

A UNIFYING POLYHEDRAL APPROXIMATION FRAMEWORK FOR CONVEX OPTIMIZATION

DIMITRI P. BERTSEKAS* AND HUIZHEN YU†

Abstract. We propose a unifying framework for polyhedral approximation in convex optimization. It subsumes classical methods, such as cutting plane and simplicial decomposition, but also includes new methods, and new versions/extensions of old methods, such as a simplicial decomposition method for nondifferentiable optimization, and a new piecewise linear approximation method for convex single commodity network flow problems. Our framework is based on an extended form of monotropic programming, a broadly applicable model, which includes as special cases Fenchel duality and Rockafellar's monotropic programming, and is characterized by an elegant and symmetric duality theory. Our algorithm combines flexibly outer and inner linearization of the cost function. The linearization is progressively refined by using primal and dual differentiation, and the roles of outer and inner linearization are reversed in a mathematically equivalent dual algorithm. We provide convergence results for the general case where outer and inner linearization are combined in the same algorithm.

Key words. cutting plane, simplicial decomposition, polyhedral approximation, convex optimization, monotropic programming

AMS subject classifications. 90C25, 90C59, 65K05

1. Introduction. We consider the problem

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^m f_i(x_i) && (1.1) \\ & \text{subject to} && (x_1, \dots, x_m) \in S, \end{aligned}$$

where (x_1, \dots, x_m) is a vector in $\mathfrak{R}^{n_1+\dots+n_m}$, with components $x_i \in \mathfrak{R}^{n_i}$, $i = 1, \dots, m$, and

$f_i : \mathfrak{R}^{n_i} \mapsto (-\infty, \infty]$ is a closed proper convex function for each i ,¹
 S is a subspace of $\mathfrak{R}^{n_1+\dots+n_m}$.

This problem has been studied recently by the first author in [Ber10], under the name extended monotropic programming. It is an extension of Rockafellar's monotropic programming framework [Roc84], where each function f_i is one-dimensional ($n_i = 1$ for all i).

*Dimitri Bertsekas is with the Dept. of Electr. Engineering and Comp. Science, M.I.T., Cambridge, Mass., 02139. His research was supported by NSF Grant ECCS-0801549, and by the Air Force Grant FA9550-10-1-0412.

†Huizhen Yu is with the Laboratory for Information and Decision Systems (LIDS), M.I.T., Cambridge, Mass., 02139. This work was done when she was with the Dept. of Computer Science and the Helsinki Institute for Information Technology (HIIT), Univ. of Helsinki, and supported in part by the Academy of Finland Grant 118653 (ALGODAN).

¹We will be using standard terminology of convex optimization, as given for example in textbooks such as Rockafellar's [Roc70], or the first author's recent book [Ber09]. Thus a closed proper convex function $f : \mathfrak{R}^n \mapsto (-\infty, \infty]$ is one whose epigraph $\text{epi}(f) = \{(x, w) \mid f(x) \leq w\}$ is a nonempty closed convex set. Its effective domain, $\text{dom}(f) = \{x \mid f(x) < \infty\}$, is the nonempty projection of $\text{epi}(f)$ on the space of x . If $\text{epi}(f)$ is a polyhedral set, then f is called polyhedral.

Note that a variety of problems can be converted to the form (1.1). For example the problem

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^m f_i(x) \\ & \text{subject to} && x \in X, \end{aligned}$$

where $f_i : \mathfrak{R}^n \mapsto (-\infty, \infty]$ are closed proper convex functions, and X is a subspace of \mathfrak{R}^n , can be converted to the format (1.1). This can be done by introducing m copies of x , i.e., auxiliary vectors $z_i \in \mathfrak{R}^n$ that are constrained to be equal, and write the problem as

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^m f_i(z_i) \\ & \text{subject to} && (z_1, \dots, z_m) \in S, \end{aligned}$$

where $S = \{(x, \dots, x) \mid x \in X\}$. A related case is the problem arising in the Fenchel duality framework,

$$\min_{x \in \mathfrak{R}^n} \{f_1(x) + f_2(Qx)\},$$

where Q is a matrix; it is equivalent to the following special case of problem (1.1):

$$\min_{(x_1, x_2) \in S} \{f_1(x_1) + f_2(x_2)\},$$

where $S = \{(x, Qx) \mid x \in \mathfrak{R}^n\}$.

Generally, any problem involving linear constraints and a convex cost function can be converted to a problem of the form (1.1). For example, the problem

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^m f_i(x_i) \\ & \text{subject to} && Ax = b, \end{aligned}$$

where A is a given matrix and b is a given vector, is equivalent to

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^m f_i(x_i) + \delta_Z(z) \\ & \text{subject to} && Ax - z = 0, \end{aligned}$$

where z is a vector of artificial variables, and δ_Z is the indicator function of the set $Z = \{z \mid z = b\}$. This is a problem of the form (1.1), where the constraint subspace is

$$S = \{(x, z) \mid Ax - z = 0\}.$$

Problems with nonlinear convex constraints, such as $g(x) \leq 0$, may be converted to the form (1.1) by introducing as additive terms in the cost corresponding indicator functions, such as $\delta(x) = 0$ for all x with $g(x) \leq 0$ and $\delta(x) = \infty$ otherwise.

An important property of problem (1.1) is that it admits an elegant and symmetric duality theory, an extension of Rockafellar's monotropic programming duality (which

in turn includes as special cases linear and quadratic programming duality). Our purpose in this paper is to develop a polyhedral approximation framework for problem (1.1), which is based on its favorable duality properties, as well as the generic duality between outer and inner linearization. In particular, we develop a general algorithm for problem (1.1) that contains as special cases the classical outer linearization (cutting plane) and inner linearization (simplicial decomposition) methods, but also includes new methods, and new versions/extensions of classical methods.

At a typical iteration, our algorithm solves an approximate version of problem (1.1), where some of the functions f_i are outer linearized, some are inner linearized, and some are left intact. Thus, in our algorithm outer and inner linearization are combined. Furthermore, their roles are reversed in the dual problem. At the end of the iteration, the linearization is refined by using the duality properties of problem (1.1).

There are several potential advantages of our method over classical cutting plane and simplicial decomposition methods (as described for example in the books [BGL09, Ber99, HiL93, Pol97]), depending on the problem's structure:

(a) The refinement process may be faster, because at each iteration, multiple cutting planes and break points are added (as many as one per function f_i). As a result, in a single iteration, a more refined approximation may result, compared with classical methods where a single cutting plane or extreme point is added. Moreover, when the component functions f_i are one-dimensional, adding a cutting plane/break point to the polyhedral approximation of f_i can be very simple, as it requires a one-dimensional differentiation or minimization for each f_i .

(b) The approximation process may preserve some of the special structure of the cost function and/or the constraint set. For example if the component functions f_i are one-dimensional, or have partially overlapping dependences, e.g.,

$$f(x_1, \dots, x_m) = f_1(x_1, x_2) + f_2(x_2, x_3) + \dots + f_{m-1}(x_{m-1}, x_m) + f_m(x_m),$$

the minimization of f by the classical cutting plane method leads to general/unstructured linear programming problems. By contrast, using our algorithm with separate outer or inner linearization of the component functions leads to linear programs with special structure, which can be solved efficiently by specialized methods, such as network flow algorithms (see Section 6.4), or interior point algorithms that can exploit the sparsity structure of the problem.

In this paper, we place emphasis on the general conceptual framework for polyhedral approximation and its convergence analysis. We do not include computational results, in part due to the fact that our algorithm contains several special cases of interest in diverse problem settings, which must be tested separately for a thorough algorithmic evaluation. However, it is clear that in at least two special cases, described in detail in Section 6, our algorithm offers distinct advantages over existing methods. These are:

(1) Simplicial decomposition methods for specially structured nondifferentiable optimization problems, where simplicial decomposition can exploit well the problem's structure (e.g., multicommodity flow problems [CaG74, FIH95, PaY84, LaP92]).

(2) Nonlinear convex single-commodity network flow problems, where the approximating subproblems can be solved with extremely fast linear network flow algorithms (see e.g., the textbooks [Roc84, AMO93, Ber98]), while the refinement of the approximation involves one-dimensional differentiation, and can be carried out very simply.

The paper is organized as follows. In Section 2 we define outer and inner linearizations, and we review the conjugacy correspondence between them, in a form which is suitable for our algorithmic purposes, while in Section 3 we review the duality theory for our problem. In Sections 4 and 5 we describe our algorithm and analyze its convergence properties, while in Section 6 we discuss various special cases, including classical methods and some generalized versions, such as new simplicial decomposition methods for minimizing a convex extended real-valued and/or nondifferentiable function f over a convex set C .

2. Outer and Inner Linearizations. In this section we define outer and inner linearizations, and we formalize their conjugacy relation and other related properties. An outer linearization of a closed proper convex function $f : \mathfrak{R}^n \mapsto (-\infty, \infty]$ is defined by a finite set of vectors Y such that for every $\tilde{y} \in Y$, we have $\tilde{y} \in \partial f(x_{\tilde{y}})$ for some $x_{\tilde{y}} \in \mathfrak{R}^n$.² It is given by

$$\underline{f}_Y(x) = \max_{\tilde{y} \in Y} \{f(x_{\tilde{y}}) + (x - x_{\tilde{y}})' \tilde{y}\}, \quad x \in \mathfrak{R}^n, \quad (2.1)$$

and it is illustrated in the left side of Fig. 2.1. The choices of $x_{\tilde{y}}$ such that $\tilde{y} \in \partial f(x_{\tilde{y}})$ may not be unique, but result in the same function $\underline{f}_Y(x)$: the epigraph of \underline{f}_Y is determined by the supporting hyperplanes to the epigraph of f with normals defined by $\tilde{y} \in Y$, and the points of support $x_{\tilde{y}}$ are immaterial. In particular, the definition (2.1) can be equivalently written as

$$\underline{f}_Y(x) = \max_{\tilde{y} \in Y} \{\tilde{y}'x - f^*(\tilde{y})\}, \quad (2.2)$$

using the relation $x'_{\tilde{y}} \tilde{y} = f(x_{\tilde{y}}) + f^*(\tilde{y})$, which is implied by $\tilde{y} \in \partial f(x_{\tilde{y}})$.

Note that $\underline{f}_Y(x) \leq f(x)$ for all x , so, as is true for any outer approximation of f , the conjugate $(\underline{f}_Y)^*$ satisfies $(\underline{f}_Y)^*(y) \geq f^*(y)$ for all y . Moreover, $(\underline{f}_Y)^*$ can be described as an inner linearization of the conjugate f^* of f , as illustrated in the right side of Fig. 2.1. Indeed we have, using Eq. (2.2),

$$\begin{aligned} (\underline{f}_Y)^*(y) &= \sup_{x \in \mathfrak{R}^n} \{y'x - \underline{f}_Y(x)\} \\ &= \sup_{x \in \mathfrak{R}^n} \left\{ y'x - \max_{\tilde{y} \in Y} \{ \tilde{y}'x - f^*(\tilde{y}) \} \right\} \\ &= \sup_{\substack{x \in \mathfrak{R}^n, \xi \in \mathfrak{R} \\ \tilde{y}'x - f^*(\tilde{y}) \leq \xi, \tilde{y} \in Y}} \{y'x - \xi\}. \end{aligned}$$

By linear programming duality, the optimal value of the linear program in (x, ξ) of the preceding equation can be replaced by the dual optimal value, and we have with a straightforward calculation

$$(\underline{f}_Y)^*(y) = \begin{cases} \inf_{\substack{\sum_{\tilde{y} \in Y} \alpha_{\tilde{y}} \tilde{y} = y, \\ \alpha_{\tilde{y}} \geq 0, \tilde{y} \in Y}} \sum_{\tilde{y} \in Y} \alpha_{\tilde{y}} f^*(\tilde{y}) & \text{if } y \in \text{conv}(Y), \\ \infty & \text{otherwise,} \end{cases} \quad (2.3)$$

²We denote by $\partial f(x)$ the set of all subgradients of f at x . By convention, $\partial f(x) = \emptyset$ for $x \notin \text{dom}(f)$. We also denote by f^* and f^{**} the conjugate of f and its double conjugate (conjugate of f^*). Two facts for a closed proper convex function $f : \mathfrak{R}^n \mapsto (-\infty, \infty]$ that we will use often are: (a) $f = f^{**}$ (the Conjugacy Theorem, see e.g., [Ber09], Prop. 1.6.1), and (b) the three conditions $y \in \partial f(x)$, $x \in \partial f^*(y)$, and $x'y = f(x) + f^*(y)$ are equivalent (the Conjugate Subgradient Theorem, see e.g., [Ber09], Prop. 5.4.3).

