Math. Program., Ser. A 86: 595–614 (1999)

Digital Object Identifier (DOI) 10.1007/s101079900101

Sebastián Ceria* · João Soares**

Convex programming for disjunctive convex optimization

Received November 27, 1996 / Revised version received June 11, 1999
Published online November 9, 1999

Abstract. Given a finite number of closed convex sets whose algebraic representation is known, we study the problem of finding the minimum of a convex function on the closure of the convex hull of the union of those sets. We derive an algebraic characterization of the feasible region in a higher-dimensional space and propose a solution procedure akin to the interior-point approach for convex programming.

1. Introduction

The literature in optimality conditions and solution methods for convex programming is extensive when the feasible set is either given in abstract form or explicitly by convex constraints. The problem that we address in this article lies in between, namely, we seek the minimum of a convex function on a closed convex set defined as the closure of the convex hull of a finite number of individual closed convex sets that have a known representation.

This mathematical program occurs in the context of disjunctive convex optimization where the infimum of a convex function is sought over the union of a finite number of individual closed convex sets. The optimal value of our mathematical problem provides a lower bound to that of the corresponding disjunctive convex program but, if the objective function is linear, then the optimal value of both programs coincides and there is at least one optimal solution that is feasible to both programs. Since any disjunctive convex program can be formulated with a linear objective function then our framework also provides a means for solving disjunctive convex programs. Examples of applications with some nonlinear element in the formulation include market product positioning, process synthesis network design, see [6] and references therein, and limited-diversification portfolio selection, see [3,5]. References of applications in the linear case that are not of the combinatorial type may be found in [12, Chapters 1,8] and [1].

S. Ceria: 417 Uris Hall, Graduate School of Business, Columbia University, New York, NY 10027, USA.
e-mail: `sebas@cumparsita.gsb.columbia.edu`

J. Soares: Departamento de Matemática da Universidade de Coimbra, 3000 Coimbra, Portugal.
e-mail: `jsoares@mat.uc.pt`

Mathematics Subject Classification (1991): 90C30, 90C11, 90C25

* Research partly supported by NSF HPCC Grant DMS 95-27124. Also affiliated with Computational Optimization Research Center (CORC), Columbia University.

** Supported by Subprograma Ciência e Tecnologia do 2^{o} Quadro comunitário de Apoio grant Praxis XXI/BD/2831/94.

In this paper we study the two central issues that arise when tackling these problems, namely, algebraic representation and solution procedure. We first provide an algebraic characterization of the set of feasible solutions in a higher-dimensional space, i.e., involving additional variables. Though this characterization is primarily nonconvex, we are able to obtain an equivalent convex representation through the use of the *perspective function* concept from convex analysis. However, the new representation does not inherit fully the differentiability properties of the original functions that characterize the individual sets and it may not have a closed-form expression at all the points of interest which, in particular, implies that we cannot directly apply a standard nonlinear programming algorithm. We propose a procedure that uses an idea akin to the interior-point approach of introducing a parameter and a barrier term to the objective function. For every value of the parameter, the resulting program is convex and amenable by standard convex programming algorithms. As the barrier parameter goes to zero, any accumulation point of the sequence of optimal points is optimal for the original program. We also show that the Lagrange multipliers associated with the parameterized programs define approximations to the dual certificates of optimality for the original program.

Our work extends a well-known special case studied by Balas [2], when the objective function is linear and every individual set is defined by a set of linear constraints. Our convexification argument was also used by Jeroslow [9] who assumed that the individual sets are bounded and contained in the first orthant. Stubbs and Merhotra [13] also used similar assumptions and argument in the formulation of the cut generation problem for convex programming with binary variables. Most of our results rely on the excellent first volume of the book by H. Urruty and C. Lemaréchal [7], from where we have tried to follow the same notation closely.

The paper is structured in the following way. In the remainder of this section we introduce some notation and recall some basic results from convex analysis. In Sect. 2 we provide a higher-dimensional algebraic characterization of the closure of the convex hull in terms of points and directions of the individual sets without requiring any special assumptions, boundedness in particular. We show that this characterization defines a nonconvex set in the higher-dimensional space unless each individual set is a singleton. We also explain the convex reformulation and relate it to the work of Balas in the linear case. In Sect. 3 we recall the abstract optimality conditions for the optimization problem, present optimality conditions involving Lagrange multipliers and show that, under a constraint qualification, they are also necessary. In Sect. 4 we show that the optimal solution may be obtained from solving a sequence of parameterized standard convex programs and show that any accumulation point of the sequence of Lagrange multipliers for some individual sets provides a dual certificate of optimality concerning the same individual sets. Finally, in Sect. 5 we draw conclusions and give directions of future research.

We denote by $\mathbb{R}^n$ the n-dimensional real space and by $\mathbb{R}^n_+$ the set of nonnegative vectors of $\mathbb{R}^n$. Elements of $\mathbb{R}^n$ can be column vectors or row vectors. The distinction should be drawn from the context, but often it can also be made from the variable identificators used, the last letters of the latin alphabet like x, y, z refer to column vectors and greek letters like α, β, γ refer to row vectors, although they may also represent scalars. A sequence of scalars is indexed by a subscript like in $\{\mu_t\}$ while a sequence of matrices or vectors is indexed by a superscript like in $\{\lambda^t\}$. The symbol

