequence of iterates generated by the proposed algorithm is an exact solution of the unconstrained minimization problem. Numerical tests emphasize the theoretical findings.

1. Introduction

Consider minimizing the following problem:

$$\min\{f(x) \mid x \in R^n\}, \tag{1.1}$$

where $f : R^n \to R$ is a nondifferentiable convex function. It is well known that many practical problems can be formulated as (1.1), for example the problem of catastrophe, ruin, vitality, data mining, and finance. A classical conceptual algorithm for solving (1.1) is the proximal point method, based on the Moreau-Yosida regulation of f [1, 2]. Implementable forms of the method can be obtained by means of a bundle technique, alternating serious steps with sequences of null steps [3, 4].

More recently, new conceptual schemes for solving (1.1) have been developed by using an approach that is somewhat different from Moreau-Yosida regularization. This is the $\mathcal{VU}$-theory introduced in [5]; see also [6, 7]. The idea is to decompose R^n into two orthogonal

subspace $\mathcal{V}$ and $\mathcal{U}$ at a point $\overline{x}$ that the nonsmoothness of f is concentrated essentially on $\mathcal{V}$, and the smoothness of f appears on $\mathcal{U}$-subspace. More precisely, for a given $\overline{g} \in \partial f(\overline{x})$, where $\partial f(\overline{x})$ denotes the subdifferential of f at $\overline{x}$ in the sense of convex analysis, then R^n can be decomposed as direct sum of two orthogonal subspaces, that is, $R^n = \mathcal{U} \oplus \mathcal{V}$, where $\mathcal{V} = \operatorname{lin}(\partial f(\overline{x}) - \overline{g})$ and $\mathcal{U} = \mathcal{V}^{\perp}$. They define the $\mathcal{U}$-Lagrangian, an approximation of the original function, which can be used to create a second-order expansion of f along certain manifolds.

Mifflin and Sagastizábal design a $\mathcal{VU}$-algorithm for convex function in [8]. This algorithm brings the iterate to the primal track with the help of bundle subroutine. Then the $\mathcal{U}$-Newton step is performed to gain superlinear decrease of the distance to solution. In order to implement this algorithm, [9] gives an algorithm which only uses the subgradients. However, this algorithm is conceptual in the sense that it needs to compute the exact function values of the objective function, which is difficult to evaluate them. For example, consider the situation of Lagrangian relaxation. The primal problem is

$$\max\{q(\xi) \mid \xi \in P,\ h(\xi) = 0\}, \tag{1.2}$$

where P is a compact subset of R^m and $q : R^m \to R$, $h : R^m \to R^n$. Lagrangian relaxation of the equality constraints in the problem leads to the following problem

$$\min_{x \in R^n} C(x), \tag{1.3}$$

where

$$C(x) = \max_{\xi \in P}\{q(\xi) + \langle x,\ h(\xi)\rangle\} \tag{1.4}$$

is the dual function. Trying to solve problem (1.2) by means of solving its dual problem (1.3) makes sense in many situations. In this case, evaluating the function value $C(x)$ and a subgradient $g(x) \in \partial C(x)$ requires solving the optimization problem (1.4) exactly. Actually, in some cases, computing exact values of $C(x)$ is unnecessary and inefficient. For this reason, some modifications of bundle methods in [9] were needed.

The paper is organized as follows. In the next section we present the approximate $\mathcal{U}$-Lagrangian based on the approximate subgradient. Then we design a conceptual Algorithm 2.6 which can deal with the approximate subgradients and approximate function values. Section 3 breaks into 3 parts. In the first part, we propose the approximate primal-dual track. The proximal bundle-type subroutine with inexact data is introduce in the second part. The third part of Section 3 is devoted to establishing an implemental Algorithm 3.5 which substitutes the approximate V-step in Algorithm 2.6 with proximal bundle subroutine. Numerical testing of the resulting Algorithm 3.5 is reported in the final section.

2. A Conceptual Approximate Decomposition Algorithm

2.1. Approximate $\mathcal{U}$-Lagrangian and Its Properties

In some cases, computing exact values of the objective function and exact subgradient is unnecessary and inefficient. For this reason, some modification of the $\mathcal{U}$-Lagrangian will be proposed in this section. We assume that $\overline{g}$ satisfies

$$f(\overline{x}+u\oplus v)\geq \widetilde{f}(\overline{x})+\langle \overline{g},u\oplus v\rangle,\quad \text{where } \widetilde{f}(\overline{x})\in\left[f(\overline{x})-\varepsilon,f(\overline{x})\right],\quad \varepsilon>0. \tag{2.1}$$

Introducing this $\overline{g}$ to [5], one can restate the definition of $\mathcal{U}$-Lagrangian and its properties as follows.

Definition 2.1. Assume (2.1). The approximate $\mathcal{U}$-Lagrangian of f, denoted by $L_{\overline{g}}(u)$, is defined as follows.

$$L_{\overline{g}}(u)=\min_{v\in\mathcal{U}}\left\{f(\overline{x}+u\oplus v)-\left\langle \overline{g}_{\mathcal{U}},v\right\rangle_{\mathcal{U}}\right\}, \tag{2.2}$$

and $W_{\overline{g}}(u)$ is defined by

$$W_{\overline{g}}(u)=\operatorname{Arg}\min_{v\in\mathcal{U}}\left\{f(\overline{x}+u\oplus v)-\left\langle \overline{g}_{\mathcal{U}},v\right\rangle_{\mathcal{U}}\right\}. \tag{2.3}$$

Theorem 2.2. *Assume* (2.1). *Then the following assertions are true:*

(i) *the function* $L_{\overline{g}}$ *defined in* (2.2) *is convex and finite everywhere;*

(ii) $W_{\overline{g}}(u)=\{w\mid \exists g\in\partial f(\overline{x}+u\oplus w): g_{\mathcal{U}}=\overline{g}_{\mathcal{U}}\}$;

(iii) $0\in W_{\overline{g}}(0)$ *and* $L_{\overline{g}}(0)=f(\overline{x})$.

