NPSOR-92-011

NAVAL POSTGRADUATE SCHOOL

Monterey, California

AD-A246 436

DTIC
ELECTE
FEB 26 1992
S B D

PROXIMAL MINIMIZATION ALGORITHMS WITH CUTTING PLANES

Siriphong Lawphongpanich

December 1991

Approved for public release; distribution is unlimited.

Prepared for:
National Science Foundation
Washington, DC

92-04571

92 2 21 023

NAVAL POSTGRADUATE SCHOOL,
MONTEREY, CALIFORNIA

Rear Admiral R. W. West, Jr.
Superintendent

Harrison Shull
Provost

This report was funded by the National Science Foundation and OM&N Direct Funding at the Naval Postgraduate School.

This report was prepared by:

SIRIPHONG LAWPHONGPANICH
Assistant Professor,
Department of Operations Research

Reviewed by:

PETER PURDUE
Professor and Chairman
Department of Operations Research

Released by:

PAUL J. MARTO
Dean of Research

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	3. DISTRIBUTION /AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NPSOR-92-011	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b. OFFICE SYMBOL (*If applicable*) OR	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c. ADDRESS (*City, State, and ZIP Code*) Monterey, CA 93943	7b. ADDRESS (*City, State, and ZIP Code*) Monterey, CA 93943
8a. NAME OF FUNDING/SPONSORING ORGANIZATION National Science Foundation 8b. OFFICE SYMBOL (*If applicable*)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER O&MN Direct Funding
8c. ADDRESS (*City, State, and ZIP Code*) 1800 Gth Street, NW, Washington, DC 20550	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (*Include Security Classification*)
Proximal Minimization Algorithms with Cutting Planes

12. PERSONAL AUTHOR(S)
Siriphong Lawphongpanich

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (*Year, month day*)	15. PAGE COUNT
Technical	FROM_____TO_____	December, 1991	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)
FIELD	GROUP	SUB-GROUP	Proximal minimization; nondifferentiable optimization; cutting planes; Lagrangian dual; convex programming

19. ABSTRACT (*Continue on reverse if necessary and identify by block number*)

This paper examines a class of proximal minimization algorithms in which the objective function of the underlying convex program is approximated by cutting planes. This class includes algorithms such as cutting plane, cutting plane with line search and bundle methods. Among these algorithms, the bundle methods can be viewed as a quadratic counterpart of the cutting plane algorithm with line search, for they both attempt to decrease the true objective function at every iteration. On the other hand, the cutting plane algorithm does not explicitly and/or directly attempt to decrease the true objective function. However, it relies on the monotonicity of the approximating function to guarantee convergence to an optimal solution. This prompts the question of whether there exists a quadratic counterpart for the cutting plane algorithm. To provide an affirmative answer, this paper constructs a new convergent algorithm which resembles, but is different from, the bundle methods. Also, to make the relationship between bundle methods and proximal minimization more concrete, this paper also supplies a convergence proof for a variant of the bundle methods which utilizes analysis common to proximal minimization.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT [x] UNCLASSIFIED/UNLIMITED [ ] SAME AS RPT. [ ] DTIC USERS	21. ABSTRACT SECURITY CLASSICIATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL S. Lawphongpanich	22b. TELEPHONE (*Include Area Code*) (408) 646-2106 2c. OFFICE SYMBOL OR/Lp

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE
S/N 0102-LF-014-6603 UNCLASSIFIED

Proximal Minimization Algorithms with Cutting Planes

Siriphong Lawphongpanich*

Department of Operations Research
Naval Postgraduate School
Monterey, California 93943

December, 1991

Abstract

This paper examines a class of proximal minimization algorithms in which the objective function of the underlying convex program is approximated by cutting planes. This class includes algorithms such as cutting plane, cutting plane with line search and bundle methods. Among these algorithms, the bundle methods can be viewed as a quadratic counterpart of the cutting plane algorithm with line search, for they both attempt to decrease the true objective function at every iteration. On the other hand, the cutting plane algorithm does not explicitly and/or directly attempt to decrease the true objective function. However, it relies on the monotonicity of the approximating function to guarantee convergence to an optimal solution. This prompts the question of whether there exists a quadratic counterpart for the cutting plane algorithm. To provide an affirmative answer, this paper constructs a new convergent algorithm which resembles, but different from, the bundle methods. Also, to make the relationship between bundle methods and proximal minimization more concrete, this paper also supplies a convergence proof for a variant of the bundle methods which utilizes analysis common to proximal minimization.

*This research was supported by grants from National Science Foundation (DDM-8814499) and the direct funding program at the Office of Naval Research and Naval Postgraduate School

KEY WORDS: Proximal Minimization, Nondifferentiable Optimization, Cutting Planes, Lagrangian Dual, Convex Programming

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

DTIC COPY INSPECTED 4

1. Introduction

This paper proposes an application of the proximal minimization algorithm for the following problem.

$$D: \qquad \mathcal{L}(u^*) = \min_{u \in U} \mathcal{L}(u)$$

where $\mathcal{L}(u)$ is convex and U is a compact subset of R^m. In particular, U is assumed to be a polyhedral of the form $\{u : Au \leq b \text{ and } u \in R^m\}$, where A is a $p \times m$ and b is a vector in R^p. To simplify the presentation and motivate applications to Lagrangian duality and variational inequalities, the objective function $\mathcal{L}(u)$ is also assumed to have the following form:

$$\mathcal{L}(u) = \max_{x \in X} \{f(x) + u \cdot g(x)\} \tag{1}$$

where X is a compact subset of R^n, $f(x)$ is a real-valued function on R^n, and $g(x)$ is a vector-valued function mapping R^n to R^m. The notation $a \cdot b$ denotes the usual dot product between two vectors, a and b.

When U is taken to be the (noncompact) set $\{u : u \geq 0 \text{ and } u \in R^m\}$, D is simply the Lagrangian dual problem of the following nonlinear program:

$$P: \qquad \begin{aligned} f(x^*) = \max_x \ & f(x) \\ \text{s.t. } g(x) &\geq 0 \\ x &\in X. \end{aligned}$$

Under an additional assumption that there exists an $\hat{x}$ such that $g(\hat{x}) > 0$, the solution to P can be obtained by solving D with $U = \{u : 0 \leq u \leq M \text{ and } u \in R^m\}$, where M is sufficiently large.

On the other hand, when $f(x) = F(x) \cdot x$ and $g(x) = F(x)$, where $F(x)$ is a continuous mapping from R^m into itself and satisfies, for some $\alpha > 0$,

$$(F(u) - F(x)) \cdot (u - x) \geq \alpha \|u - x\|^2, \quad \forall u, x \in U,$$

then

$$\mathcal{L}(u) = \max_{x \in U} \{-F(x) \cdot (x - u)\}$$

and D becomes

$$\min_{u\in U}\max_{x\in U}\{-F(x)\cdot(x-u)\} = -\max_{u\in U}\min_{x\in U}\{F(x)\cdot(x-u)\}.$$

Hearn et al. (1984) referred to the problem on the right as the *dual* of the formulation based on the gap function for the following variational inequality:

$$\text{Find } u^* \in U \text{ such that} \quad F(u^*)\cdot(x-u^*) \geq 0, \quad \forall\, x \in U.$$

For the remainder, it is convenient to simply refer to $\mathcal{L}(u)$ as the dual function.

