September 2009 **Report LIDS - 2820**

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 and error bounds for the general case where outer and inner linearization are combined in the same algorithm.

† 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.

‡ Huizhen Yu is with the Dept. of Computer Science and the Helsinki Institute for Information Technology (HIIT), Univ. of Helsinki, Finland.
Her research was supported in part by the Academy of Finland grant 118653 (ALGODAN).

1. INTRODUCTION

We consider the problem

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

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

$f_i : \Re^{n_i} \mapsto (-\infty,\infty]$ is a closed proper convex function for each i,†

S is a subspace of $\Re^{n_1+\cdots+n_m}$.

This problem has been studied recently by the first author in [Ber08], 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).

Note that the problem

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

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

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

where S is the subspace

$$S = \big\{(x,\dots,x) \mid x \in X\big\}.$$

A related case is the problem arising in the Fenchel duality framework,

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

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

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

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

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

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

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

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

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

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 scalar, 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 scalar, or have partially overlapping dependences, e.g.,

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

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 approximation of the components 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.

The paper is organized as follows. In Sections 2 and 3 we briefly review background material: duality (Section 2), and the conjugacy correspondence between outer and inner linearizations (Section 3). 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.

2. DUALITY

In this section we review some aspects of the duality theory associated with problem (1.1). For more details, including conditions that guarantee strong duality, we refer to [Ber08]. To derive the appropriate dual problem, we introduce auxiliary vectors $z_i \in \Re^{n_i}$ and we convert problem (1.1) to the equivalent form

$$\begin{aligned}&\text{minimize} \quad \sum_{i=1}^m f_i(z_i)\\ &\text{subject to} \quad z_i = x_i,\quad i=1,\ldots,m,\qquad (x_1,\ldots,x_m)\in S.\end{aligned} \tag{2.1}$$

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

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

The dual function is

$$
\begin{aligned}
q(\lambda) &= \inf_{(x_1,\ldots,x_m)\in S,\, z_i\in\Re^{n_i}} L(x_1,\ldots,x_m,z_1,\ldots,z_m,\lambda_1,\ldots,\lambda_m)\\
&= \inf_{(x_1,\ldots,x_m)\in S} \sum_{i=1}^{m} \lambda_i x_i + \sum_{i=1}^{m} \inf_{z_i\in\Re^{n_i}} \big\{ f_i(z_i) - \lambda_i' z_i \big\} \\
&= \begin{cases} \sum_{i=1}^{m} q_i(\lambda_i) & \text{if } (\lambda_1,\ldots,\lambda_m)\in S^{\perp},\\ -\infty & \text{otherwise,} \end{cases}
\end{aligned}
\tag{2.2}
$$

where

$$
q_i(\lambda_i) = \inf_{z_i\in\Re^{n_i}} \big\{ f_i(z_i) - \lambda_i' z_i \big\}, \qquad i = 1,\ldots,m,
$$

and $S^{\perp}$ is the orthogonal subspace of S.

The dual problem is

$$
\begin{aligned}
&\text{maximize} && \sum_{i=1}^{m} q_i(\lambda_i)\\
&\text{subject to} && (\lambda_1,\ldots,\lambda_m)\in S^{\perp}.
\end{aligned}
$$

Note that q_i can be written as

$$
q_i(\lambda_i) = -\sup_{z_i\in\Re^{n_i}} \big\{ \lambda_i' z_i - f_i(z_i) \big\},
$$

so $-q_i$ is equal to $f_i^{\star}$, the conjugate of f_i. Thus, with a change of sign to convert maximization to minimization, the dual problem becomes

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

and has the same form as the primal. Furthermore, since the functions f_i are assumed closed proper and convex, we have $f_i^{\star\star} = f_i$, where $f_i^{\star\star}$ is the conjugate of $f_i^{\star}$ (see e.g., [Ber09], Prop. 1.6.1), so when the dual problem is dualized, it yields the primal problem, and the duality is fully symmetric.

We denote by f_{opt} and q_{opt} the optimal primal and dual values, and we assume that strong duality holds ($q_{opt} = f_{opt}$). By viewing the equivalent problem (2.1) as a convex programming problem with equality constraints, it can be seen that 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

$$
q_i(\lambda_i^{opt}) = \inf_{x_i\in\Re^{n_i}} \big\{ f_i(x_i) - x_i' \lambda_i^{opt} \big\}, \qquad i = 1,\ldots,m;
$$

cf. Eq. (2.2). We thus obtain the following.

Proposition 2.1: (Optimality Conditions) We have $-\infty < q_{opt} = f_{opt} < \infty$, and $x^{opt} = (x_1^{opt}, \ldots, x_m^{opt})$ and $\lambda^{opt} = (\lambda_1^{opt}, \ldots, \lambda_m^{opt})$ are optimal primal and dual solutions, respectively, of problem (1.1) if and only if

$$x^{opt} \in S, \qquad \lambda^{opt} \in S^{\perp}, \qquad x_i^{opt} \in \arg\min_{x_i \in \Re^n} \big\{ f_i(x_i) - x_i' \lambda_i^{opt} \big\}, \quad i = 1, \ldots, m. \tag{2.3}$$

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

$$\lambda_i^{opt} \in \partial f_i(x_i^{opt}), \qquad x_i^{opt} \in \partial f_i^{\star}(\lambda_i^{opt}), \tag{2.4}$$

where we generically use $\partial f(x)$ to denote the subdifferential of a function f at a vector x.

3. CONJUGACY OF INNER AND OUTER LINEARIZATIONS

We will now review the well-known conjugacy relation between outer and inner linearization, as a first step towards a general polyhedral approximation framework. Note that from the definition of the conjugate, it follows that for any closed convex functions $f : \Re^n \mapsto (-\infty, \infty]$ and $\underline{f} : \Re^n \mapsto (-\infty, \infty]$, with $\underline{f}(x) \leq f(x)$ for all $x \in \Re^n$, we have $f^{\star}(\lambda) \leq \underline{f}^{\star}(\lambda)$ for all $\lambda \in \Re^n$. Thus if the function f is approximated by a lower bounding function $\underline{f}$ in the primal problem, its conjugate $f^{\star}$ is approximated by the upper bounding function $\underline{f}^{\star}$ in the dual problem. What we will show below is the particular way in which an outer linearization of f corresponds to an inner linearization of the conjugate $f^{\star}$ and reversely.

Consider an outer linearization of the epigraph of f defined by a finite set of vectors Λ and corresponding hyperplanes that support the epigraph of f at points $x_{\tilde{\lambda}}$ such that $\tilde{\lambda} \in \partial f(x_{\tilde{\lambda}})$ for each $\tilde{\lambda} \in \Lambda$. It is given by

$$F(x) = \max_{\tilde{\lambda} \in \Lambda} \big\{ f(x_{\tilde{\lambda}}) + (x - x_{\tilde{\lambda}})' \tilde{\lambda} \big\}, \qquad x \in \Re^n; \tag{3.1}$$

cf. Fig. 3.1. We will verify that the conjugate $F^{\star}$ of the outer linearization F can be described as an inner linearization of the conjugate $f^{\star}$ of f.