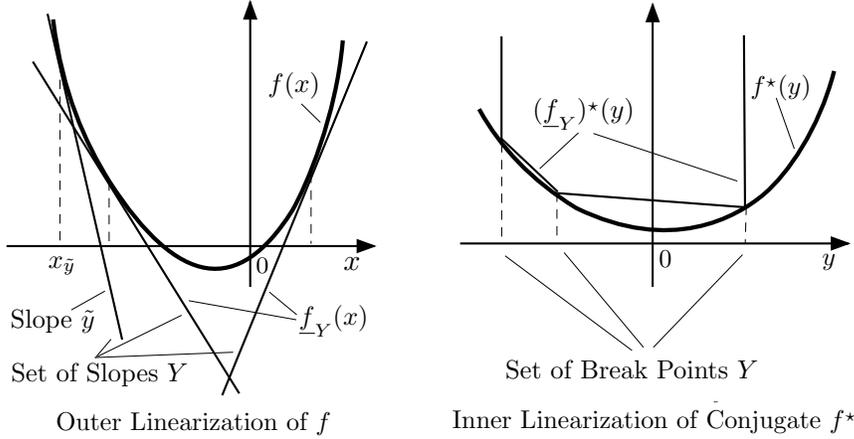


Fig. 2.1: Illustration of the conjugate $(\underline{f}_Y)^*$ of an outer linearization \underline{f}_Y of a convex function f defined by a finite set of “slopes” $\tilde{y} \in Y$ and corresponding points $x_{\tilde{y}}$ such that $\tilde{y} \in \partial f(x_{\tilde{y}})$ for all $\tilde{y} \in Y$. It is an inner linearization of the conjugate f^* of f , a piecewise linear function whose break points are $\tilde{y} \in Y$.

where $\alpha_{\tilde{y}}$ is the dual variable of the constraint $\tilde{y}'x - f^*(\tilde{y}) \leq \xi$.

From this formula, it can be seen that $(\underline{f}_Y)^*$ is a piecewise linear approximation of f^* with domain

$$\text{dom}((\underline{f}_Y)^*) = \text{conv}(Y),$$

and “break points” at $\tilde{y} \in Y$ with values equal to the corresponding values of f^* . In particular, as indicated in Fig. 2.1, the epigraph of $(\underline{f}_Y)^*$ is the convex hull of the union of the vertical halflines corresponding to $\tilde{y} \in Y$:

$$\text{epi}((\underline{f}_Y)^*) = \text{conv}\left(\bigcup_{\tilde{y} \in Y} \{(\tilde{y}, w) \mid f^*(\tilde{y}) \leq w\}\right).$$

In what follows, by an outer linearization of a closed proper convex function f defined by a finite set Y , we will mean the function \underline{f}_Y given by Eq. (2.1), while by an inner linearization of its conjugate f^* , we will mean the function $(\underline{f}_Y)^*$ given by Eq. (2.3). Note that not all sets Y define conjugate pairs of outer and inner linearizations via Eqs. (2.1) and (2.3), respectively, within our framework: it is necessary that for every \tilde{y} there exists $x_{\tilde{y}}$ such that $\tilde{y} \in \partial f(x_{\tilde{y}})$, or equivalently that $\partial f^*(\tilde{y}) \neq \emptyset$ for all $\tilde{y} \in Y$. By exchanging the roles of f and f^* , we also obtain dual definitions and statements. For example, for a finite set X to define an inner linearization \bar{f}_X of a closed proper convex function f as well as an outer linearization $(\bar{f}_X)^* = (\underline{f^*})_X$ of its conjugate f^* , it is necessary that $\partial f(\tilde{x}) \neq \emptyset$ for all $\tilde{x} \in X$.

3. Duality. In this section we review some aspects of the duality theory associated with problem (1.1). In particular, we will show that a dual problem has the form

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^m f_i^*(\lambda_i) && (3.1) \\ & \text{subject to} && \lambda = (\lambda_1, \dots, \lambda_m) \in S^\perp, \end{aligned}$$

where f_i^* is the conjugate of f_i , and S^\perp is the orthogonal subspace of S . Thus the dual problem has the same form as the primal problem (1.1). Furthermore, since the functions f_i are assumed closed proper and convex, we have $f_i^{**} = f_i$, where f_i^{**} is the conjugate of f_i^* , so when the dual problem is dualized, it yields the primal problem, and the duality is fully symmetric.

To derive the dual problem, we introduce auxiliary vectors $z_i \in \mathfrak{R}^{n_i}$ and we convert problem (1.1) to the equivalent form

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^m f_i(z_i) && (3.2) \\ & \text{subject to} && z_i = x_i, \quad i = 1, \dots, m, && (x_1, \dots, x_m) \in S. \end{aligned}$$

We then assign a multiplier vector $\lambda_i \in \mathfrak{R}^{n_i}$ to the constraint $z_i = x_i$, thereby obtaining the Lagrangian function

$$L(x_1, \dots, x_m, z_1, \dots, z_m, \lambda_1, \dots, \lambda_m) = \sum_{i=1}^m (f_i(z_i) + \lambda_i'(x_i - z_i)).$$

The dual function is

$$\begin{aligned} q(\lambda) &= \inf_{(x_1, \dots, x_m) \in S, z_i \in \mathfrak{R}^{n_i}} L(x_1, \dots, x_m, z_1, \dots, z_m, \lambda_1, \dots, \lambda_m) \\ &= \inf_{(x_1, \dots, x_m) \in S} \sum_{i=1}^m \lambda_i' x_i + \sum_{i=1}^m \inf_{z_i \in \mathfrak{R}^{n_i}} \{f_i(z_i) - \lambda_i' z_i\} \\ &= \begin{cases} -\sum_{i=1}^m f_i^*(\lambda_i) & \text{if } \lambda = (\lambda_1, \dots, \lambda_m) \in S^\perp, \\ -\infty & \text{otherwise,} \end{cases} \end{aligned} \quad (3.3)$$

where

$$f_i^*(\lambda_i) = \sup_{z_i \in \mathfrak{R}^{n_i}} \{\lambda_i' z_i - f_i(z_i)\}$$

is the conjugate of f_i . Thus the dual problem is to maximize $q(\lambda)$ over $\lambda \in S^\perp$, which, with a change of sign to convert maximization to minimization, takes the form (3.1).

We denote by f_{opt} the optimal value of the primal problem (1.1) and by f_{opt}^* the optimal value of the dual problem (3.1). We assume that strong duality holds ($-f_{opt}^* = f_{opt}$). By viewing the equivalent problem (3.2) as a convex programming problem with equality constraints, we may apply standard theory and obtain conditions that guarantee that $-f_{opt}^* = f_{opt}$ [for conditions beyond the standard, that exploit the special structure of problem (1.1), we refer to [Ber10], which shows among others that strong duality holds if each function f_i is either real-valued or is polyhedral]. Also, x^{opt} and λ^{opt} form an optimal primal and dual solution pair if and only if they satisfy the standard primal feasibility, dual feasibility, and Lagrangian optimality conditions (see e.g., Prop. 5.1.5 of [Ber99]). The latter condition is satisfied if and only if x_i^{opt} attains the infimum in the equation

$$-f_i^*(\lambda_i^{opt}) = \inf_{x_i \in \mathfrak{R}^{n_i}} \{f_i(x_i) - \lambda_i^{opt} x_i\}, \quad i = 1, \dots, m;$$

cf. Eq. (3.3). We thus obtain the following.

PROPOSITION 3.1 (Optimality Conditions). *We have $-\infty < -f_{opt}^* = f_{opt} < \infty$, and $x^{opt} = (x_1^{opt}, \dots, x_m^{opt})$ and $\lambda^{opt} = (\lambda_1^{opt}, \dots, \lambda_m^{opt})$ are optimal primal and dual solutions, respectively, of problem (1.1) if and only if*

$$x^{opt} \in S, \quad \lambda^{opt} \in S^\perp, \quad x_i^{opt} \in \arg \min_{x_i \in \mathbb{R}^{n_i}} \{f_i(x_i) - x_i' \lambda_i^{opt}\}, \quad i = 1, \dots, m. \quad (3.4)$$

Note that by the Conjugate Subgradient Theorem (Prop. 5.4.3 in [Ber09]), the condition $x_i^{opt} \in \arg \min_{x_i \in \mathbb{R}^{n_i}} \{f_i(x_i) - x_i' \lambda_i^{opt}\}$ of the preceding proposition is equivalent to either one of the following two subgradient conditions:

$$\lambda_i^{opt} \in \partial f_i(x_i^{opt}), \quad x_i^{opt} \in \partial f_i^*(\lambda_i^{opt}). \quad (3.5)$$

Our polyhedral approximation algorithm, to be introduced shortly, involves the solution of problems of the form (1.1), where f_i are either the original problem functions or polyhedral approximations thereof, and may require the simultaneous determination of both primal and dual optimal solutions x^{opt} and λ^{opt} . This can be done in a number of ways, depending on the convenience afforded by the problem's character. One way is to use a specialized algorithm that takes advantage of the problem's special structure to simultaneously find a primal solution of the equivalent problem (3.2) as well as a dual solution/multiplier. An example is when the functions f_i are themselves polyhedral (possibly through linearization), in which case problem (3.2) is a linear program whose primal and dual optimal solutions can be obtained by linear programming algorithms such as the simplex method. Another example, which involves a favorable special structure, is monotropic programming and network optimization (see the discussion of Section 6.4).

If we use an algorithm that finds only an optimal primal solution x^{opt} , we may still be able to obtain an optimal dual solution through the optimality conditions of Prop. 3.1. In particular, given $x^{opt} = (x_1^{opt}, \dots, x_m^{opt})$, we may find $\lambda^{opt} = (\lambda_1^{opt}, \dots, \lambda_m^{opt})$, either through the differentiation $\lambda_i^{opt} \in \partial f_i(x_i^{opt})$ [cf. Eq. (3.5)] or through the equivalent optimization

$$\lambda_i^{opt} \in \arg \max_{\lambda_i \in \mathbb{R}^{n_i}} \{\lambda_i' x_i^{opt} - f_i^*(\lambda_i)\}.$$

However, neither of these two conditions are sufficient for optimality of λ^{opt} because the condition $\lambda^{opt} \in S^\perp$ must also be satisfied as per Prop. 3.1 (unless each f_i is differentiable at x_i^{opt} , in which case λ^{opt} is unique). Thus, the effectiveness of this approach may depend on the special structure of the problem at hand (see Section 6.2 for an example).

4. Generalized Polyhedral Approximation. We will now introduce our algorithm, referred to as *generalized polyhedral approximation* or GPA, whereby problem (1.1) is approximated by using inner and/or outer linearization of some of the functions f_i . The optimal primal and dual solution pair of the approximate problem is then used to construct more refined inner and outer linearizations. The algorithm uses a fixed partition of the index set $\{1, \dots, m\}$,

$$\{1, \dots, m\} = I \cup \bar{I} \cup \bar{\bar{I}},$$

which determines the functions f_i that are outer approximated (set \bar{I}) and the functions f_i that are inner approximated (set $\bar{\bar{I}}$). We assume that at least one of the sets \bar{I} and $\bar{\bar{I}}$ is nonempty.

For $i \in \underline{I}$, given a finite set Λ_i such that $\partial f_i^*(\tilde{\lambda}) \neq \emptyset$ for all $\tilde{\lambda} \in \Lambda_i$, we consider the outer linearization of f_i corresponding to Λ_i , and denote it by

$$\underline{f}_{i,\Lambda_i}(x_i) = \max_{\tilde{\lambda} \in \Lambda_i} \{\tilde{\lambda}'x_i - f_i^*(\tilde{\lambda})\}.$$

Equivalently, as noted in Section 2 [cf. Eqs. (2.1) and (2.2)], we have

$$\underline{f}_{i,\Lambda_i}(x_i) = \max_{\tilde{\lambda} \in \Lambda_i} \{f_i(x_{\tilde{\lambda}}) + (x_i - x_{\tilde{\lambda}})' \tilde{\lambda}\},$$

where for each $\tilde{\lambda} \in \Lambda_i$, $x_{\tilde{\lambda}}$ is such that $\tilde{\lambda} \in \partial f_i(x_{\tilde{\lambda}})$.

For $i \in \bar{I}$, given a finite set X_i such that $\partial f_i(\tilde{x}) \neq \emptyset$ for all $\tilde{x} \in X_i$, we consider the inner linearization of f_i corresponding to X_i , and denote it by $\bar{f}_{i,X_i}(x_i)$:

$$\bar{f}_{i,X_i}(x_i) = \begin{cases} \min_{\{\alpha_{\tilde{x}} | \tilde{x} \in X_i\} \in C(x_i, X_i)} \sum_{\tilde{x} \in X_i} \alpha_{\tilde{x}} f_i(\tilde{x}) & \text{if } x_i \in \text{conv}(X_i), \\ \infty & \text{otherwise,} \end{cases}$$

where $C(x_i, X_i)$ is the set of all vectors with components $\alpha_{\tilde{x}}$, $\tilde{x} \in X_i$, satisfying

$$\sum_{\tilde{x} \in X_i} \alpha_{\tilde{x}} \tilde{x} = x_i, \quad \sum_{\tilde{x} \in X_i} \alpha_{\tilde{x}} = 1, \quad \alpha_{\tilde{x}} \geq 0, \quad \forall \tilde{x} \in X_i,$$

[cf. Eq. (2.3)]. As noted in Section 2, this is the function whose epigraph is the convex hull of the union of the halflines $\{(\tilde{x}, w) \mid f_i(\tilde{x}) \leq w\}$, $\tilde{x} \in X_i$ (cf. Fig. 2.1).