Theorem 2.3. *Assume* (2.1) *and* $W_{\overline{g}}(u)\neq\emptyset$. *Then one has that*

$$\partial L_{\overline{g}}(u)=\left\{u^{*}\mid u^{*}\oplus\overline{g}_{\mathcal{U}}\in\partial f(\overline{x}+u\oplus w)\right\}. \tag{2.4}$$

Remark 2.4. Assume (1.2). If $\varepsilon=0$, then the approximate $\mathcal{U}$-Lagrangian in this paper is exactly the $\mathcal{U}$-Lagrangian in [5].

2.2. Approximate Decomposition Algorithm Frame

In order to give an approximate decomposition algorithm frame, we restate the definition of Hessian matrix in [5] as follows.

Definition 2.5. Assume that f is finite, $\overline{x}$ is fixed, and $\overline{g}$ satisfies (2.1). We say that f has at $\overline{x}$ a $\mathcal{U}$-Hessian $H_{\mathcal{U}}f(\overline{x})$ associated with $\overline{g}$ if $L_{\overline{g}}(u)$ has a generalized Hessian at 0, setting

$$H_{\mathcal{U}}f(\overline{x})=HL_{\overline{g}}(0). \tag{2.5}$$

Assume (2.1), we investigate an approximate decomposition algorithm frame based on the definition of the approximate $\mathcal{U}$-Lagrangian.

Algorithm 2.6.

Step 0. Initiation

$x_0,\ \eta > 0,\ 0 < \tau < 1,\ \varepsilon^0 > 0,\ g^0$ satisfies for all $x,\ f(x) \geq \widetilde{f}(x_0) + \langle g^0, x - x_0 \rangle$, where $\widetilde{f}(x_0) \in [f(x_0) - \varepsilon^0, f(x_0)]$ and $k = 0$.

Step 1. Stop if $\|g^k\| \leq \eta$.

Step 2. Approximate $\mathcal{V}$-step.

Compute an optimal solution $\delta v \in \mathcal{V}$ satisfying

$$\delta v \in \operatorname{Arg}\min_{\delta v \in \mathcal{V}} f(x_k + 0 \oplus \delta v). \tag{2.6}$$

Set $x' = x_k + 0 \oplus \delta v$.

Step 3. $\mathcal{U}$-step. Make a Newton step in x'.

Compute that g^k satisfies for all $x,\ f(x) \geq \widetilde{f}(x') + \langle g^k, x - x' \rangle$, where $\widetilde{f}(x') \in [f(x') - \varepsilon^k, f(x')]$, such that

$$g^k_{\mathcal{V}} = 0, \qquad g^k_{\mathcal{U}} \in \partial L_{g^k}\left(\left(x' - \overline{x}\right)_{\mathcal{U}}\right). \tag{2.7}$$

Compute the solution $\delta u \in \mathcal{U}$ satisfying

$$g^k_{\mathcal{U}} + H_{\mathcal{U}} f(\overline{x}) \delta u = 0. \tag{2.8}$$

Set $x_{k+1} = x' + \delta u \oplus 0 = x_k + \delta u \oplus \delta v$ and $\varepsilon^{k+1} = \tau \varepsilon^k$.

Step 4. Update-set

Compute that g^{k+1} satisfies for all $x,\ f(x) \geq \widetilde{f}(x_{k+1}) + \langle g^{k+1}, x - x_{k+1} \rangle$, where $\widetilde{f}(x_{k+1}) \in [f(x_{k+1}) - \varepsilon^{k+1}, f(x_{k+1})]$. Set $k = k + 1$ and go to Step 1.

Theorem 2.7. *Assume* (2.1) *and f has a positive definite $\mathcal{U}$-Hessian at $\overline{x}$, a minimizer of f. Then the iterate points $\{x_k\}$ constructed by Algorithm 2.6 satisfy*

$$\|x_{k+1} - \overline{x}\| = o(\|x_k - \overline{x}\|). \tag{2.9}$$

Remark 2.8. If $\{\varepsilon^k\}_{k=0}^{\infty} \equiv 0$, this algorithm is the same as Algorithm 4.5 in [5]. However, it only uses the approximate objective function values which make the algorithm easier to implement.

3. Approximate Decomposition Algorithm

Since the Algorithm 2.6 in Section 2 relies on knowing the subspaces $\mathcal{U}$ and $\mathcal{V}$ and converges only locally, it needs significant modification. In [10], Mifflin and Sagastizábal show that a proximal point sequence follows primal track near a minimizer. This opens the way for defining a $\mathcal{VU}$-algorithm where $\mathcal{V}$-steps are replaced by proximal steps. In addition, the proximal step can be estimated with a bundle technique which also can approximate the unknown $\mathcal{U}$ and $\mathcal{V}$ subspaces as a computational byproduct. Therefore, they establish Algorithm 6 in [8] by combing the bundle subroutine with the $\mathcal{VU}$-space decomposition method. However, this algorithm needs the exact function values and exact subgradients, which is expensive to compute. Therefore, the study of using approximate values instead of the exact ones is deserving.

3.1. Approximate Primal-Dual Track

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

$$p_\mu(x) := \arg\min_{p\in R^n}\left\{ f(p) + \frac{1}{2}\mu\|p-x\|^2 \right\}, \quad x \in R^n, \tag{3.1}$$

where $\|\cdot\|$ stands for the Euclidean norm. It has the property: $g_\mu(x) := \mu(x - p_\mu(x)) \in \partial f(p_\mu(x))$.

Similarly to the definition of primal track, we define the approximate primal track.

