A DOUBLY INEXACT INTERIOR PROXIMAL BUNDLE METHOD
FOR CONVEX OPTIMIZATION

KOUHEI HARADA†

Abstract. We propose a version of bundle method for minimizing a non-smooth convex function. Our bundle method have three features. In bundle method, we approximate the objective with some cutting planes and minimize the model with some stabilizing term. Firstly, it allows inexactness in this minimization. At the same time, evaluations of the objective and the subgradient are also required to generate the cutting planes. Secondly, inexact evaluation is also allowed. In this sense, our algorithm is doubly inexact. Thirdly, the stabilizing term are not assumed to be quadratic. We use interior proximal distance which was developed by Auslender and Teboulle, as the stabilizing term. We show that it globally converges to an optimal solution. Furthermore, it is also applicable to Lagrange relaxation problems. We show that both primal feasibility and optimality are asymptotically achieved.

Key words. bundle method, Lagrangian relaxation, inexact oracle, primal recovery, proximal distance

AMS subject classifications. 90C25, 90C30

1. Introduction.
In this paper, we present an extension of proximal bundle method for constrained non-smooth convex programming problems

\[
\text{minimize } f(x), \ x \in X,
\]

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \) is a proper convex, lower-semicontinuous, and possibly non-smooth function, \( X \) is a convex open set of the Euclidean space \( \mathbb{R}^n \) and \( \overline{X} \) denotes the closure of \( X \). We also assume \( \partial f \) is bounded on the bounded subset of \( X \).

In proximal bundle method, we have to generate cutting planes which are tangent to the objective \( f \). The aim of this evaluation is to approximate the objective \( f \). Minimizing this approximated objective with stabilizing term which is in general quadratic, we obtain a trial point. The validity of the trial point is judged by some criteria. If the trial point is valid, we move on to the point. Otherwise, we stay the same point. Whether the trial point is valid or not, we generate a cutting plane again at the trial point, and repeating the procedure until the stopping criteria is satisfied. This is a classical manner of the proximal bundle method [13, 17, 9, 27].

Our algorithm differs from classical bundle method in three points. Firstly, it allows inexact minimization in searching a trial point. Secondly, it also allows inexact evaluation in generating a cutting plane. Therefore, this is not always tangent to the objective \( f \). However, we assume that the generated cutting plane always support the objective \( f \). In the context of Lagrangian relaxation, it implies that Lagrangian relaxation problems do not have to be solved precisely. Thirdly, the stabilizing term is not limited to be quadratic. Interior proximal distance which was developed by Auslender [3] is allowed as the stabilizing term with some additional assumptions. The extension of proximal bundle method has been widely researched, and our research is inspired by their works. We classify them from three points of features which we mentioned above, model minimization, evaluation and stabilizing term.

Firstly, we focus on model minimization. On this point, almost all research re-
quire exact model minimization and little attention has been given to inexact model minimization. Inexact model minimization has been only considered in Kiwiel [20], as far as we know. They consider Bregman-distance function as the stabilizing term, and proved its convergence. However, exact evaluation is assumed.

Secondly, we consider evaluation which is often called oracle. Contrary to model minimization, many researches has been established on this point. In general, bundle methods which cover this problem are often called inexact or approximate bundle method. On the other hand, we have to realize that the degree of inexactness are different. An early work can be seen in Kiwiel [16]. The assumptions considered in Kiwiel [18] and Solodov [28] are similar to ours. Furthermore, stability analysis is considered in [28], but this is not the subject of this paper. Hintermuller [12] deals with inexact subgradient, but the evaluation of the objective \( f \) should be exact. More generalized inexactness were considered in Kiwiel [21] and Oliveira [10]. In [21], both objective and subgradient are independently inexact, and upper inexactness for objective is allowed. Detailed classification of inexact evaluation is mentioned in [21, Section 1]. Furthermore, uncontrollable inexactness is considered in [10]. They use noise attenuation technique to cope with inexact oracle mainly derived from stochastic objective. Although stochastic programming is not the subject of this paper, the degree of inexactness assumed at [10] is most generalized to our knowledge. Unfortunately, all these researches have been done only for quadratic stabilizing term. In our algorithm, we proved similar assumption at [18, 28] is possible with inexact model minimization and non-quadratic stabilizing term.

Thirdly, we consider the extension of stabilizing term, it has been proved that some distance-like functions are also possible to bundle method. For example, Bregman function by Kiwiel [20], \( \varphi \)-divergence with homogeneous kernel by Auslender [2, Section 3] and Van [25], and symmetric type distance by Frangioni [11]. The relations among distance-like functions are deeply compared in [11, Section 9]. In our algorithm, we adopt an interior proximal distance function developed by [3, Section 2] as a stabilizing term with some additional assumptions. From this point of view, our algorithm is an extension of [2, Section 3], [25] and partially include [20].

In this paper, we also focus on an application to Lagrangian relaxation method. It is well known that optimizing a dual function is a typical example of non-smooth convex optimization. Nevertheless, to apply bundle method to this problem, we not only have to obtain a dual optimal solution, but also have to obtain a primal feasible and optimal solution [23, 22]. This problem was solved by Kiwiel [18] in quadratic stabilizing term. In this paper, we prove that both primal feasibility and optimality are asymptotically achieved despite extended stabilizing term and doubly inexactness are assumed.

This paper is organized as follows. In section 2, we give the framework of the algorithm, doubly inexact interior proximal bundle method. In section 3, we introduce interior proximal distance and some additional assumptions with some examples [3, Section 3]. In section 4, we prove the globally convergence property of our algorithm. The proof in this section is an extension of [20, 2]. In section 5, we consider an application of this algorithm to Lagrangian relaxation method. In section 6, it is shown that inexactness stems from model minimization is controllable. In section 7, we briefly conclude this paper and consider some possibilities for further research.

We use the following standard notation which are mainly derived from [26]. For all closed convex set \( S \), \( \delta_S \) denote the indicator function of \( S \). For a proper convex and lower semicontinuous function \( F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \), its effective domain is defined by \( \text{dom}F = \{x|F(x) < +\infty\} \). For all \( \epsilon \geq 0 \), \( \epsilon \)-subdifferential of a function \( F \) at \( x \)
is defined by $\partial F(x) = \{ s \in \mathbb{R}^n \mid F(y) \geq F(x) + \langle s, y - x \rangle - \epsilon, \forall y \in \mathbb{R}^n \}$. For all $\alpha \in \mathbb{R}$, level set of $F$ with respect to $\alpha$ is defined by $L_F(\alpha) = \{ x \mid F(x) \leq \alpha \}$. We call a function $F$ is level-bounded if $L_F(\alpha)$ is bounded for every $\alpha \in \mathbb{R}$. The domain of $\partial F = \partial_0 F$ is defined by $\text{dom} \partial F = \{ x \mid \partial F(x) \neq \emptyset \}$. $\partial_1 d(x,y)$ denote the gradient map of a function $d(x,y)$ with respect to the first variable $x$. If $\partial_1 d(x,y)$ is a singleton, it is written as $\nabla_1 d(x,y)$ [3]. The set of $n$-vectors with nonnegative(positive) components is denoted by $\mathbb{R}^n_+ (\mathbb{R}^n_{++})$.

Let $\gamma$ be a piecewise affine function given by

$$\varphi(x) = \max_{i \in I} \{ \langle g^i, x \rangle + \beta_i \}.$$ 

Then $\gamma$-subdifferential of $\varphi$ at $x$ is given as follows.

$$\partial_{\gamma} \varphi(x) = \left\{ \sum_{i \in I} \theta_i g^i \mid \sum_{i \in I} \theta_i = 1, \theta \geq 0, \sum_{i \in I} \theta_i \alpha_i(x) \leq \gamma \right\},$$

$$\alpha_i(x) = \varphi(x) - \langle g^i, x \rangle - \beta_i.$$
The validity of the trial point $y^{k+1}$ is confirmed by descent test. It judges if the
goal decrease is sufficient or not. Letting

$$\delta_k := \hat{f}(x^k) - \varphi^k(y^{k+1}),$$

and a positive constant $m \in (0, 1)$, descent test is defined as

$$\hat{f}(x_k) - \hat{f}(y^{k+1}) \geq m\delta_k.$$

If the test returns true, $y^{k+1}$ is adopted to the next point. Otherwise, we stay the
same point. After the descent test, regardless of its result (accepted or not), we have
a chance to reconstruct bundle. This phase is often called as aggregation. From (2.4)
and Lemma 2.1, there exists $\theta^k \in \mathbb{R}_+^n$ such that

$$s^{k+1} = \sum_{i \in I_k} \theta^k_i g^i, \sum_{i \in I_k} \theta^k_i = 1, \sum_{i \in I_k} \theta^k \alpha_i(y^{k+1}) \leq \gamma_k.$$  

Then, the following active indices are defined.

$$\hat{I}_k = \{i \in I_k \mid \theta^k_i > 0\}.$$  

The bundle $I_{k+1}$ at the next iteration have to satisfy $I_{k+1} \supseteq \hat{I}_k \cup \{k + 1\}$. The aim
of the aggregation is to save storage and solve SMMP rapidly.