At the typical iteration of the algorithm, we have for each $i \in \underline{I}$, a finite set Λ_i such that $\partial f_i^*(\tilde{\lambda}) \neq \emptyset$ for all $\tilde{\lambda} \in \Lambda_i$, and for each $i \in \bar{I}$, a finite set X_i such that $\partial f_i(\tilde{x}) \neq \emptyset$ for all $\tilde{x} \in X_i$. The iteration is as follows.

Typical Iteration of GPA Algorithm:

Step 1: (Approximate Problem Solution) Find a primal and dual optimal solution pair $(\hat{x}_1, \dots, \hat{x}_m, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$ of the problem

$$\begin{aligned} & \text{minimize} && \sum_{i \in \underline{I}} f_i(x_i) + \sum_{i \in \underline{I}} \underline{f}_{i,\Lambda_i}(x_i) + \sum_{i \in \bar{I}} \bar{f}_{i,X_i}(x_i) && (4.1) \\ & \text{subject to} && (x_1, \dots, x_m) \in S, \end{aligned}$$

where $\underline{f}_{i,\Lambda_i}$ and \bar{f}_{i,X_i} are the outer and inner linearizations of f_i corresponding to X_i and Λ_i , respectively.

Step 2: (Enlargement and Test for Termination) Enlarge the sets X_i and Λ_i using the following differentiation process (see Fig. 4.1):

- (a) For $i \in \underline{I}$, we add $\tilde{\lambda}_i$ to the corresponding set Λ_i , where $\tilde{\lambda}_i \in \partial f_i(\hat{x}_i)$.
- (b) For $i \in \bar{I}$, we add \tilde{x}_i to the corresponding set X_i , where $\tilde{x}_i \in \partial f_i^*(\hat{\lambda}_i)$.

If there is no strict enlargement, i.e., for all $i \in \underline{I}$ we have $\tilde{\lambda}_i \in \Lambda_i$, and for all $i \in \bar{I}$ we have $\tilde{x}_i \in X_i$, the algorithm terminates. Otherwise, we proceed to the next iteration, using the enlarged sets Λ_i and X_i .

We will show shortly that when the algorithm terminates, then

$$(\hat{x}_1, \dots, \hat{x}_m, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$$

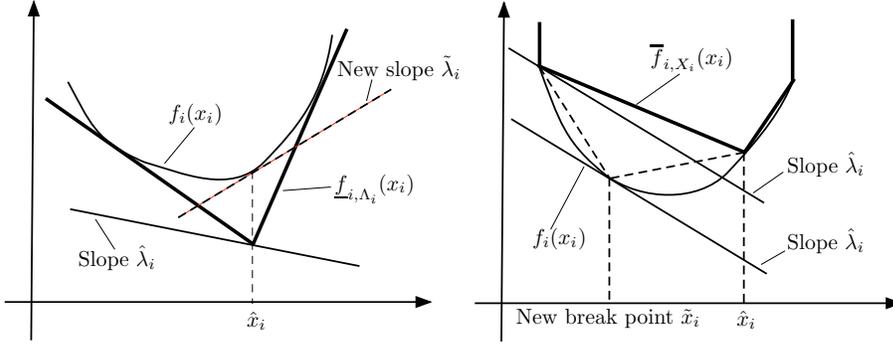


Fig. 4.1: Illustration of the enlargement step in the polyhedral approximation method, after we obtain a primal and dual optimal solution pair $(\hat{x}_1, \dots, \hat{x}_m, \hat{\lambda}_1, \dots, \hat{\lambda}_m)$. The enlargement step on the left [finding $\tilde{\lambda}_i$ with $\tilde{\lambda}_i \in \partial f_i(\hat{x}_i)$] is also equivalent to finding $\tilde{\lambda}_i$ satisfying $\hat{x}_i \in \partial f_i^*(\tilde{\lambda}_i)$ [cf. the Conjugate Subgradient Theorem, (Prop. 5.4.3 in [Ber09])]. The enlargement step on the right [finding \tilde{x}_i with $\tilde{x}_i \in \partial f_i^*(\hat{\lambda}_i)$] is also equivalent to finding \tilde{x}_i satisfying $\hat{\lambda}_i \in \partial f_i(\tilde{x}_i)$.

is a primal and dual optimal solution pair of the original problem. Note that we implicitly assume that at each iteration, there exists a primal and dual optimal solution pair of problem (4.1). Furthermore, we assume that the enlargement step can be carried out, i.e., that $\partial f_i(\hat{x}_i) \neq \emptyset$ for all $i \in \underline{I}$ and $\partial f_i^*(\hat{\lambda}_i) \neq \emptyset$ for all $i \in \bar{I}$. Sufficient assumptions may need to be imposed on the problem to guarantee that this is so.

There are two prerequisites for the method to be effective:

(1) The (partially) linearized problem (4.1) must be easier to solve than the original problem (1.1). For example, problem (4.1) may be a linear program, while the original may be nonlinear (cf. the cutting plane method, to be discussed in Section 6.1); or it may effectively have much smaller dimension than the original (cf. the simplicial decomposition method, to be discussed in Section 6.2).

(2) Finding the enlargement vectors ($\tilde{\lambda}_i$ for $i \in \underline{I}$, and \tilde{x}_i for $i \in \bar{I}$) must not be too difficult. Note that if the differentiation $\tilde{\lambda}_i \in \partial f_i(\hat{x}_i)$ for $i \in \underline{I}$, and $\tilde{x}_i \in \partial f_i^*(\hat{\lambda}_i)$ for $i \in \bar{I}$ is not convenient for some of the functions (e.g., because some of the f_i or the f_i^* are not available in closed form), we may calculate $\tilde{\lambda}_i$ and/or \tilde{x}_i via the equivalent relations

$$\hat{x}_i \in \partial f_i^*(\tilde{\lambda}_i), \quad \hat{\lambda}_i \in \partial f_i(\tilde{x}_i);$$

(cf. Prop. 5.4.3 of [Ber09]). This amounts to solving optimization problems. For example, finding \tilde{x}_i such that $\hat{\lambda}_i \in \partial f_i(\tilde{x}_i)$ is equivalent to solving the problem

$$\begin{aligned} & \text{maximize} \quad \{\hat{\lambda}_i' x_i - f_i(x_i)\} \\ & \text{subject to} \quad x_i \in \mathfrak{R}^{n_i}, \end{aligned} \quad (4.2)$$

and may be nontrivial (cf. Fig. 4.1).

The facility of solving the linearized problem (4.1) and carrying out the subsequent enlargement step may guide the choice of functions that are inner or outer linearized.

We note that in view of the symmetry of duality, the GPA algorithm may be applied to the dual of problem (1.1):

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^m f_i^*(\lambda_i) && (4.3) \\ & \text{subject to} && (\lambda_1, \dots, \lambda_m) \in S^\perp, \end{aligned}$$

where f_i^* is the conjugate of f_i . Then the inner (or outer) linearized index set \bar{I} of the primal becomes the outer (or inner, respectively) linearized index set of the dual. At each iteration, the algorithm solves the dual of the approximate version of problem (4.1),

$$\begin{aligned} & \text{minimize} && \sum_{i \in I} f_i^*(\lambda_i) + \sum_{i \in \underline{I}} (\underline{f}_{i, \Lambda_i})^*(\lambda_i) + \sum_{i \in \bar{I}} (\bar{f}_{i, X_i})^*(\lambda_i) && (4.4) \\ & \text{subject to} && (\lambda_1, \dots, \lambda_m) \in S^\perp, \end{aligned}$$

where the outer (or inner) linearization of f_i^* is the conjugate of the inner (or respectively, outer) linearization of f_i (cf. Section 2). The algorithm produces mathematically identical results when applied to the primal or the dual, as long as the roles of outer and inner linearization are appropriately reversed. The choice of whether to apply the algorithm in its primal or its dual form is simply a matter of whether calculations with f_i or with their conjugates f_i^* are more or less convenient. In fact, when the algorithm makes use of both the primal solution $(\hat{x}_1, \dots, \hat{x}_m)$ and the dual solution $(\hat{\lambda}_1, \dots, \hat{\lambda}_m)$ in the enlargement step, the question of whether the starting point is the primal or the dual becomes moot: it is best to view the algorithm as applied to the pair of primal and dual problems, without designation of which is primal and which is dual.

Now let us show the optimality of the primal and dual solution pair obtained upon termination of the algorithm. We will use two basic properties of outer approximations. The first is that for a closed proper convex function f and any x ,

$$\underline{f} \leq f, \quad \underline{f}(x) = f(x) \quad \implies \quad \partial \underline{f}(x) \subset \partial f(x). \quad (4.5)$$

The second is that for an outer linearization \underline{f}_Λ of f and any x ,

$$\tilde{\lambda} \in \Lambda, \quad \tilde{\lambda} \in \partial f(x) \quad \implies \quad \underline{f}_\Lambda(x) = f(x). \quad (4.6)$$

The first property follows from the definition of subgradients, whereas the second property follows from the definition of \underline{f}_Λ .

PROPOSITION 4.1 (Optimality at Termination). *If the GPA algorithm terminates at some iteration, the corresponding primal and dual solutions, $(\hat{x}_1, \dots, \hat{x}_m)$ and $(\hat{\lambda}_1, \dots, \hat{\lambda}_m)$, form a primal and dual optimal solution pair of problem (1.1).*

Proof. From Prop. 3.1, and the definition of $(\hat{x}_1, \dots, \hat{x}_m)$ and $(\hat{\lambda}_1, \dots, \hat{\lambda}_m)$ as a primal and dual optimal solution pair of the approximate problem (4.1), we have

$$(\hat{x}_1, \dots, \hat{x}_m) \in S, \quad (\hat{\lambda}_1, \dots, \hat{\lambda}_m) \in S^\perp.$$

We will show that upon termination, we have for all i

$$\hat{\lambda}_i \in \partial f_i(\hat{x}_i), \quad (4.7)$$

which by Prop. 3.1 implies the desired conclusion. Since $(\hat{x}_1, \dots, \hat{x}_m)$ and $(\hat{\lambda}_1, \dots, \hat{\lambda}_m)$ are a primal and dual optimal solution pair of problem (4.1), Eq. (4.7) holds for all $i \notin \underline{I} \cup \bar{I}$ (cf. Prop. 3.1). We will complete the proof by showing that it holds for all $i \in \underline{I}$ (the proof for $i \in \bar{I}$ follows by a dual argument).

Indeed, let us fix $i \in \underline{I}$ and let $\tilde{\lambda}_i \in \partial f_i(\hat{x}_i)$ be the vector generated by the enlargement step upon termination. We must have $\tilde{\lambda}_i \in \Lambda_i$, since there is no strict enlargement upon termination. Since $\underline{f}_{i, \Lambda_i}$ is an outer linearization of f_i , by Eq. (4.6), the fact $\tilde{\lambda}_i \in \Lambda_i, \tilde{\lambda}_i \in \partial f_i(\hat{x}_i)$ implies

$$\underline{f}_{i, \Lambda_i}(\hat{x}_i) = f_i(\hat{x}_i),$$

which in turn implies by Eq. (4.5) that

$$\partial \underline{f}_{i, \Lambda_i}(\hat{x}_i) \subset \partial f_i(\hat{x}_i).$$

By Prop. 3.1, we also have $\hat{\lambda}_i \in \partial \underline{f}_{i, \Lambda_i}(\hat{x}_i)$, so $\hat{\lambda}_i \in \partial f_i(\hat{x}_i)$. \square

5. Convergence Analysis. Generally, convergence results for polyhedral approximation methods, such as the classical cutting plane methods, are of two types: finite convergence results that apply to cases where the original problem has polyhedral structure, and asymptotic convergence results that apply to nonpolyhedral cases. Our subsequent convergence results conform to these two types.

We first derive a finite convergence result, assuming that the problem has a certain polyhedral structure, and care is taken to ensure that the corresponding enlargement vectors $\tilde{\lambda}_i$ are chosen from a finite set of extreme points, so there can be at most a finite number of strict enlargements. We assume that:

(a) All outer linearized functions f_i are real-valued and polyhedral, i.e., for all $i \in \underline{I}$, f_i is of the form

$$f_i(x_i) = \max_{\ell \in L_i} \{a'_{i\ell} x_i + b_{i\ell}\}$$

for some finite sets of vectors $\{a_{i\ell} \mid \ell \in L_i\}$ and scalars $\{b_{i\ell} \mid \ell \in L_i\}$.

(b) The conjugates f_i^* of all inner linearized functions are real-valued and polyhedral, i.e., for all $i \in \bar{I}$, f_i^* is of the form

$$f_i^*(\lambda_i) = \max_{\ell \in M_i} \{c'_{i\ell} \lambda_i + d_{i\ell}\}$$

for some finite sets of vectors $\{c_{i\ell} \mid \ell \in M_i\}$ and scalars $\{d_{i\ell} \mid \ell \in M_i\}$. (This condition is satisfied if and only if f_i is a polyhedral function with compact effective domain.)