Definition 3.1. For any $\varepsilon > 0$, $\mu = \mu(x) > 0$, we say that $\Theta(u) = \overline{x} + u \oplus v(u)$ is an approximate primal track leading to $\overline{x}$, a minimizer of f, if for all $u \in R^{\dim \mathcal{U}}$ small enough, it satisfies the following:

(i) $\Theta(u_\mu(x)) = \overline{x} + u_\mu(x) \oplus v(u_\mu(x))$, where $u_\mu(x) = (p_\mu(x) - \overline{x})_{\mathcal{U}}$;

(ii) $v : R^{\dim \mathcal{U}} \to R^{\dim \mathcal{V}}$ is a C^2-function satisfying $\overline{V}v \in W_{\overline{g}}(u)$ for all $\overline{g}$ satisfies (2.1);

(iii) the Jacobian $J\Theta(u)$ is a basis matrix for $\mathcal{V}(\Theta(u))^\perp$;

(iv) the particular $\mathcal{U}$-Lagrangian $L_0(u)$ is a C^2-function.

Accordingly, we have the approximate dual track denoted by $\Gamma(u)$ corresponding to the approximate primal track. More precisely,

$$\Gamma(u) = \arg\min\left\{ \|g\|^2 : f(x) \geq \widetilde{f}(\Theta(u)) + \langle g, x - \Theta(u)\rangle \right\}. \tag{3.2}$$

In fact, if $\varepsilon = 0$, the approximate primal-dual track is exactly the primal-dual track shown in [8].

The next lemma addresses that making an approximate $\mathcal{V}$-step in Algorithm 2.6 essentially amounts to finding a corresponding approximate primal track point.

Lemma 3.2. *Let $\Theta(u_\mu(x))$ be an approximate primal track leading to $\overline{x}$, a minimizer of f, and let $\overline{H} := \nabla^2 L_0(0)$. Then for all u sufficiently small $\Theta(u_\mu(x))$ is the unique minimizer of f on the affine set $\Theta(u_\mu(x)) + \mathcal{V}(\Theta(u_\mu(x)))$.*

Proof. Since $J\Theta(u)$ is a basis for $\mathcal{U}(\Theta(u))^{\perp}$, Theorem 3.4 in [10] with $\mathcal{B}_{\mathcal{U}}(u) = J\Theta(u)$ gives the result. □

3.2. The Proximal Bundle-Type Subroutine with Inexact Data

Throughout this section, we make the following assumption: at each given point $x \in R^n$, and for $\varepsilon \geq 0$, we can find some $\widetilde{f}(x) \in R$ and $g(x) \in R^n$ satisfying

$$\begin{aligned} \widetilde{f}(x) &\in \left[f(x) - \varepsilon, f(x)\right], \\ f(\zeta) &\geq f(x) + \left\langle g(x), \zeta - x \right\rangle - \overline{\varepsilon}, \quad \forall \zeta \in R^{\mathrm{n}}, \end{aligned} \tag{3.3}$$

where $\overline{\varepsilon} = f(x) - \widetilde{f}(x)$. At the same time, it can be ensured that

$$\widetilde{f}_j \longrightarrow \widetilde{f}_k \quad \text{if } y^j \longrightarrow x_k, \tag{3.4}$$

where $\widetilde{f}_j \in [f(y^j) - \varepsilon^k, f(y^j)]$ and $\widetilde{f}_k \in [f(x_k) - \varepsilon^k, f(x_k)]$ for given $\varepsilon^k > 0$. The condition (3.3) means that $g(x) \in \partial_{\overline{\varepsilon}} f(x)$. This setting is realistic in many applications; see [11].

The bundle method includes two phases. (i) The first phase makes use of the information $(x_k, \widetilde{f}_k, g^k)$, $(x_{k-1}, \widetilde{f}_{k-1}, g^{k-1}), \ldots, (x_1, \widetilde{f}_1, g^1)$ in bundles to establish a polyhedral approximation of f at the actual iterate x_k. (ii) Due to the kinky structure of f, the model is possibly not precise for approximation f. Then, more information around the actual iterate x_k is mobilized to obtain a more reliable model. Feature (i) leads to the following approximation of f at x_k. Let I_k denote the index set at the kth iteration with each $j \in I_k$ representing $(y^j, \widetilde{f}_j, g^j)$, where $\widetilde{f}_j$ and g^j satisfy

$$\begin{aligned} \widetilde{f}_j &\in \left[f(y^j) - \varepsilon^k, f(y^j)\right], \\ f(x) &\geq \widetilde{f}_j + \left\langle g^j, x - y^j \right\rangle, \quad \forall x \in R^n, \end{aligned} \tag{3.5}$$

for given $\varepsilon^k > 0$. From the choices of $\widetilde{f}_j$ and g^j, we have that, for all $x \in R^n$ and for all $j \in I_k$,

$$f(x) \geq \widetilde{f}_j + \left\langle g^j, x - y^j \right\rangle = \widetilde{f}_k + g^{jT}(x - x_k) + \varepsilon^k - \alpha_{k,j}, \tag{3.6}$$

where

$$\alpha_{k,j} = \alpha\left(x_k, y^j\right) = \widetilde{f}_k - \widetilde{f}_j - g^{jT}\left(x_k - y^j\right) + \varepsilon^k. \tag{3.7}$$

On the basis of the above observation, we attempt to explore the possibility of utilizing the approximate subgradient and approximate function values instead of the exact ones. We approximate f at x_k from below by a piecewise linear convex function φ of the form:

$$\varphi(z) := \widetilde{f}_k + \max_{j \in I_k}\left\{ g^{jT}(z - x_k) - \alpha_{k,j} \right\} + \varepsilon^k. \tag{3.8}$$