Indeed, we have

$$\begin{aligned} F^{\star}(\lambda) &= \sup_{x \in \Re^n} \big\{ \lambda' x - F(x) \big\} \\ &= \sup_{x \in \Re^n} \left\{ \lambda' x - \max_{\tilde{\lambda} \in \Lambda} \big\{ f(x_{\tilde{\lambda}}) + (x - x_{\tilde{\lambda}})' \tilde{\lambda} \big\} \right\} \\ &= \sup_{\substack{x \in \Re^n,\, \xi \in \Re \\ f(x_{\tilde{\lambda}}) + (x - x_{\tilde{\lambda}})' \tilde{\lambda} \leq \xi,\ \tilde{\lambda} \in \Lambda}} \{ \lambda' x - \xi \}. \end{aligned}$$

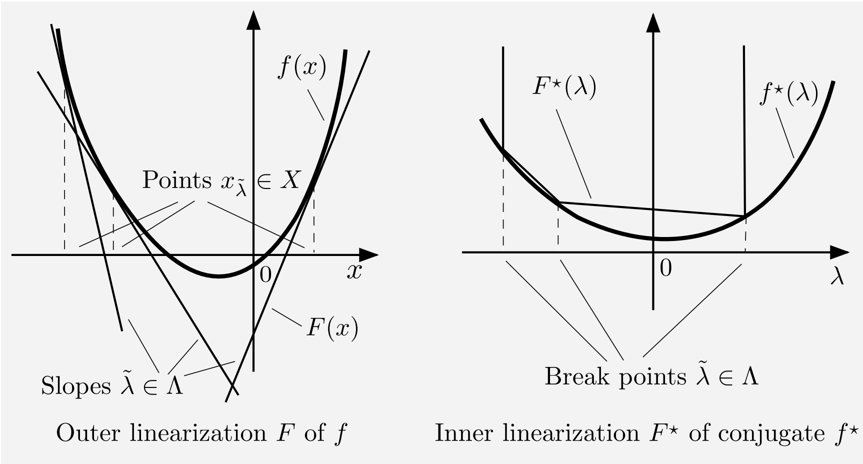

Figure 3.1. Illustration of the conjugate $F^\star$ of an outer linearization F of a convex function f defined by a finite set of "slopes" Λ and corresponding set X, consisting of points $x_{\tilde{\lambda}}$ such that $\tilde{\lambda} \in \partial f(x_{\tilde{\lambda}})$, $\tilde{\lambda} \in \Lambda$. It is a piecewise linear, inner linearization of the conjugate $f^\star$ of f. Its break points are the "slopes" $\tilde{\lambda}$ of the supporting planes $\tilde{\lambda} \in \Lambda$.

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

$$F^\star(\lambda) = \inf_{\substack{\sum_{\tilde{\lambda}\in\Lambda} \alpha_{\tilde{\lambda}}\tilde{\lambda}=\lambda,\ \sum_{\tilde{\lambda}\in\Lambda} \alpha_{\tilde{\lambda}}=1 \\ \alpha_{\tilde{\lambda}}\geq 0,\ \tilde{\lambda}\in\Lambda}} \sum_{\tilde{\lambda}\in\Lambda} \alpha_{\tilde{\lambda}}\big(f(x_{\tilde{\lambda}}) - x_{\tilde{\lambda}}'\tilde{\lambda}\big),$$

where $\alpha_{\tilde{\lambda}}$ is the dual variable of the constraint $f(x_{\tilde{\lambda}}) + (x - x_{\tilde{\lambda}})'\tilde{\lambda} \leq \xi$. Since the hyperplanes defining F are supporting $\text{epi}(f)$, we have

$$x_{\tilde{\lambda}}'\tilde{\lambda} - f(x_{\tilde{\lambda}}) = f^\star(\tilde{\lambda}), \qquad \tilde{\lambda} \in \Lambda,$$

so we obtain

$$F^\star(\lambda) = \begin{cases} \inf_{\substack{\sum_{\tilde{\lambda}\in\Lambda} \alpha_{\tilde{\lambda}}\tilde{\lambda}=\lambda,\ \sum_{\tilde{\lambda}\in\Lambda} \alpha_{\tilde{\lambda}}=1 \\ \alpha_{\tilde{\lambda}}\geq 0,\ \tilde{\lambda}\in\Lambda}} \sum_{\tilde{\lambda}\in\Lambda} \alpha_{\tilde{\lambda}} f^\star(\tilde{\lambda}) & \text{if } \lambda \in \text{conv}(\Lambda), \\ \infty & \text{otherwise.} \end{cases} \tag{3.2}$$

Thus, $F^\star$ is a piecewise linear (inner) linearization of $f^\star$ with domain

$$\text{dom}(F^\star) = \text{conv}(\Lambda),$$

and "break points" at $\tilde{\lambda} \in \Lambda$ with values equal to the corresponding values of $f^\star$. In particular, the epigraph of $F^\star$ is the convex hull of the union of the vertical halflines corresponding to $\tilde{\lambda} \in \Lambda$:

$$\text{epi}(F^\star) = \text{conv}\Big(\big\{\{(\tilde{\lambda}, w_{\tilde{\lambda}}) \mid f^\star(\tilde{\lambda}) \leq w_{\tilde{\lambda}}\} \mid \tilde{\lambda} \in \Lambda\big\}\Big)$$

(see Fig. 3.1).

Note that the inner linearization $F^\star$ is determined by the set Λ and is independent of the points $x_{\tilde\lambda}$, $\tilde\lambda \in \Lambda$ (which are not uniquely defined). This indicates that the same is true for its conjugate F, and indeed, since

$$f(x_{\tilde\lambda}) - \tilde\lambda' x_{\tilde\lambda} = -f^\star(\tilde\lambda),$$

from Eq. (3.1) we obtain

$$F(x) = \max_{\tilde\lambda \in \Lambda} \big\{ \tilde\lambda' x - f^\star(\tilde\lambda) \big\}. \tag{3.3}$$

However, not every function of the above form qualifies as an outer linearization within our framework: it is necessary that for every $\tilde\lambda \in \Lambda$ there exists $x_{\tilde\lambda}$ such that $\tilde\lambda \in \partial f(x_{\tilde\lambda})$, or equivalently that $\partial f^\star(\tilde\lambda) \neq \emptyset$ for all $\tilde\lambda \in \Lambda$. Similarly, not every function of the form (3.2) qualifies as an inner linearization within our framework: it is necessary that $\partial f^\star(\tilde\lambda) \neq \emptyset$ for all $\tilde\lambda \in \Lambda$.