(c) The vectors $\tilde{\lambda}_i$ and \tilde{x}_i added to the polyhedral approximations at each iteration correspond to the hyperplanes defining the corresponding functions f_i and f_i^* , i.e., $\tilde{\lambda}_i \in \{a_{i\ell} \mid \ell \in L_i\}$ and $\tilde{x}_i \in \{c_{i\ell} \mid \ell \in M_i\}$.

Let us also recall that in addition to the preceding conditions, we have assumed that the steps of the algorithm can be executed, and that in particular, a primal and dual optimal solution pair of problem (4.1) can be found at each iteration.

PROPOSITION 5.1 (Finite Termination in the Polyhedral Case). *Under the preceding polyhedral assumptions the GPA algorithm terminates after a finite number of iterations with a primal and dual optimal solution pair of problem (1.1).*

Proof. At each iteration there are two possibilities: either the algorithm terminates and by Prop. 4.1, $(\hat{x}, \hat{\lambda})$ is an optimal primal and dual pair for problem (1.1), or the approximation of one of the functions f_i , $i \in \underline{I} \cup \bar{I}$, will be refined/enlarged strictly. Since the vectors added to Λ_i , $i \in \underline{I}$, and X_i , $i \in \bar{I}$, belong to the finite sets $\{a_{i\ell} \mid \ell \in L_i\}$ and $\{c_{i\ell} \mid \ell \in M_i\}$, respectively, there can be only a finite number of strict enlargements, and convergence in a finite number of iterations follows. \square

5.1. Asymptotic Convergence Analysis: Pure Cases. We will now derive asymptotic convergence results for nonpolyhedral problem cases. We will first consider the cases of pure outer linearization and pure inner linearization, which are comparatively simple. We will subsequently discuss the mixed case, which is more complex.

PROPOSITION 5.2. *Consider the pure outer linearization case of the GPA algorithm ($\bar{I} = \emptyset$), and let \hat{x}^k be the solution of the approximate primal problem at the k th iteration, and $\tilde{\lambda}_i^k$, $i \in \underline{I}$, be the vectors generated at the corresponding enlargement step. Then if $\{\hat{x}^k\}_{\mathcal{K}}$ is a convergent subsequence such that the sequences $\{\tilde{\lambda}_i^k\}_{\mathcal{K}}$, $i \in \underline{I}$, are bounded, the limit of $\{\hat{x}^k\}_{\mathcal{K}}$ is primal optimal.*

Proof. For $i \in \underline{I}$, let $\underline{f}_{i, \Lambda_i^k}$ be the outer linearization of f_i at the k th iteration. For all $x \in S$, and k, ℓ with $\ell < k$, we have

$$\sum_{i \notin \underline{I}} f_i(\hat{x}_i^k) + \sum_{i \in \underline{I}} (f_i(\hat{x}_i^\ell) + (\hat{x}_i^k - \hat{x}_i^\ell)' \tilde{\lambda}_i^\ell) \leq \sum_{i \notin \underline{I}} f_i(\hat{x}_i^k) + \sum_{i \in \underline{I}} \underline{f}_{i, \Lambda_i^k}(\hat{x}_i^k) \leq \sum_{i=1}^{\ell} f_i(x_i),$$

where the first inequality follows from the definition of $\underline{f}_{i, \Lambda_i^k}$, and the second inequality follows from the optimality of \hat{x}^k for the k th approximate problem. Consider a subsequence $\{\hat{x}^k\}_{\mathcal{K}}$ that converges to $\bar{x} \in S$ and is such that the sequences $\{\tilde{\lambda}_i^k\}_{\mathcal{K}}$, $i \in \underline{I}$, are bounded. We will show that \bar{x} is optimal. Indeed, taking limit as $\ell \rightarrow \infty$, $k \in \mathcal{K}$, $\ell \in \mathcal{K}$, $\ell < k$, in the preceding relation, and using the closedness of f_i , which implies that

$$f_i(\bar{x}_i) \leq \liminf_{k \rightarrow \infty, k \in \mathcal{K}} f_i(\hat{x}_i^k), \quad \forall i,$$

we obtain that $\sum_{i=1}^m f_i(\bar{x}_i) \leq \sum_{i=1}^m f_i(x_i)$ for all $x \in S$, so \bar{x} is optimal. \square

Exchanging the roles of primal and dual, we obtain a convergence result for the pure inner linearization case.

PROPOSITION 5.3. *Consider the pure inner linearization case of the GPA algorithm ($\underline{I} = \emptyset$), and let $\hat{\lambda}^k$ be the solution of the approximate dual problem at the k th iteration, and \tilde{x}_i^k , $i \in \bar{I}$, be the vectors generated at the corresponding enlargement step. Then if $\{\hat{\lambda}^k\}_{\mathcal{K}}$ is a convergent subsequence such that the sequences $\{\tilde{x}_i^k\}_{\mathcal{K}}$, $i \in \bar{I}$, are bounded, the limit of $\{\hat{\lambda}^k\}_{\mathcal{K}}$ is dual optimal.*

5.2. Asymptotic Convergence Analysis: Mixed Case. We will next consider the mixed case, where some of the component functions are outer linearized while some others are inner linearized. We will establish a convergence result for GPA under some reasonable assumptions. We first show a general result about outer approximations of convex functions.

PROPOSITION 5.4. *Let $g : \mathfrak{R}^n \mapsto (-\infty, \infty]$ be a closed proper convex function, and let $\{\hat{x}^k\}$ and $\{\tilde{\lambda}^k\}$ be sequences such that $\tilde{\lambda}^k \in \partial g(\hat{x}^k)$ for all $k \geq 0$. Let $g_k, k \geq 1$, be outer approximations of g such that*

$$g(x) \geq g_k(x) \geq \max_{i=0, \dots, k-1} \{g(\hat{x}^i) + (x - \hat{x}^i)' \tilde{\lambda}^i\}, \quad \forall x \in \mathfrak{R}^n, \quad k = 1, \dots \quad (5.1)$$

Then if $\{\hat{x}^k\}_{\mathcal{K}}$ is a subsequence that converges to some \bar{x} with $\{\tilde{\lambda}^k\}_{\mathcal{K}}$ being bounded, we have

$$g(\bar{x}) = \lim_{k \rightarrow \infty, k \in \mathcal{K}} g(\hat{x}^k) = \lim_{k \rightarrow \infty, k \in \mathcal{K}} g_k(\hat{x}^k).$$

Proof. Since $\tilde{\lambda}^k \in \partial g(\hat{x}^k)$, we have

$$g(\hat{x}^k) + (\bar{x} - \hat{x}^k)' \tilde{\lambda}^k \leq g(\bar{x}), \quad k = 0, 1, \dots$$

Taking limsup of the left-hand side along \mathcal{K} , and using the boundedness of $\tilde{\lambda}^k, k \in \mathcal{K}$, we have

$$\limsup_{k \rightarrow \infty, k \in \mathcal{K}} g(\hat{x}^k) \leq g(\bar{x}),$$

and since by the closedness of g , we also have

$$\liminf_{k \rightarrow \infty, k \in \mathcal{K}} g(\hat{x}^k) \geq g(\bar{x}),$$

it follows that

$$g(\bar{x}) = \lim_{k \rightarrow \infty, k \in \mathcal{K}} g(\hat{x}^k). \quad (5.2)$$

Combining this equation with the fact $g_k \leq g$, we obtain

$$\limsup_{k \rightarrow \infty, k \in \mathcal{K}} g_k(\hat{x}^k) \leq \limsup_{k \rightarrow \infty, k \in \mathcal{K}} g(\hat{x}^k) = g(\bar{x}). \quad (5.3)$$

Using Eq. (5.1), we also have for any $k, \ell \in \mathcal{K}$ such that $k > \ell$,

$$g_k(\hat{x}^k) \geq g(\hat{x}^\ell) + (\hat{x}^k - \hat{x}^\ell)' \tilde{\lambda}^\ell.$$

By taking liminf of both sides along \mathcal{K} and using the boundedness of $\tilde{\lambda}^\ell, \ell \in \mathcal{K}$ and Eq. (5.2), we have

$$\liminf_{k \rightarrow \infty, k \in \mathcal{K}} g_k(\hat{x}^k) \geq \liminf_{\ell \rightarrow \infty, \ell \in \mathcal{K}} g(\hat{x}^\ell) = g(\bar{x}). \quad (5.4)$$

From Eqs. (5.3) and (5.4), we obtain $g(\bar{x}) = \lim_{k \rightarrow \infty, k \in \mathcal{K}} g_k(\hat{x}^k)$. \square

We now relate the optimal value and the solutions of an inner and outer-approximated problem to those of the original problem, and we characterize these relations in terms of the local function approximation errors of the approximate problem. This result will then be combined with the preceding proposition to establish asymptotic convergence of the GPA algorithm. For notational simplicity, let us consider just two

component functions g_1 and g_2 , with outer approximation of g_1 and inner approximation of g_2 . We denote by v^* the corresponding optimal value, and assume that there is no duality gap:

$$v^* = \inf_{(y_1, y_2) \in S} \{g_1(y_1) + g_2(y_2)\} = \sup_{(\mu_1, \mu_2) \in S^\perp} \{-g_1^*(\mu_1) - g_2^*(\mu_2)\}.$$

The analysis covers the case with more than two component functions, as will be seen shortly.

PROPOSITION 5.5. *Let v be the optimal value of an approximate problem*

$$\inf_{(y_1, y_2) \in S} \{\underline{g}_1(y_1) + \bar{g}_2(y_2)\},$$

where $\underline{g}_1 : \mathfrak{R}^{n_1} \mapsto (-\infty, \infty]$ and $\bar{g}_2 : \mathfrak{R}^{n_2} \mapsto (-\infty, \infty]$ are closed proper convex functions such that $\underline{g}_1(y_1) \leq g_1(y_1)$ for all y_1 and $g_2(y_2) \leq \bar{g}_2(y_2)$ for all y_2 . Assume that the approximate problem has no duality gap, and let $(\hat{y}_1, \hat{y}_2, \hat{\mu}_1, \hat{\mu}_2)$ be a primal and dual optimal solution pair. Then

$$\bar{g}_2^*(\hat{\mu}_2) - g_2^*(\hat{\mu}_2) \leq v^* - v \leq g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1), \quad (5.5)$$

and (\hat{y}_1, \hat{y}_2) and $(\hat{\mu}_1, \hat{\mu}_2)$ are ϵ -optimal for the original primal and dual problems, respectively, with

$$\epsilon = (g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1)) + (g_2^*(\hat{\mu}_2) - \bar{g}_2^*(\hat{\mu}_2)).$$

Proof. Since $(\hat{y}_1, \hat{y}_2) \in S$ and $(\hat{\mu}_1, \hat{\mu}_2) \in S^\perp$, we have

$$-g_1^*(\hat{\mu}_1) - g_2^*(\hat{\mu}_2) \leq v^* \leq g_1(\hat{y}_1) + g_2(\hat{y}_2).$$

Using $g_2 \leq \bar{g}_2$ and $g_1^* \leq \underline{g}_1^*$ (since $\underline{g}_1 \leq g_1$) as well as the optimality of $(\hat{y}_1, \hat{y}_2, \hat{\mu}_1, \hat{\mu}_2)$ for the approximate problem, we also have

$$\begin{aligned} g_1(\hat{y}_1) + g_2(\hat{y}_2) &\leq g_1(\hat{y}_1) + \bar{g}_2(\hat{y}_2) \\ &= \underline{g}_1(\hat{y}_1) + \bar{g}_2(\hat{y}_2) + g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1) \\ &= v + g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1), \\ -g_1^*(\hat{\mu}_1) - g_2^*(\hat{\mu}_2) &\geq -\underline{g}_1^*(\hat{\mu}_1) - g_2^*(\hat{\mu}_2) \\ &= -\underline{g}_1^*(\hat{\mu}_1) - \bar{g}_2^*(\hat{\mu}_2) + \bar{g}_2^*(\hat{\mu}_2) - g_2^*(\hat{\mu}_2) \\ &= v + \bar{g}_2^*(\hat{\mu}_2) - g_2^*(\hat{\mu}_2). \end{aligned}$$

Combining the preceding three relations, we obtain Eq. (5.5). Combining Eq. (5.5) with the last two relations, we obtain

$$\begin{aligned} g_1(\hat{y}_1) + g_2(\hat{y}_2) &\leq v^* + v - v^* + g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1) \\ &\leq v^* + (g_2^*(\hat{\mu}_2) - \bar{g}_2^*(\hat{\mu}_2)) + (g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1)), \\ -g_1^*(\hat{\mu}_1) - g_2^*(\hat{\mu}_2) &\geq v^* + v - v^* + \bar{g}_2^*(\hat{\mu}_2) - g_2^*(\hat{\mu}_2) \\ &\geq v^* - (g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1)) - (g_2^*(\hat{\mu}_2) - \bar{g}_2^*(\hat{\mu}_2)), \end{aligned}$$

which implies that (\hat{y}_1, \hat{y}_2) and $(\hat{\mu}_1, \hat{\mu}_2)$ are ϵ -optimal for the original primal and dual problems, respectively. \square

We now specialize the preceding proposition to deal with the GPA algorithm in the general case with multiple component functions, and with both inner and outer linearization. Let $y_1 = (x_i)_{i \in \underline{I}}$, $y_2 = (x_i)_{i \in \underline{I} \cup \bar{I}}$, and

$$g_1(y_1) = \sum_{i \in \underline{I}} f_i(x_i), \quad g_2(y_2) = \sum_{i \in \underline{I}} f_i(x_i) + \sum_{i \in \bar{I}} f_i(x_i). \quad (5.6)$$

For the dual variables, let $\mu_1 = (\lambda_i)_{i \in \underline{I}}$, $\mu_2 = (\lambda_i)_{i \in \underline{I} \cup \bar{I}}$. Then the original primal problem corresponds to $\inf_{(y_1, y_2) \in S} \{g_1(y_1) + g_2(y_2)\}$, and the dual problem corresponds to $\inf_{(\mu_1, \mu_2) \in S^\perp} \{g_1^*(\mu_1) + g_2^*(\mu_2)\}$.