Since (3.8) becomes more and more crude if an approximation of f is farther away from x_k, we add the proximal term $(1/2)\mu\|p - x_k\|^2$, $\mu > 0$, to it. To approximate an proximal point, we solve the first quadratic programming subproblem

$$\begin{aligned} \min \quad & r + \frac{1}{2}\mu\|p - x_k\|^2 \\ \text{s.t.} \quad & r \geq \widetilde{f}_k + g^{jT}(p - x_k) - \alpha_{k,j} + \varepsilon^k, \quad j \in I_k. \end{aligned} \tag{3.9}$$

Its corresponding dual problem is

$$\begin{aligned} \min \quad & \frac{1}{2\mu}\left\|\sum_{j\in I_k}\lambda_j g^j\right\|^2 + \sum_{j\in I_k}\lambda_j\alpha_{k,j} \\ \text{s.t.} \quad & \sum_{j\in I_k}\lambda_j = 1, \quad \lambda_j \geq 0, \quad j \in I_k. \end{aligned} \tag{3.10}$$

Let $(\widehat{r}, \widehat{p})$ and $\widehat{\lambda} = (\widehat{\lambda}_1, \ldots, \widehat{\lambda}_{|I_k|})$ denote the optimal solution of (3.9) and (3.10), then it is easily seen that

$$\widehat{r} = \varphi(\widehat{p}), \quad \widehat{p} = x_k - \frac{1}{\mu}\widehat{g}, \quad \text{where } \widehat{g} := \sum_{j\in I_k}\widehat{\lambda}_j g^j. \tag{3.11}$$

In addition, $\widehat{\lambda}_j = 0$ for all $j \in I_k$ such that $r > \widetilde{f}_k + g^{jT}(\widehat{p} - x_k) - \alpha_{k,j} + \varepsilon^k$ and

$$\begin{aligned} \varphi(\widehat{p}) &= \widetilde{f}_k + \sum_{j\in I_k}\widehat{\lambda}_j\left[g^{jT}(\widehat{p} - x_k) - \alpha_{k,j} + \varepsilon^k\right] \\ &= \widetilde{f}_k - \sum_{j\in I_k}\widehat{\lambda}_j\alpha_{k,j} + \varepsilon^k - \frac{1}{\mu}\|\widehat{g}\|^2. \end{aligned} \tag{3.12}$$

The vector $\widehat{p}$ is an estimate of an approximate proximal point. Hence, it approximates an approximate primal track point when the latter exists. To proceed further we let $y^{j_+} := \widehat{p}$ and compute $\widetilde{f}(\widehat{p})$, $\widehat{e} := \widetilde{f}(\widehat{p}) - \varphi(\widehat{p}) = \widetilde{f}(\widehat{p}) - \widehat{r}$, and g^{j+} satisfying $f(z) \geq \widetilde{f}(\widehat{p}) + \langle g^{j+}, z - \widehat{p}\rangle$, for all $z \in R^n$.

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

$$\widehat{I}_k := \left\{j \in I_k : \widehat{r} = \widetilde{f}_k + g^{jT}(\widehat{p} - x_k) - \alpha_{k,j} + \varepsilon^k\right\} \cup \{j_+\}. \tag{3.13}$$

The second quadratic programming problem is

$$\begin{aligned} \min \quad & r + \frac{1}{2}\|p - x_k\|^2 \\ \text{s.t.} \quad & r \geq g^{jT}(p - x_k), \quad j \in \widehat{I}_k. \end{aligned} \tag{3.14}$$

It has a dual problem

$$\begin{aligned} \min \quad & \frac{1}{2}\left\| \sum_{j\in \widehat{I}_k} \lambda_j g^j \right\|^2 \\ \text{s.t.} \quad & \sum_{j\in \widehat{I}_k} \lambda_j = 1, \quad \lambda_j \geq 0,\ j \in \widehat{I}_k. \end{aligned} \tag{3.15}$$

Similar to (3.11), the respective solutions, denoted by $(\overline{r}, \overline{p})$ and $\overline{\lambda}$, satisfy

$$\overline{r} = \varphi(\overline{p}), \qquad \overline{p} - x_k = -\widehat{s}, \quad \text{where } \widehat{s} = \sum_{j\in \widehat{I}_k} \overline{\lambda}_j g^j. \tag{3.16}$$

Given $\sigma \in (0, 1/2]$, the proximal bundle subprocedure is terminated and $\widehat{p}$ is declared to be an approximation of $p_\mu(x_k)$ if

$$\widehat{e} \leq \frac{\sigma}{\mu}\|\widehat{s}\|^2. \tag{3.17}$$

Otherwise, I_k above is replaced by $\widehat{I}_k$, and new iterate data are computed by solving updated subproblems (3.9) and (3.14). This update, appending $(\alpha_{k,j_+}, g^{j_+})$ to active data at (3.9), ensures convergence to a minimizing point $\overline{x}$ in case of nontermination.

Remark 3.3. From the talking above, the following results are true:

(i) $\widehat{s} = \arg\min\{\|s\|^2 : s \in \operatorname{co}\{g^j : j \in \widehat{I}_k\}\}$;

(ii) since $p_\mu(x)$ is an approximate primal track point $\Theta(u_\mu(x))$ approximated by $\widehat{p}$ and $\operatorname{co}\{g^j : j \in \widehat{I}_k\}$ approximates $\operatorname{co}\{g : f(x) \geq \widetilde{f}(\Theta(u)) + \langle g, x - \Theta(u)\rangle\}$, from (3.2) the corresponding $\Gamma(u_\mu(x))$ is estimated by $\widehat{s}$;

(iii) we can obtain the $\widehat{U}$ by means of the following iteration.

Let

$$\widehat{I}_k^{\text{act}} := \left\{ j \in \widehat{I}_k : \overline{r} = g^{jT}(\overline{p} - x) \right\}. \tag{3.18}$$

Then, from (3.16), $\overline{r} = -g^{jT}\widehat{s}$, $j \in \widehat{I}_k^{\text{act}}$, so

$$\left(g^j - g^l\right)^T \widehat{s} = 0, \tag{3.19}$$

for all such j and for a fixed $l \in \widehat{I}_k^{\text{act}}$. Define a full column rank matrix $\widehat{V}$ by choosing the largest number of indices j satisfying (3.19) such that the corresponding vectors $g^j - g^l$ 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 $\widehat{V}^T$ with $\widehat{U} = I$ if $\widehat{V}$ is vacuous.