To solve D, the proximal minimization algorithm (see, e.g., Bertsekas and Tsitsiklis, 1990, Martinet, 1970, and Rockafellar, 1976) generates a sequence of points in U by the iteration

$$u^{k+1} = \arg\min_{u\in U}\{\mathcal{L}(u) + \frac{1}{2c_k}\|u-u^k\|^2\} \qquad k = 1,2,\ldots \tag{2}$$

where u^1 is a starting point, $\|\cdot\|$ denotes the Euclidean norm and c_k is a sequence of positive numbers with

$$\liminf_{k\to\infty} c_k > 0.$$

Although the above iterative process converges to an optimal solution of D, there is a concern regarding its practicality. Bertsekas and Tsitsiklis (1990) pointed out in their book that the proximal minimization algorithm requires solutions to a sequence of problems instead of just one problem. When $\mathcal{L}(u)$ is nondifferentiable, this concern is more acute. Adding the 'proximal' term $\frac{1}{2c_k}\|u-u^k\|^2$ only makes the objective function of the problem in (2) strictly convex. So, when $\mathcal{L}(u)$ is nondifferentiable, the objective function in (2) is still nondifferentiable and solving a sequence of nondifferentiable, but strictly convex, does not appear as attractive as solving only one nondifferentiable problem that may not be strictly convex.

To make proximal minimization more amenable to D, this paper approximates $\mathcal{L}(u)$ in (2) by the following function:

$$\mathcal{L}(u) \approx L^k(u) = \max_{i=1,\ldots,k}\{f(x^i) + u\cdot g(x^i)\}$$

where $x^i \in X$. When x^i is chosen appropriately, $L^k(u)$ is simple a maximum of a finite number of hyperplanes tangential to $\mathcal{L}(u)$. These hyperplanes are generally known as *cuts* or *cutting planes.*

To unify the above scheme with other algorithms that use cutting planes, this paper describes in the next section a generic algorithm which combines cutting planes with proximal minimization. From this generic algorithm, three algorithms from the literature can be derived; they are the cutting plane algorithm, the cutting plane algorithm with line search and the family of bundle methods. Among these algorithms, the bundle methods can be viewed as a quadratic counterpart of the cutting plane algorithm with line search or vice versa, i.e., the latter is a linear counterpart for the former. This prompts the question of whether there exists a quadratic counterpart for the (plain) cutting plane algorithm. The results in this paper provide an affirmative response to the question.

For the remaining, Section 2 formally states the generic algorithm and derives from it the three algorithms in the literature. Also derived is the new algorithm which is a quadratic counterpart of the cutting plane algorithm. Section 3 provides convergent results for the new algorithm. To establish a closer relationship between proximal minimization and bundle methods, Section 4 provides a convergence proof for a simple version of the latter which is different from those in the literature and uses analysis common to proximal minimization. Finally, Section 5 concludes the paper.

2. Classification of Algorithms

To classify and establish relationships among algorithms, we first state a generic algorithm and then show how it can be specialized to the four algorithms. Three of the four exist in the literature and the last is new and shown to be a quadratic counterpart of the cutting plane algorithm.

A GENERIC ALGORITHM

Step 1: Select $u^1 \in U$. Set $k = 1, v^1 = u^1$ and

$$x^1 = \arg\max_{x \in X} \{f(x) + u^1 \cdot g(x)\}.$$

Step 2: Solve the master problem

$$u^{k+1} = \arg\min_{u \in U} \{L^k(u) + \frac{1}{2c_k}\|u - v^k\|^2\}.$$

If v^k also solves the problem, stop and v^k solves D.

Step 3: Solve the subproblem

$$x^{k+1} = \arg\max_{x \in X} \{f(x) + u^{k+1} \cdot g(x)\}$$

Note that $\mathcal{L}(u^{k+1}) = f(x^{k+1}) + u^{k+1} \cdot g(x^{k+1})$.

Step 4: Derive $v^{k+1} \in U$ from u^{k+1} and v^k using some process and/or criteria (see discussion below). Set $k = k + 1$ and return to Step 1.

Note that the (master) problem in Step 2 is slightly different from the one in equation (2) of the previous section. The 'prox-center' in the proximal term is v^k for the master problem and it is u^k for the problem in (2). In addition, the master problem in Step 2 can be stated as

$$\begin{array}{lllll} MP: & \min\ w & + & \frac{1}{2c_k}\|u - v^k\|^2 & \\ & \text{s.t.} & & & \\ & w & \geq & f(x^i) + u \cdot g(x^i) & i = 1, \ldots, k \\ & Au & \leq & b & \end{array}$$

where the first k constraints are generally referred to as *cuts* or *cutting planes.* The dual of MP can be written as

$$
\begin{array}{llrcl}
MD: & \max -\frac{c_k}{2}\|G\pi + A^t\lambda\|^2 & + & (G^t v^k + \hat{f}) \cdot \pi + (Av^k - b) \cdot \lambda & \\
& \text{s.t.} & \sum_{i=1}^{k} \pi_i & = & 1 \\
& & \pi_i & \geq & 0, \quad i = 1, \ldots, k \\
& & \lambda_j & \geq & 0 \quad j = 1, \ldots, p
\end{array}
$$

where $\hat{f}$ denotes a vector in R^k with $f(x^i)$ as its components, G denotes a $m \times k$ matrix with $g(x^i)$ as its columns, π_i are the dual variables corresponding to the cutting plane constraints and λ_j are dual variables corresponding to the constraints defined by the matrix A. In any case, both the master problem and its dual can be solved in a finite number of iterations. Pang (1983) and Lin and Pang (1987) reviewed a large number of algorithms applicable to both MP and MD. More specifically, Kiwiel (1991) designed a dual algorithm to solve MP and Bertsekas (1982) proposed an efficient algorithm designed especially for convex programming problems with simple constraints such as those in MD.

Below, we describe four specializations of the generic algorithm. They are the cutting plane algorithm, the cutting plane algorithm with line searches, the bundle methods and a new algorithm called the proximal minimization algorithm with cutting planes.

The cutting plane (CP) algorithm: The generic algorithm reduces to the CP algorithm when $c_k = \infty$ and $v^{k+1} = u^{k+1}$ $\forall$ k. First, setting $c_k = \infty$ makes the proximal term vanishes from the objective function of the master problem in Step 2, thereby reducing it to the following linear program:

$$
\begin{array}{llrcl}
ML: & \min w & & & \\
& s.t. & & & \\
& & w & \geq & f(x^i) + u \cdot g(x^i) \quad i = 1, \ldots, k \\
& & Au & \leq & b
\end{array}
$$

Without the proximal term and always setting $v^{k+1} = u^{k+1}$, the variable v becomes superfluous and can be eliminated from the algorithm entirely. This reduces Step 4 to simply increment k by one. It can be shown that the stopping rule in

Step 2 of the generic algorithm is equivalent to the one typical for the CP algorithm which is to stop when $\mathcal{L}(u^{k+1}) = L^k(u^{k+1})$ in Step 3.