Consider the approximate problem

$$\inf_{(x_1, \dots, x_m) \in S} \left\{ \sum_{i \in \underline{I}} \underline{f}_{i, \Lambda_i^k}(x_i) + \sum_{i \in \bar{I}} \bar{f}_{i, X_i^k}(x_i) + \sum_{i \in \underline{I}} f_i(x_i) \right\}$$

at the k th iteration of the GPA algorithm, where $\underline{f}_{i, \Lambda_i^k}$ and \bar{f}_{i, X_i^k} are the outer and inner linearizations of f_i for $i \in \underline{I}$ and $i \in \bar{I}$, respectively, at the k th iteration. We can write this problem as

$$\inf_{(y_1, y_2) \in S} \{ \underline{g}_{1,k}(y_1) + \bar{g}_{2,k}(y_2) \},$$

where

$$\underline{g}_{1,k}(y_1) = \sum_{i \in \underline{I}} \underline{f}_{i, \Lambda_i^k}(x_i), \quad \bar{g}_{2,k}(y_2) = \sum_{i \in \bar{I}} \bar{f}_{i, X_i^k}(x_i) + \sum_{i \in \underline{I}} f_i(x_i) \quad (5.7)$$

are outer and inner approximations of g_1 and g_2 , respectively. Let $(\hat{x}^k, \hat{\lambda}^k)$ be a primal and dual optimal solution pair of the approximate problem, and $(\hat{y}_1^k, \hat{y}_2^k, \hat{\mu}_1^k, \hat{\mu}_2^k)$ be the same pair expressed in terms of the components $y_i, \mu_i, i = 1, 2$. Then,

$$g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k) = \sum_{i \in \underline{I}} \left(f_i(\hat{x}_i^k) - \underline{f}_{i, \Lambda_i^k}(\hat{x}_i^k) \right), \quad (5.8)$$

$$\bar{g}_{2,k}^*(\hat{\mu}_2^k) - g_2^*(\hat{\mu}_2^k) = \sum_{i \in \bar{I}} \left((\bar{f}_{i, X_i^k})^*(\hat{\lambda}_i^k) - f_i^*(\hat{\lambda}_i^k) \right). \quad (5.9)$$

By Prop. 5.5, with v_k being the optimal value of the k th approximate problem and with $v^* = f_{opt}$, we have

$$\bar{g}_{2,k}^*(\hat{\mu}_2^k) - g_2^*(\hat{\mu}_2^k) \leq f_{opt} - v_k \leq g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k),$$

i.e.,

$$\sum_{i \in \bar{I}} \left((\bar{f}_{i, X_i^k})^*(\hat{\lambda}_i^k) - f_i^*(\hat{\lambda}_i^k) \right) \leq f_{opt} - v_k \leq \sum_{i \in \underline{I}} \left(f_i(\hat{x}_i^k) - \underline{f}_{i, \Lambda_i^k}(\hat{x}_i^k) \right), \quad (5.10)$$

and $(\hat{y}_1^k, \hat{y}_2^k)$ and $(\hat{\mu}_1^k, \hat{\mu}_2^k)$ (equivalently, \hat{x}^k and $\hat{\lambda}^k$) are ϵ_k -optimal for the original primal and dual problems, respectively, with

$$\begin{aligned} \epsilon_k &= (g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k)) + (g_2^*(\hat{\mu}_2^k) - \bar{g}_{2,k}^*(\hat{\mu}_2^k)) \\ &= \sum_{i \in \underline{I}} \left(f_i(\hat{x}_i^k) - \underline{f}_{i, \Lambda_i^k}(\hat{x}_i^k) \right) + \sum_{i \in \bar{I}} \left(f_i^*(\hat{\lambda}_i^k) - (\bar{f}_{i, X_i^k})^*(\hat{\lambda}_i^k) \right). \end{aligned} \quad (5.11)$$

Equations (5.10)-(5.11) show that for the approximate problem at an iteration of the GPA algorithm, the suboptimality of its solutions and the difference between its optimal value and f_{opt} can be bounded in terms of the function approximation errors at the solutions generated by the GPA algorithm.³

We will now derive an asymptotic convergence result for the GPA algorithm in the general case, by combining Eqs. (5.10)-(5.11) with properties of outer approximations and Prop. 5.4 in particular. Here we implicitly assume that primal and dual solutions of the approximate problems exist, and that the enlargement steps can be carried out.

PROPOSITION 5.6. *Consider the GPA algorithm. Let $(\hat{x}^k, \hat{\lambda}^k)$ be a primal and dual optimal solution pair of the approximate problem at the k th iteration, and let $\tilde{\lambda}_i^k, i \in \underline{I}$ and $\tilde{x}_i^k, i \in \bar{I}$ be the vectors generated at the corresponding enlargement step. Suppose that there exist convergent subsequences $\{\hat{x}_i^k\}_{\mathcal{K}}, i \in \underline{I}$, $\{\hat{\lambda}_i^k\}_{\mathcal{K}}, i \in \bar{I}$, such that the sequences $\{\tilde{\lambda}_i^k\}_{\mathcal{K}}, i \in \underline{I}$, $\{\tilde{x}_i^k\}_{\mathcal{K}}, i \in \bar{I}$, are bounded. Then:*

(a) *The subsequence $\{(\hat{x}^k, \hat{\lambda}^k)\}_{\mathcal{K}}$ is asymptotically optimal in the sense that*

$$\lim_{k \rightarrow \infty, k \in \mathcal{K}} \sum_{i=1}^m f_i(\hat{x}_i^k) = f_{opt}, \quad \lim_{k \rightarrow \infty, k \in \mathcal{K}} \sum_{i=1}^m f_i^*(\hat{\lambda}_i^k) = -f_{opt}.$$

In particular, any limit point of the sequence $\{(\hat{x}^k, \hat{\lambda}^k)\}_{\mathcal{K}}$ is a primal and dual optimal solution pair of the original problem.

(b) *The sequence of optimal values v_k of the approximate problems converges to the optimal value f_{opt} as $k \rightarrow \infty$.*

Proof. (a) We use the definitions of (y_1, y_2, μ_1, μ_2) , $(\hat{y}_1^k, \hat{y}_2^k, \hat{\mu}_1^k, \hat{\mu}_2^k)$, and $g_1, g_2, \bar{g}_{1,k}, \underline{g}_{2,k}$ as given in the discussion preceding the proposition. Let v_k be the optimal value of the k th approximate problem and let $v^* = f_{opt}$. As shown earlier, by Prop. 5.5, we have

$$\bar{g}_{2,k}^*(\hat{\mu}_2^k) - \underline{g}_{2,k}^*(\hat{\mu}_2^k) \leq v^* - v_k \leq g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k), \quad k = 0, 1, \dots, \quad (5.12)$$

and $(\hat{y}_1^k, \hat{y}_2^k)$ and $(\hat{\mu}_1^k, \hat{\mu}_2^k)$ are ϵ_k -optimal for the original primal and dual problems, respectively, with

$$\epsilon_k = (g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k)) + (g_2^*(\hat{\mu}_2^k) - \bar{g}_{2,k}^*(\hat{\mu}_2^k)). \quad (5.13)$$

³It is also insightful to express the error in approximating the conjugates, $f_i^*(\hat{\lambda}_i^k) - (\bar{f}_{i, X_i^k})^*(\hat{\lambda}_i^k)$, $i \in \bar{I}$, as the error in approximating the respective functions f_i . We have for $i \in \bar{I}$,

$$\bar{f}_{i, X_i^k}(\tilde{x}_i^k) + (\bar{f}_{i, X_i^k})^*(\hat{\lambda}_i^k) = \hat{\lambda}_i^{k'} \tilde{x}_i^k, \quad f_i(\tilde{x}_i^k) + f_i^*(\hat{\lambda}_i^k) = \hat{\lambda}_i^{k'} \tilde{x}_i^k,$$

where \tilde{x}_i^k is the enlargement vector at the k th iteration, so by subtracting the first relation from the second,

$$\begin{aligned} f_i^*(\hat{\lambda}_i^k) - (\bar{f}_{i, X_i^k})^*(\hat{\lambda}_i^k) &= \bar{f}_{i, X_i^k}(\tilde{x}_i^k) - \left(f_i(\tilde{x}_i^k) + (\hat{x}_i^k - \tilde{x}_i^k)' \hat{\lambda}_i^k \right) \\ &= \left(\bar{f}_{i, X_i^k}(\tilde{x}_i^k) - f_i(\tilde{x}_i^k) \right) + \left(f_i(\tilde{x}_i^k) - f_i(\hat{x}_i^k) - (\hat{x}_i^k - \tilde{x}_i^k)' \hat{\lambda}_i^k \right). \end{aligned}$$

The right-hand side involves the sum of two function approximation error terms at \tilde{x}_i^k : the first term is the inner linearization error, and the second term is the linearization error obtained by using $f_i(\tilde{x}_i^k)$ and the single subgradient $\hat{\lambda}_i^k \in \partial f_i(\tilde{x}_i^k)$. Thus the estimates of f_{opt} and ϵ_k in Eqs. (5.10) and (5.11) can be expressed solely in terms of the inner/outer approximation errors of f_i as well as the linearization errors at various points.

Under the stated assumptions, we have by Prop. 5.4,

$$\begin{aligned} \lim_{k \rightarrow \infty, k \in \mathcal{K}} \underline{f}_{i, \Lambda_i^k}(\hat{x}_i^k) &= \lim_{k \rightarrow \infty, k \in \mathcal{K}} f_i(\hat{x}_i^k), \quad i \in I, \\ \lim_{k \rightarrow \infty, k \in \mathcal{K}} (\bar{f}_{i, X_i^k})^*(\hat{\lambda}_i^k) &= \lim_{k \rightarrow \infty, k \in \mathcal{K}} f_i^*(\hat{\lambda}_i^k), \quad i \in \bar{I}, \end{aligned}$$

where we obtained the first relation by applying Prop. 5.4 to f_i and its outer linearizations $\underline{f}_{i, \Lambda_i^k}$, and the second relation by applying Prop. 5.4 to f_i^* and its outer linearizations $(\bar{f}_{i, X_i^k})^*$. Using the definitions of $g_1, g_2, \bar{g}_{1,k}, \underline{g}_{2,k}$ [cf. Eqs. (5.6)-(5.9)], this implies

$$\begin{aligned} \lim_{k \rightarrow \infty, k \in \mathcal{K}} \left(g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k) \right) &= \lim_{k \rightarrow \infty, k \in \mathcal{K}} \sum_{i \in I} \left(f_i(\hat{x}_i^k) - \underline{f}_{i, \Lambda_i^k}(\hat{x}_i^k) \right) = 0, \\ \lim_{k \rightarrow \infty, k \in \mathcal{K}} \left(g_2^*(\hat{\mu}_2^k) - \bar{g}_{2,k}^*(\hat{\mu}_2^k) \right) &= \lim_{k \rightarrow \infty, k \in \mathcal{K}} \sum_{i \in \bar{I}} \left(f_i^*(\hat{\lambda}_i^k) - (\bar{f}_{i, X_i^k})^*(\hat{\lambda}_i^k) \right) = 0, \end{aligned}$$

so from Eqs. (5.12) and (5.13),

$$\lim_{k \rightarrow \infty, k \in \mathcal{K}} v_k = v^*, \quad \lim_{k \rightarrow \infty, k \in \mathcal{K}} \epsilon_k = 0,$$

proving the first statement in part (a). This, combined with the closedness of the sets S, S^\perp and the functions f_i, f_i^* , implies the second statement in part (a).