Theorem 3.4. *At the kth iteration, the above proximal bundle subprocedure satisfies the following:*

(i) $g^j \in \partial_{\bar{\varepsilon}+\hat{e}} f(\hat{p})$, $j \in \hat{I}_k$, *where* $\bar{\varepsilon} = f(\hat{p}) - \tilde{f}(\hat{p})$ *and* $\tilde{f}(\hat{p}) \in [f(\hat{p}) - \varepsilon^k, f(\hat{p})]$;

(ii) $\hat{s} \in \partial_{\bar{\varepsilon}+\hat{e}} f(\hat{p})$ *and* $\hat{g} \in \partial_{\bar{\varepsilon}+\hat{e}} f(\hat{p})$;

(iii) $\mu\|\hat{p} - p_\mu(x_k)\|^2 \le \bar{\varepsilon} + \hat{e}$;

(iv) $\|\hat{s}\| \le \|\hat{g}\|$, *where* $\hat{g} = \mu(x_k - \hat{p})$;

(v) *for any parameter* $m \in (0,1)$, (3.17) *implies*

$$\tilde{f}(\hat{p}) - \tilde{f}_k \le -\frac{m}{2\mu}\|\hat{g}\|^2 + \varepsilon^k. \tag{3.20}$$

Proof. (i) Since g^{j+} satisfies $f(z) \ge \tilde{f}(\hat{p}) + \langle g^{j+}, z - \hat{p}\rangle$ and $\hat{e} = \tilde{f}(\hat{p}) - \varphi(\hat{p}) \ge 0$, g^{j+} satisfies

$$\begin{aligned} f(z) &\ge \tilde{f}(\hat{p}) + \left\langle g^{j+}, z - \hat{p}\right\rangle - \hat{e} \\ &= f(\hat{p}) + \left\langle g^{j+}, z - \hat{p}\right\rangle - \hat{e} - \bar{\varepsilon}, \end{aligned} \tag{3.21}$$

where $\bar{\varepsilon} = f(\hat{p}) - \tilde{f}(\hat{p})$, so the result of item (i) holds for $j = j_+$. From the definition of $\hat{p}$, $\hat{r}$, and $\hat{I}_k$ we have that for all $j \ne j_+$ in $\hat{I}_k$

$$\varphi(\hat{p}) = \hat{r} = \tilde{f}_k + g^{jT}(\hat{p} - x_k) - \alpha_{k,j} + \varepsilon^k, \tag{3.22}$$

so for all such j,

$$\hat{e} = \tilde{f}(\hat{p}) - \varphi(\hat{p}) = \tilde{f}(\hat{p}) - \tilde{f}_k - g^{jT}(\hat{p} - x_k) + \alpha_{k,j} - \varepsilon^k. \tag{3.23}$$

In addition,

$$f(z) \ge \tilde{f}_k + g^{jT}(z - x_k) - \alpha_{k,j} + \varepsilon^k, \quad z \in R^n. \tag{3.24}$$

Adding $0 = \hat{e} - \hat{e}$ to this inequality gives

$$\begin{aligned} f(z) &\ge \tilde{f}_k + g^{jT}(z - x_k) + \varepsilon^k - \alpha_{k,j} + \hat{e} - \hat{e} \\ &= \tilde{f}(\hat{p}) + g^{jT}(z - \hat{p}) - \hat{e} \\ &= f(\hat{p}) + g^{jT}(z - \hat{p}) - \hat{e} - \bar{\varepsilon}, \quad \forall z \in R^n, \end{aligned} \tag{3.25}$$

which means that $g^j \in \partial_{\bar{\varepsilon}+\hat{e}} f(\hat{p})$ for $j_+ \ne j \in \hat{I}_k$.

(ii) Multiplying each inequality in (3.25) by its corresponding multiplier $\bar{\lambda}_j \ge 0$ and summing these inequalities, we have

$$\sum_{j\in\hat{I}_k}\bar{\lambda}_j f(z) \ge \sum_{j\in\hat{I}_k}\bar{\lambda}_j f(\hat{p}) + \sum_{j\in\hat{I}_k}\bar{\lambda}_j g^{jT}(z - \hat{p}) - \sum_{j\in\hat{I}_k}\bar{\lambda}_j \hat{e} - \sum_{j\in\hat{I}_k}\bar{\lambda}_j \bar{\varepsilon}, \quad \forall z \in R^n. \tag{3.26}$$

Using the definition of $\widehat{s}$ from (3.16) and the fact that $\sum_{j\in\widehat{I}_k}\overline{\lambda}_j = 1$ gives

$$f(z) \geq f(\widehat{p}) + \widehat{s}^T(z-\widehat{p}) - \widehat{e} - \overline{\varepsilon}, \quad \forall z \in R^n, \tag{3.27}$$

which means that $\widehat{s} \in \partial_{\overline{\varepsilon}+\widehat{e}} f(\widehat{p})$. In a similar manner, this time using the multipliers $\widehat{\lambda}_j$ that solve dual problem (3.10) and define $\widehat{g}$ in (3.11), together with $\widehat{\lambda}_{j_+} := 0$, obtains the result.