We may now state the algorithm, doubly inexact interior proximal bundle method (DIIPBM),
in detail.

**Algorithm 1 DIIPBM (doubly inexact interior proximal bundle method)**

**Step 0. Initialize**

Let $m \in (0, 1)$, $0 < \lambda_L \leq \lambda_U$, $\epsilon_{opt} > 0$ sufficiently small and $c_{max} \in \mathbb{N}$
sufficiently large. Set $x_0 \in X$, $k = 0$, $c = 0$ and $\lambda_0 \in [\lambda_L, \lambda_U]$. Evaluate
$f(x_0), g_0 \in \partial f(x_0)$ and let $\varphi^0(x) = f(x_0) + \langle g_0, x - x_0 \rangle$. Set $I_0 = \{0\}.$

**Step 1. Yielding a next trial point (Solving SMMP)**

Solve SMMP approximately and obtain a trial point $y^{k+1}$ and search direction
$s^{k+1}$ such that (2.4) hold. We also obtain active indices set $\hat{I}_k$ as (2.9).

**Step 2. Evaluation (Generating a cutting plane)**

Evaluate $f(y^{k+1}), g^k \in \partial f(y^{k+1}).$

**Step 3. Descent Test**

If $f(x^k) - f(y^{k+1}) \geq m\delta_k$, then $x^{k+1} = y^{k+1}$, $\lambda_{k+1} = [\lambda_L, \lambda_U]$ and set $c = 0.$
Goto Step 5 [Decrease Step]

Otherwise, $x^{k+1} = x^k$, $\lambda_{k+1} = \lambda_k$ and increment $c$ as $c = c + 1$. If $c = 0$, we
memorize the index $k$ as $\hat{k}$. Goto Step 4 [Null Step]

**Step 4. Exact evaluation for many null steps**

If $c$ reaches $c_{max}$, evaluate $f(x^{k+1}), g_{k+1} \in \partial f(x^{k+1})$ again exactly.

**Step 5. Stopping criteria**

If $\max\{\|y^{k+1} - x^k\|, \|\delta_k\|, \|\gamma_k\|, \|\eta_k\|\} \leq \epsilon_{opt}$, exit.

**Step 6. Aggregation**

Prepare $I_{k+1} \supseteq \hat{I}_k \cup \{k + 1\}$ and $\varphi^{k+1}$.

**Step 7. Index updating**

Update $k = k + 1$. Goto Step 1.

In our algorithm, following assumptions are required for the non-negative se-
quences $\{\eta_k\}$ and $\{\gamma_k\}$.  

4
(A1) \( \sum_{k=0}^{\infty} \eta_k < +\infty; \)
(A2) \( \sum_{k=0}^{\infty} \gamma_k < +\infty. \)

Before we finish this section, we comment the difference in inexact oracle between [18, 28] and ours. As we mentioned at section 1, our assumption is similar to [18, 28], because only lower inexactness are allowed, and the inexactness of the objective and subgradient are not independent. In these points, our assumptions are equal to theirs. The difference is whether the inexactness is controllable or not. The inexactness is assumed to be controllable in [18, 28]. To the contrary, we don’t prepare any rule to control \( \eta_k \) except a special case at STEP4. The role of STEP4 is that we only need an exact evaluation at \( \bar{k} \in \mathbb{N} \) such that only null-steps occur for all \( k \) larger than \( \bar{k} \). The necessity of the exact oracle is explained in Section 4.2. Unfortunately, we cannot predict whether only null steps occur or not beforehand. Thus, we have to evaluate again exactly if \( \bar{k} \) is a candidate which is judged by comparing \( c_{max} \) and \( c \). Except the special case, even \( \eta_{k+1} > \eta_k \) at some \( k \) might be possible as far as (A1) hold.

Our assumption is more practically useful in the context of Lagrangian relaxation, because we cannot measure if the obtained inexact solution is \( \eta \)-optimal or not without an exact solution. Inexact evaluation is useful mainly because we can avoid exact evaluation.

On the other hand, in terms of inexact model minimization, we can control \( \gamma_k \) which is explained in Section 6.

3. Interior proximal distances. In our algorithm, some assumptions are required for distance-like function \( d \). We assume that \( d \) have to be an interior proximal distance which was developed by Auslender [3]. It should be mentioned that they don’t use the notation interior proximal distance, but we use the notation to distinguish it with other general proximal distances. In this section, we introduce its definition and some additional properties which are also necessary for our algorithm. The contents of this section is mainly based on [3, Section 2, 3].

We call a function \( d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\} \) as a interior proximal distance with respect to an open nonempty convex set \( X \subset \mathbb{R}^n \) if following properties hold [3, Definition 2.1].

**Definition 3.1.** For every \( z \in X \), following properties are satisfied.
(P1) \( d(\cdot, z) \) is proper, lower-semicontinuous, strictly convex and \( C^1 \) over \( X \);
(P2) \( \text{dom}d(\cdot, z) \subset \bar{X} \) and \( \text{dom}\partial_1 d(\cdot, z) = X \);
(P3) \( \lim_{\|x\| \to \infty} d(x, z) = +\infty \);
(P4) \( d(z, z) = 0 \);

Note that the definition is slightly different from the original. In addition to original definition, strict convexity of \( d(\cdot, z) \) is assumed in our case. In accordance with Auslender [3], we denote \( d \in \mathcal{D}(X) \) if a function \( d \) satisfies (P1) to (P4). Furthermore, a interior proximal distance \( d \) should satisfy the following properties [3, Definition 2.2].

**Definition 3.2.** Let \( H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\} \) be a function with respect to \( d \). \( H \) is also finite valued and positive on \( \bar{X} \times \bar{X} \). For all \( z \in \bar{X} \), \( H(z, \cdot) \) is level-bounded. Furthermore, for every \( x, y \in X \), following properties are satisfied.
(P5) \( H(x, x) = 0 \);
(P6) \( H(z, x) - H(z, y) \geq \langle z - y, \nabla_1 d(y, x) \rangle \);
(P7) \( \lim_{\|x\| \to \infty} H(z, x) = +\infty \);
(P8) Suppose the sequence \( \{x^k\} \) is bounded and \( x^k \in X \) for all \( k \), For all \( x \in \bar{X} \), \( \lim_{k \to +\infty} H(x, x^k) = 0 \iff \lim_{k \to +\infty} x^k = x \).
It should be remarked that these conditions are most restricted version among mentioned in [3]. We denote the pair \((d, H) \in \mathcal{F}_+(\overline{X})\) if (P1) to (P8) hold for \((d, H)\) in accordance with [3].

In our algorithm, we also need the additional following assumptions which are required to prove the convergence.

\textbf{(AD1)} There exists two non-negative functions \(P, Q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}\) such that,
\[
P(z, x) - P(y, x) \geq \langle z - y, \nabla d(y, x) \rangle + Q(z, y) \quad \text{for all } x, y, z \in X;
\]

\textbf{(AD2)} Suppose \(Q(x^k, x^{k+1}) \to 0, x^k \in X\) for all \(k\) and the sequence \(\{x^k\}\) is bounded, then \(\|x^k - x^{k+1}\| \to 0\);

\textbf{(AD3)} Suppose the sequence \(\{x^k\}\) is bounded and \(x^k \in X\) for all \(k\),

For all \(x \in \overline{X}\), \(\lim_{k \to +\infty} Q(x, x^k) = 0 \iff \lim_{k \to +\infty} x^k = x\).

We denote \((d, H, P, Q) \in \mathcal{A}(X)\) if \((d, H) \in \mathcal{F}_+(\overline{X})\) and (AD1) to (AD3) hold for \((d, P, Q)\).

Following lemma is useful to prepare a new interior proximal distance for a direct product.

\textbf{Lemma 3.3.} Suppose \(X \subset \mathbb{R}^{n_X}\) and \(Y \subset \mathbb{R}^{n_Y}\) are open nonempty convex sets, \((d_X, H_X, P_X, Q_X) \in \mathcal{A}(X)\) and \((d_Y, H_Y, P_Y, Q_Y) \in \mathcal{A}(Y)\). \((d, H, P, Q) = (d_X + d_Y, H_X + H_Y, P_X + P_Y, Q_X + Q_Y) \in \mathcal{A}(X \times Y)\) hold.

\textbf{Proof.} Just a confirmation. \(\square\)

We briefly introduce some examples of interior proximal distance [3, Section 3].