(b) The preceding argument has shown that $\{v_k\}_{\mathcal{K}}$ converges to v^* , so there remains to show that the entire sequence $\{v_k\}$ converges to v^* . For any ℓ sufficiently large, let k be such that $k < \ell$ and $k \in \mathcal{K}$. We can view the approximate problem at the k th iteration as an approximate problem for the problem at the ℓ th iteration with $\bar{g}_{2,k}$ being an inner approximation of $\bar{g}_{2,\ell}$ and $\underline{g}_{1,k}$ being an outer approximation of $\underline{g}_{1,\ell}$. Then, by Prop. 5.5,

$$\bar{g}_{2,k}^*(\hat{\mu}_2^k) - \bar{g}_{2,\ell}^*(\hat{\mu}_2^k) \leq v_\ell - v_k \leq \underline{g}_{1,\ell}(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k).$$

Since $\lim_{k \rightarrow \infty, k \in \mathcal{K}} v_k = v^*$, to show that $\lim_{\ell \rightarrow \infty} v_\ell = v^*$, it is sufficient to show that

$$\lim_{\substack{k, \ell \rightarrow \infty, \\ k < \ell, k \in \mathcal{K}}} \left(\bar{g}_{2,\ell}^*(\hat{\mu}_2^k) - \bar{g}_{2,k}^*(\hat{\mu}_2^k) \right) = 0, \quad \lim_{\substack{k, \ell \rightarrow \infty, \\ k < \ell, k \in \mathcal{K}}} \left(\underline{g}_{1,\ell}(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k) \right) = 0. \quad (5.14)$$

Indeed, since $\bar{g}_{2,k}^* \leq \bar{g}_{2,\ell}^* \leq g_2^*$ for all k, ℓ with $k < \ell$, we have

$$0 \leq \bar{g}_{2,\ell}^*(\hat{\mu}_2^k) - \bar{g}_{2,k}^*(\hat{\mu}_2^k) \leq g_2^*(\hat{\mu}_2^k) - \bar{g}_{2,k}^*(\hat{\mu}_2^k),$$

and as shown earlier, by Prop. 5.4 we have under our assumptions

$$\lim_{k \rightarrow \infty, k \in \mathcal{K}} \left(g_2^*(\hat{\mu}_2^k) - \bar{g}_{2,k}^*(\hat{\mu}_2^k) \right) = 0.$$

Thus we obtain

$$\lim_{\substack{k, \ell \rightarrow \infty, \\ k < \ell, k \in \mathcal{K}}} \left(\bar{g}_{2,\ell}^*(\hat{\mu}_2^k) - \bar{g}_{2,k}^*(\hat{\mu}_2^k) \right) = 0,$$

which is the first relation in Eq. (5.14). The second relation in Eq. (5.14) follows with a similar argument. The proof is complete. \square

Proposition 5.6 implies in particular that if the sequences of generated cutting planes (or break points) for the outer (or inner, respectively) linearized functions are bounded, then every limit point of the generated sequence of the primal and dual optimal solution pairs of the approximate problems is an optimal primal and dual solution pair of the original problem.

Proposition 5.6 also implies that in the pure inner linearization case ($\underline{I} = \emptyset$), under the assumptions of the proposition, the sequence $\{\hat{x}^k\}$ is asymptotically optimal for the original primal problem, and in particular, any limit point of $\{\hat{x}^k\}$ is a primal optimal solution of the original problem. This is because by part (b) of the preceding proposition and the property of inner approximations:

$$\sum_{i \in \underline{I}} f_i(\hat{x}_i^k) + \sum_{i \in \bar{I}} f_i(\hat{x}_i^k) \leq \sum_{i \in \underline{I}} f_i(\hat{x}_i^k) + \sum_{i \in \bar{I}} \bar{f}_i(\hat{x}_i^k) = v_k \rightarrow f_{opt}, \quad \text{as } k \rightarrow \infty.$$

This strengthens the conclusion of Prop. 5.3. The conclusion of Prop. 5.2 can be similarly strengthened.

6. Special Cases. In this section we apply the GPA algorithm to various types of problems, and we show that when properly specialized, it yields the classical cutting plane and simplicial decomposition methods, as well as new nondifferentiable versions of simplicial decomposition. We will also indicate how in the special case of a monotropic programming problem, the GPA algorithm can offer substantial advantages over the classical methods.

6.1. Application to Classical Cutting Plane Methods. Consider the problem

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && x \in C, \end{aligned} \tag{6.1}$$

where $f : \mathfrak{R}^n \mapsto \mathfrak{R}$ is a real-valued convex function, and C is a closed convex set. It can be converted to the problem

$$\begin{aligned} & \text{minimize} && f_1(x_1) + f_2(x_2) \\ & \text{subject to} && (x_1, x_2) \in S, \end{aligned} \tag{6.2}$$

where

$$f_1 = f, \quad f_2 = \delta_C, \quad S = \{(x_1, x_2) \mid x_1 = x_2\}, \tag{6.3}$$

with δ_C being the indicator function of C . Note that both the original and the approximate problems have primal and dual solution pairs of the form $(\hat{x}, \hat{x}, \hat{\lambda}, -\hat{\lambda})$ [to satisfy the constraints $(x_1, x_2) \in S$ and $(\lambda_1, \lambda_2) \in S^\perp$].

One possibility is to apply the GPA algorithm to this formulation with outer linearization of f_1 and no inner linearization:

$$\underline{I} = \{1\}, \quad \bar{I} = \emptyset.$$

Using the notation of the original problem (6.1), at the typical iteration, we have a finite set of subgradients Λ of f and corresponding points $x_{\tilde{\lambda}}$ such that $\tilde{\lambda} \in \partial f(x_{\tilde{\lambda}})$ for each $\tilde{\lambda} \in \Lambda$. The approximate problem is equivalent to

$$\begin{aligned} & \text{minimize} && \underline{f}_\Lambda(x) \\ & \text{subject to} && x \in C, \end{aligned} \tag{6.4}$$

where

$$\underline{f}_\Lambda(x) = \max_{\tilde{\lambda} \in \Lambda} \{f(x_{\tilde{\lambda}}) + \tilde{\lambda}'(x - x_{\tilde{\lambda}})\}. \quad (6.5)$$

According to the GPA algorithm, if \hat{x} is an optimal solution of problem (6.4) [so that $(\hat{x}, \hat{\lambda})$ is an optimal solution of the approximate problem], we enlarge Λ by adding any $\tilde{\lambda}$ with $\tilde{\lambda} \in \partial f(\hat{x})$. The vector \hat{x} can also serve as the primal vector $x_{\tilde{\lambda}}$ that corresponds to the new dual vector $\tilde{\lambda}$ in the new outer linearization (6.5). We recognize this as the classical cutting plane method (see e.g., [Ber99], Section 6.3.3). Note that in this method it is not necessary to find a dual optimal solution $(\hat{\lambda}, -\hat{\lambda})$ of the approximate problem.

Another possibility that is useful when C either is nonpolyhedral, or is a complicated polyhedral set, can be obtained by outer-linearizing f and either outer- or inner-linearizing δ_C . For example, suppose we apply the GPA algorithm to the formulation (6.2)-(6.3) with

$$\underline{I} = \{1\}, \quad \bar{I} = \{2\}.$$

Then, using the notation of problem (6.1), at the typical iteration we have a finite set Λ of subgradients of f , corresponding points $x_{\tilde{\lambda}}$ such that $\tilde{\lambda} \in \partial f(x_{\tilde{\lambda}})$ for each $\tilde{\lambda} \in \Lambda$, and a finite set $X \subset C$. We then solve the polyhedral program

$$\begin{aligned} & \text{minimize} && \underline{f}_\Lambda(x) \\ & \text{subject to} && x \in \text{conv}(X), \end{aligned} \quad (6.6)$$

where $\underline{f}_\Lambda(x)$ is given by Eq. (6.5). The set Λ is enlarged by adding any $\tilde{\lambda}$ with $\tilde{\lambda} \in \partial f(\hat{x})$, where \hat{x} solves the polyhedral problem (6.6) [and can also serve as the primal vector that corresponds to the new dual vector $\tilde{\lambda}$ in the new outer linearization (6.5)]. The set X is enlarged by finding a dual optimal solution $(\hat{\lambda}, -\hat{\lambda})$, and by adding to X a vector \tilde{x} that satisfies $\tilde{x} \in \partial f_2^*(-\hat{\lambda})$, or equivalently, solves the problem

$$\begin{aligned} & \text{minimize} && \hat{\lambda}'x \\ & \text{subject to} && x \in C, \end{aligned}$$

[cf. Eq. (4.2)]. By Prop. 3.1, the vector $\hat{\lambda}$ must be such that $\hat{\lambda} \in \partial \underline{f}_\Lambda(\hat{x})$ and $-\hat{\lambda} \in \partial \bar{f}_{2,X}(\hat{x})$ [equivalently $-\hat{\lambda}$ must belong to the normal cone of the set $\text{conv}(X)$ at \hat{x} ; see [Ber09], p. 185]. It can be shown that one may find such $\hat{\lambda}$ while solving the polyhedral program (6.6) by using standard methods, e.g., the simplex method.

6.2. Generalized Simplicial Decomposition. We will now describe the application of the GPA algorithm with inner linearization to the problem

$$\begin{aligned} & \text{minimize} && f(x) + h(x) \\ & \text{subject to} && x \in \mathfrak{R}^n, \end{aligned} \quad (6.7)$$

where $f : \mathfrak{R}^n \mapsto (-\infty, \infty]$ and $h : \mathfrak{R}^n \mapsto (-\infty, \infty]$ are closed proper convex functions. This is a simplicial decomposition approach that descends from the original proposal of Holloway [Hol74] (see also [Hoh77]), where the function f is required to be real-valued and differentiable, and h is the indicator function of the closed convex set C . In addition to our standing assumption of no duality gap, we assume that $\text{dom}(h)$

contains a point in the relative interior of $\text{dom}(f)$; this guarantees that the problem is feasible and also ensures that some of the steps of the algorithm (to be described later) can be carried out.

A straightforward simplicial decomposition method that can deal with nondifferentiable cost functions is to apply the GPA algorithm with

$$f_1 = f, \quad f_2 = h, \quad S = \{(x, x) \mid x \in \mathbb{R}^n\},$$

while inner linearizing both functions f and h . The linearized approximating subproblem to be solved at each GPA iteration is a linear program whose primal and dual optimal solutions may be found by several alternative methods, including the simplex method. Let us also note that the case of a nondifferentiable real-valued convex function f and a polyhedral set C has been dealt with an approach different from ours, using concepts of ergodic sequences of subgradients and a conditional subgradient method by Larsson, Patriksson, and Stromberg (see [Str97, LPS98]).

In this section we will focus on the GPA algorithm for the case where the function h is inner linearized, while the function f is left intact. This is the case where simplicial decomposition has traditionally found important specialized applications, particularly with h being the indicator function of a closed convex set. As in Section 6.1, the primal and dual optimal solution pairs have the form $(\hat{x}, \hat{x}, \hat{\lambda}, -\hat{\lambda})$. We start with some finite set $X_0 \subset \text{dom}(h)$ such that X_0 contains a point in the relative interior of $\text{dom}(f)$, and $\partial h(\tilde{x}) \neq \emptyset$ for all $\tilde{x} \in X_0$. After k iterations, we have a finite set X_k such that $\partial h(\tilde{x}) \neq \emptyset$ for all $\tilde{x} \in X_k$, and we use the following three steps to compute vectors \tilde{x}^k and an enlarged set $X_{k+1} = X_k \cup \{\tilde{x}^k\}$ to start the next iteration (assuming the algorithm does not terminate):

- (1) **Solution of approximate primal problem:** We obtain

$$\hat{x}^k \in \arg \min_{x \in \mathbb{R}^n} \{f(x) + H_k(x)\}, \quad (6.8)$$

where H_k is the polyhedral/inner linearization function whose epigraph is the convex hull of the union of the halflines $\{(\tilde{x}, w) \mid h(\tilde{x}) \leq w\}$, $\tilde{x} \in X_k$. The existence of a solution \hat{x}^k of problem (6.8) is guaranteed by a variant of Weierstrass' Theorem ([Ber09], Prop. 3.2.1; the minimum of a closed proper convex function whose domain is bounded is attained), because $\text{dom}(H_k)$ is the convex hull of a finite set. The latter fact provides also the main motivation for simplicial decomposition: the vector x admits a relatively low-dimensional representation, which can be exploited to simplify the solution of problem (6.8). This is particularly so if f is real-valued and differentiable, but there are interesting cases where f is extended real-valued and nondifferentiable, as will be discussed later in this section.

- (2) **Solution of approximate dual problem:** We obtain a subgradient $\hat{\lambda}^k \in \partial f(\hat{x}^k)$ such that

$$-\hat{\lambda}^k \in \partial H_k(\hat{x}^k). \quad (6.9)$$

The existence of such a subgradient is guaranteed by standard optimality conditions, applied to the minimization in Eq. (6.8), since H_k is polyhedral and its domain, X_k , contains a point in the relative interior of the domain of f ; cf. [Ber09], Prop. 5.4.7(3). Note that by the optimality conditions (3.4)-(3.5) of Prop. 3.1, $(\hat{\lambda}^k, -\hat{\lambda}^k)$ is an optimal solution of the dual approximate problem.