(iii) Since $g_\mu \in \partial f(p_\mu(x_k))$, we have

$$f(p_\mu(x_k)) \leq f(\widehat{p}) - g_\mu(x_k)^T(\widehat{p} - p_\mu(x_k)). \tag{3.28}$$

From (ii): $\widehat{g} \in \partial_{\overline{\varepsilon}+\widehat{e}} f(\widehat{p})$, we get

$$f(p_\mu(x_k)) \geq f(\widehat{p}) + \widehat{g}^T(p_\mu(x_k) - \widehat{p}) - \overline{\varepsilon} - \widehat{e}. \tag{3.29}$$

Therefore,

$$f(\widehat{p}) + \widehat{g}^T(p_\mu(x_k) - \widehat{p}) - \overline{\varepsilon} - \widehat{e} \leq f(p_\mu(x_k)) \leq f(\widehat{p}) - g_\mu(x_k)^T(\widehat{p} - p_\mu(x_k)), \tag{3.30}$$

that is, $(\widehat{g} - g_\mu(x_k))^T(p_\mu(x_k) - \widehat{p}) - \overline{\varepsilon} - \widehat{e} \leq 0$. Then, since the expression for $\widehat{g}$ from (3.11) written in the form

$$\widehat{g} = -\mu(\widehat{p} - x_k), \tag{3.31}$$

combined with $g_\mu(x_k) = \mu(x_k - p_\mu(x_k))$ implies that $\widehat{g} - g_\mu(x_k) = \mu(p_\mu(x_k) - \widehat{p})$, we obtain item (iii).

(iv) From (3.10), (3.11), (3.31), and the definition of $\widehat{I}_k$, we have that $\mu(x_k - \widehat{p}) = \widehat{g}$ is in the convex hull of $\{g^j,\ j \in \widehat{I}_k\}$. We obtain the result by virtue of the minimum norm property of $\widehat{s}$.

(v) Since $\sigma \leq 1/2$ and $m \in (0,1)$, we have $\sigma \leq 1 - (m/2)$. Thus if (3.17) holds then $\widehat{e} \leq [(1-(m/2))/\mu]\|\widehat{s}\|^2$. Together with the definition of $\widehat{e}$, (3.12) and the nonnegativity of $\widehat{\lambda}_j\alpha_j$ gives

$$\begin{aligned}
\widetilde{f}(\widehat{p}) - \widetilde{f}_k &= \widehat{e} + \varphi(\widehat{p}) - \widetilde{f}_k \\
&= \widehat{e} - \sum_{j\in I_k}\widehat{\lambda}_j\alpha_{k,j} - \frac{1}{\mu}\|\widehat{g}\|^2 + \varepsilon^k \\
&\leq \widehat{e} - \frac{1}{\mu}\|\widehat{g}\|^2 + \varepsilon^k \\
&\leq \left[\frac{(1-(m/2))}{\mu}\right]\|\widehat{s}\|^2 - \frac{1}{\mu}\|\widehat{g}\|^2 + \varepsilon^k.
\end{aligned} \tag{3.32}$$

Finally, combing this inequality with item (iv) gives (3.20). □

3.3. Approximate Decomposition Algorithm and Convergence Analysis

Substituting the approximate $\mathcal{V}$-step in Algorithm 2.6 with proximal bundle-type subroutine, we present an approximate decomposition algorithm as follows. Afterwards a detailed convergence analysis is given. The main statement comprises the fact that each cluster point of the sequence of iterates generated by the algorithm is an optimal solution.

Algorithm 3.5. Choose a starting point $p_0 \in R^n$ and positive parameters η, ε^0, $\underline{\mu}$, τ, and m with $\tau < 1$, $m < 1$.

Step 0. Compute g^0 satisfying $f(z) \geq \widetilde{f}(p_0) + \langle g^0, z - p_0 \rangle$, where $\widetilde{f}(p_0) \in [f(p_0) - \varepsilon^0, f(p_0)]$. 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 $\|s_k\|^2 \leq \eta$.

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 approximate $\mathcal{U}$-Newton step by solving the linear system

$$H_k \Delta u_k = -U_k^T s_k. \tag{3.33}$$

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

Step 4. Choose $\mu_{k+1} > \underline{\mu}$, $\sigma_{k+1} \in (0, 1/2]$, initialize I_k, and run the bundle subprocedure with $x = x'_{k+1}$. Compute recursively, and set $(e'_{k+1}, p'_{k+1}, s'_{k+1}, U'_{k+1}) := (\widehat{e}, \widehat{p}, \widehat{s}, \widehat{U})$.

Step 5. If

$$\widetilde{f}(p'_{k+1}) - \widetilde{f}(p_k) \leq -\frac{m}{2\mu_{k+1}} \left\| s'_{k+1} \right\|^2 + \varepsilon^{k+1}, \tag{3.34}$$

then set

$$\left(x_{k+1}, e_{k+1}, p_{k+1}, s_{k+1}, U_{k+1}, \varepsilon^{k+1} \right) := \left(x'_{k+1}, e'_{k+1}, p'_{k+1}, s'_{k+1}, U'_{k+1}, \tau\varepsilon^k \right). \tag{3.35}$$

Otherwise, execute a line search

$$x_{k+1} := \arg\min \left\{ \widetilde{f}(p_k), \widetilde{f}(p'_{k+1}) \right\}, \tag{3.36}$$

reinitialize I_k, and rerun the bundle subroutine with $x = x_{k+1}$, to find new values for $(\widehat{e}, \widehat{p}, \widehat{s}, \widehat{U})$, then set $(e_{k+1}, p_{k+1}, s_{k+1}, U_{k+1}, \varepsilon^{k+1}) = (\widehat{e}, \widehat{p}, \widehat{s}, \widehat{U}, \tau\varepsilon^k)$.

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

Remark 3.6. In this algorithm, $\{\varepsilon^k\}_{k=0}^{\infty} \downarrow 0$. If $\{\varepsilon^k\}_{k=0}^{\infty} \equiv 0$, this algorithm is the same as Algorithm 6 in [8]. However, this algorithm uses proximal bundle-type subroutine which can deal with the approximate subgradients and the approximate function values.