4. GENERALIZED POLYHEDRAL APPROXIMATION

We will now describe an algorithm 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 $\{1,\ldots,m\}$,

$$\{1,\ldots,m\} = I \cup \underline{I} \cup \bar{I},$$

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

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

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

Equivalently, as noted in Section 3 [cf. Eqs. (3.1) and (3.3)], we have

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

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 $X_i \subset \mathrm{dom}(f_i)$ 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

$$\bar{f}_{i,X_i}(x_i) = \begin{cases} \min_{\substack{\sum_{\tilde x \in X_i} \alpha_{\tilde x} \tilde x = x_i, \\ \sum_{\tilde x \in X_i} \alpha_{\tilde x} = 1,\ \alpha_{\tilde x} \geq 0,\ \tilde x \in X_i}} \sum_{\tilde x \in X_i} \alpha_{\tilde x} f_i(\tilde x) & \text{if } x_i \in \mathrm{conv}(X_i), \\ \infty & \text{otherwise.} \end{cases}$$

As noted in Section 3, this is the function whose epigraph is the convex hull of the union of the halflines $\big\{(\tilde{x}, w) \mid f_i(\tilde{x}) \le w\big\}$, $\tilde{x} \in X_i$ (cf. Fig. 3.1).

At the typical iteration of the algorithm, we have for each $i \in \underline{I}$, a finite set Λ_i such that $\partial f_i^\star(\tilde{\lambda}) \neq \emptyset$ 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 $\tilde{x} \in X_i$. We find a primal and dual optimal solution pair $(\hat{x}_1, \ldots, \hat{x}_m, \hat{\lambda}_1, \ldots, \hat{\lambda}_m)$ of the problem

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

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. Then, we 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 $\tilde{x}_i \in \partial f_i^\star(\hat{\lambda}_i)$.

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, and we will show in the subsequent proposition that $(\hat{x}_1, \ldots, \hat{x}_m, \hat{\lambda}_1, \ldots, \hat{\lambda}_m)$ is a primal and dual optimal solution pair of the original problem. Otherwise, we proceed to the next iteration, using the enlarged sets Λ_i and X_i.

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 $\partial f_i^\star(\hat{\lambda}_i) \neq \emptyset$ for all $i \in \bar{I}$. Sufficient assumptions may need to be imposed on the problem to guarantee that this is so.

We refer to the preceding algorithm as the *generalized polyhedral approximation* or GPA algorithm. Note 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); or it may effectively have much smaller dimension than the original (cf. the simplicial decomposition method, to be discussed in Section 6).

(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 $\tilde{x}_i \in \partial f_i^\star(\hat{\lambda}_i)$ for $i \in \bar{I}$ is not convenient for some of the functions (e.g., because some of the f_i or the $f_i^\star$ are not available in closed form), we may calculate $\tilde{\lambda}_i$ and/or $\tilde{x}_i$ via the equivalent relations

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

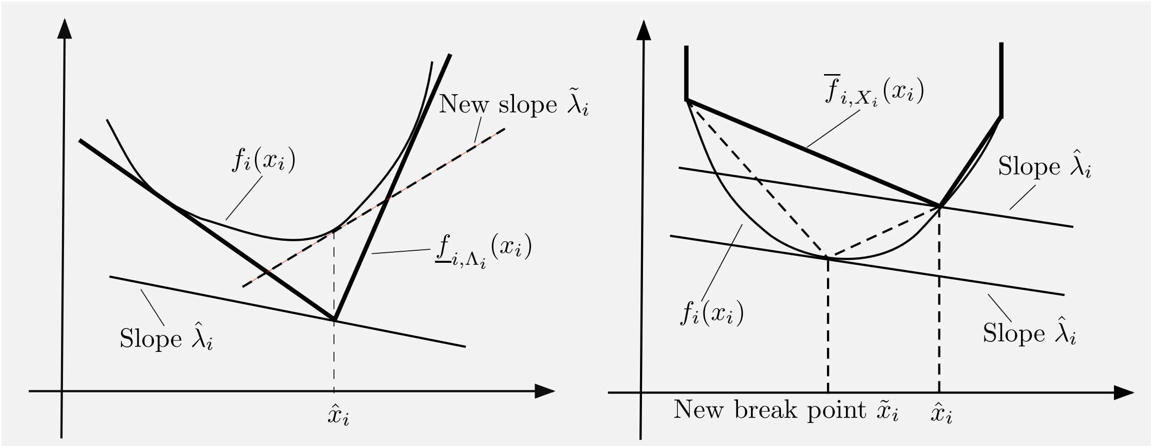

Figure 4.1. Illustration of the enlargement step in the polyhedral approximation method, after we obtain a primal and dual optimal solution pair $(\hat x_1,\dots,\hat x_m,\hat\lambda_1,\dots,\hat\lambda_m)$. The enlargement step on the left [finding $\tilde\lambda_i$ with $\tilde\lambda_i \in \partial f_i(\hat x_i)$] is also equivalent to finding $\tilde\lambda_i$ satisfying $\hat x_i \in \partial f_i^\star(\tilde\lambda_i)$ [cf. the Conjugate Subgradient Theorem, (Prop. 5.4.3 in [Ber09])], or equivalently, solving the optimization problem

$$\begin{aligned}&\text{maximize} \quad \left\{\lambda_i'\hat x_i - f_i^\star(\lambda_i)\right\}\\ &\text{subject to} \;\; \lambda_i \in \Re^{n_i}.\end{aligned}$$

The enlargement step on the right [finding $\tilde x_i$ with $\tilde x_i \in \partial f_i^\star(\hat\lambda_i)$] is also equivalent to finding $\tilde x_i$ satisfying $\hat\lambda_i \in \partial f_i(\tilde x_i)$, or equivalently, solving the optimization problem

$$\begin{aligned}&\text{maximize} \quad \left\{\hat\lambda_i' x_i - f_i(x_i)\right\}\\ &\text{subject to} \;\; x_i \in \Re^{n_i}.\end{aligned}$$

(cf. the Conjugate Subgradient Theorem, 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

$$\begin{aligned}&\text{maximize} \quad \left\{\hat\lambda_i' x_i - f_i(x_i)\right\}\\ &\text{subject to} \;\; x_i \in \Re^{n_i},\end{aligned} \tag{4.2}$$

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):

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

where $f_i^\star$ 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),

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

where the outer (or inner) linearization of $f_i^\star$ is the conjugate of the inner (or respectively, outer) linearization of f_i. 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 $f_i^\star$ are more or less convenient. In fact, when the algorithm makes use of both the primal solution $(\hat{x}_1,\ldots,\hat{x}_m)$ and the dual solution $(\hat{\lambda}_1,\ldots,\hat{\lambda}_m)$ 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 $\underline{f}$,

$$\underline{f} \le f, \quad \underline{f}(x) = f(x) \qquad \Longrightarrow \qquad \partial \underline{f}(x) \subset \partial f(x). \tag{4.5}$$

The second is that for an outer linearization $\underline{f}_\Lambda$ of f,

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

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

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

Proof: From Prop. 2.1 and the definition of $(\hat{x}_1,\ldots,\hat{x}_m)$ and $(\hat{\lambda}_1,\ldots,\hat{\lambda}_m)$ as a primal and dual optimal solution pair of the approximate problem (4.1), we have

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

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

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

which by Prop. 2.1 implies the desired conclusion.