The CP algorithm was first introduced by Cheney and Goldstein (1959) and Kelly (1960). Dantzig and Wolfe (1960) developed a related algorithm called the *column generation* technique in the context of decomposing large scale linear programs. Column generation was later generalized to solve Lagrangian dual problems for mathematical programs (see, Dantzig, 1963 and Magnanti et al.,1976) and was given the name *generalized linear programming* technique. Regardless of the terminology, it is well known (see, e.g., Dantzig, 1963, Magnanti et al., 1976 and Zangwill, 1969) that the convergence of the CP algorithm follows from the monotonicity of the sequence $\{w^k\}$ or $\{L^{k-1}(u^k)\}$. However, the corresponding sequence of dual function values, $\{\mathcal{L}(u^k)\}$ is not necessarily monotonic. Therefore, the CP algorithm is a variant of the generic algorithm which has a *linear* master problem and *does not attempt to descend the dual function.*

The cutting plane algorithm with Line Search (CPLS): In an effort to force the CP algorithm to descent the dual function, Hearn and Lawphongpanich (1989b & 1990) added a line search step. CPLS can be obtained from the generic algorithm by setting $c_k = \infty$ for all k and, in Step 3, letting

$$v^{k+1} = \arg \min_{0 \le \lambda \le \lambda_{up}} \{\mathcal{L}(v^k + \lambda(u^{k+1} - v^k))\}$$

where $\lambda_{up} = \max\{\lambda : v^k + \lambda(u^{k+1} - v^k) \in U\}$. Thus, v^{k+1} minimizes $\mathcal{L}(u)$ along the direction $d^k = u^{k+1} - v^k$. Hearn and Lawphongpanich (1989a) showed that, if $\mathcal{L}(u)$ is differentiable at v^k, then d^k is a descent direction and $\mathcal{L}(v^{k+1}) < \mathcal{L}(v^k)$. Therefore, CPLS is a variant of the generic algorithm which has a *linear* master problem and *attempts to descend the dual function*, i.e., a descent is guaranteed whenever the dual function is differentiable at the current iterate, v^k.

The bundle methods: As in CPLS, the main thrust of the bundle methods, first introduced by Lemarechal (1974, 1975) and Wolfe (1975), is to generate a monotonic sequence of dual function values. From the generic algorithm, one can obtain a version of the bundle methods by setting $c_k < \infty$ for all k and, in Step 3, letting

$$v^{k+1} = \begin{cases} u^{k+1} & \text{if } \mathcal{L}(u^{k+1}) + m(L^k(v^k) - L^k(u^{k+1})) \le \mathcal{L}(v^k) \\ v^k & \text{otherwise} \end{cases}$$

where $m \in (0,1)$. Other methods for determining v^{k+1} exist and they can be found in, e.g., Auslender (1987), Fukushima (1984), Gaudioso and Monaco (1982), Kiwiel (1985 & 1989), Lemarechal (1989) and Mifflin (1977). Also, note that updating v^{k+1} is in essence choosing the prox-center for the next iteration.

Several authors (e.g., Fukushima, 1984, Kiwiel, 1989 and Lemarechal, 1991) have observed the similarity between proximal minimization and bundle methods. However, it is interesting that the developments of the two types of algorithms appear different. In an effort to unify the development of bundle methods and proximal minimization, Section 4 provides a convergence proof for a simple method for updating v^{k+1} which is different from. but related to, the one shown above.

When $v^{k+1} = u^{k+1}$, the kth iteration is called a 'serious' step. Otherwise (i.e., $v^{k+1} = v^k$), it is called a 'null' step. So, after every serious step, the dual function decreases and bundle methods change the prox-center. Since $c_k < \infty$, the master problem for the bundle methods is quadratic (see problem MP or MD). Therefore, any bundle method can be considered as a quadratic counterpart of CPLS since it has a *quadratic* master problem and *attempts to descend the dual function*, in that it decreases the dual function at every *serious* step. To emphasize the fact that bundle methods are variants of the generic algorithm, we also refer to them as *proximal minimization algorithms with subgradient bundles* (PMSB).

A Proximal minimization with cutting planes (PMCP): Setting $c_k < \infty$ and *always* letting $v^{k+1} = u^{k+1}$ in Step 3 produces a variant of the generic algorithm which has a *quadratic* master problem and *does not attempt to descend the dual function.* Note the PMCP is similar to the bundle methods because both have a quadratic master problem; however, it is different because it changes the prox-center after *every* iteration instead of after a *serious* iteration. In the framework of the generic algorithm, PMCP is a quadratic counterpart of the cutting plane algorithm, for they both do not attempt to descend the dual function and one has a linear master problem and the other, quadratic.

As mentioned earlier, the convergence of the CP algorithm does not require any monotonicity of the dual function values. On one hand, it is curious that an algorithm can converge without any attempt to decrease the dual function directly. On the other hand, the convergence of the CP algorithm confirms that decreases in the cutting plane approximating function sufficiently insures that the dual function eventually converges (not necessarily in a monotonic manner) to

the optimal value. The convergence proof for PMCP in the next section further corroborates this hypothesis.

Table 1 below summarizes the relationships among the four algorithms which use cutting planes to approximate the objective function. Recall that the phrase '*attempt to descend the dual function*' is to indicate that, although none of the four algorithms guarantees a decrease in the dual function at *every* iteration, some make an attempt to decrease the function in each one. In particular, the bundle methods only yield a decrease at every serious step and CPLS yields one whenever the dual function is differentiable. Nevertheless, all is proven to converge to a solution of D.

	Attempt to Descend the Dual Function	
Master Problem	Yes	No
Linear ($c_k = \infty$)	CPLS	CP
Quadratic ($c_k < \infty$)	Bundle methods or PMSB	PMCP

Table 1: Classes of algorithms which use cutting planes

3. Convergence of PMCP

Below, we restate more concisely the generic algorithm as specialized to the proximal minimization algorithm with cutting planes.

A PROXIMAL MINIMIZATION ALGORITHM
WITH CUTTING PLANES (PMCP)

Step 1: Select $u^1 \in U$. Set $k = 1$ and

$$x^1 = \arg\max_{x \in X} \{f(x) + u^1 \cdot g(x)\}.$$

Step 2: Solve the master problem

$$u^{k+1} = \arg\min_{u \in U} \{L^k(u) + \frac{1}{2c_k}\|u - u^k\|^2\}.$$

If $u^{k+1} = u^k$, stop and u^k is an optimal solution.

Step 3: Solve the subproblem

$$x^{k+1} = \arg\max_{x \in X} \{f(x) + u^{k+1} \cdot g(x)\}$$

Increment k by 1 and go to Step 1.

First, note that since v^{k+1} always equals to u^{k+1} the variable v is not needed and has been eliminated from the above algorithm. Then, recall that in Step 2 $L^k(u)$ is convex and defined previously as

$$L^k(u) = \max_{i=1,\ldots,k} \{f(x^i) + u \cdot g(x^i)\}.$$

In Step 3, x^{k+1} satisfies

$$L^{k+1}(u^{k+1}) = f(x^{k+1}) + u^{k+1} \cdot g(x^{k+1}) = \mathcal{L}(u^{k+1}). \tag{3}$$

The theorem below validates the stopping rule in Step 2.