\textbf{3.1. Euclidean distance.} If \(X = \mathbb{R}^n, d(x, y) = \frac{1}{2} \|x - y\|^2\) is, of course, interior proximal distance. Letting \(d(x, y) = H(x, y) = \frac{1}{2} \|x - y\|^2\), \((d, H)\) satisfies (P1) to (P8). Since the following three term property,
\[
\frac{1}{2} \|z - x\|^2 = \frac{1}{2} \|z - y\|^2 + \frac{1}{2} \|y - x\|^2 + \langle z - y, y - x \rangle,
\]
hold, (AD1) hold by \(P(x, y) = Q(x, y) = \frac{1}{2} \|x - y\|^2\). (AD2),(AD3) are obvious.

\textbf{3.2. Bregman proximal distances.} Let \(h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}\) be a proper, lsc, and strictly convex functions with \(\text{dom} h = X\) and \(\text{dom} \nabla h = X\), strictly convex and continuous on \(\text{dom} h\), \(C^1\) on \(\text{int} \text{dom} h = X\).

The \(D\)-function \(D_h(x, y)\) with respect to function \(h\) is defined by
\[
D_h(x, y) := h(x) - [h(y) + \langle \nabla h(y), x - y \rangle].
\]

We call \(h\) is a Bregman function with zone \(X\) if the following conditions hold [7].

\textbf{(B1)} \(\text{dom} h = \overline{X}\);

\textbf{(B2)} For all \(x \in X\), \(D_h(x, \cdot)\) is level-bounded on \(\text{int(dom} h)\);

\textbf{(B3)} For all \(y \in X\), \(D_h(\cdot, y)\) is level-bounded;

\textbf{(B4)} (P8) hold for \(H = D_h\).

Letting \(d(x, y) = H(x, y) = D_h(x, y)\), \(D\)-function satisfy (P1) to (P8) if \(h\) is a Bregman function with zone \(X\) [3].

Here are examples of Bregman functions with zone \(\mathbb{R}_+\).

- \(h_1(x) = x \log x [6]\),
- \(h_2(x) = (px - x^p)/(1 - p)\) with \(p \in (0, 1) [29]\).

For every \(q = 1, 2\), if \(h(x) = \sum_{j=1}^n h_q(x)\), we have \(D_h(x, y) = \sum_{j=1}^n D_{h_q}(x, y)\).

Hence, we can construct a separable interior proximal distance in \(\mathbb{R}_+^n\) using Bregman
functions with zone $\mathbb{R}_+$ as a component. It should be noted that if $h(x) = \|x\|^2/2$, $D_h(x, y)$ become a quadratic function. Hence it is a Bregman function with zone $\mathbb{R}^n$

It is well known that $D$-functions satisfy the three point identity [8]

$$D_h(x, y) = D_h(z, y) + D_h(y, x) + [z - y, \nabla_1 D_h(y, x)], \forall x, y \in X, \forall z \in \text{dom} h.$$ 

Therefore, (AD1) is recovered by letting $P(x, y) = Q(x, y) = D_h(x, y)$. (AD2) follows from [19, Lemma 2.10]. Since $D_h = H = Q$, (AD3) is obvious.

3.3. $\varphi$-divergence with regularization. We introduce another class of distance, $\varphi$-divergence. Let $\varphi : \mathbb{R} \to [0, \infty]$. We call $\varphi$ be a $\varphi$-divergence kernel if the following properties hold.

1. $\varphi$ is an lower-semicontinuous, convex and proper function.
2. $\text{dom} \varphi \subset \mathbb{R}_+$ and $\text{dom} \partial \varphi = \mathbb{R}_+$.
3. $\varphi$ is strictly convex on $\mathbb{R}_+$.
4. $\varphi(1) = \varphi'(1) = 0$.

In addition to the properties above, if $\varphi$ satisfies the following property:

$$\varphi''(1) \left(1 - \frac{1}{t}\right) \leq \varphi'(t) \leq \varphi''(1) \log t, \forall t > 0.$$ (3.1)

We call them class $\Phi_1$. The other subclass of $\Phi$ is denoted by $\Phi_2$ in which (3.1) is replaced by

$$\varphi''(1) \left(1 - \frac{1}{t}\right) \leq \varphi'(t) \leq \varphi''(1)(t - 1), \forall t > 0.$$ (3.2)

Corresponding to the classes $\Phi_r$, with $r = 1, 2$, we define a $\varphi$-divergence function by

$$d_\varphi(x, y) = \sum_{i=1}^n (y_i^r) \varphi\left(\frac{x_i}{y_i^2}\right).$$

For $\varphi \in \Phi_1$, it has been proved that $d(x, y) = d_\varphi(x, y), H(x, y) = \varphi''(1) \sum_{j=1}^n D_{h_1}(x_j, y_j)$ satisfies (P1) to (P7) [30]. But this is not sufficient to cover (P8). To hold (P8), letting $d(x, y) = d_\varphi(x, y) + \frac{\sigma}{2} \|x - y\|^2$, $H(x, y) = \varphi''(1) \sum_{j=1}^n D_{h_1}(x_1, y_j) + \frac{\sigma}{2} \|x - y\|^2$.

As shown in [1], one can verify that this also covers (P1) to (P8). Then, it can be shown that $P(x, y) = d(x, y), Q(x, y) = \frac{\sigma}{2} \|x - y\|^2$ satisfies (AD1) to (AD3).

At the same time, for $\varphi \in \Phi_2$, it has been proved by Auslender [4] that $d(x, y) = d_\varphi(x, y) + \frac{\sigma}{2} \|x - y\|^2$, $H(x, y) = (\sigma - 1) \|x - y\|^2$ covers (P1) to (P8) with $\sigma > 1$. In this case, it is easy to show that $P(x, y) = d(x, y), Q(x, y) = \frac{\sigma}{2} \|x - y\|^2$ covers (AD1) to (AD3).

For more detailed explanation of $\varphi$-divergence, see [29, 15, 14].

4. Convergence. In this section, we show the global convergence of DIIPBM. To prove the convergence, we assume $\epsilon_{opt} = 0$ throughout this section. We have three possibilities about the behavior of the algorithm as follows:

Case1: Infinitely many decrease steps
Case2: Finitely decrease steps and infinitely many null steps
Case3: Finitely decrease steps and stops

We show that in every case, the algorithm converges to an optimal point.
4.1. Decrease Step. In this section, we consider (Case 1). Let \( K \) be a set of the indices of decrease steps. In this subsection, we are interested in only decrease steps. Therefore, we renumber the indices \( \kappa \in K \) consecutively as 0, 1, 2, \ldots for simplicity.

One difficulty of this renumbering is STEP 4 of the algorithm. After renumbering, notation \( f(x^k) \) is vague. More precisely, suppose \( \kappa \in K, \kappa + p \in K, \kappa + 1, \ldots, \kappa + p - 1 \notin K, p \) is larger than \( c_{\max} \) and \( k, k + p \) are renumbered to \( k, k + 1 \) respectively. From STEP 4, we have \( f(x^\kappa) \leq f(x^{\kappa + p - 1}) = f(x^\kappa) \), before renumbering. Hence, after renumbering, \( f(x^k) = f(x^k) \) from the viewpoint of \( k + 1 \). On the other hand \( f(x^k) = f(x^k) - \eta_k \) is valid from the viewpoint of \( k \).

To distinguish these two, we use the notation \( f(x^k) \) an original value, and simply \( f(x^k) \) as a replaced value after rearrangement. Let \( K_0 \) as a subset of \( \mathbb{N} \) which consists of \( k \) such that STEP 4(replacement) occurs between \( k \) and \( k + 1 \). Let \( \delta_k \) as

\[
\delta_k := \begin{cases} 
\tilde{f}(x^k) - \varphi^k(x^{k+1}), & k \notin K_0, \\
\tilde{f}(x^k) - \varphi^k(x^{k+1}), & k \in K_0.
\end{cases}
\]

Considering only decrease step, following property hold throughout this subsection.

(4.1) \[
m\delta_k \leq \begin{cases} 
\tilde{f}(x^k) - \tilde{f}(x^{k+1}), & k \notin K_0, \\
\tilde{f}(x^k) - \tilde{f}(x^{k+1}), & k \in K_0.
\end{cases}
\]

In standard bundle method, \( \delta_k \geq 0 \) is clear, but in our case the statement is not always true. We start with the following lemma.

**Lemma 4.1.** Let \( \{\delta_k\}, \{\eta_k\}, \{\gamma_k\} \) and \( \{\tilde{f}(x_k)\} \) be the sequences produced by Algorithm DIIPBM. Suppose infinitely many decrease steps occur, the following hold.

(i) \( \delta_k + \eta_k + \gamma_k \geq 0 \).
(ii) \( \tilde{f}(x_k) \) converges to some \( f^* \).
(iii) \( \delta_k + \eta_k + \gamma_k \downarrow 0 \).