Since $(\hat{x}_1, \ldots, \hat{x}_m)$ and $(\hat{\lambda}_1, \ldots, \hat{\lambda}_m)$ 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. 2.1). 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

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

By Prop. 2.1, we also have $\hat{\lambda}_i \in \partial \underline{f}_{i,\Lambda_i}(\hat{x}_i)$, so $\hat{\lambda}_i \in \partial f_i(\hat{x}_i)$. **Q.E.D.**

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

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

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 $f_i^\star$, 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. **Q.E.D.**

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 kth 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 all $x \in S$, and k, ℓ with $\ell < k$, since $\tilde{\lambda}_i^\ell \in \partial f_i(\hat{x}_i^\ell)$ for all $i \in \underline{I}$, we have

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

where for $i \in \underline{I}$, $\underline{f}_{i,\Lambda_i^k}$ is the outer linearization of f_i at the kth iteration. Let $\{\hat{x}^k\}_\mathcal{K}$ converge to $\bar{x} \in S$ and be such that the sequences $\{\tilde{\lambda}_i^k\}_\mathcal{K}$, $i \in \underline{I}$, are bounded. Taking limit as $\ell \to \infty$, $k \in \mathcal{K}$, $\ell \in \mathcal{K}$, $\ell < k$, in the preceding relation, and using the closedness of f_i, which implies that

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

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. **Q.E.D.**

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 kth 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 $g : \Re^n \mapsto (-\infty, \infty]$ 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. Consider the outer linearizations of g defined by

$$g_k(x) = \max_{i=0,\ldots,k} \big\{g(\hat{x}^i) + (x - \hat{x}^i)'\tilde{\lambda}^i\big\}, \qquad k = 0, 1, \ldots.$$

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

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

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

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

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

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

and since by the closedness of g, we also have

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

it follows that

$$g(\bar{x}) = \lim_{k\to\infty,\, k\in\mathcal{K}} g(\hat{x}^k). \tag{5.1}$$

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

$$\limsup_{k\to\infty,\, k\in\mathcal{K}} g_k(\hat{x}^k) \le \limsup_{k\to\infty,\, k\in\mathcal{K}} g(\hat{x}^k) = g(\bar{x}). \tag{5.2}$$

Using the definition of g_k, we also have for any $k,\, \ell \in \mathcal{K}$ with $k > \ell$,

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

By taking liminf of both sides along $\mathcal{K}$ and using the boundedness of $\tilde{\lambda}^k, k \in \mathcal{K}$ and Eq. (5.1), we have

$$\liminf_{k\to\infty,\, k\in\mathcal{K}} g_k(\hat{x}^k) \ge \liminf_{\ell\to\infty,\, \ell\in\mathcal{K}} g(\hat{x}^\ell) = g(\bar{x}). \tag{5.3}$$

From Eqs. (5.2) and (5.3), we obtain $g(\bar{x}) = \lim_{k\to\infty,\, k\in\mathcal{K}} g_k(\hat{x}^k)$. **Q.E.D.**

We now show the following fact, which is useful for bounding the suboptimality of the solutions of an inner and outer-approximated problem. For simplicity of presentation, we present the result in terms of 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:

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

The analysis entails the case with more than two component functions, as will be seen when we subsequently apply the result in estimating the suboptimality of the solutions and establishing asymptotic convergence of the GPA algorithm.

Proposition 5.5: Let v be the optimal value of an approximate problem

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

where $\underline{g}_1 : \Re^{n_1} \mapsto (-\infty \to \infty]$ and $\bar{g}_2 : \Re^{n_2} \mapsto (-\infty \to \infty]$ are closed proper convex functions such that $\underline{g}_1(y_1) \le g_1(y_1)$ for all y_1 and $g_2(y_2) \le \bar{g}_2(y_2)$ for all y_2, and let $(\hat{y}_1, \hat{y}_2, \hat{\mu}_1, \hat{\mu}_2)$ be a primal and dual optimal solution pair of the approximate problem. Then

$$\bar{g}_2^\star(\hat{\mu}_2) - g_2^\star(\hat{\mu}_2) \le v^* - v \le g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1). \tag{5.4}$$

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

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

Proof: We have

$$\begin{aligned} v^* &= \inf_{y_1} \left\{ g_1(y_1) + \inf_{y_2} \left\{ g_2(y_2) + \delta_S(y_1, y_2) \right\} \right\} \\ &\le g_1(\hat{y}_1) + \inf_{y_2} \left\{ g_2(y_2) + \delta_S(\hat{y}_1, y_2) \right\} \\ &\le g_1(\hat{y}_1) + \inf_{y_2} \left\{ \bar{g}_2(y_2) + \delta_S(\hat{y}_1, y_2) \right\} \\ &= g_1(\hat{y}_1) + \bar{g}_2(\hat{y}_2) + \delta_S(\hat{y}_1, \hat{y}_2) \\ &= v + \Big(g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1) \Big), \end{aligned}$$

where the second inequality follows from $g_2 \le \bar{g}_2$, and the second equality follows from the primal optimality of $(\hat{y}_1, \hat{y}_2)$ for the approximate problem. We also have

$$\begin{aligned} v^* &= \sup_{\mu_2} \left\{ -g_2^\star(\mu_2) + \sup_{\mu_1} \left\{ -g_1^\star(\mu_1) - \delta_{S^\perp}(\mu_1, \mu_2) \right\} \right\} \\ &\ge -g_2^\star(\hat{\mu}_2) + \sup_{\mu_1} \left\{ -g_1^\star(\mu_1) - \delta_{S^\perp}(\mu_1, \hat{\mu}_2) \right\} \\ &\ge -g_2^\star(\hat{\mu}_2) + \sup_{\mu_1} \left\{ -\underline{g}_1^\star(\mu_1) - \delta_{S^\perp}(\mu_1, \hat{\mu}_2) \right\} \\ &= -g_2^\star(\hat{\mu}_2) - \underline{g}_1^\star(\hat{\mu}_1) - \delta_{S^\perp}(\hat{\mu}_1, \hat{\mu}_2) \\ &= v + \left(\bar{g}_2^\star(\hat{\mu}_2) - g_2^\star(\hat{\mu}_2) \right), \end{aligned}$$

where the second inequality follows from $g_1^\star \le \underline{g}_1^\star$ (since $\underline{g}_1 \le g_1$), and the second equality follows from the dual optimality of $(\hat{\mu}_1, \hat{\mu}_2)$ for the approximate problem. Combining the preceding two relations, we obtain Eq. (5.4).