Theorem 1. *If $u^{k+1} = u^k$, then u^{k+1} is an optimal solution to problem D.*

Proof: Consider the cutting plane representation of the master problem at the kth iteration.

$$\begin{aligned} \min\ w \ &+ \ \frac{1}{2c_k}\|u - u^k\|^2 \\ \text{s.t.} \\ w \ &\geq \ f(x^i) + u \cdot g(x^i), \quad i = 1, \ldots, k \\ Au \ &\leq \ b \end{aligned}$$

Then, (w^{k+1}, u^{k+1}), where $w^{k+1} = L^k(u^{k+1})$, is an optimal solution. Since $u^{k+1} = u^k$, it follows from (3) that

$$w^{k+1} = L^k(u^{k+1}) = L^k(u^k) = \mathcal{L}(u^k).$$

In addition, the KKT conditions are necessary at (w^{k+1}, u^k) and there must exist vector π and λ satisfying the following equations:

$$\begin{aligned} \sum_{i \in I'} g(x^i)\pi_i + \sum_{j \in J'} a^j \lambda_j \ &= \ 0 \\ \sum_{i \in I'} \pi_i \ &= \ 1 \\ \pi_i,\ \lambda_j \ &\geq \ 0 \quad \forall\, i \in I' \text{ and } j \in J' \end{aligned}$$

where

$$\begin{aligned} a^j \ &= \ \text{the } j^{th} \text{ row of matrix A,} \\ I' \ &= \ \{i : w^{k+1} = f(x^i) + u^k \cdot g(x^i) \text{ for } i = 1, \ldots, k\} \text{ and} \\ J' \ &= \ \{j : a^j \cdot u^k = b_j \text{ for } j = 1, \ldots, p\}. \end{aligned}$$

Since $w^{k+1} = \mathcal{L}(u^k)$, $g(x^i)$, $\forall\, i \in I'$, are subgradients of $\mathcal{L}(u^k)$ and

$$H(g(x^i) : i \in I') \subseteq \partial\mathcal{L}(u^k)$$

where $H(\cdot)$ denotes a convex hull. Thus, the KKT conditions can be written more compactly as

$$0 \in \partial L(u^k) + \sum_{j \in J'} a^j \lambda_j.$$

However, this is the KKT condition for problem D. Since $\mathcal{L}(u)$ is convex and U is a polyhedron, the condition is sufficient and the proof is complete. □

By the above theorem, if PMCP stops after a finite number of iterations, it must stop at an optimal solution. When PMCP generates an infinite sequence, it is sufficient to show that PMCP converges to an optimal solution for the case: $c_k = c > 0\ \forall k$. (This is true because of the assumption that $\liminf_{k\to\infty} c_k > 0$.) To do so, define the following:

$X^\infty = \{x^1, x^2, x^3, \ldots\}$, i.e., the set of x^i generated by Steps 1 and 3 of PMCP.

$[X^\infty]$ = the closure of X^∞. Note that $[X^\infty] \subseteq X$.

$L^\infty(u) = \max_{x\in[X^\infty]}\{f(x) + u \cdot g(x)\}$.

From the above description, it is clear that

$$L^k(u) \le L^\infty(u) \le \mathcal{L}(u) \qquad \text{for } k = 1, 2, \ldots$$

where the first inequality follows from the fact that $\{x^i : i = 1, \ldots, k\} \subseteq [X^\infty]$ and the second inequality from the fact that $[X^\infty] \subseteq X$. Observe also that for any k

$$L^{k+j}(u^k) = \mathcal{L}(u^k) = f(x^k) + u^k \cdot g(x^k) \quad \forall\, j = 0, 1, 2, \ldots \tag{4}$$

Similarly, since $x^k \in [X^\infty]$, the following must hold

$$L^\infty(u^k) = \mathcal{L}(u^k) \quad \forall\, k < \infty. \tag{5}$$

Moreover, $\{L^k(u)\}_k$ is a sequence of continuous convex function which converges pointwise to $L^\infty(u)$. However, since $\{L^k(u)\}_k$ is also monotonic, it must also converge uniformly to $L^\infty(u)$ (see, Theorem 7.13 in Rudin, 1976).

To prove convergence and obtain a solution to D, define a sequence $\{z^k\}_k$ as follows: let $z^1 = \mathcal{L}(u^1)$ and for $k = 1, 2, 3, \ldots$ let

$$z^{k+1} = \begin{cases} \mathcal{L}(u^{k+1}) & \text{if } \mathcal{L}(u^{k+1}) + \frac{m}{2c}\|u^{k+1} - u^k\|^2 \le z^k \\ z^k & \text{otherwise} \end{cases}$$

where $m \in (0, 1)$. Also, we have from (4) that $\mathcal{L}(u^{k+1}) = L^{k+1}(u^{k+1})$. So, computing z^k requires no extra effort. Next, construct an index set $\mathcal{K}$ as follows

$$\mathcal{K} = \{k : z^{k+1} = \mathcal{L}(u^{k+1}).\}$$

In words, $\mathcal{K}$ is the index set of iterations in which there is a sufficient decrease in the dual function, i.e., by an amount $\frac{m}{2c}\|u^{k+1}-u^k\|^2$. The next two results address the convergence of PMCP which $\mathcal{K}$ is an infinite set.

Lemma 2. *Let $\mathcal{K}$ be an infinite set. If a subsequence $\{u^k\}_{k\in K}$ converges to u^∞ for some $K \subseteq \mathcal{K}$, then $\{u^{k+1}\}_{k\in K}$ also converges to u^∞.*

Proof: Consider the sequence $\{z^k\}_k$. By definition, it is a nonincreasing sequence which is bounded below by $\mathcal{L}(u^*)$. Thus, $\{z^k\}_k$ must converge. Since $K \subseteq \mathcal{K}$, the following must hold for all $k \in K$

$$\begin{aligned} z^{k+1} + \frac{m}{2c}\|u^{k+1}-u^k\|^2 &\leq z^k \\ \frac{m}{2c}\|u^{k+1}-u^k\|^2 &\leq z^k - z^{k+1} \end{aligned}$$

Taking the limit as $k \to \infty$ and $k \in K$ yields that

$$\lim_{k\in K} \frac{m}{2c}\|u^{k+1}-u^k\|^2 = 0.$$

Since both m and c are positive, $\{u^{k+1}\}_{k\in K}$ and $\{u^k\}_{k\in K}$ must have a common limit point, u^∞. □

Theorem 3. *If $\mathcal{K}$ is an infinite set, then every limit point of the sequence $\{u^k\}_{k\in\mathcal{K}}$ is a solution to D.*

Proof: Let u^* be a solution to D and $\lim_{k\in K} u^k = u^\infty$ for some $K \subseteq \mathcal{K}$. Since u^{k+1} solves the master problem in Step 2, the following must hold

$$L^k(u^{k+1}) + \tfrac{1}{2c}\|u^{k+1}-u^k\|^2 \leq L^k(u) + \tfrac{1}{2c}\|u-u^k\|^2 \;\; \forall\, u \in U \;\&\; k. \tag{6}$$

For any $\alpha \in (0,1)$, setting $u = \alpha u^* + (1-\alpha)u^{k+1}$ in (6) gives

$$\begin{aligned} L^k(u^{k+1}) + \tfrac{1}{2c}\|u^{k+1}-u^k\|^2 &\leq L^k(\alpha u^* + (1-\alpha)u^{k+1}) + \\ &\quad \tfrac{1}{2c}\|\alpha u^* + (1-\alpha)u^{k+1} - u^k\|^2 \\ L^k(u^{k+1}) + \tfrac{1}{2c}\|u^{k+1}-u^k\|^2 &\leq \alpha L^k(u^*) + (1-\alpha)L^k(u^{k+1}) + \\ &\quad \frac{1}{2c}\|\alpha(u^*-u^k) + (1-\alpha)(u^{k+1}-u^k)\|^2 \\ L^k(u^{k+1}) + \tfrac{1}{2c}\|u^{k+1}-u^k\|^2 &\leq \alpha L^k(u^*) + (1-\alpha)L^k(u^{k+1}) + \\ &\quad \frac{1}{2c}(\|\alpha(u^*-u^k)\| + \|(1-\alpha)(u^{k+1}-u^k)\|)^2 \end{aligned}$$