(3) **Enlargement:** We obtain \tilde{x}^k such that

$$-\hat{\lambda}^k \in \partial h(\tilde{x}^k),$$

and form $X_{k+1} = X_k \cup \{\tilde{x}^k\}$.

Our earlier assumptions guarantee that steps (1) and (2) can be carried out. Regarding the enlargement step (3), we note that it is equivalent to finding

$$\tilde{x}^k \in \arg \min_{x \in \mathbb{R}^n} \{x' \hat{\lambda}^k + h(x)\}, \quad (6.10)$$

and that this is a linear program in the important special case where h is polyhedral. The existence of its solution must be guaranteed by some assumption, such as coercivity of h .

Let us first assume that f is real-valued and differentiable, and discuss a few special cases:

(a) When h is the indicator function of a bounded polyhedral set C and $X_0 = \{x^0\}$, we can show that the method reduces to the classical simplicial decomposition method [Hol74, Hoh77], which finds wide application in specialized problem settings, such as optimization of multicommodity flows (see e.g., [CaG74, FIH95, PaY84, LaP92]). At the typical iteration of the GPA algorithm, we have a finite set of points $X \subset C$. We then solve the problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in \text{conv}(X), \end{aligned} \quad (6.11)$$

[cf. step (1)]. If $(\hat{x}, \hat{\lambda}, -\hat{\lambda})$ is a corresponding optimal primal and dual solution pair, we enlarge X by adding to X any \tilde{x} with $-\hat{\lambda}$ in the normal cone of C at \tilde{x} [cf. step (3)]. This is equivalent to finding \tilde{x} that solves the optimization problem

$$\begin{aligned} &\text{minimize} && \hat{\lambda}'x \\ &\text{subject to} && x \in C, \end{aligned} \quad (6.12)$$

[cf. Eqs. (4.2) and (6.10); we assume that this problem has a solution, which is guaranteed if C is bounded]. The resulting method, illustrated in Fig. 6.1, is identical to the classical simplicial decomposition method, and terminates in a finite number of iterations.

(b) When h is a general closed proper convex function, the method is illustrated in Fig. 6.2. Since f is assumed differentiable, step (2) yields $\hat{\lambda}^k = \nabla f(\hat{x}^k)$. The method is closely related to the preceding/classical simplicial decomposition method (6.11)-(6.12) applied to the problem of minimizing $f(x) + w$ subject to $(x, w) \in \text{epi}(h)$. In the special case where h is a polyhedral function, it can be shown that the method terminates finitely, assuming that the vectors $(\tilde{x}^k, h(\tilde{x}^k))$ obtained by solving the corresponding linear program (6.10) are extreme points of $\text{epi}(h)$.

Generalized Simplicial Decomposition - Extended Real-Valued/Nondifferentiable Case. Let us now consider the general case of problem (6.7) where f is extended real-valued and nondifferentiable, and apply our simplicial decomposition algorithm, thereby obtaining a new method. Recall that the optimal primal and dual solution pair $(\hat{x}^k, \hat{\lambda}^k, -\hat{\lambda}^k)$ of problem (6.8) must satisfy $\hat{\lambda}^k \in \partial f(\hat{x}^k)$ and

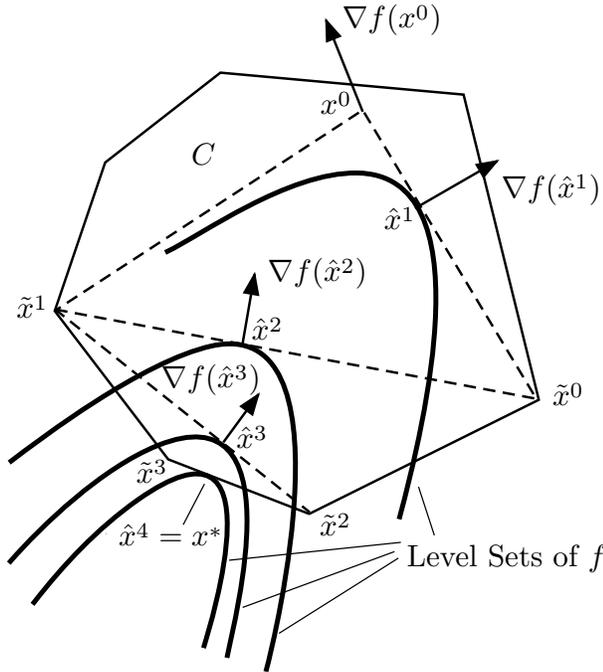


Fig. 6.1: Successive iterates of the classical simplicial decomposition method in the case where f is differentiable and C is polyhedral. For example, the figure shows how given the initial point x^0 and the calculated extreme points \tilde{x}^0 , \tilde{x}^1 , we determine the next iterate \hat{x}^2 as a minimizing point of f over the convex hull of $\{x^0, \tilde{x}^0, \tilde{x}^1\}$. At each iteration, a new extreme point of C is added, and after four iterations, the optimal solution is obtained.

$-\hat{\lambda}^k \in \partial H_k(\hat{x}^k)$ [cf. condition (6.9) of step (2)]. When h is the indicator function of a set C , the latter condition is equivalent to $-\hat{\lambda}^k$ being in the normal cone of $\text{conv}(X_k)$ at \hat{x}^k (cf. [Ber09], p. 185); see Fig. 6.3. If in addition C is polyhedral, the method terminates finitely, assuming that the vector \tilde{x}^k obtained by solving the linear program (6.10) is an extreme point of C (cf. Fig. 6.3). The reason is that in view of Eq. (6.9), the vector \tilde{x}^k does not belong to X_k (unless \hat{x}^k is optimal), so X_{k+1} is a strict enlargement of X_k . In the more general case where h is a closed proper convex function, the convergence of the method is covered by Prop. 5.3.

Let us now address the calculation of a subgradient $\hat{\lambda}^k \in \partial f(\hat{x}^k)$ such that $-\hat{\lambda}^k \in \partial H_k(\hat{x}^k)$ [cf. Eq. (6.9)]. This may be a difficult problem when f is nondifferentiable at \hat{x}^k , as it may require knowledge of $\partial f(\hat{x}^k)$ as well as $\partial H_k(\hat{x}^k)$. However, in special cases, $\hat{\lambda}^k$ may be obtained simply as a byproduct of the minimization (6.8). We discuss cases where h is the indicator of a closed convex set C , and the nondifferentiability and/or the domain of f are expressed in terms of differentiable functions.

Consider first the case where

$$f(x) = \max \{f_1(x), \dots, f_r(x)\},$$

where f_1, \dots, f_r are convex differentiable functions. Then the minimization (6.8)

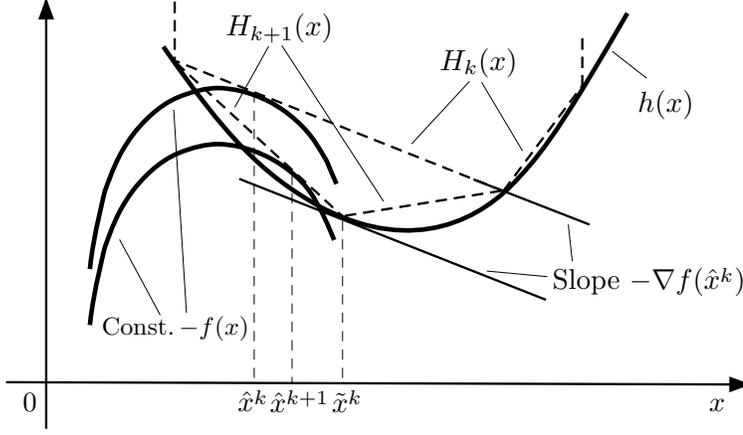


Fig. 6.2: Illustration of successive iterates of the generalized simplicial decomposition method in the case where f is differentiable. Given the inner linearization H_k of h , we minimize $f + H_k$ to obtain \hat{x}^k (graphically, we move the graph of $-f$ vertically until it touches the graph of H_k). We then compute \tilde{x}^k as a point at which $-\nabla f(\hat{x}^k)$ is a subgradient of h , and we use it to form the improved inner linearization H_{k+1} of h . Finally, we minimize $f + H_{k+1}$ to obtain \hat{x}^{k+1} (graphically, we move the graph of $-f$ vertically until it touches the graph of H_{k+1}).

takes the form

$$\begin{aligned} & \text{minimize } z & (6.13) \\ & \text{subject to } f_j(x) \leq z, \quad j = 1, \dots, r, \quad x \in \text{conv}(X_k). \end{aligned}$$

According to standard optimality conditions, the optimal solution (\hat{x}^k, z^*) together with dual optimal variables $\mu_j^* \geq 0$, satisfies the Lagrangian optimality condition

$$(\hat{x}^k, z^*) \in \arg \min_{x \in \text{conv}(X_k), z \in \mathbb{R}} \left\{ \left(1 - \sum_{j=1}^r \mu_j^* \right) z + \sum_{j=1}^r \mu_j^* f_j(x) \right\},$$

and the complementary slackness conditions $f_j(\hat{x}^k) = z^*$ if $\mu_j^* > 0$. Thus, since $z^* = f(\hat{x}^k)$, we must have

$$\sum_{j=1}^r \mu_j^* = 1, \quad \mu_j^* \geq 0, \quad \text{and } \mu_j^* > 0 \implies f_j(\hat{x}^k) = f(\hat{x}^k), \quad j = 1, \dots, r, \quad (6.14)$$

and

$$\left(\sum_{j=1}^r \mu_j^* \nabla f_j(\hat{x}^k) \right)' (x - \hat{x}^k) \geq 0, \quad \forall x \in \text{conv}(X_k). \quad (6.15)$$

From Eq. (6.14) it follows that the vector

$$\hat{\lambda}^k = \sum_{j=1}^r \mu_j^* \nabla f_j(\hat{x}^k) \quad (6.16)$$

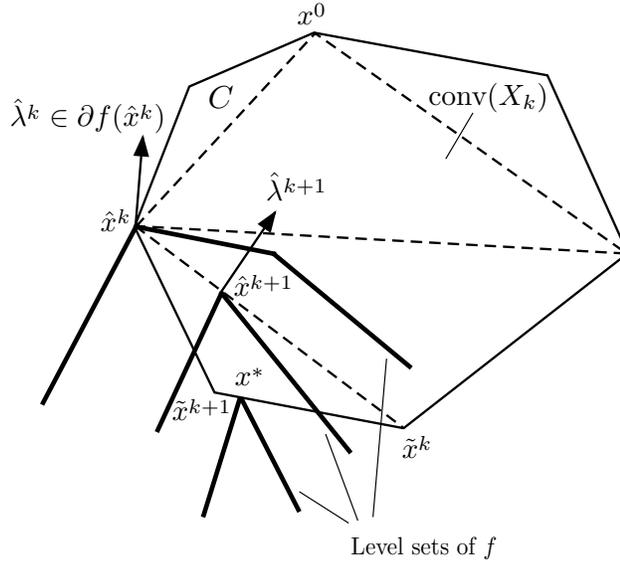


Fig. 6.3: Illustration of the generalized simplicial decomposition method for the case where f is nondifferentiable and h is the indicator function of a polyhedral set C . For each k , we compute a subgradient $\hat{\lambda}^k \in \partial f(\hat{x}^k)$ such that $-\hat{\lambda}^k$ lies in the normal cone of $\text{conv}(X_k)$ at \hat{x}^k , and we use it to generate a new extreme point \tilde{x}^k of C .

is a subgradient of f at \hat{x}^k (cf. [Ber09], p. 199). Furthermore, from Eq. (6.15), it follows that $-\hat{\lambda}^k$ is in the normal cone of $\text{conv}(X_k)$ at \hat{x}^k , so $-\hat{\lambda}^k \in \partial H_k(\hat{x}^k)$ as required by Eq. (6.9).

In conclusion, $\hat{\lambda}^k$ as given by Eq. (6.16), is such that $(\hat{x}^k, \hat{x}^k, \hat{\lambda}^k, -\hat{\lambda}^k)$ is an optimal primal and dual solution pair of the approximating problem (6.8), and furthermore, it is a suitable subgradient of f at \hat{x}^k for determining a new extreme point \tilde{x}^k via problem (6.10) or equivalently problem (6.12).

We next consider a more general problem where there are additional inequality constraints defining the domain of f . This is the case where f is of the form

$$f(x) = \begin{cases} \max \{f_1(x), \dots, f_r(x)\}, & \text{if } g_i(x) \leq 0, \quad i = 1, \dots, p, \\ \infty & \text{otherwise,} \end{cases} \quad (6.17)$$

with f_j and g_i being convex differentiable functions. Applications of this type include multicommodity flow problems with side constraints (the inequalities $g_i(x) \leq 0$, which are separate from the network flow constraints that comprise the set C ; cf. [Ber98], Chapter 8, [LaP99]). The case where $r = 1$ and there are no side constraints is important in a variety of communication, transportation, and other resource allocation problems, and is one of the principal successful applications of simplicial decomposition; see e.g., [FIH95]. Side constraints and nondifferentiabilities in this context are often eliminated using barrier, penalty, or augmented Lagrangian functions, but this can be awkward and restrictive. Our approach allows a more direct treatment.