**Proof.** From (4.1) (2.2) and (2.4), we have

\[
\delta_k + \eta_k \geq \tilde{f}(x^k) - \varphi^k(x^{k+1}) = [f(x^k) - \varphi^k(x^k)] + [\varphi^k(x^k) - \varphi^k(x^{k+1})] \\
\geq \langle s^{k+1}, x^k - x^{k+1} \rangle - \gamma_k.
\]

Hence, we have

\[
\delta_k + \eta_k + \gamma_k \geq \lambda_{k+1}^{-1} \langle \nabla_1 d(x^{k+1}, x^k), x^{k+1} - x^k \rangle \geq \lambda_{k+1}^{-1} d(x^{k+1}, x^k) \geq 0.
\]

Therefore (i) is proved.

From (4.2), for all \( k \in K \), we have

\[
f(x^k) - \tilde{f}(x^{k+1}) \geq m\delta_k.
\]

Therefore, we have

(4.3) \[
\hat{f}(x^k) - \tilde{f}(x^{k+1}) + (1 + m)\eta_k + m\gamma_k \geq m(\delta_k + \eta_k + \gamma_k).
\]

Let \( F(x^k) := \hat{f}(x^k) - (1 + m) \sum_{i=0}^{k-1} \eta_i - m \sum_{i=0}^{k-1} \gamma_i \), (4.3) yield

\[
F(x^k) - F(x^{k+1}) = \hat{f}(x^k) - \tilde{f}(x^{k+1}) + (1 + m)\eta_k + m\gamma_k \\
\geq m(\delta_k + \eta_k + \gamma_k).
\]
Since \( \delta_k + \eta_k + \gamma_k \geq 0 \) from (i), \( \{F(x^k)\} \) is a decreasing sequence. Furthermore, \( \eta_k \downarrow 0 \), function \( f \) is lower bounded, and by (A1)(A2), There exists a \( F^* \) such that \( F(x^k) \downarrow F^* \). Therefore, there exists \( f^* \in \mathbb{R} \) such that

\[
\tilde{f}(x^k) \to f^* = F^* - \sum_{k}^{\infty} \eta_k - \sum_{k}^{\infty} \gamma_k.
\]

We established (ii).

Summing inequalities (4.3), we have

\[
\tilde{f}(x_0) - f^* + (1 + m) \sum_{k=0}^{\infty} \eta_k + m \sum_{k=0}^{\infty} \gamma_k \geq m \sum_{k=0}^{\infty} (\delta_k + \eta_k + \gamma_k).
\]

It follows from (A1)(A2) and (ii) that the left hand side of the inequation is bounded. Hence, we have \( \gamma_k + \delta_k + \eta_k \downarrow 0 \). (iii) is proved.

To prove the global convergence, we use the following result proved by Auslender [3, Theorem 2.2].

**Theorem 4.2.** Given \((d,H) \in \mathcal{F}(X), \lambda_k > 0\) and \( \epsilon_k \geq 0 \) for all \( k \). Let \( \{x^k\} \in X \) be a sequence such that \( s^k \in \partial \epsilon_k f(x^k), s^k + \lambda_k^{-1} \nabla_d(x^k,x^{k-1}) = 0 \).

Then, suppose the following conditions

\[
\sum_{k=0}^{+\infty} \lambda_k \to +\infty, \sum_{k=0}^{+\infty} \lambda_k \epsilon_k < +\infty, \sum_{k=0}^{+\infty} \epsilon_k < +\infty,
\]

are satisfied, the sequence \( \{x^k\} \) converges to an optimal solution.

To apply this result, we define \( \epsilon_k \) as follows:

\[
\epsilon_k := (1 - m)\delta_{k-1} + \eta_k + \gamma_k.
\]

**Proposition 4.3.** Let \( \{x^k\}, \{\epsilon_k\} \) and \( \{s^k\} \) be sequences produced by Algorithm DIIPBM. Suppose infinitely many decrease steps occur, the following hold.

(i) \( \epsilon_k \downarrow 0 \).

(ii) \( \sum_{k=0}^{+\infty} \epsilon_k < +\infty \).

(iii) \( s^k \in \partial \epsilon_k f(x^k) \).

**Proof.** From (4.1) and (4.2), we have

\[
f(x^k) - \varphi^{k-1}(x^k) = \left[ f(x^k) - f(x^{k-1}) \right] + \left[ f(x^{k-1}) - \varphi^{k-1}(x^k) \right] = \begin{cases} 
\tilde{f}(x^k) - \tilde{f}(x^{k-1}) + \left[ f(x^{k-1}) - \varphi^{k-1}(x^k) \right] + \eta_k, k \notin K_0 \\
\tilde{f}(x^k) - f(x^{k-1}) + \left[ f(x^{k-1}) - \varphi^{k-1}(x^k) \right] + \eta_k, k \in K_0 
\end{cases}
\]

\[
\leq (1 - m)\delta_{k-1} + \eta_k.
\]

In view of (2.2), we have

\[
(1 - m)\delta_{k-1} + \eta_k \geq 0.
\]
Thus, we've proved (iii).

Meanwhile, since we have \( \delta_k + \eta_k + \gamma_k \downarrow 0, \eta_k \downarrow 0 \) and \( \gamma_k \downarrow 0 \), it implies \( \delta_k \to 0 \). Then, \( (1 - m)\delta_{k-1} + \eta_k + \gamma_k \to 0 \) and (i) is proved. (ii) follows from \((A1)(A2)\) and \((4.4)\).

To prove (iii), in view of \((2.2)\), \((2.4)\) and \((4.6)\), we have

\[
\begin{align*}
\ell(x) & \geq \varphi^{k-1}(x) \\
& \geq \varphi^{k-1}(x^k) + \langle s^k, x - x^k \rangle - \gamma_k \\
& \geq [-f(x^k) + \varphi^{k-1}(x^k)] + f(x^k) + \langle s^k, x - x^k \rangle - \gamma_k \\
& \geq f(x^k) + \langle s^k, x - x^k \rangle - \epsilon_k.
\end{align*}
\]

Thus, we’ve proved (iii).

**Theorem 4.4.** Suppose infinitely many decrease steps occur in Algorithm DI-IPBM, the sequence \( \{x^k\} \) globally converges to an optimal solution.

**Proof.** Since Proposition 4.3 hold, we can apply Theorem 4.2. Third Condition follows from Proposition 4.3 (ii). First condition follows from \( 0 < \lambda_L \leq \lambda_k \). Second condition follows from \( 0 < \lambda_k \leq \lambda_U \) and third condition.

The validity of the STEP5(stopping criteria) is confirmed as follows.

**Proposition 4.5.** Suppose infinitely many decrease steps occur in Algorithm DI-IPBM, the stopping criteria at STEP5 hold asymptotically. More specifically, we hold the following:

\[
\lim_{k \to +\infty} (y^{k+1} - x^k) = 0, \quad \lim_{k \to +\infty} \delta_k = 0, \quad \lim_{k \to +\infty} \gamma_k = 0, \quad \lim_{k \to +\infty} \eta_k = 0
\]

**Proof.** \( \gamma_k \to 0, \eta_k \to 0 \) is obvious from \((A1)(A2)\). In view of Lemma 4.1, \( \delta_k \to 0 \) is also obvious. Since \( \delta_k = \delta^k - \varphi^k(y^{k+1}) = [\langle \nabla d(y^{k+1}, x^k), y^{k+1} - x^k \rangle] + [\varphi^k(x^k) - \varphi^k(y^{k+1})], \)

\( \delta_k \to 0 \) and \( \eta_k \to 0 \) implies \( [\varphi^k(x^k) - \varphi^k(y^{k+1})] \to 0 \). At the same time, by substituting \((x, y, z) = (x^k, y^{k+1}, x^k)\) to \((AD1)\), we have

\[
\langle \nabla d(y^{k+1}, x^k), y^{k+1} - x^k \rangle \geq P(y^{k+1}, x^k) + Q(x^k, y^{k+1})
\]

Hence, from \((2.4)\), we have

\[
\varphi^k(x^k) - \varphi^k(y^{k+1}) + \gamma_k \geq \lambda_k^{-1} \langle \nabla d(y^{k+1}, x^k), y^{k+1} - x^k \rangle \\
\geq P(y^{k+1}, x^k) + Q(x^k, y^{k+1}) \\
\geq 0
\]

Therefore, \( [\varphi^k(x^k) - \varphi^k(y^{k+1})] \to 0 \) and \( \gamma_k \to 0 \) yield \( Q(x^k, y^{k+1}) \to 0 \). From \((AD2)\), we obtain \( \lim_{k \to +\infty} \|y^{k+1} - x^k\| = 0 \). The proof is complete.