$$
\begin{aligned}
\alpha L^k(u^{k+1}) &\leq \alpha L^k(u^*) - \tfrac{1}{2c}\|u^{k+1} - u^k\|^2 + \\
&\quad \frac{1}{2c}(\|\alpha(u^* - u^k)\| + \|(1-\alpha)(u^{k+1} - u^k)\|)^2 \\
\alpha L^k(u^{k+1}) &\leq \alpha \mathcal{L}(u^*) - \tfrac{1}{2c}\|u^{k+1} - u^k\|^2 + \\
&\quad \frac{1}{2c}(\|\alpha(u^* - u^k)\| + \|(1-\alpha)(u^{k+1} - u^k)\|)^2 \qquad (7)
\end{aligned}
$$

where the second inequality follows from convexity of $L^k(u)$, the third from triangular inequality and the last from the fact that $L^k(u^*) \leq \mathcal{L}(u^*)$. Since $L^j(u)$ is continuous for all $j = 1, 2, \ldots$ and, from Lemma 2, $\|u^{k+1} - u^k\| \to 0$ for $k \in K$, there must exist, for any $\varepsilon > 0$, a sufficiently large k_1 such that for any j

$$|L^j(u^{k+1}) - L^j(u^k)| \leq \varepsilon, \quad \forall\, k \in K \text{ and } k > k_1, \text{ or,}$$

$$L^j(u^k) - \varepsilon \leq L^j(u^{k+1}) \leq L^j(u^k) + \varepsilon, \quad \forall\, k \in K \text{ and } k > k_1.$$

Setting $j = k$ and using (3), i.e., $L^k(u^k) = \mathcal{L}(u^k)$, yield the following

$$\mathcal{L}(u^k) - \varepsilon \leq L^k(u^{k+1}) \leq \mathcal{L}(u^k) + \varepsilon, \quad \forall\, k \in K \text{ and } k > k_1.$$

Combine the left inequality with (7) to obtain that

$$
\begin{aligned}
&\alpha(\mathcal{L}(u^k) - \varepsilon) \leq \\
&\quad \alpha\mathcal{L}(u^*) - \tfrac{1}{2c}\|u^{k+1} - u^k\|^2 + \\
&\quad \frac{1}{2c}(\|\alpha(u^* - u^k)\| + \|(1-\alpha)(u^{k+1} - u^k)\|)^2, \quad \forall\, k \in K \text{ and } k > k_1
\end{aligned}
$$

Take the limit as $k \to \infty$ and $k \in K$ and obtain

$$
\begin{aligned}
\alpha(\mathcal{L}(u^\infty) - \varepsilon) &\leq \alpha\mathcal{L}(u^*) + \frac{1}{2c}\|\alpha(u^* - u^\infty)\|^2 \\
\mathcal{L}(u^\infty) - \varepsilon &\leq \mathcal{L}(u^*) + \frac{\alpha}{2c}\|u^* - u^\infty\|^2 \\
\mathcal{L}(u^\infty) - \mathcal{L}(u^*) &\leq \frac{\alpha}{2c}\|(u^* - u^\infty)\|^2 + \varepsilon. \qquad (8)
\end{aligned}
$$

Since (8) holds for any $\alpha \in (0,1)$ and ε can be chosen arbitrarily small, it must be true that

$$\mathcal{L}(u^\infty) - \mathcal{L}(u^*) = 0,$$

Thus, u^∞ is a solution to D.□

An immediate consequence of Theorem 3 is that the entire sequence $\{u^k\}_{k\in\mathcal{K}}$ converges to the optimal solution when D has a unique solution (see, e.g., Bazaraa and Shetty, 1979).

Consider now the case when $\mathcal{K}$ is finite. Define $\ell = \max\{k : k \in \mathcal{K}\} + 1$. Then,

$$z^k = z^\ell \quad \forall\, k \geq \ell, \text{ and}$$

$$\mathcal{L}(u^{k+1}) + \frac{m}{2c}\|u^{k+1} - u^k\|^2 > z^k = z^\ell. \quad \forall\, k \geq \ell \tag{9}$$

Lemma 4. *Let $\mathcal{K}$ be a finite set and ℓ be as defined above. Then,*

$$\liminf_{k\geq\ell} \|u^{k+1} - u^k\|^2 = 0.$$

Proof: Assume otherwise, i.e., there exists a $\delta > 0$ such that

$$\liminf_{k\geq\ell} \|u^{k+1} - u^k\|^2 \geq \delta. \tag{10}$$

In other words, for a sufficiently large $k_1 \geq \ell$,

$$\|u^{k+1} - u^k\|^2 \geq \frac{3}{4}\delta, \quad \forall k > k_1 \tag{11}$$

From Theorem 3, setting $u = u^k$ in (6) produces the following

$$L^k(u^{k+1}) + \tfrac{1}{2c}\|u^{k+1} - u^k\|^2 \leq L^k(u^k) \qquad \forall\, k. \tag{12}$$

Since $\{L^k(u)\}_k$ converges uniformly to $L^\infty(u)$, there must exist for every $\varepsilon \in (0, \frac{\delta}{2})$ a sufficiently large k_2 such that

$$|L^k(u) - L^\infty(u)| \leq \frac{\varepsilon}{4c} \quad \forall\, k > k_2 \text{ and } u \in U, \text{ or}$$

$$L^\infty(u) - \frac{\varepsilon}{4c} \leq L^k(u) \leq L^\infty(u) + \frac{\varepsilon}{4c} \quad \forall\, k > k_2 \text{ and } u \in U. \tag{13}$$

Combining (12) and (13) yields

$$\begin{aligned} L^\infty(u^{k+1}) - \frac{\varepsilon}{4c} + \tfrac{1}{2c}\|u^{k+1} - u^k\|^2 &\leq L^\infty(u^k) + \frac{\varepsilon}{4c} \quad \forall k > k_2 \\ L^\infty(u^{k+1}) + \frac{1}{2c}\{\|u^{k+1} - u^k\|^2 - \varepsilon\} &\leq L^\infty(u^k) \quad \forall k > k_2 \end{aligned}$$

Using (5), we must have that

$$\mathcal{L}(u^{k+1}) + \frac{1}{2c}\{\|u^{k+1} - u^k\|^2 - \varepsilon\} \leq \mathcal{L}(u^k) \quad \forall k > k_2 \tag{14}$$

However, (11) and (14) imply that the subsequence $\{\mathcal{L}(u^k)\}_{k \geq \bar{k}}$, where $\bar{k} = \max(k_1, k_2)$, is a monotonically decreasing sequence and bounded below by $\mathcal{L}(u^*)$. Therefore, $\{\mathcal{L}(u^k)\}_{k \geq \bar{k}}$ must converge and

$$\begin{aligned} \lim_{k \geq \bar{k}} \frac{1}{2c}\{\|u^{k+1} - u^k\|^2 - \varepsilon\} &= \lim_{k \geq \bar{k}} (\mathcal{L}(u^{k+1}) - \mathcal{L}(u^k)) = 0 \\ \lim_{k \geq \bar{k}} \|u^{k+1} - u^k\|^2 &= \varepsilon. \end{aligned}$$

Since ε can be chosen arbitrarily small, this contradicts (10). □

The above lemma implies that there exists a $K \subseteq \{k : k \geq \ell\}$ such that

$$\lim_{k \in K} \|u^{k+1} - u^k\|^2 = 0.$$

Since U is compact and $u^k \in U$ for all k, there must also exists a $K' \subseteq K$ such that $\{u^k\}_{k \in K'}$ converges to, say, u^∞. As a consequence, $\{u^{k+1}\}_{k \in K'}$ must converge to u^∞ as well.