Since

$$g_1(\hat{y}_1) + g_2(\hat{y}_2) \le g_1(\hat{y}_1) + \bar{g}_2(\hat{y}_2) \le \underline{g}_1(\hat{y}_1) + \bar{g}_2(\hat{y}_2) + g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1) = v + g_1(\hat{y}_1) - \underline{g}_1(\hat{y}_1),$$

and from Eq. (5.4), $v \le v^* + g_2^\star(\hat{\mu}_2) - \bar{g}_2^\star(\hat{\mu}_2)$, we obtain

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

By a similar argument, we have

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

This completes the proof. **Q.E.D.**

An immediate consequence of the preceding proposition is estimates of f_{opt} and the suboptimality of the solutions of the GPA algorithm in the general case with both inner and outer linearization. Let $(\hat{x}^k, \hat{\lambda}^k)$ be a primal and dual optimal solution pair of the approximate problem at the kth iteration of the GPA algorithm. We apply Prop. 5.5 with the following identifications, in which we write the variables and the sum of the functions corresponding to the set $\underline{I}$ or $\bar{I}$ as one variable and function, respectively. Specifically, let $y_1 = (x_i)_{i \in \underline{I}}$, $y_2 = (x_i)_{i \in I \cup \bar{I}}$, and

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

The original primal problem can be written as $\inf_{(y_1,y_2) \in S} \big\{g_1(y_1) + g_2(y_2)\big\}$. The dual variables are correspondingly defined as $\mu_1 = (\lambda_i)_{i \in \underline{I}}$, $\mu_2 = (\lambda_i)_{i \in I \cup \bar{I}}$. We denote by $\underline{g}_{1,k}$ and $\bar{g}_{2,k}$ the outer and inner approximations of g_1 and g_2, respectively, that are obtained after k iterations of the GPA algorithm, and we denote by $\underline{g}_{1,k}^\star$ and $\bar{g}_{2,k}^\star$ the corresponding conjugates. Note that this only changes the notation and the algorithm remains unaffected. In terms of the original notation, we have

$$\underline{g}_{1,k}(y_1) = \sum_{i \in \underline{I}} \underline{f}_{i,\Lambda_i^k}(x_i), \qquad \bar{g}_{2,k}^\star(\mu_2) = \sum_{i \in I} f_i^\star(\lambda_i) + \sum_{i \in \bar{I}} \bar{f}_{i,X_i^k}^\star(\lambda_i),$$

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 kth iteration. The solution $(\hat{x}^k, \hat{\lambda}^k)$ of the approximate problem corresponds to $(\hat{y}_1^k, \hat{y}_2^k, \hat{\mu}_1^k, \hat{\mu}_2^k)$ in the simplified notation. Then, with v_k being the optimal value of the kth approximate problem and with $v^* = f_{opt}$, by Prop. 5.5, we have

$$\bar{g}_{2,k}^\star(\hat{\mu}_2^k) - g_2^\star(\hat{\mu}_2^k) \le f_{opt} - v_k \le g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k), \tag{5.5}$$

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

$$\epsilon_k = \big(g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k)\big) + \big(g_2^\star(\hat{\mu}_2^k) - \bar{g}_{2,k}^\star(\hat{\mu}_2^k)\big). \tag{5.6}$$

Equivalently, substituting the definitions of $g_{1,k}$ and $\bar{g}^\star_{2,k}$ in the above expressions, we have

$$\sum_{i\in\bar{I}}\left(\bar{f}^\star_{i,X_i^k}(\hat{\lambda}_i^k)-f_i^\star(\hat{\lambda}_i^k)\right)\leq f_{opt}-v_k\leq\sum_{i\in\underline{I}}\left(f_i(\hat{x}_i^k)-\underline{f}_{i,\Lambda_i^k}(\hat{x}_i^k)\right), \tag{5.7}$$

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

$$\epsilon_k=\sum_{i\in\underline{I}}\left(f_i(\hat{x}_i^k)-\underline{f}_{i,\Lambda_i^k}(\hat{x}_i^k)\right)+\sum_{i\in\bar{I}}\left(f_i^\star(\hat{\lambda}_i^k)-\bar{f}^\star_{i,X_i^k}(\hat{\lambda}_i^k)\right). \tag{5.8}$$

This gives us an estimate of f_{opt} based on the function approximation error at the solutions generated by the GPA algorithm, as well as an estimate of the suboptimality of these solutions.

The above estimates can be equivalently expressed solely in terms of the functions f_i and their approximations by rewriting the terms involving the conjugates. We have for $i\in\bar{I}$,

$$\bar{f}_{i,X_i^k}(\hat{x}_i^k)+\bar{f}^\star_{i,X_i^k}(\hat{\lambda}_i^k)=\hat{\lambda}_i^{k\prime}\hat{x}_i^k,\qquad f_i(\tilde{x}_i^k)+f_i^\star(\hat{\lambda}_i^k)=\hat{\lambda}_i^{k\prime}\tilde{x}_i^k,$$

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

$$f_i^\star(\hat{\lambda}_i^k)-\bar{f}^\star_{i,X_i^k}(\hat{\lambda}_i^k)=\bar{f}_{i,X_i^k}(\hat{x}_i^k)-\left(f_i(\tilde{x}_i^k)+(\hat{x}_i^k-\tilde{x}_i^k)'\hat{\lambda}_i^k\right). \tag{5.9}$$

The right-hand side can be viewed as the sum of two function approximation error terms at $\hat{x}_i^k$: one is the inner linearization error

$$\bar{f}_{i,X_i^k}(\hat{x}_i^k)-f_i(\hat{x}_i^k)$$

and the other is the linearization error

$$f_i(\hat{x}_i^k)-\left(f_i(\tilde{x}_i^k)+(\hat{x}_i^k-\tilde{x}_i^k)'\hat{\lambda}_i^k\right)$$

obtained by using the single subgradient $\hat{\lambda}_i^k\in\partial f_i(\tilde{x}_i^k)$. Correspondingly, we can express the above estimates of f_{opt} and ϵ_k solely in terms of the inner/outer approximation errors of f_i as well as the linearization errors at various points.

We will now derive an asymptotic convergence result for the GPA algorithm in the general case with both inner and outer linearization. Here we implicitly assume that the primal and dual solutions of the approximate problems and enlargement vectors exist. Our convergence analysis will combine the above estimates with properties of outer approximations and Prop. 5.4 in particular.

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 kth 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

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

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 $k \to \infty$.

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 kth approximate problem and let $v^* = f_{opt}$. As shown earlier [cf., Eqs. (5.5) and (5.6)] and repeated here, by Prop. 5.5, we have

$$\bar{g}_{2,k}^\star(\hat{\mu}_2^k) - g_2^\star(\hat{\mu}_2^k) \le v^* - v_k \le g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k), \qquad k = 0, 1, \ldots, \tag{5.10}$$

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

$$\epsilon_k = \big(g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k)\big) + \big(g_2^\star(\hat{\mu}_2^k) - \bar{g}_{2,k}^\star(\hat{\mu}_2^k)\big). \tag{5.11}$$

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

$$\lim_{k\to\infty,\, k\in\mathcal{K}} \underline{f}_{i,\Lambda_i^k}(\hat{x}_i^k) = \lim_{k\to\infty,\, k\in\mathcal{K}} f_i(\hat{x}_i^k), \quad i \in \underline{I}, \qquad \lim_{k\to\infty,\, k\in\mathcal{K}} \bar{f}_{i,X_i^k}^\star(\hat{\lambda}_i^k) = \lim_{k\to\infty,\, k\in\mathcal{K}} f_i^\star(\hat{\lambda}_i^k), \quad i \in \bar{I},$$