Theorem 3.7. *One of the following two cases is true:*

(i) *if the proximal bundle procedure in Algorithm 3.5 does not terminate, that is, if* (3.17) *never hold, then the sequence of $\widehat{p}$-values converges to $p_\mu(x)$ and $p_\mu(x)$ is a minimizer of f;*

(ii) *if the procedure terminates with $\widehat{s} = 0$, then the corresponding $\widehat{p}$ equals $p_\mu(x)$ and is a minimizer of f.*

Proof. By ([12], Prop. 4.3), if this procedure does not terminate then it generates an infinite sequence of $\widehat{e}$-values and $\overline{\varepsilon}$-value converging to zero. Since (3.17) does not hold, the sequence of $\|\widehat{s}\|$-values also converges to 0. Thus, item (iii) in Theorem 3.4 implies that $\{\widehat{p}\} \to p_\mu(x)$. And Theorem 3.4 (ii) gives

$$f(z) \geq f(\widehat{p}) + \widehat{s}^T(z - \widehat{p}) - (\widehat{e} + \widehat{\varepsilon}), \quad z \in R^n. \tag{3.37}$$

By the continuity of f, this becomes

$$f(z) \geq f(p_\mu(x)), \quad z \in R^n. \tag{3.38}$$

The termination case with $\widehat{s} = 0$ follows in a similar manner, since (3.17) implies $\widehat{e} = 0$ in this case. □

The next theorem establishes the convergence of Algorithm 3.5, and the proof of which is similar to Theorem 9 in [8].

Theorem 3.8. *Suppose that the sequence $\{\mu_k\}$ in Algorithm 3.5 is bounded above by $\overline{\mu}$. Then the following hold:*

(i) *the sequence $\{\widetilde{f}(p_k)\}$ is decreasing and either $\{\widetilde{f}(p_k)\} \to -\infty$ or $\{\|s_k\|\}$ and $\{e_k\}$ both converge to 0;*

(ii) *if f is bounded from below, then any accumulation point of $\{p_k\}$ is a minimizers of f.*

Proof. In this paper, the inequalities of (3.15), (3.16), and (3.17) in [8] become

$$\widetilde{f}(p_{k+1}) - \widetilde{f}(p_k) \leq -\frac{m}{2\mu_{k+1}}\|s_{k+1}\|^2, \tag{3.39}$$

$$f(z) \geq f(p_k) + s_k^T(z - p_k) - \varepsilon_p^k - e_k, \quad \forall z \in R^n, \text{ where } \varepsilon_p^k = f(p_k) - \widetilde{f}(p_k), \tag{3.40}$$

$$e_k \leq \frac{\sigma_k}{\mu_k}\|s_k\|^2, \tag{3.41}$$

since $\|s_k\| \neq 0$, (3.39) implies that $\{\widetilde{f}(p_k)\}$ is decreasing. Suppose $\{\widetilde{f}(p_k)\} \nrightarrow -\infty$. Then summing (3.39) over k and using the fact that $m/2\mu_k \geq m/2\overline{\mu}$ for all k implies that $\{\|s_k\|\} \to 0$. Then (3.41) with $\sigma_k \leq 1/2$ and $\mu_k \geq \underline{\mu} > 0$ implies that $\{e_k\} \to 0$, which establishes (i).

Now suppose f is bounded from below and $\overline{p}$ is any accumulation point of $\{p_k\}$. Then, because $\{\|s_k\|\}$, $\{e_k\}$, and $\{\varepsilon_p^k\}$ converge to 0, (3.40) together with the continuity of f implies that $f(\overline{p}) \le f(z)$ for all $z \in R^n$ and (ii) is proved. □

4. An Illustrated Numerical Example

We test some examples in this section to validate the effectiveness of the proposed algorithm. The platform is Matlab R2009a, Intel Pentium(R) Dual CPU T2390 1.87 GHz. All test examples are of the form

$$f = \max_{i \in I} f_i, \tag{4.1}$$

where I is finite and each f_i is C^2 on R^n.

For our runs we used the following examples:

(i) F2d: the objective function is given in [8], defined for $x \in R^n$ by

$$\text{F2d}(x) := \max\left\{ \frac{1}{2}\left(x_1^2 + x_2^2\right) - x_2,\ x_2 \right\}; \tag{4.2}$$

(ii) F3d-$\mathcal{U}v$: 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 $x \in R^3$

$$\text{F3d-}\mathcal{U}v(x) := \max\left\{ \frac{1}{2}\left(x_1^2 + x_2^2 + 0.1x_3^2\right) - e^T x - \beta_1^v,\ x_1^2 - 3x_1 - \beta_2^v,\ x_2 - \beta_3^v,\ x_2 - \beta_4^v \right\}, \tag{4.3}$$

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 Table 1, we show some relevant data for the problems, including the dimensions of $\mathcal{V}$ and $\mathcal{U}$, the (known) optimal values and solutions, and the starting points.

We calculate an ε-subgradient at x by using the method in [13]: $g_\varepsilon(x) = \lambda g(x) + (1 - \lambda)g(x_1)$, where $g(x)$ is a subgradient at x and $g(x_1)$ is a subgradient at a point x_1 such that $0 < \alpha_0(x, x_1) = f(x) - f(x_1) - g(x_1)^T(x - x_1) \le \varepsilon$. Here $x_1 \in B_r(x) = \{z \in R^n \mid \|z - x\| \le r\}$ and $\lambda \in [0, 1]$ are randomly chosen. The approximate function value $\tilde{f}(x)$ is randomly taken out from the interval $[f(x) - \varepsilon, f(x)]$. The radius r is adjusted iteratively in the following way: If we find the linearization error $\alpha_0(x, x_1) > \varepsilon$ then r is reduced by a multiple smaller than one. On the other hand, if $\alpha_0(x, x_1)$ is significantly smaller than ε, then r is increased by a multiple greater than one. When $s_k = \hat{s}$ in the algorithm, then $\mathcal{U}$-Hessian at x is computed in the following form: $H_k = U_k^T(\sum_{j \in \hat{\mathcal{B}}} \overline{\lambda}_j H^j)U_k$, where $H^j = \nabla^2 f_{i_j}(y^j)$, i_j is an active index such that $f_{i_j}(y^j) = f(y^j)$, $\overline{\lambda}_j$ correspond to $\hat{s}$ via (3.16).