**4.2. Null Step.** In this subsection, we consider (Case2). Let \( \bar{k} \) be the last index of decrease step. Throughout this subsection, the following properties hold.

\[
\begin{align*}
\eta_k &= 0, \\
f(x^\bar{k}) - \bar{f}(y^{\bar{k}+1}) &< m(f(x^\bar{k}) - \varphi^k(y^{\bar{k}+1})), \forall k \geq \bar{k}.
\end{align*}
\]

Note that \((4.7)\) stems from STEP4 of DIIPBM.
To prove the convergence, we define \( l^k(x) \), \( \bar{l}^k(x) \) and \( e_k \) as follows:

\[ l^k(x) := \varphi^{k-1}(y^k) + (s^k, x - y^k), \]

\[ \bar{l}^k(x) := l^k(x) + \lambda_{k-1}^{-1} P(x, x^k). \]

\[ e_k := f(y^k) - \varphi^{k-1}(y^k). \]

In ordinary bundle method, \( \varphi^k \geq l^k \) hold. However, in our algorithm, some technical modifications are necessary.

**Lemma 4.6.** Let \( \{y^k\} \) and \( \{\gamma_k\} \) be sequences generated in Algorithm DIIPBM. Suppose infinitely many null steps occur, the following hold.

(i) For all \( x \in \mathbf{X} \), \( \varphi^{k+1}(x) \geq l^{k+1}(x) - \gamma_k \).

(ii) \( y^k \) is a minimizer of the function \( \bar{l}^k(y^k) \).

(iii) \( l^k(y^k) \) converges to some \( l^* \in \mathbb{R} \).

(iv) \( Q(y^{k+1}, y^k) \to 0 \).

(v) The sequence \( \{y^k\} \) is bounded.

**Proof.** From (2.2) and (2.6), for every \( i \in \hat{I}_k \), we have

\[ \varphi^{k+1}(x) \geq \bar{f}(y^i) + \langle g^i, x - y^i \rangle \\
= \langle g^i, x - y^{k+1} \rangle + \bar{f}(y^i) + \langle g^i, y^{k+1} - y^i \rangle \\
= \langle g^i, x - y^{k+1} \rangle + (\varphi^k(y^{k+1}) - \alpha_i(y^{k+1})). \]

Therefore, we have

\[ \theta_k^i \varphi^{k+1}(x) = \theta_k^i [\langle g^i, x - y^{k+1} \rangle + (\varphi^k(y^{k+1}) - \alpha_i(y^{k+1}))]. \]

Summing these inequalities for \( i \in \hat{I}_k \), (2.5), (2.8) and (4.9) implies,

\[ \varphi^{k+1}(x) = \varphi^k(y^{k+1}) + \sum_{i \in \hat{I}_k} \theta_k^i g^i, x - y^{k+1} \sum_{i \in \hat{I}_k} \theta_k^i \alpha_i \\
\geq \varphi^k(y^{k+1}) + (s^{k+1}, x - y^{k+1}) - \gamma_k \\
\geq l^{k+1}(x) - \gamma_k. \]

We’ve proved (i). Note that in this proof \( \varphi^{k+1} \) is evaluated by cutting planes which are not derived from \( \hat{I}_{k+1} \) but \( \hat{I}_k \). From (4.10) and (P8), we have

\[ \bar{l}^k(x) - \bar{l}^k(y^k) = \langle s^k, x - y^k \rangle + \lambda_{k-1}^{-1} [P(x, x^k) - P(y^k, x^k)] \geq \lambda_{k-1}^{-1} Q(x, y^k) \geq 0. \]

(4.12) Therefore, (ii) is proved.

From (4.10), (4.9), (i) and (ii), we have

\[ \bar{l}^{k+1}(y^{k+1}) = l^{k+1}(y^{k+1}) + \lambda_{k-1}^{-1} P(y^{k+1}, x^k) \\
= \varphi^k(y^{k+1}) + \lambda_{k-1}^{-1} P(y^{k+1}, x^k) \\
\geq l^k(y^{k+1}) + \lambda_{k-1}^{-1} P(y^{k+1}, x^k) - \gamma_{k-1} \\
= \bar{l}^k(y^{k+1}) - \gamma_{k-1} \\
\geq \bar{l}^k(y^k) - \gamma_{k-1}. \]
Let $L^k(y^k) := \hat{l}^k(y^k) + \sum_{i=k}^{k-2} \gamma_i$, it follows that the sequence $\{L^k(y^k)\}$ is an non-decreasing sequence.

$$L^{k+1}(y^{k+1}) - L^k(y^k) = \hat{l}^{k+1}(y^{k+1}) - \hat{l}^k(y^k) + \gamma_k \geq 0.$$ 

Let $\gamma_0$ is a sufficiently large positive value which greater than $\gamma_k$ for all $k \geq \bar{k}$, i.e. Therefore, from (i), (ii) and (2.2), for all $k \geq \bar{k}$, we have

$$\hat{l}^k(y^k) \leq \hat{l}^k(x^k) \leq \varphi^k(x^k) + \gamma_{k-1} \leq \varphi^k(x^k) + \gamma \leq f(x^k) + \gamma.$$ 

This inequality and (A2) imply that the sequence $\{L^k(y^k)\}$ is bounded. Therefore, from (i), (ii) and (2.2), we also have

$$f(x^k) + \gamma \geq \varphi^k(x^k) + \gamma \geq \varphi^k(x^k) + \gamma_k \geq \hat{l}^k(x^k).$$ 

Therefore, we have

$$\lambda_{U}^{-1} Q(x^k, y^k) \leq f(x^k) + \gamma - t^* + \epsilon.$$ 

Since $Q(x^k, \cdot)$ is level bounded, the sequence $\{y^k\}$ is bounded. Now we established (v).

**Proposition 4.7.** Let $\{y^k\}$ and $\{\gamma_k\}$ be sequences generated in Algorithm DI-IPBM. Suppose infinitely many null steps occur, the following hold.

(i) $0 \in \partial u + \gamma_{k-1} f(y^k) + \lambda_k^{-1} d(y^k, x^k)$.

(ii) $y^k$ converges to $y^*$ which minimizes $f(x) + \lambda_k^{-1} d(x, x^k)$.

**Proof.** From (2.2) $c_k \geq 0$ is obvious. We’d like to show $c_k \downarrow 0$. First, from the mean value theorem, we have

$$\exists y^k \in [y^k, y^{k+1}], \exists c_k \in \partial f(y^k), f(y^{k+1}) - f(y^k) = c_k \| y^{k+1} - y^k \|.$$ 

In view of Lemma 4.6(iv) and (v), $\{y^k\}$ is a bounded and $Q(y^{k+1}, y^k) \to 0$. Hence, (AD2) implies $\|y^{k+1} - y^k\| \to 0$. Since we assume of $\partial f$ is bounded in $[y^k, y^{k+1}]$, $c_k$ is bounded. Therefore, we have

$$f(y^{k+1}) - f(y^k) \to 0.$$
On the other hand, we have
\[
f(y^{k+1}) - f(y^k) \geq \varphi^k(y^{k+1}) - f(y^k)
\]
\[
\geq (g^k, y^{k+1} - y^k) - \eta_k
\]
\[
\geq - \|g^k\| |y^{k+1} - y^k| - \eta_k.
\]
Since \( f(y^{k+1}) - f(y^k) \to 0 \), \( |y^{k+1} - y^k| \to 0 \), \( \eta_k \downarrow 0 \) and \( \|g^k\| \) is bounded, we also have
\[
\varphi^k(y^{k+1}) - f(y^k) \to 0.
\]
Then, we have
\[
(4.13) \quad \epsilon_k = f(y^k) - \varphi^{-1}(y^k) = [f(y^k) - f(y^{k+1})] + [f(y^{k+1}) - \varphi^{-1}(y^k)] \downarrow 0.
\]
At the same time, from (2.2) and (2.4), we have
\[
f(x) \geq \varphi^{-1}(x)
\]
\[
\geq \varphi^{-1}(y^k) + (s^k, x - y^k) - \gamma_{k-1}
\]
\[
= f(y^k) + (s^k, x - y^k) - \epsilon_k - \gamma_{k-1}.
\]
We proved (i).
Since \( d(\cdot, x^k) \) is level bounded and \( f(\cdot) \) is lower bounded, there exist some \( y^* \) which minimizes \( f(x) + \lambda_{k}^{-1}d(x, x^k) \). Hence, we have
\[
f(x) \geq f(y^*) - \lambda_{k}^{-1}(\nabla d(y^*, x^k), x - y^*).
\]
Substituting \( x = y^k \), in view of (P8), we have
\[
f(y^k) - f(y^*) \geq -\lambda_{k}^{-1}(\nabla d(y^*, x^k), y^k - y^*)
\]
\[
\geq -\lambda_{k}^{-1} [P(y^k, x^k) - P(y^*, x^k) - Q(y^k, y^*)].
\]
Meanwhile, from the result of part(ii), we have
\[
f(x) \geq f(y^k) + (s^k, x - y^k) - \epsilon_k - \gamma_{k-1}.
\]
In contrast, substituting \( x = y^* \), we also have
\[
f(y^*) - f(y^k) + \epsilon_k + \gamma_{k-1} \geq -\lambda_{k}^{-1} [P(y^*, x^k) - P(y^k, x^k) - Q(y^*, y^k)].
\]
Adding these two inequalities, we have
\[
Q(y^k, y^*) + Q(y^*, y^k) \leq \lambda_{k}(\epsilon_k + \gamma_{k-1}).
\]
Since \( \epsilon_k + \gamma_{k-1} \to 0 \), we have \( Q(y^k, y^*) \to 0 \), In view of (P7), it follows that \( y^k \) converges to \( y^* \). (ii) is proved.