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 $f_i^\star$ and its outer linearizations $\bar{f}_{i,X_i^k}^\star$. This implies

$$\lim_{k\to\infty,\, k\in\mathcal{K}} \Big(g_1(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k)\Big) = \lim_{k\to\infty,\, k\in\mathcal{K}} \sum_{i\in\underline{I}} \Big(f_i(\hat{x}_i^k) - \underline{f}_{i,\Lambda_i^k}(\hat{x}_i^k)\Big) = 0,$$

$$\lim_{k\to\infty,\, k\in\mathcal{K}} \Big(g_2^\star(\hat{\mu}_2^k) - \bar{g}_{2,k}^\star(\hat{\mu}_2^k)\Big) = \lim_{k\to\infty,\, k\in\mathcal{K}} \sum_{i\in\bar{I}} \Big(f_i^\star(\hat{\lambda}_i^k) - \bar{f}_{i,X_i^k}^\star(\hat{\lambda}_i^k)\Big) = 0,$$

so that from Eqs. (5.10) and (5.11),

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

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

(b) We note that 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 kth iteration as an approximate problem for the problem at the ℓth 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,

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

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

$$\lim_{\substack{k,\ell\to\infty,\\ k<\ell,\ k\in\mathcal{K}}} \Big(\bar{g}_{2,\ell}^\star(\hat{\mu}_2^k) - \bar{g}_{2,k}^\star(\hat{\mu}_2^k)\Big) = 0, \qquad \lim_{\substack{k,\ell\to\infty,\\ k<\ell,\ k\in\mathcal{K}}} \Big(\underline{g}_{1,\ell}(\hat{y}_1^k) - \underline{g}_{1,k}(\hat{y}_1^k)\Big) = 0. \tag{5.12}$$

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

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

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

$$\lim_{k\to\infty,\, k\in\mathcal{K}} \Big(g_2^\star(\hat{\mu}_2^k) - \bar{g}_{2,k}^\star(\hat{\mu}_2^k)\Big) = 0.$$

Thus we obtain

$$\lim_{\substack{k,\ell\to\infty,\\ k<\ell,\ k\in\mathcal{K}}} \Big(\bar{g}_{2,\ell}^\star(\hat{\mu}_2^k) - \bar{g}_{2,k}^\star(\hat{\mu}_2^k)\Big) = 0,$$

which is the first relation in Eq. (5.12). The second relation in Eq. (5.12) follows with a similar argument. The proof is complete. **Q.E.D.**

The preceding proposition 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.

The preceding proposition also implies that in the pure inner linearization case, the sequence $\{\hat{x}^k\}$ is asymptotic optimal for the original primal problem, and particularly, 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:

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

This strengthens the conclusion of Prop. 5.3. The conclusion of Prop. 5.2 may 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 a new nondifferentiable version 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 $f : \Re^n \mapsto \Re$ 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

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

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 $(\lambda_1, \lambda_2) \in S^\perp$].

Let us apply the GPA algorithm to this formulation with outer linearization of f_1 and no inner linearization:

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

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

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

and for each $\tilde{\lambda} \in \Lambda$, $x_{\tilde{\lambda}}$ is such that $\tilde{\lambda} \in \partial f(x_{\tilde{\lambda}})$. According to the GPA algorithm, if $\hat{x}$ is an optimal solution of problem (6.4) [so that $(\hat{x}, \hat{x})$ 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.

A variant of the classical cutting plane method, which 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

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

Then, using the notation of the original 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

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

where

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

As before, 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 $\tilde{x} \in \partial f_2^\star(-\hat{\lambda})$, or equivalently, solves the problem

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

[cf. Eq. (4.2)]. By Prop. 2.1, 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 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

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

where $f : \Re^n \mapsto (-\infty, \infty]$ and $h : \Re^n \mapsto (-\infty, \infty]$ are closed proper convex functions. This is the problem of the Fenchel duality context, and it contains as a special case the problem (6.1), where f is real-valued and h is the indicator function of the closed convex set C. The corresponding framework is defined by

$$f_1 = f, \qquad f_2 = h, \qquad S = \big\{ (x, x) \mid x \in \Re^n \big\}.$$

We will apply the GPA algorithm, where the function h is inner linearized, while the function f is left intact. As in the preceding example, the primal and dual optimal solution pairs have the form $(\hat{x}, \hat{x}, \hat{\lambda}, -\hat{\lambda})$.

We start with some finite set $X_0 \subset \mathrm{dom}(f)$ such that $\partial h(\tilde{x}) \neq \emptyset$ 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 $x_k, \tilde{x}_k$, 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

$$x_k \in \arg\min_{x \in \Re^n} \big\{f(x) + H_k(x)\big\}, \tag{6.8}$$

where H_k is the polyhedral/inner linearization function whose epigraph is the convex hull of the union of the rays $\big\{(\tilde{x}, w) \mid h(\tilde{x}) \leq w\big\}$, $\tilde{x} \in X_k$. The existence of a solution x_k of problem (6.8) is guaranteed by Weierstrass' Theorem, since $\mathrm{dom}(H_k)$ is the convex hull of the finite set X_k.

(2) **Solution of approximate dual problem:** We obtain a subgradient $g_k \in \partial f(x_k)$ such that

$$-g_k \in \partial H_k(x_k). \tag{6.9}$$

The existence of such a subgradient is guaranteed by standard optimality conditions, applied to the minimization in Eq. (6.8), under mild conditions [X_k should contain a point in the relative interior of the domain of f; cf., [Ber09], Prop. 5.4.7(3)]. Note that by the optimality conditions (2.3)-(2.4) of Prop. 2.1, $(g_k, -g_k)$ is an optimal solution of the dual approximate problem.

(3) **Enlargement:** We obtain $\tilde{x}_k$ such that

$$-g_k \in \partial h(\tilde{x}_k),$$

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

We assume that f and h are such that the steps (1)-(3) above can be carried out (conditions for this will be discussed shortly). Note that the enlargement step (3) is equivalent to finding

$$\tilde{x}_k \in \arg\min_{x \in \Re^n} \big\{g_k'(x - x_k) + h(x)\big\}, \tag{6.10}$$

and that this is a linear program in the important special case where h is polyhedral.

Let us first assume that f is 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]. At the typical iteration of this method, we have a finite set of points $X \subset C$. We then solve the problem

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

According to the GPA algorithm, if $(\hat{x}, \hat{x}, \hat{\lambda}, -\hat{\lambda})$ is an optimal primal and dual solution pair, we enlarge X by adding to X any $\tilde{x}$ with $\tilde{x} \in \partial f_2^\star(-\hat{\lambda})$. This is equivalent to finding $\tilde{x}$ that solves the optimization problem

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

[cf. Eq. (4.2); we assume that this problem has a solution, which is guaranteed if C is bounded]. The resulting method, illustrated in Fig. 6.1, is the classical simplicial decomposition method, and terminates in a finite number of iterations. In this context, step (1) corresponds to the minimization (6.11), step (2) simply yields $g_k = \nabla f(x_k)$, and step (3), as implemented in Eq. (6.10), corresponds to solution of the linear program (6.12) that generates a new extreme point of C.