Theorem 5. *If $\mathcal{K}$ is finite, then u^ℓ is a solution to D.*

Proof: Based on the preceding discussion, there must exist a $K \subseteq \{k : k \geq \ell\}$ such that the following conditions hold

1. $\lim_{k \in K} \|u^{k+1} - u^k\|^2 = 0$.
2. $\{u^k\}_{k \in K} \to u^\infty$.
3. $\{u^{k+1}\}_{k \in K} \to u^\infty$.

From Theorem 3, setting $u = \alpha u^* + (1 - \alpha)u^{k+1}$ in (6) for any $\alpha \in (0, 1)$ gives

$$\begin{aligned} L^k(u^{k+1}) + \tfrac{1}{2c}\|u^{k+1} - u^k\|^2 \leq\ & L^k(\alpha u^* + (1 - \alpha)u^{k+1}) + \\ & \tfrac{1}{2c}\|\alpha u^* + (1 - \alpha)u^{k+1} - u^k\|^2. \end{aligned}$$

Using the same argument as in Theorem 3 with the index set K, it can be shown that

$$\mathcal{L}(u^{\infty}) = \mathcal{L}(u^{*}). \tag{15}$$

Similarly, setting $u = \alpha u^{\ell} + (1-\alpha)u^{k+1}$ in (6) for any $\alpha \in (0,1)$ gives

$$\begin{aligned} L^k(u^{k+1}) + \tfrac{1}{2c}\|u^{k+1} - u^k\|^2 \;\leq\; & L^k(\alpha u^{\ell} + (1-\alpha)u^{k+1}) + \\ & \tfrac{1}{2c}\|\alpha u^{\ell} + (1-\alpha)u^{k+1} - u^k\|^2, \end{aligned}$$

and by the same reasoning it must follow that

$$\mathcal{L}(u^{\infty}) - \mathcal{L}(u^{\ell}) \leq 0 \text{ or } \mathcal{L}(u^{\infty}) \leq \mathcal{L}(u^{\ell}). \tag{16}$$

However, from (9) it is true that

$$\mathcal{L}(u^{k+1}) + \frac{m}{2c}\|u^{k+1} - u^k\|^2 > z^{\ell}. \quad \forall\, k \geq \ell$$

Take the limit as $k \to \infty$ and $k \in K$ and invoke the continuity of $\mathcal{L}(u)$ to obtain that

$$\mathcal{L}(u^{\infty}) \geq z^{\ell} = \mathcal{L}(u^{\ell}) \tag{17}$$

Combining (15), (16) and (17) yields

$$\mathcal{L}(u^{\infty}) = \mathcal{L}(u^{*}) = \mathcal{L}(u^{\ell}).$$

So, u^{ℓ} must be a solution to D.□

In addition to the above convergence results, if $f(x)$ and $g(x)$ are linear functions and X is a bounded polyhedral, then x^{k+1} in Step 3 can be restricted to extreme points of X, for which there are finitely many. In which case, there must exist a sufficiently large ℓ such that $L^k(u) = L^{\ell}(u)\ \forall\ k > \ell$ and $u \in U$. However, this implies that after ℓ iterations PMCP reduces to the application of the proximal minimization algorithm to the following linear program:

$$\begin{aligned} &\min\ w \\ &\text{s.t.} \\ &\quad w \;\geq\; f(x^i) + u \cdot g(x^i) \qquad i = 1, \ldots, \ell \\ &\quad Au \;\leq\; b \end{aligned}$$

Then, it follows from Exercise 4.3 in Bertsekas and Tsitsiklis (1989) that PMCP terminates finitely when D is the dual of a linear program, or equivalently, $\mathcal{L}(u)$ is piecewise linear.

4. A Bundle Method

Below, we describe a particular variant of the bundle methods which uses a different scheme for updating v^{k+1} in Step 4 of the generic algorithm. For later reference, we call this variant *a proximal minimization algorithm with subgradient bundles* (PMSB). One intention of this section is to present a convergent proof for PMSB which uses analysis similar to that of PMCP, thereby making the relationship between bundle methods and proximal minimization more concrete. Also, it should be noted that some variants of the bundle methods require a line search step (see, e.g., Fukushima, 1984, Gaudioso and Monaco, 1982 and Kiwiel, 1985). However, PMSB as stated below does not require any line search.

A PROXIMAL MINIMIZATION ALGORITHM
WITH SUBGRADIENT BUNDLES (PMSB)

Step 1: Select $u^1 \in U$ and m such that $0 < m < 1$. Set $k = 1, v^1 = u^1$ and

$$x^1 = \arg\max_{x \in X}\{f(x) + u^1 \cdot g(x)\}.$$

Step 2: Solve the master problem

$$u^{k+1} = \arg\min_{u \in U}\{L^k(u) + \frac{1}{2c_k}\|u - v^k\|^2\}.$$

if $u^{k+1} = v^k$, stop and v^k is an optimal solution.

Step 3: Solve the subproblem

$$x^{k+1} = \arg\max_{x \in X}\{f(x) + u^{k+1} \cdot g(x)\}$$

Note that $\mathcal{L}(u^{k+1}) = f(x^{k+1}) + u^{k+1} \cdot g(x^{k+1})$.

Step 4: Set

$$v^{k+1} = \begin{cases} u^{k+1} & \text{if } \mathcal{L}(u^{k+1}) + \frac{m}{2c_k}\|u^{k+1} - v^k\|^2 \leq \mathcal{L}(v^k) \\ v^k & \text{otherwise} \end{cases} \tag{18}$$

Recall that, when $v^{k+1} = u^{k+1}$, iteration k is called a 'serious' step or iteration. Otherwise, ($v^{k+1} = v^k$), it is called a 'null' step. In addition, the updating formula for v^{k+1} in Step 4 is also related to the one in Section 2 (see also Lemarechal, 1991) which is

$$v^{k+1} = \begin{cases} u^{k+1} & \text{if } \mathcal{L}(u^{k+1}) + m(L^k(v^k) - L^k(u^{k+1})) \leq \mathcal{L}(v^k) \\ v^k & \text{otherwise} \end{cases} \tag{19}$$

To obtain the relationship, observe that since u^{k+1} is a solution to the master problem

$$\begin{aligned} L^k(u^{k+1}) + \tfrac{1}{2c_k}\|u^{k+1} - v^k\|^2 &\leq L^k(v^k) \\ \tfrac{1}{2c_k}\|u^{k+1} - v^k\|^2 &\leq L^k(v^k) - L^k(u^{k+1}) \\ m\tfrac{1}{2c_k}\|u^{k+1} - v^k\|^2 &\leq m(L^k(v^k) - L^k(u^{k+1})) \\ \mathcal{L}(u^{k+1}) + m\tfrac{1}{2c_k}\|u^{k+1} - v^k\|^2 &\leq \mathcal{L}(u^{k+1}) + m(L^k(v^k) - L^k(u^{k+1})) \end{aligned}$$

So, the updating formula (19) implies (18).

When PMSB terminates finitely, Theorem 1 in the previous section still guarantees that v^k is an optimal solution to D. Below are convergence results for the case when the algorithm generates an infinite sequence. As in Section 3, it is assumed without loss of generality that $c_k = c > 0$, $\forall\, k$, and let

$$\mathcal{K} = \{k : v^{k+1} = u^{k+1}\}.$$

So, $\mathcal{K}$ is the index set for the serious steps (iterations).