**Theorem 4.8.** Suppose infinitely many null steps occur in Algorithm DIIPBM, the sequence \( \{y^k\} \) converges to an optimal solution \( y^* \).* Furthermore, \( y^* \) is equal to \( x^k \).
Proof. First, we prove that \( x^k \) is an optimal solution of the original problem. In view of (4.8), we have
\[
f(x^k) - \tilde{f}(y^{k+1}) < m[f(x^k) - \varphi^k(y^{k+1})], \forall k \geq \bar{k}.
\]
Since \( \eta_k \downarrow 0 \) and \( e_k \downarrow 0 \), we have
\[
f(x^k) - f(y^*) \leq m[f(y^k) - f(y^*)].
\]
It implies \( f(x^k) \leq f(y^*) \). Therefore, we have
\[
f(x^k) + \lambda^{-1}_k - d(x^k, x^k) \leq f(y^*) + \lambda^{-1}_k d(y^*, x^k).
\]
It follows that \( x^k \) is also a minimizer of \( f(x) + \lambda^{-1}_k - d(x, x^k) \). In view of (P4), we have \( \nabla_1d(x^k, x^k) = 0 \). This yield \( 0 \in \partial f(x^k) \). Then, \( x^k \) is a minimizer of \( f(x) \). On the other hand, since \( y^* \) is a minimizer of \( f(x) + \lambda^{-1}_k - d(x, x^k) \), we have
\[
f(x^k) + \lambda^{-1}_k d(x^k, x^k) = f(y^*) + \lambda^{-1}_k d(y^*, x^k).
\]
Therefore, we have \( f(y^*) = f(x^k) \) and \( d(y^*, x^k) = 0 \). This implies \( f(y^*) \) is also an optimal solution of the original problem. From (P1), \( d(\cdot, x^k) \) is strictly convex, and it is also non-negative. This implies \( y^* = x^k \). We complete the proof.

It should be mentioned that Theorem 4.8 does not hold without (4.7), i.e. STEP4. Without STEP4, \( f(x^k) \) is replaced by \( f(x^k) \).

Proposition 4.9. Suppose infinitely many null steps occur in Algorithm DI-IPBM, the stopping criteria at STEP5 hold asymptotically. More specifically, the following hold.

\[
\lim_{k \to +\infty} (y^{k+1} - x^k) = 0, \lim_{k \to +\infty} \delta_k = 0, \lim_{k \to +\infty} \gamma_k = 0, \lim_{k \to +\infty} \eta_k = 0.
\]

Proof. \( y^{k+1} - x^k \to 0 \) is obvious from Theorem 4.8. \( \gamma_k \to 0, \eta_k \to 0 \) is immediate from (A1)(A2). In terms of \( \delta_k \), from (4.11), we have
\[
\delta_k = f(x^k) - \varphi^k(y^{k+1}) = [f(x^k) - f(y^{k+1})] + e_{k+1}.
\]
In view of (P2), \( y^{k+1} \in X \) for all \( k \geq \bar{k} \). Then, lower semicontinuity of \( f \) implies
\[
\lim_{k \to +\infty} f(y^{k+1}) = f(\lim_{k \to +\infty} y^{k+1}) = f(y^*) = f(x^k).
\]
Therefore, from (4.14) and (4.13), we have \( \delta_k \to 0 \). We complete the proof.

The argument described in this subsection is a generalization of [2, Theorem 3.1]. The difference is that we have to consider inexactness which come from \( \eta_k \) and \( \gamma_k \). Furthermore, our proof covers more generalized stabilizing term.

4.3. Finite Step. In this subsection, we consider (Case3). When only finite steps occur, it is fairly easy to show the convergence property.

Proposition 4.10. In algorithm DI-IPBM, suppose STEP5 is satisfied at some finite \( k \), i.e. \( y^{k+1} - x^k = 0, \delta_k = 0, \gamma_k = 0, \eta_k = 0 \) hold, then \( x^k \) is a optimal solution.
Proof. From \( y^{k+1} = x^k \), \( \gamma_k = 0 \) and (2.4), \( y^{k+1} \) is a minimizer of \( \varphi^k \). Therefore, from (2.2), \( \delta_k = 0 \) and \( \eta_k = 0 \), we have
\[
f(x) \geq \varphi^k(x) \geq \varphi^k(y^{k+1}) = \hat{f}(x^k) = f(x^k).
\]
This implies that \( x^k \) is an optimal solution.

5. Application to Lagrange Relaxation. In this section, we consider the sufficiently smooth convex, potentially large, programming problem
\[
\begin{align*}
\text{maximize} & \quad c(z), \quad z \in Z, \\
\text{subject to} & \quad u_I(z) \leq 0, \\
& \quad u_E(z) = 0.
\end{align*}
\]
where \( c : \mathbb{R}^p \to \mathbb{R} \) is concave, \((u_I)_j : \mathbb{R}^p \to \mathbb{R} \) is convex for all \( j \in \{1, \cdots , n_I\} \), 
\((u_E)_j : \mathbb{R}^p \to \mathbb{R} \) is affine for all \( j \in \{n_I+1, \cdots , n_I+n_E\} \), \( Z \) is a closed convex subset of \( \mathbb{R}^p \), and \( n = n_I + n_E \). Let \( x = (x_I, x_E) \in \mathbb{R}^{n_I} \times \mathbb{R}^{n_E} \) as the Lagrange multiplier of the problem and \( u(z) = (u_I(z), u_E(z)) \) for simplicity, we can define Lagrange function \( L(z, x) \) as follows:
\[
L(z, x) := c(z) - \langle x, u(z) \rangle.
\]
Therefore, minimizing the dual function \( \max_{z \in Z} L(z, x) \) is a typical example of our original problem. In this section, we denote
\[
f(x) := \max_{z \in Z} L(z, x).
\]
In minimizing \( f(x) \), we use DHPBM explained in section 2. Let \( \hat{z}^k \) be an \( \eta_k \)-optimal \( f(x^k) \), i.e.
\[
\hat{f}(x^k) = L(\hat{z}^k, x^k).
\]
We assume that \( \hat{z}^k \) is obtained by approximately solving the Lagrange relaxation problem, and exact value of \( \eta_k \geq 0 \) is unknown. Nevertheless, we also assume that we can set \( \eta_k = 0 \) if necessary (STEP 4).

Firstly, we show an easy lemma that an approximate solution of the Lagrangian relaxation problem not only gives \( \eta_k \)-optimal \( f(x^k) \) but also gives \( \eta_k \)-subgradient.

**Lemma 5.1.** Suppose \( \hat{z}^k \) gives an \( \eta_k \)-optimal \( f(x^k) \), the following hold.
\[
-u(\hat{z}^k) = g^k \in \partial_{\eta_k} f(x^k).
\]
Proof. From (5.1) and (5.2), for all \( x \in \mathbb{R}^{n_I} \times \mathbb{R}^{n_E} \), we have
\[
\begin{align*}
f(x) \geq & \quad L(\hat{z}^k, x) \\
= & \quad L(\hat{z}^k, x^k) + \langle -u(\hat{z}^k), x - x^k \rangle \\
= & \quad f(x^k) + \langle -u(\hat{z}^k), x - x^k \rangle - \eta_k.
\end{align*}
\]

To apply DHPBM for this problem, we also need a stabilizing term which satisfies \((d, H, P, Q) \in A(\mathbb{R}^{n_I} \times \mathbb{R}^{n_E})\). From Lemma 3.3, \((d_I, H_I, P_I, Q_I) \in A(\mathbb{R}^{n_I})\) and \((d_E, H_E, P_E, Q_E) \in A(\mathbb{R}^{n_E})\) produce \((d, H, P, Q) = (d_I + d_E, H_I + H_E, P_I + P_E, Q_I + Q_E) \in A(\mathbb{R}^{n_I} \times \mathbb{R}^{n_E})\).