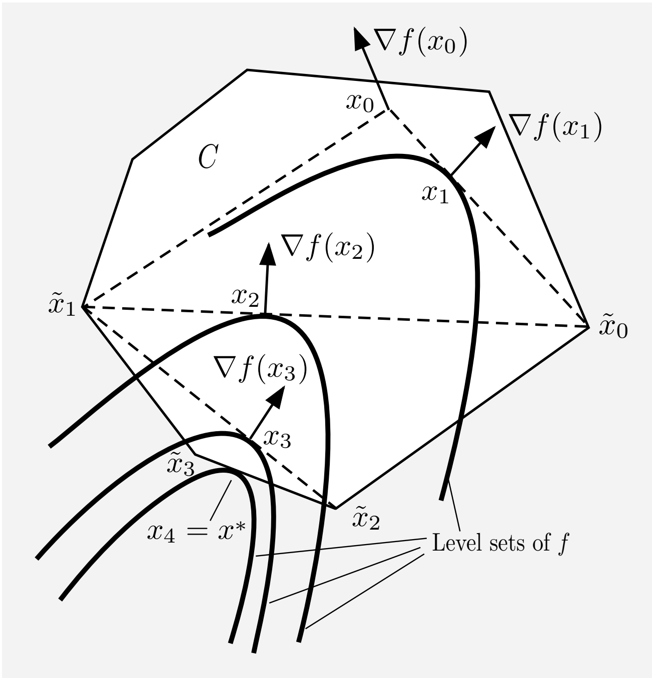

Figure 6.1. Successive iterates of the classical simplicial decomposition method in the case where f is differentiable and C is polyhedral. For example, the figure shows how given the initial point x_0 and the calculated extreme points $\tilde{x}_0$, $\tilde{x}_1$, we determine the next iterate x_2 as 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 X is added, and after four iterations, the optimal solution is obtained.

(b) When h is a general closed proper convex function, the method is illustrated in Fig. 6.2. The existence of a solution x_k to problem (6.8) is guaranteed by the compactness of $\text{conv}(X_k)$ and Weierstrass' Theorem, while step (2) yields $g_k = \nabla f(x_k)$. The existence of a solution to problem (6.10) must be guaranteed by some assumption such as coercivity of h. 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 $x \in X$ and $(x, w) \in \text{epi}(h)$. In the special case where h is a polyhedral function, it can be shown that the method terminates finitely, assuming that the vectors $\big(\tilde{x}_k, h(\tilde{x}_k)\big)$ obtained by solving the corresponding linear program (6.10) are extreme points of epi(h).

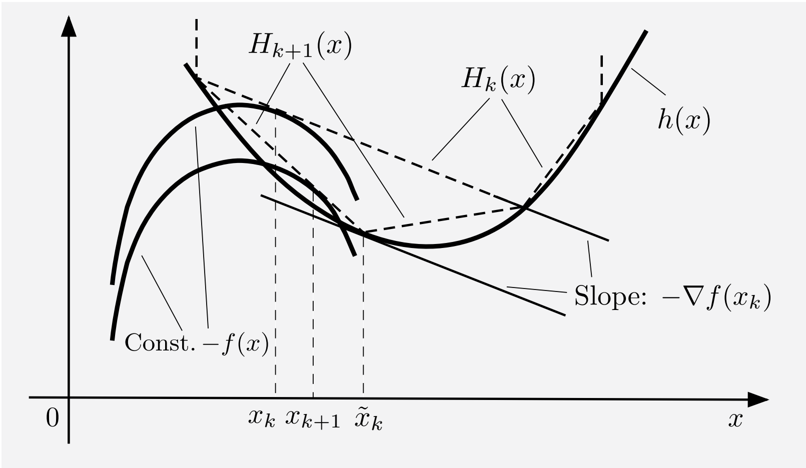

Figure 6.2. Illustration of successive iterates of the generalized simplicial decomposition method in the case where f is differentiable. Given the inner linearization H_k of h, we minimize $f + H_k$ to obtain x_k (graphically, we move the graph of $-f$ vertically until it touches the graph of H_k).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

Generalized Simplicial Decomposition - Nondifferentiable Case

Let us now consider the case where f is extended real-valued and nondifferentiable. For this case there are no simplicial decomposition algorithms in the literature, to our knowledge. When h is the indicator function of a polyhedral set C, the condition (6.9) of step (2) is that $-g_k$ is in the normal cone of conv(X_k) at x_k (cf. [Ber09], p. 185). The method is illustrated for this case in Fig. 6.3. It terminates finitely, assuming that the vector $\tilde{x}_k$ obtained by solving the linear program (6.10) is an extreme point of C. The reason is that in view of Eq. (6.9), the vector $\tilde{x}_k$ does not belong to X_k (unless 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 $g_k \in \partial f(x_k)$ such that $-g_k \in \partial H_k(x_k)$ [cf. Eq. (6.9)]. This may be a difficult problem as it may require knowledge of $\partial f(x_k)$ as well as $\partial H_k(x_k)$. However, in special cases, g_k may be obtained simply as a byproduct of the minimization (6.8). As an example, consider the common case where h is the indicator of a closed convex set C, and

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

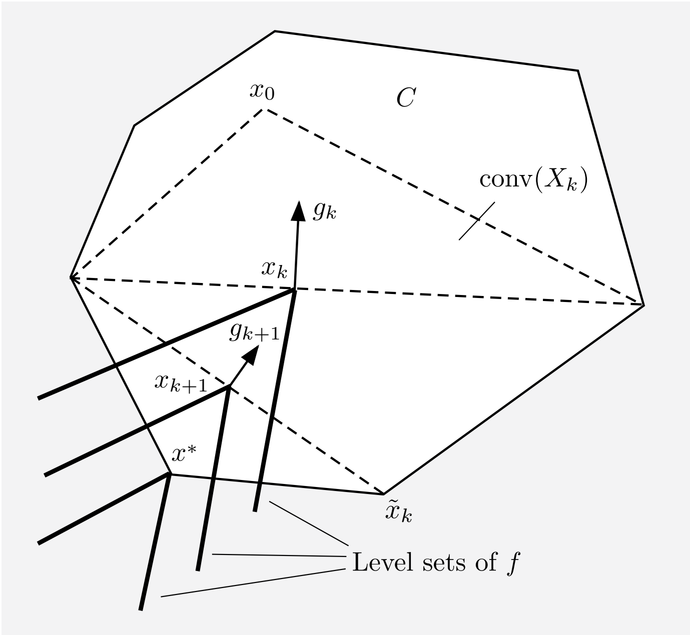

Figure 6.3. Illustration of the generalized simplicial decomposition method for the case where f is nondifferentiable and h is the indicator function of a polyhedral set C. For each k, we compute a subgradient $g_k \in \partial f(x_k)$ such that $-g_k$ lies in the normal cone of $\text{conv}(X_k)$ at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

where $f_1, \ldots, f_r$ are differentiable functions. Then the minimization (6.8) takes the form

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

According to standard optimality conditions, the optimal solution (x_k, z^*) together with dual optimal variables $\mu_j^* \geq 0$, satisfies

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

It follows that

$$\sum_{j=1}^r \mu_j^* = 1, \qquad \mu_j^* \geq 0 \text{ and } \mu_j^* > 0 \ \Rightarrow \ f_j(x_k) = f(x_k), \quad j = 1, \ldots, r, \tag{6.14}$$

and

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

From Eq. (6.14) it follows that the vector

$$g_k = \sum_{j=1}^r \mu_j^* \nabla f_j(x_k) \tag{6.16}$$

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

In conclusion, g_k as given by Eq. (6.16), is a suitable subgradient of f at x_k for determining a new extreme point $\tilde{x}_k$ via problem (6.10). Note an important advantage of this method over potential competitors in the case where C is polyhedral: it involves solution of linear programs of the form (6.10) to generate new extreme points of C, and low-dimensional nonlinear programs of the form (6.13). When each f_j is twice differentiable, the latter programs can be solved by fast Newton-like methods, such as sequential quadratic programming (see e.g., [Ber82], [Ber99], [NoW06]).