Lemma 6. *Let $\mathcal{K}$ be an infinite set. If a subsequence $\{v^k\}_{k\in K}$ converges to v^∞ where $K \subseteq \mathcal{K}$, then $\{v^{k+1}\}_{k\in K}$ also converges to v^∞.*

Proof: Note that the sequence $\{\mathcal{L}(v^k)\}_k$ is a nonincreasing sequence which is bounded below by $\mathcal{L}(u^*)$. So, $\{\mathcal{L}(v^k)\}_k$ must converge. Since $K \subseteq \mathcal{K}$, the following must hold

$$\mathcal{L}(v^{k+1}) + \frac{m}{2c}\|v^{k+1} - v^k\|^2 \leq \mathcal{L}(v^k) \quad \forall\, k \in K$$

Following the same argument in Lemma 2, it can be shown that

$$0 = \lim_{k\in K} \frac{m}{2c}\|v^{k+1} - v^k\|^2.$$

Since both c and m are positive, $\{v^{k+1}\}_{k\in K}$ must converge to v^∞.□

Theorem 7. *If the cardinality of $\mathcal{K}$ is infinite, then every limit point of the sequence $\{v^k\}_{k\in\mathcal{K}}$ is a solution to D.*

Proof: Let $\{v^k\}_{k\in K}$ be a convergent subsequence where $K \subseteq \mathcal{K}$. Since $v^{k+1} = u^{k+1}\ \forall\, k \in K$ and u^{k+1} is optimal to the master problem, the following must hold

$$L^k(v^{k+1}) + \tfrac{1}{2c}\|v^{k+1} - v^k\|^2 \le L^k(u) + \tfrac{1}{2c}\|u - v^k\|^2, \quad \forall\, u \in U \;\&\; k \in K.$$

Using the same analysis as in Theorem 3 and the result for Lemma 6, it can be shown that $\{v^k\}_{k\in K}$ converges to u^*.□

When the cardinality of $\mathcal{K}$ is finite, define as before $\ell = \max\{k : k \in \mathcal{K}\} + 1$. So, every iteration $k \ge \ell$ must be a null step and the master problem in Step 2 must have the form:

$$u^{k+1} = \arg\max\{L^k(u) + \tfrac{1}{2c}\|u - v^\ell\|^2\}, \qquad \forall\, k \ge$$

Next, let

$$\begin{aligned} F^k(u) &= L^k(u) + \tfrac{1}{2c}\|u - v^\ell\|^2, \text{ and} \\ F^\infty(u) &= L^\infty(u) + \tfrac{1}{2c}\|u - v^\ell\|^2. \end{aligned}$$

Then, $F^k(u)$ *epi-converges* to $F^\infty(u)$ since $L^k(u)$ pointwise and monotonically converges to $L^\infty(u)$. (For the definition and properties of *epi-convergence*, see, e.g., the appendix in Wets, 1989.) In addition, it follows from Theorem A.2 of Wets(1989) that if $\{u^{k+1}\}_{k\in K} \to u^\infty$ for some $K \subseteq \{1, 2, 3, \ldots\}$, then

$$u^\infty = \arg\min_{u\in U} F^\infty(u).$$

Furthermore, since $L^k(\cdot)$ also uniformly converges to $L^\infty(\cdot)$, there must exists, for any $\varepsilon > 0$, a sufficiently large k_1 such that

$$|L^k(u) - L^\infty(u)| < \varepsilon \quad \forall\, u \in U \text{ and } k > k_1,$$

and by setting $u = u^{k+1}$

$$\begin{aligned} |L^k(u^{k+1}) - L^\infty(u^{k+1})| &\le \varepsilon \quad \forall\, k > k_1 \\ |L^k(u^{k+1}) - \mathcal{L}(u^{k+1})| &\le \varepsilon \quad \forall\, k > k_1 \\ \mathcal{L}(u^{k+1}) - \varepsilon \le L^k(u^{k+1}) &\le \mathcal{L}(u^{k+1}) + \varepsilon \quad \forall\, k > k_1 \end{aligned} \tag{20}$$

where the middle inequality follows from (5).

Theorem 8. *If the cardinality of $\mathcal{K}$ is finite, then v^ℓ is a solution to D.*

Proof: Since U is a compact set, there must exist a set $K \subseteq \{k : k \geq \ell\}$ such that $\{u^{k+1}\}_{k \in K}$ converges to u^∞.

If $u^\infty = v^\ell$, then the above observation concerning the epi-convergence of $F^k(u)$ implies that

$$\begin{aligned}
\mathcal{L}(v^\ell) &= L^\infty(v^\ell) \\
&= L^\infty(u^\infty) \\
&= \min_{u \in U}\{L^\infty(u) + \tfrac{1}{2c}\|u - v^\ell\|^2\} \\
&\leq \min_{u \in U}\{\mathcal{L}(u) + \tfrac{1}{2c}\|u - v^\ell\|^2\} \\
&\leq \mathcal{L}(v^\ell)
\end{aligned}$$

where the first equation follows from (5), the third from the observation that $u^\infty = \arg\min_{u \in U} F^\infty(u)$, the fourth from the fact that $L^\infty(u) \leq \mathcal{L}(u)\ \forall\ u \in U$, and the last from the fact that v^ℓ is an element of U. Thus,

$$\mathcal{L}(v^\ell) = \min_{u \in U}\{\mathcal{L}(u) + \tfrac{1}{2c}\|u - v^\ell\|^2\}.$$

However, this implies that v^ℓ solves D.

Assume that $u^\infty \neq v^\ell$. Let $\delta = \|u^\infty - v^\ell\|^2$. Then, there must exists a sufficiently large k_2 such that

$$\|u^{k+1} - v^\ell\|^2 \geq \frac{\delta}{2} \quad \forall\ k \geq k_2\ \&\ k \in K \tag{21}$$

However, since u^{k+1} solves the master problem in Step 2 with v^ℓ as its prox-center, it must be true that

$$L^k(u^{k+1}) + \tfrac{1}{2c}\|u^{k+1} - v^\ell\|^2 \leq L^k(v^\ell) = \mathcal{L}(v^\ell), \quad \forall\ k \geq \ell,$$

where the equality follows from (4) in Section 3. For any $\varepsilon > 0$, let k_1 be as in (20) so that

$$\mathcal{L}(u^{k+1}) - \varepsilon + \tfrac{1}{2c}\|u^{k+1} - v^\ell\|^2 \leq \mathcal{L}(v^\ell), \quad \forall\ k \geq \max(\ell, k_1).$$

Set $\varepsilon = \frac{(1-m)\delta}{4c}$ and obtain

$$\mathcal{L}(u^{k+1}) - \frac{1}{2c}(\|u^{k+1} - v^\ell\|^2 - (1-m)\frac{\delta}{2}) \leq \mathcal{L}(v^\ell), \quad \forall\, k \geq \max(\ell, k_1).$$

Then, for any $k \geq \max(\ell, k_1, k_2)$ and $k \in K$, (21) implies that

$$\begin{aligned}
\mathcal{L}(v^\ell) &\geq \mathcal{L}(u^{k+1}) + \frac{1}{2c}(\|u^{k+1} - v^\ell\|^2 - (1-m)\frac{\delta}{2}) \\
&\geq \mathcal{L}(u^{k+1}) + \frac{1}{2c}(\|u^{k+1} - v^\ell\|^2 - (1-m)\|u^{k+1} - v^\ell\|^2) \\
&\geq \mathcal{L}(u^{k+1}) + \frac{m}{2c}\|u^{k+1} - v^\ell\|^2
\end{aligned}$$

However, this implies that there must be a serious step after iteration ℓ which is a contradiction. Thus, every convergent subsequence of $\{u^{k+1}\}_{k\geq\ell}$ converges to v^ℓ which is a solution to D. However, this implies that the sequence $\{u^{k+1}\}_{k\geq\ell}$ converges to v^ℓ as well. □

5. Conclusion

This paper presents a generic algorithm in the framework of proximal minimization. It is shown that this generic algorithm can be specialized to four different algorithms; they are the cutting plane algorithm, the cutting plane algorithm with line searches, the bundle methods (or proximal minimization with subgradient bundles, PMSB) and proximal minimization with cutting planes (PMCP). The first three can be found in the current literature; however, the last one is new.