Unfortunately, these conditions are not sufficient for Lagrangian relaxation. From Theorem 4.4, Theorem 4.8 and Proposition 4.10, we’ve already known that the sequence \( \{x^k\} \) converges to some \( x^* \). For this convergent sequence \( \{x^k\} \), following assumptions are required.
If \((x^*_I)_j = 0\), \(\limsup_{k \to +\infty} \nabla_1 d_I(x^k_I, x^{k-1}_I)_j \leq 0\).
If \((x^*_I)_j > 0\), \(\lim_{k \to +\infty} \nabla_1 d_I(x^k_I, x^{k-1}_I)_j = 0\).

(AL2) \(\lim_{k \to +\infty} \nabla_1 d_E(x^k_E, x^{k-1}_E) = 0\).

5.1. Primal Recovery. In attacking Lagrange Relaxation problems, our aim is to obtain an optimal solution of the primal problem, and obtaining optimal dual solution is not enough. Just solving a Lagrange relaxation problem for optimal dual \(x^*\) does not always give primal optimal solution. For more detailed discussion, we recommend [22].

In this subsection, we show that primal feasibility and primal optimality are asymptotically satisfied with the help (AL1)(AL2).

**Theorem 5.2.** Let approximate feasible solution \(z^k := \sum_{i \in I_k} \theta^k_i \tilde{z}_i\), applying DI-IPBM to problem \((P)\), primal feasibility and primal optimality are satisfied asymptotically. More specifically, the following hold.

(i) \(\limsup_{k \to \infty} u_I(z^k) \leq 0\), \(\lim_{k \to \infty} u_E(z^k) = 0\);
(ii) \(\lim_{k \to \infty} c(z^k) \geq \sup_{u(z) \leq 0, z \in Z} c(z)\).

**Proof.** Let \(g^k = (g^k_I, g^k_E)\) and \(s^k = (s^k_I, s^k_E)\) for the sake of simplicity, from Lemma 5.1, (2.8) and convexity of \(u_I\), we have

\[
 u_I(z^k) = u_I\left(\sum_{i \in I_k} \theta^k_i \tilde{z}_i\right) \leq \sum_{i \in I_k} \theta^k_i u_I(\tilde{z}_i) = -\sum_{i \in I_k} \theta^k_i g^i_I = -s^{k+1}_I.
\]

From (AL1) and \(0 < \lambda_L \leq \lambda_k \leq \lambda_U\), we have

\[
 (5.4) \quad u_I(z^k) \leq \limsup_{k \to +\infty} \lambda_k^{-1} \nabla_1 d(x^{k+1}, x^k) \leq 0.
\]

Meanwhile, by a similar argument, we have

\[
 u_E(z^k) = u_E\left(\sum_{i \in I_k} \theta^k_i \tilde{z}_i\right) = \sum_{i \in I_k} \theta^k_i u_E(\tilde{z}_i) = -\sum_{i \in I_k} \theta^k_i g^i_E = -s^{k+1}_E.
\]

From (AL2) and \(0 < \lambda_L \leq \lambda_k \leq \lambda_U\), we have

\[
 \lim_{k \to +\infty} s^{k+1}_E = 0.
\]

The proof of (i) has finished.
From (5.3), (2.6), (2.8) and (2.7), we have
\[
c(z^k) = c\left(\sum_{i \in I_k} \theta_i^k \hat{z}_i\right)
\geq \sum_{i \in I_k} \theta_i^k c(\hat{z}_i)
= \sum_{i \in I_k} \theta_i^k [\hat{f}(x^i) + \langle g^i, x^i \rangle]
= \sum_{i \in I_k} \theta_i^k [\hat{f}(x^i) - \langle g^i, x^k - x^i \rangle] + \sum_{i \in I_k} \theta_i^k \langle g^i, x^k \rangle
\geq \varphi_k(x^{k+1}) - \alpha_k(x^{k+1}) - \langle s^{k+1}, x^k \rangle
\geq f(x^k) - \delta_k - \eta_k - \gamma_k - \langle s^{k+1}, x^k \rangle
= \sup_{u(z) \leq 0, z \in Z} c(z) - \delta_k - \eta_k - \gamma_k - \langle s^{k+1}, x^k \rangle - \langle s_E^{k+1}, x_E^k \rangle.
\]
Final inequality follows from weak duality theorem. From Proposition 4.10, Proposition 4.5 and Proposition 4.9, we have \(\delta_k \to 0, \gamma_k \to 0, \eta_k \to 0\). If \((x_j^k)_j = 0\), then (AL1) and (5.4) implies \((s_j^{k+1})_j(x_j^k)_j \to 0\). If \((x_j^k)_j > 0\), then we have \((s_j^{k+1})_j(x_j^k)_j \to 0\) since \(x^*\) is obviously bounded. Hence, we have \((s_j^{k+1}, x_j^k) \to 0\). \(\langle s_E^{k+1}, x_E^k \rangle \to 0\) follows from (AL2). (ii) is now proved.

5.2. Interior proximal distance examples. In section 3, we gave some interior proximal distance examples which satisfies assumptions (AD1) to (AD3). However, if you would like to apply DIPBM to the Lagrangian Relaxation problems, additional assumptions (AL1) and (AL2) are required. We consider whether the examples referred at section 3 are suitable for \((d_I, d_E)\).

For quadratic functions, let \(d_E(x_E, y_E) = \frac{1}{2} \|x - y\|^2\), (AL2) clearly hold.

For Bregman functions, let \(d_I(x_I, y_I) = D_h(x_I, y_I)\) and \(h\) be a Bregman function with zone \(\mathbb{R}^n_+\). Suppose \(\text{dom} \nabla h = \mathbb{R}^n_+\), by similar argument from [19, Theorem 8.4], \(\nabla_1 d(x_j^k, x_j^{k-1}) = \nabla D_h(x_j^k, x_j^{k-1}) = \nabla h(x^k) - \nabla h(x^{k-1}) \to 0\). Therefore (AL1) hold. Unfortunately, this condition is too strong to hold for the two examples \(h_1(x), h_2(x)\) introduced at Section 3. Nevertheless, we can confirm that (AL1) hold for \(h_1(x), h_2(x)\). For \(h_1(x) = x \log x\), we have
\[
\nabla D_{h_1}(x_j^k, x_j^{k-1}) = \nabla h_1(x_j^k) - \nabla h_1(x_j^{k-1}) \leq \frac{x_j^k}{x_j^{k-1}} - 1.
\]
If \(x_j^+ > 0\), since \(x_j^k \in \text{dom} \nabla h_1 = \mathbb{R}^n_+\), \(\nabla h_1(x_j^k) - \nabla h_1(x_j^{k-1}) \to 0\). If \(x_j^* = 0\), we have \(\limsup_{k \to \infty} x_j^k / x_j^{k-1} \leq 1\). Suppose it doesn’t hold, we have \(\limsup_{k \to \infty} x_j^k / x_j^{k-1} > 1\), therefore we can choose a strictly increasing subsequence. It clearly contradict to \(x_j^* = 0\). Hence \(\limsup_{k \to \infty} \nabla_1 d_I(x_j^k, x_j^{k-1}) \leq 0\). For \(h_2(x) = (px - x^p)/(1 - p)\) with \(p \in (0, 1)\), we have
\[
\nabla D_{h_2}(x_j^k, x_j^{k-1}) = \nabla h_2(x_j^k) - \nabla h_2(x_j^{k-1}) = \frac{p}{1 - p} \left[ -1 + \left( \frac{x_j^k}{x_j^{k-1}} \right)^{1-p} \right] (x_j^k)^{p-1}.
\]
If \( x_j^* > 0 \), similarly \( \nabla h(x_j^k) - \nabla h(x_j^{k-1}) \to 0 \) hold. If \( x_j^* = 0 \), we have \( \limsup_{k \to \infty} (x_j^k/x_j^{k-1})^{1-p} \leq 1 \) from the similar argument of \( h_1(x) \). Therefore, we have \( \limsup_{k \to \infty} \nabla d_I(x_j^k, x_j^{k-1}) \leq 0 \).