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

$$\begin{aligned} &\text{minimize} \;\; f_1^\star(\lambda) + f_2^\star(-\lambda) \\ &\text{subject to} \;\; \lambda \in \Re^n, \end{aligned}$$

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

$$\begin{aligned} &\text{minimize} \;\; f_1^\star(\lambda) + \bar{f}_{2,X_k}^\star(-\lambda) \\ &\text{subject to} \;\; \lambda \in \Re^n, \end{aligned} \tag{6.17}$$

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

Note that if g_k is a solution of problem (6.17), the associated subgradient of $f_2^\star$ at $-g_k$ is the vector $\tilde{x}_k$ generated by the enlargement step (6.10), as shown in Fig. 6.4. The ordinary cutting plane method, described in the beginning of this section, is obtained as the special case where $f_1^\star(\lambda) \equiv 0$.

Whether the primal implementation, based on solution of problem (6.8), or the dual implementation, based on solution of problem (6.17), is preferable depends on the structure of the functions f and h. When f (and hence also $f_1^\star$) is not polyhedral, the dual implementation may not be attractive, because it requires the n-dimensional nonlinear optimization (6.17) 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.

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

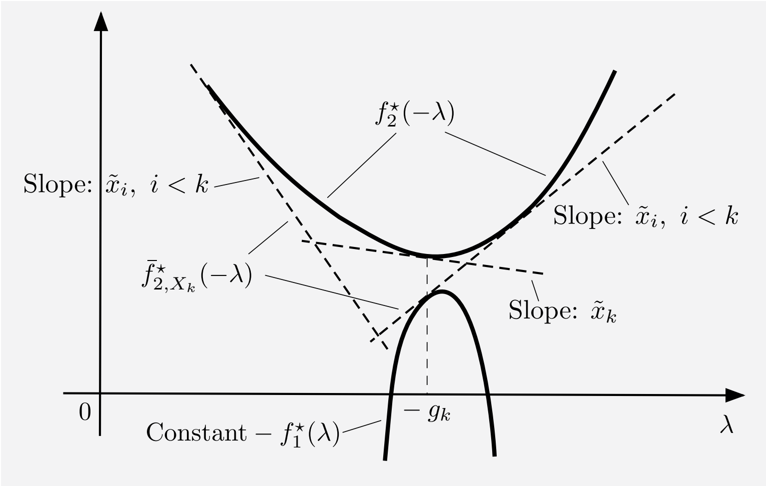

Figure 6.4. Illustration of the cutting plane implementation of the generalized simplicial decomposition method. The ordinary cutting plane method, described in the beginning of Section 6, is obtained as the special case where $f_1^\star(x) \equiv 0$. In this case, f is the indicator function of the set consisting of the just the origin, and the primal problem is to evaluate $h(0)$.

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 $x = \big\{x_a \mid a \in \mathcal{A}\big\}$ 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 on 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 approximation cutting plane and simplicial decomposition methods. 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.

8. REFERENCES

[BGL09] Bonnans, J. F., Gilbert, J. C., Lemaréchal, C., and Sagastizábal, S. C., 2009. Numerical Optimization: Theoretical and Practical Aspects, Springer, N. Y.

[Ber82a] Bertsekas, D. P., 1982. Constrained Optimization and Lagrange Multiplier Methods, Academic Press, N. Y.; republished by Athena Scientific, Belmont, MA, 1997.

[Ber98] Bertsekas, D. P., 1998. Network Optimization: Continuous and Discrete Models, Athena Scientific, Belmont, MA.

[Ber99] Bertsekas, D. P., 1999. Nonlinear Programming: 2nd Edition, Athena Scientific, Belmont, MA.

[Ber08] Bertsekas, D. P., 2008. "Extended Monotropic Programming and Duality," JOTA, Vol. 139, pp. 209-225.

[Ber09] Bertsekas, D. P., 2009. Convex Optimization Theory, Athena Scientific, Belmont, MA.

[GoP79] Gonzaga, C., and Polak, E., 1979. "On Constraint Dropping Schemes and Optimality Functions for a Class of Outer Approximations Algorithms," SIAM J. on Control and Optimization, Vol. 17, pp.477-493.

[HLV87] Hearn, D. W., Lawphongpanich, S., and Ventura, J. A., 1987. "Restricted Simplicial Decomposition: Computation and Extensions," Math. Programming Studies, Vol. 31, pp. 119-136.

[HiL93] Hiriart-Urruty, J.-B., and Lemarechal, C., 1993. Convex Analysis and Minimization Algorithms, Vols. I and II, Springer-Verlag, Berlin and N. Y.

[Hoh77] Hohenbalken, B. von, 1977. "Simplicial Decomposition in Nonlinear Programming," Math. Programming, Vol. 13, pp. 49-68.

[Hol74] Holloway, C. A., 1974. "An Extension of the Frank and Wolfe Method of Feasible Directions," Math. Programming, Vol. 6, pp. 14-27.

[Mey79] Meyer, R. R., 1979. "Two-Segment Separable Programming," Management Science, Vol. 29, pp. 385-395.

[NoW99] Nocedal, J., and Wright, S. J., 1999. Numerical Optimization, Springer, N. Y.

[Pol97] Polak, E., 1997. Optimization: Algorithms and Consistent Approximations, Springer-Verlag, N. Y.

[Roc70] Rockafellar, R. T., 1970. Convex Analysis, Princeton Univ. Press, Princeton, N. J.

[Roc84] Rockafellar, R. T., 1984. Network Flows and Monotropic Optimization, Wiley, N. Y.; republished by Athena Scientific, Belmont, MA, 1998.

[TsB00] Tseng, P., and Bertsekas, D. P., 2000. "An Epsilon-Relaxation Method for Separable Convex Cost Generalized Network Flow Problems," Math. Progamming, Vol. 88, pp. 85-104.

[Tse01] Tseng, P., 2001. "An Epsilon Out-of-Kilter Method for Monotropic Programming," Math. Oper. Res., Vol. 26, pp. 221-233.

[VeH93] Ventura, J. A., and Hearn, D. W., 1993. "Restricted Simplicial Decomposition for Convex Constrained Problems," Math. Programming, Vol. 59, pp. 71-85.