Besides the obvious relationship that all four algorithms can be derived from the generic algorithm, other relationships based on the master problem and convergence behavior are also established. Convergence proofs for PMSB and PMCP are also given.

References

[1] Auslender, A., (1987), 'Numerical Methods for Nondifferentiable Convex Optimization,' *Mathematical Programming Study*, 30, 102-126.

[2] Bazaraa, M. S. and Shetty, C. M., (1979), *Nonlinear Programming: Theory and Algorithms*, Wiley, New York, New York.

[3] Bertsekas, D. P., (1982), 'Projected Newton Methods for Optimization Problems with Simple Constraints,' *SIAM Journal of Control and Optimization*, 20, 221-246.

[4] Bertsekas, D. P.. and Tsitsiklis, J. N., (1989), *Parallel and Distributed Computation,* Prentice Hall, Englewood Cliffs, New Jersey.

[5] Cheney, E. W., and Goldstein, A. A., (1959), 'Newton's Method for Convex Programming and Chebyshev Approximation,' *Numerische Mathematik*, 1, 253-268.

[6] Dantzig, G. B., (1963), *Linear Programming and Extensions,* Princeton University Press, Princeton, New Jersey.

[7] Dantzig, D. B., and Wolfe, P., (1960), 'Decomposition Principles for Linear Programs,' *Operations Research*, 8, 101-111.

[8] Fukushima, M., (1984), 'A Descent Algorithm for Nonsmooth Convex Programming,' *Mathematical Programming*, 30, 163-175.

[9] Gaudioso, M. and Monaco, M. F., (1982), 'A Bundle Type Approach to the Unconstrained Minimization of Convex Nonsmooth Functions,' *Mathematical Programming*, 23, 216-226.

[10] Hearn, D. W., Lawphongpanich, S., and Nguyen, S., (1984), 'Convex Programming Formulations of the Asymmetric Traffic Assignment Problem,' *Transportation Research*, 18b, 357-365.

[11] Hearn, D. W. and Lawphongpanich, S., (1989a), 'Lagrangian Dual Ascent by Generalized Linear Programming,' *Operations Research Letters*, 8, 189-196.

[12] Hearn, D. W. and Lawphongpanich, S., (1989b), 'Generalized Linear Programming with Line Search,' *Proceedings of the 28th IEEE Conference on Decision and Control.*

[13] Hearn, D. W. and Lawphongpanich, S., (1990), 'A Dual Ascent Algorithm for Traffic Assignment Problems,' *Transportation Research*, 24, 423-430.

[14] Kelly, J. E., (1960), 'The Cutting Plane Method for Solving Convex Programs,' *Journal of SIAM*, VIII, 703-712.

[15] Kiwiel, K. C., (1985), *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics 1133, Springer, Berlin, Germany.

[16] Kiwiel, K. C., (1989), 'Proximity Control in Bundle Methods for Convex Nondifferentiable Minimization,' *Mathematical Programming*, 46, 105-122.

[17] Kiwiel, K. C., (1991), 'A Dual Method for Solving Certain Positive Semidefinite Quadratic Programming Problems,' *SIAM Journal on Scientific and Statistic Computing*, to appear.

[18] Lemarechal, C., (1974), 'An Algorithm for Minimizing Convex Functions,' *Proceedings of IFIP'74 Congress*, J. L. Rosenfeld (Ed.), North Holland, Amsterdam, The Netherlands, 552-556.

[19] Lemarechal, C., (1975), 'An Extension of Davidon Methods to Nondifferentiable Problems,' *Mathematical Programming Study*, 3, 95-109.

[20] Lemarechal, C., (1989), 'Nondifferentiable Optimization,' in *Optimization*, G. L. Nemhauser, A.H.G. Rinnooy Kan and M. J. Todd (Eds.), North Holland, New York, New York, 529-572.

[21] Lemarechal, C., (1991), 'Lagrangian Decomposition and Nonsmooth Optimization: Bundle Algorithm, Prox Iteration, Augmented Lagrangian,' *Proceedings to the 1991 Conference on Nonsmooth Optimization: Methods and Applications at ERICE(Trapani), Sicily*, Giannessi, F. (Ed.).

[22] Lin, Y. Y., and Pang, J.-S., (1987), 'Iterative Methods for Large Convex Quadratic Programs: A Survey,' *SIAM Journal of Control and Optimization*, 25, 383-411.

[23] Magnanti, T. L., Shapiro, J. F., and Wagner, M. H., (1976), 'Generalized Linear Programming Solves the Dual,' *Management Science*, 22, 1195-1203.

[24] Martinet, B., (1970), 'Regularisation d'Inequations Variationnelles par Approximation Successives,' *Rev. Francaise Inf. Rech. Oper.*, 4, 154-159.

[25] Mifflin, R., (1977), 'An Algorithm for Constrained Optimization with Semismooth Functions,' *Mathematics of Operations Research*, 2, 191-207.

[26] Pang, J.-S., (1983), 'Methods for Quadratic Programming: A Survey,' *Computers and Chemical Engineering*, 7, 583-594.

[27] Rockafellar, R. T., (1976), 'Monotone Operators and the Proximal Point Algorithm,' *SIAM Journal of Control and Optimization*, 14, 877-898.

[28] Rudin, W., (1976), *Principles of Mathematical Analysis,* McGraw Hill, New York, New York.

[29] Wets, R. J.-B., (1989), 'Stochastic Programming,' in *Optimization*, G. L. Nemhauser, A.H.G. Rinnooy Kan and M. J. Todd (Eds.), North Holland, New York, New York, 573-629.

[30] Wolfe, P., (1975), 'A Method of Conjugate Subgradient for Minimizing Nondifferentiable Functions,' *Mathematical Programming Study*, 3, 145-173.

[31] Zangwill, W. I., (1969), *Nonlinear Programming*, Prentice Hall, Englewood Cliffs, New Jersey.

INITIAL DISTRIBUTION LIST

1. Library (Code 52) 2
   Naval Postgraduate School
   Monterey, CA 93943-5000

2. Defense Technical Information Center 2
   Cameron Station
   Alexandria, VA 22314

3. Office of Research Administration (Code 81) 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

4. Prof. Peter Purdue 1
   Code OR-Pd
   Naval Postgraduate School
   Monterey, CA 93943-5000

5. Prof. Siriphong Lawphongpanich 20
   Code OR-Lp
   Naval Postgraduate School
   Monterey, CA 93943-5000