The parameters have values $\eta = 1.0 \times 10^{-4}$, $\varepsilon^0 = 1.0 \times 10^{-4}$, $m = 1.0 \times 10^{-1}$, $\tau = 1.0 \times 10^{-1}$, and U_0 equal to the $n \times n$ identity matrix. As for σ_k, μ_k, one can refer to [8].

Table 1: Problem data.

Problem	dim	dim-$\mathcal{V}$	dim-$\mathcal{U}$	x^*	fx^*	x^0
F2d	2	1	1	(0 0)	0	(0.9, 1.9)
F3d-$U3$	3	0	3	(0 1 10)	0	(100, 34, −90)
F3d-$U2$	3	1	2	(0 0 10)	0	(100, 33, −90)
F3d-$U1$	3	2	1	*(0 −1.741657 10)*	*−1.741657*	(100, 33, −100)
F3d-$U0$	3	3	0	*(1.5 −0.442695 9.962006)*	*−0.25*	(101, 33, −100)

The italic data in Table 1 is calculated by our algorithms.

Table 2: Numerical results of UV-decomposition algorithm with inexact data.

		Algorithm 6	Algorithm 3.5
F2d	$\#f/g$	28	34
	x	(0.000 0.000)	(0.000 0.000)
	$\lvert fx - fx^*\rvert$	0.000E0	1.753E-10
F3d-$U3$	$\#f/g$	11	32
	x	(0.000 1.000 10.000)	(0.000 1.000 10.000)
	$\lvert fx - fx^*\rvert$	0.000E0	2.248E-13
F3d-$U2$	$\#f/g$	32	43
	x	(0.000 −0.000 10.000)	(0.000 0.000 10.000)
	$\lvert fx - fx^*\rvert$	0.000E0	1.275E-10
F3d-$U1$	$\#f/g$	****	33
	x	****	(−0.0000 −1.742 10.000)
	$\lvert fx - fx^*\rvert$	****	7.50E-15
F3d-$U0$	$\#f/g$	31	44
	x	(1.500 −0.425 10.000)	(1.500 −0.443 9.962)
	$\lvert fx - fx^*\rvert$	8.771E-11	2.908E-11

Table 2 shows the results of Algorithm 3.5 for these examples, compared with Algorithm 6 in [8]. Number of f/g denotes the number of evaluation of the function and subgradient (ε-subgradient) in Algorithm 6 and Algorithm 3.5. x is the calculated solution, $|fx - fx^*|$ stands for the difference between the function values at x and x^*.

It is shown in Table 2 that we obtain quite accurate solutions by Algorithm 3.5 with inexact data costing a slightly more evaluation number than that with exact data. One noticeable exceptional occurs in the example F3d-$U1$; it seems that the decomposition algorithm is sensible with exact data, but is more stable when applying inexact data (function values and subgradients). This favorable results demonstrate that it is suitable to use approximate decomposition algorithm to solve (1.1) numerically.

Acknowledgment

The research is supported by the National Natural Science Foundation of China under Project no. 11171049 and no. 11171138.

References

[1] J. J. Moreau, "Proximitéet dualité dans un espace hilbertien," *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965.

[2] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM Journal on Control and Optimization*, vol. 14, no. 5, pp. 877–898, 1976.
[3] A. Auslender, "Numerical methods for nondifferentiable convex optimization," *Mathematical Programming Study*, no. 30, pp. 102–126, 1987.
[4] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algo-Rithms*, vol. 2, Springer, 1993.
[5] C. Lemaréchal, F. Oustry, and C. Sagastizábal, "The $\mathcal{U}$-lagrangian of a convex function," *Transactions of the American Mathematical Society*, vol. 352, no. 2, pp. 711–729, 2000.
[6] R. Mifflin and C. Sagastizábal, "$\mathcal{VU}$-decomposition derivatives for convex max-functions," in *Ill-Posed Variational Problems and Regularization Techniques*, vol. 477 of *Lecture Notes in Economics and Mathematical Systems*, pp. 167–186, Springer, Berlin, Germany, 1999.
[7] R. Mifflin and C. Sagastizábal, "On $\mathcal{VU}$-theory for functions with primal-dual gradient structure," *SIAM Journal on Optimization*, vol. 11, no. 2, pp. 547–571, 2000.
[8] R. Mifflin and C. Sagastizábal, "A $\mathcal{VU}$-algorithm for convex minimization," *Mathematical Programming B*, vol. 104, no. 2-3, pp. 583–608, 2005.
[9] Y. Lu, L.-P. Pang, X.-J. Liang, and Z.-Q. Xia, "An approximate decomposition algorithm for convex minimization," *Journal of Computational and Applied Mathematics*, vol. 234, no. 3, pp. 658–666, 2010.
[10] R. Mifflin and C. Sagastizábal, "Proximal points are on the fast track," *Journal of Convex Analysis*, vol. 9, no. 2, pp. 563–579, 2002.
[11] M. V. Solodov, "On approximations with finite precision in bundle methods for nonsmooth optimization," *Journal of Optimization Theory and Applications*, vol. 119, no. 1, pp. 151–165, 2003.
[12] R. Correa and C. Lemaréchal, "Convergence of some algorithms for convex minimization," *Mathematical Programming B*, vol. 62, no. 2, pp. 261–275, 1993.
[13] M. Hintermüller, "A proximal bundle method based on approximate subgradients," *Computational Optimization and Applications*, vol. 20, no. 3, pp. 245–266, 2001.