As in the preceding case, we introduce additional dual variables $\nu_i^* \geq 0$ for the constraints $g_i(x) \leq 0$, and we write the Lagrangian optimality and complementary

slackness conditions. Then Eq. (6.15) takes the form

$$\left(\sum_{j=1}^r \mu_j^* \nabla f_j(\hat{x}^k) + \sum_{i=1}^p \nu_i^* \nabla g_i(\hat{x}^k) \right)' (x - \hat{x}^k) \geq 0, \quad \forall x \in \text{conv}(X_k),$$

and it can be shown that the vector $\hat{\lambda}^k = \sum_{j=1}^r \mu_j^* \nabla f_j(\hat{x}^k) + \sum_{i=1}^p \nu_i^* \nabla g_i(\hat{x}^k)$ is a subgradient of f at \hat{x}^k , while $-\hat{\lambda}^k \in \partial H_k(\hat{x}^k)$ as required by Eq. (6.9).

Note an important advantage that our method has over potential competitors in the case where C is polyhedral: it involves a solution of linear programs of the form (6.10), to generate new extreme points of C , and a solution of typically low-dimensional nonlinear programs, such as (6.13) and its more general version for the case (6.17). The latter programs have low dimension as long as the set X_k has a relatively small number of points. When all the functions f_j and g_i are twice differentiable, these programs can be solved by fast Newton-like methods, such as sequential quadratic programming (see e.g., [Ber82, Ber99, NoW99]). We finally note that as k increases, it is natural to apply schemes for dropping points of X_k to bound its cardinality, similar to the restricted simplicial decomposition method [HLV87, VeH93]. Such extensions of the algorithm are currently under investigation.

6.3. Dual/Cutting Plane Implementation. We now provide a dual implementation of the preceding generalized simplicial decomposition method, as applied to problem (6.7). It yields an outer linearization/cutting plane-type of method, which is mathematically equivalent to generalized simplicial decomposition. The dual problem is

$$\begin{aligned} & \text{minimize} && f_1^*(\lambda) + f_2^*(-\lambda) \\ & \text{subject to} && \lambda \in \mathfrak{R}^n, \end{aligned}$$

where f_1^* and f_2^* are the conjugates of f and h , respectively. The generalized simplicial decomposition algorithm (6.8)-(6.10) can alternatively be implemented by replacing f_2^* by a piecewise linear/cutting plane outer linearization, while leaving f_1^* unchanged, i.e., by solving at iteration k the problem

$$\begin{aligned} & \text{minimize} && f_1^*(\lambda) + (\bar{f}_{2,X_k})^*(-\lambda) \\ & \text{subject to} && \lambda \in \mathfrak{R}^n, \end{aligned} \tag{6.18}$$

where $(\bar{f}_{2,X_k})^*$ is an outer linearization of f_2^* (the conjugate of H_k).

Note that if $\hat{\lambda}^k$ is a solution of problem (6.18), the vector \tilde{x}^k generated by the enlargement step (6.10) is a subgradient of $f_2^*(\cdot)$ at $-\hat{\lambda}^k$, or equivalently $-\tilde{x}^k$ is a subgradient of the function $f_2^*(-\cdot)$ at $\hat{\lambda}^k$, as shown in Fig. 6.4. The ordinary cutting plane method, described in the beginning of Section 6.1, is obtained as the special case where $f_2^*(-\cdot)$ is the function to be outer linearized, and $f_1^*(\cdot)$ is the indicator function of C [so $f_1^*(\lambda) \equiv 0$ if $C = \mathfrak{R}^n$].

Whether the primal implementation, based on solution of problem (6.8), or the dual implementation, based on solution of problem (6.18), is preferable depends on the structure of the functions f and h . When f (and hence also f_1^*) is not polyhedral, the dual implementation may not be attractive, because it requires the n -dimensional nonlinear optimization (6.18) at each iteration, as opposed to the typically low-dimensional optimization (6.8). In the alternative case where f is polyhedral, both methods require the solution of linear programs.

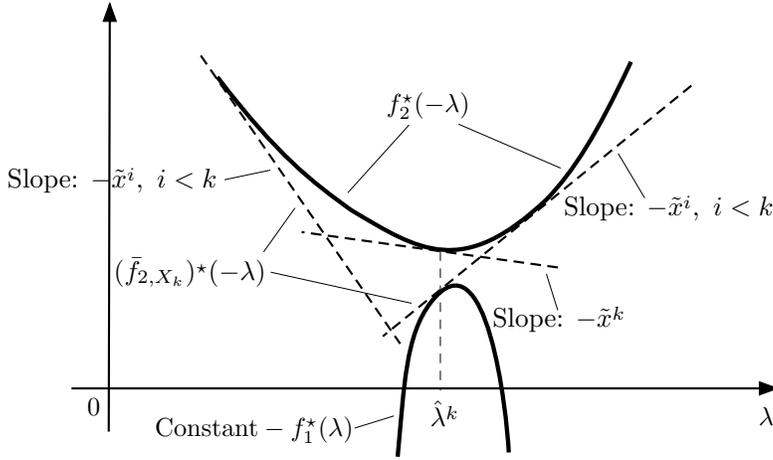


Fig. 6.4: Illustration of the cutting plane implementation of the generalized simplicial decomposition method.

6.4. Network Optimization and Monotropic Programming. Consider a directed graph with set of nodes \mathcal{N} and set of arcs \mathcal{A} . The single commodity network flow problem is to minimize a cost function

$$\sum_{a \in \mathcal{A}} f_a(x_a),$$

where f_a is a scalar closed proper convex function, and x_a is the flow of arc $a \in \mathcal{A}$. The minimization is over all flow vectors $x = \{x_a \mid a \in \mathcal{A}\}$ that belong to the circulation subspace S of the graph (sum of all incoming arc flows at each node is equal to the sum of all outgoing arc flows). This is a monotropic program that has been studied in many works, including the textbooks [Roc84] and [Ber98].

The GPA method that uses inner linearization of all the functions f_a that are nonlinear is attractive relative to the classical cutting plane and simplicial decomposition methods, because of the favorable structure of the corresponding approximate problem:

$$\begin{aligned} & \text{minimize} && \sum_{a \in \mathcal{A}} \bar{f}_{a, X_a}(x_a) \\ & \text{subject to} && x \in S, \end{aligned}$$

where for each arc a , \bar{f}_{a, X_a} is the inner approximation of f_a , corresponding to a finite set of break points $X_a \subset \text{dom}(f_a)$. By suitably introducing multiple arcs in place of each arc, we can recast this problem as a linear minimum cost network flow problem that can be solved using very fast polynomial algorithms. These algorithms, simultaneously with an optimal primal (flow) vector, yield a dual optimal (price differential) vector (see e.g., [Ber98], Chapters 5-7). Furthermore, because the functions f_a are scalar, the enlargement step is very simple.

Some of the preceding advantages of GPA method with inner linearization carry over to general monotropic programming problems ($n_i = 1$ for all i), the key idea being that the enlargement step is typically very simple. Furthermore, there are effective

algorithms for solving the associated approximate primal and dual problems, such as out-of-kilter methods [Roc84, Tse01], and ϵ -relaxation methods [Ber98, TsB00].

7. Conclusions. We have presented a unifying framework for polyhedral approximation in convex optimization. From a theoretical point of view, the framework allows the coexistence of inner and outer approximation as dual operations within the approximation process. From a practical point of view, the framework allows flexibility in adapting the approximation process to the special structure of the problem. Several specially-structured classes of problems have been identified where our methodology extends substantially the classical polyhedral approximation algorithms, including simplicial decomposition methods for extended real-valued and/or nondifferentiable cost functions, and nonlinear convex single-commodity network flow problems. In our methods, there is no provision for dropping cutting planes and break points from the current approximation. Schemes that can do this efficiently have been proposed for classical methods (see, e.g., [GoP79, Mey79, HLV87, VeH93]), and their extensions to our framework is an important subject for further research.

REFERENCES

- [AM093] AHUJA, K., MAGNANTI, T. L., ORLIN, J. B., *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall, Englewood Cliffs, N. J., 1993.
- [BGL09] BONNANS, J. F., GILBERT, J. C., LEMARÉCHAL, C., AND SAGASTIZÁBAL, S. C., *Numerical Optimization: Theoretical and Practical Aspects*, Springer, N. Y., 2009.
- [Ber82] BERTSEKAS, D. P., *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, N. Y., 1982; republished by Athena Scientific, Belmont, MA, 1997.
- [Ber98] BERTSEKAS, D. P., *Network Optimization: Continuous and Discrete Models*, Athena Scientific, Belmont, MA, 1998.
- [Ber99] BERTSEKAS, D. P., *Nonlinear Programming*, 2nd Edition, Athena Scientific, Belmont, MA, 1999.
- [Ber09] BERTSEKAS, D. P., *Convex Optimization Theory*, Athena Scientific, Belmont, MA, 2009.
- [Ber10] BERTSEKAS, D. P., *Extended Monotropic Programming and Duality*, Lab. for Information and Decision Systems Report 2692, MIT, 2010; a version appeared in JOTA, 139 (2008), pp. 209-225.
- [CaG74] CANTOR, D. G., AND GERLA, M., *Optimal Routing in a Packet Switched Computer Network*, IEEE Trans. on Computers, 23 (1974), pp. 1062-1069.
- [FlH95] FLORIAN, M. S., AND HEARN, D., *Network Equilibrium Models and Algorithms*, Handbooks in OR and MS, Ball, M. O., Magnanti, T. L., Monma, C. L., and Nemhauser, G. L., (Eds.), Vol. 8, North-Holland, Amsterdam, 1995, pp. 485-550.
- [GoP79] GONZAGA, C., AND POLAK, E., *On Constraint Dropping Schemes and Optimality Functions for a Class of Outer Approximations Algorithms*, SIAM J. on Control and Optimization, 17 (1979), pp.477-493.
- [HLV87] HEARN, D. W., LAWPHONGPANICH, S., AND VENTURA, J. A., *Restricted Simplicial Decomposition: Computation and Extensions*, Math. Programming Studies, 31 (1987), pp. 119-136.
- [HiL93] HIRIART-URRUTY, J.-B., AND LEMARECHAL, C., *Convex Analysis and Minimization Algorithms*, Vols. I and II, Springer-Verlag, Berlin and N. Y., 1993.
- [Hoh77] HOHENBALKEN, B. VON, *Simplicial Decomposition in Nonlinear Programming*, Math. Programming, 13 (1977), pp. 49-68.
- [Hol74] HOLLOWAY, C. A., *An Extension of the Frank and Wolfe Method of Feasible Directions*, Math. Programming, 6 (1974), pp. 14-27.
- [LPS98] LARSSON, T., PATRICKSSON, M., AND STROMBERG, A.-B., *Ergodic Convergence in Subgradient Optimization*, Optimization Methods and Software, 9 (1998), pp. 93-120.
- [LaP92] LARSSON, T., AND PATRICKSSON, M., *Simplicial Decomposition with Disaggregated Representation for the Traffic Assignment Problem*, Transportation Science, 26 (1982), pp. 4-17.
- [LaP99] LARSSON, T., AND PATRICKSSON, M., *Side Constrained Traffic Equilibrium Models - Analysis, Computation and Applications*, Transportation Research, 33 (1999), pp. 233-264.

- [Mey79] MEYER, R. R., *Two-Segment Separable Programming*, Management Science, 29 (1979), pp. 385-395.
- [NoW99] NOCEDAL, J., AND WRIGHT, S. J., *Numerical Optimization*, Springer, N. Y., 1999.
- [PaY84] PANG, J.-S. AND YU, C.-S., *Linearized Simplicial Decomposition Methods for Computing Traffic Equilibria on Networks*, Networks, 14 (1984), pp. 427-438.
- [Pol97] POLAK, E., *Optimization: Algorithms and Consistent Approximations*, Springer-Verlag, N. Y., 1997.
- [Roc70] ROCKAFELLAR, R. T., *Convex Analysis*, Princeton Univ. Press, Princeton, N. J., 1970.
- [Roc84] ROCKAFELLAR, R. T., *Network Flows and Monotropic Optimization*, Wiley, N. Y., 1984; republished by Athena Scientific, Belmont, MA, 1998.
- [Str97] STROMBERG, A.-B., *Conditional Subgradient Methods and Ergodic Convergence in Non-smooth Optimization*, Ph.D. Thesis, Univ. of Linkoping, Sweden, 1997.
- [TsB00] TSENG, P., AND BERTSEKAS, D. P., *An Epsilon-Relaxation Method for Separable Convex Cost Generalized Network Flow Problems*, Math. Programming, 88 (2000), pp. 85-104.
- [Tse01] TSENG, P., *An Epsilon Out-of-Kilter Method for Monotropic Programming*, Math. Oper. Res., 26 (2001), pp. 221-233.
- [VeH93] VENTURA, J. A., AND HEARN, D. W., *Restricted Simplicial Decomposition for Convex Constrained Problems*, Math. Programming, 59 (1993), pp. 71-85.