For \( \varphi \)-divergence functions with regularization, we can prove that (AL1) hold for both classes \( \Phi_r \) with \( r = 1, 2 \). For \( \varphi \in \Phi_1 \), let \( d_I(x, y) = d_\varphi(x, y) + \frac{1}{2} \|x - y\|^2 \), we have

\[
(\nabla d_I(x_j^k, x_j^{k-1}))_j = \varphi'(\frac{x_j^k}{x_j^{k-1}}) + (x_j^k - x_j^{k-1}) \\
\leq \varphi''(1)(\log x_j^k - \log x_j^{k-1}) + (x_j^k - x_j^{k-1}) \\
\leq \varphi''(1)\left(\frac{x_j^k}{x_j^{k-1}} - 1\right) + (x_j^k - x_j^{k-1}).
\]

If \( x_j^* > 0, x_j^k/x_j^{k-1} \to 1 \) hold, and \( (\nabla d_I(x_j^k, x_j^{k-1}))_j \to 0 \) follows. If \( x_j^* = 0 \), from the similar argument of \( h_1(x) \) at Bregman functions, we have \( \limsup_{k \to \infty} \nabla d_I(x_j^k, x_j^{k-1}) \leq 0 \). For \( \varphi \in \Phi_2 \), let \( d_I(x, y) = d_\varphi(x, y) + \frac{\sigma}{2} \|x - y\|^2 \), we have

\[
(\nabla d_I(x_j^k, x_j^{k-1}))_j = x_j^{k-1}\varphi'(\frac{x_j^k}{x_j^{k-1}}) + (x_j^k - x_j^{k-1}) \\
\leq (\varphi''(1) + 1)(x_j^k - x_j^{k-1}).
\]

If \( x_j^* > 0, (\nabla d_I(x_j^k, x_j^{k-1}))_j \to 0 \) hold since \( x_j^k/x_j^{k-1} \to 1 \). If \( x_j^* = 0, x_j^k - x_j^{k-1} \to 0 \) implies \( \limsup_{k \to \infty} \nabla d_I(x_j^k, x_j^{k-1}) \leq 0 \).

Let us summarize the result. Quadratic distance meets \( d_E \). For \( d_I \), both Bregman proximal distance and \( \varphi \)-divergence with regularization are applicable.

As we have already seen, the difficulty occurs at \( (x_j^*)_j = 0 \). From a practical point of view, however, this is not a too serious problem, because it means that the \( j \)th index is not active at an optimal solution. Therefore, if the nonactive inequalities cause the problem, we can just omit them and this doesn’t change the problem. Any constraint which is expected to be inactive should not be relaxed in Lagrangian relaxation.

5.3. Decomposition. An advantage of the Lagrangian relaxation method is it can decompose the problem [5, Chapter 5]. Suppose \( z = (z_1, \cdots, z_R), Z = Z_1 \times \cdots \times Z_R, z_r \in Z_r : \text{convex} \) for all \( r \in \{1, \cdots, R\} \) and all functions \( c, h_j \) are decomposable with \( z_r \), the primal problem become

\[
\begin{align*}
\text{maximize} & \quad c(z) = \sum_{r=1}^{R} c_r(z_r), \quad z_r \in Z_r, \\
\text{subject to} & \quad u_I(z) = \sum_{r=1}^{R} (u_I)_r(z_r) \leq 0, \\
& \quad u_E(z) = \sum_{r=1}^{R} (u_E)_r(z_r) = 0.
\end{align*}
\]
Hence, we can solve Lagrange relaxation problem separately with respect to $r$. Namely,

$$L(z, x) = c(z) - \langle x, u(z) \rangle = \sum_{r=1}^{R} [c_r(z_r) - \langle x, u_r(z_r) \rangle].$$

In inexact evaluation, of course, the advantage is invariant. To obtain $\tilde{f}(x_k), g_k$,
we need $\tilde{z} = (\tilde{z}_1, \ldots, \tilde{z}_R)$ which can be obtained independently maximizing $c_r(z_r) - \langle x, u_r(z_r) \rangle$ approximately.

6. Approximate stabilized model minimization. In this section, we deal with how to obtain an approximate solution of the SMMP. Unfortunately, not all kinds of approximations are suitable for our algorithm. From [13, Chapter XI, Theorem 3.1.1], we have

$$(6.1) \quad \partial \gamma_k(x^{k+1}) + \lambda_k^{-1} \nabla_1 d(x^{k+1}, x^k) \subset \partial \gamma_k[\varphi^k + \lambda_k^{-1} d(\cdot, x^k)](x^{k+1}).$$

Therefore, obtaining an approximate solution for SMMP which satisfies (2.4) is not an obvious problem.

Furthermore, in Theorem 5.2, we proved that primal recovery is achieved for $z^k = \sum_{i \in I_k} \theta^k_i \tilde{z}_i$. However, it doesn’t mention how to obtain $\theta^k$ which satisfy (2.8).

To consider these problems, we replace SMMP with a slack variable $r \in \mathbb{R}$.

$$\minimize \quad r + \lambda_k^{-1} d(x, x^k),$$

subject to

$$r \geq \tilde{f}(x^i) + \langle g^i, x - x^i \rangle.$$  

(6.2)

Letting $\theta^k$ as a dual value of (6.2), KKT condition of this problem is given by

$$\sum_{i \in I_k} \theta^k_i = 1,$$

$$\theta^k_i \geq 0,$$

$$\sum_{i \in I_k} \theta^k_i g^i + \lambda_k^{-1} \nabla_1 d(x, x^k) = 0,$$

$$\theta^k_i \alpha_i(x) = 0.$$  

(6.3)

If we relax the complementarity condition (6.3) as

$$0 \leq \theta_i \alpha_i(x) \leq \mu_i$$

We can see that the primal and dual solution $(\theta^{k+1}, \theta^k)$ satisfies (2.4), (2.8) with $\gamma_k = \sum_{i \in I_k} \mu_i$. Hence, one way to overcome the problem is to develop an algorithm which solves SMMP approximately in terms of complementarity condition. An attractive candidate is to solve an unconstrained optimization problem with log-barrier function which is defined as follows:

$$\minimize \quad r + \lambda_k^{-1} d(x, x^k) - \mu \sum_{i \in I_k} \log \left( r - [\tilde{f}(x^i) + \langle g^i, x - x^i \rangle] \right).$$  

(6.4)

We denote this unconstrained problem as Modified SMMP. The validity of Modified SMMP is confirmed as follows.
Proposition 6.1. If \((r^k, y^k)\) is an optimal solution of Modified SMMP, (2.8) hold with

\[
\gamma_k = \mu |I_k|, \quad \theta_i = \frac{\mu}{r^k - [\bar{f}(x^i) + \langle g^i, x - x^i \rangle].}
\]

Proof. Let the objective function as \(F(r, x)\). We have

\[
\nabla_r F(r, x) = 1 - \sum_{i \in I_k} \frac{1}{r - [f(x^i) + \langle g^i, x - x^i \rangle]},
\]

\[
\nabla_x F(r, x) = \lambda_k^{-1} \nabla_1 d(x, x^k) - \mu \sum_{i \in I_k} \frac{-g^i}{r - [f(x^i) + \langle g^i, x - x^i \rangle]}.\]

From the optimality condition, we have

\[
\sum_{i \in I_k} \theta_i = 1,
\]

\[
\lambda_k^{-1} \nabla_1 d(y^k, x^k) + \sum_{i \in I_k} \theta_i g^i = 0.\]

\(\theta_i \geq 0\) is obvious from the definition. Since \(r^k \geq \varphi_k(x^i)\) for all \(i \in I_k\), we have

\[
\sum_{i \in I_k} \theta_i \alpha(y^k) \leq \sum_{i \in I_k} \theta_i [r - [\bar{f}(x^i) + \langle g^i, x - x^i \rangle]] = \mu |I_k| = \gamma_k.
\]

We complete the proof.

A great advantage of Modified SMMP is that we can control \(\gamma_k\) before solving the problem. On the other hand, a drawback is that all \(i \in I_k\) become active. To overcome the problem, an ad-hoc approach is to remove the indices on which \(\alpha_i(y^{k+1})\) is large enough, and solve again.

Let \(\mu = 1\) and replace \(r\) as \(\log(\bar{f} - r)\) in the objective such that \(\bar{f}\) is a constant which is larger than \(\varphi_k(x^k)\), we incidentally recover proximal analytic center cutting plane method (Proximal ACCPM) [24]. Unlike Modified SMMP, a drawback of Proximal ACCPM is that we cannot control \(\gamma_k\) before solving the problem.

7. Conclusions and further research. We developed a new version of bundle method which allows both inexact model minimization and inexact oracle. In this method, generalized stabilizing term is also considered. Its global convergence has proved. Furthermore, it is applicable to Lagrangian relaxation method, since primal recovery has been confirmed. Roughly speaking, our paper generalizes [20] and [2, Section 3] in the spirit of [18].

For further research, we hope that more generalized stabilizing term will be acceptable to this framework. In particular, extended quadratic function \(d(x, y) = \frac{1}{2} \|x - y\|^2 + \delta_{\mathbb{R}_+^n}(x)\) is desired to be included.

Another natural subject is the treatment of inexact model minimization. To allow inexact model minimization, we assume (2.4). However, as mentioned at (6.1), it doesn’t allow all kind of inexactness. We hope more generalized inexactness will be included in this framework.

REFERENCES