$\| \cdot \|$ denotes any *norm* in the primal space $\mathbb{R}^n$ and $\|\alpha\|_* \equiv \max_{\|x\| \leq 1} \alpha x$ defines the corresponding *dual norm*, also called *conjugate*, on the dual space. A convex function $f: \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ is convex according to the modern concept that allows f to take ∞ values, see [7]. The effective domain of the function f, denoted dom (f), is the set of points $x \in \mathbb{R}^n$ for which $f(x) < +\infty$. A function f is closed when all its level sets are closed. Every convex function f admits a unique closed extension (cl f) which results from redefining f at points $x \notin$ dom (f) in such a way that for all level sets

$$L_\delta(\text{cl } f) \equiv \left\{x \in \mathbb{R}^n : (\text{cl } f)(x) \leq \delta\right\} = \text{cl } \left(\left\{x \in \mathbb{R}^n : f(x) \leq \delta\right\}\right) \equiv \text{cl } (L_\delta(f)).$$

The set $\partial f(x)$ denotes the subdifferential of the convex function f at $x \in \mathbb{R}^n$. This set is defined on the dual space and hence, it is composed of row vectors. When $\partial f(x)$ is singleton then its unique element is called the gradient of f at x, denoted $\nabla f(x)$, and again it is a row vector. If f is continuous at x^* then $\partial f(x^*)$ is nonempty and compact ([7, Theorem 6.2.2]). Furthermore, if $\{x^k\}$ is a sequence converging to x^* then $\lim \partial f(x^k) \subseteq \partial f(x^*)$ ([7, Theorem 6.2.4]). If $f = [f_i]_{i=1}^m$ is a vector of convex functions f_i then $\partial f(x) \subseteq \mathbb{R}^{m \times n}$ denotes the set of matrices S whose rows are the subgradients of the functions f_i at $x \in \mathbb{R}^n$. If f is a closed convex function then the function denoted f'_∞ is the recession function of f and the function $\tilde{f}$ is the perspective of f. A rigorous definition and some results with these functions are presented in the appendix.

Given some set $P \subseteq \mathbb{R}^n$, we use the standard topological concepts like the *closure*, denoted cl, the *relative interior*, denoted ri, and the *convex hull*, denoted conv . If P is a nonempty closed convex set, the *recession cone* of P, denoted P_∞, is the set of vectors $d \in \mathbb{R}^n$ such that for any $x \in P$ we have that $x + td \in P$, for every $t \in \mathbb{R}_+$. A (algebraic) *representation* of P is of the form

$$P = \left\{x \in \mathbb{R}^n : G(x) \leq 0\right\}, \tag{1}$$

where $G: \mathbb{R}^n \to \mathbb{R}^m$ is vector mapping whose components are closed convex functions. Given two sets $A, B \subseteq \mathbb{R}^n$, we call the *(Minkowski) sum* of A with B the set $A + B$ made of every possible sums of one element in A with one element in B. The symbol "$\subseteq$" means "contained in", while the symbol "$\subset$" means "strictly contained in". The symbol "$\equiv$" denotes a defining equality. The symbol Δ_p, or Δ_I, denotes the simplex polytope in $\mathbb{R}^p$, or $\mathbb{R}^{|I|}$.

2. Higher-dimensional characterization

Consider a closed convex set $P \subseteq \mathbb{R}^n$ defined by

$$P \equiv \text{cl conv } (K), \quad K \equiv \bigcup_{i=1}^{p} K^i, \tag{2}$$

where every set K^i is a closed convex set having the following representation

$$K^i \equiv \left\{x \in \mathbb{R}^n : G^i(x) \leq 0\right\}, \tag{3}$$

and $G^i: \mathbb{R}^n \to \mathbb{R}^{m_i}$ is a vector mapping whose components are closed convex functions.

In order to provide a higher-dimensional characterization of P, we will at first assume that the set K is either bounded below or above, i.e., there exists some $\bar{x} \in \mathbb{R}^n$ such that $\bar{x} \le x$, for every $x \in K$, or $\bar{x} \ge x$, for every $x \in K$. As we will see later, the assumption is not needed but makes the proof simpler.

Proposition 1. *Let P be given by (2), $I \equiv \{i : K^i \neq \emptyset\}$ and assume that the set K is bounded below or above. Then,*

$$P = \operatorname{conv}\left(\bigcup_{i\in I} K^i\right) + \sum_{i\in I} K^i_\infty. \tag{4}$$

Proof. We shall prove the mutual inclusion. In order to prove the "$\subseteq$" part, let $x \in P$. By definition, $x = \lim_t x^t$, where $x^t \in \operatorname{conv}(K)$ and after some algebra we can refine this characterization to $x = \lim_t \sum_{i\in I} \lambda_i^t \left(u^i\right)^t$, where $(u^i)^t \in K^i$ and $\lambda^t \equiv (\lambda_i^t) \in \Delta_I$. Since Δ_I is compact, there exists some infinite index set T_1 such that the subsequence $\{\lambda^t\}_{t\in T_1}$ converges. Hence,

$$\lim_{t\in T_1} \lambda^t = \hat{\lambda} \in \Delta_I.$$

For any $i \in I$, the sequence $\{\lambda_i^t \left(u^i\right)^t\}_{t\in T_1}$ is bounded, or otherwise, since the set K is bounded below or above, the sequence $\{x^t\}$ would not converge. Hence, there exists an index sequence $T_2 \subseteq T_1$ such that the sequence $\{\lambda_i^t \left(u^i\right)^t\}_{t\in T_2}$ converges.

For those $i \in I$ such that $\hat{\lambda}_i > 0$, the sequence $\{\left(u^i\right)^t\}_{t\in T_2}$ is bounded, or otherwise the sequence $\{\lambda_i^t \left(u^i\right)^t\}_{t\in T_2}$ would not converge. Hence, there exists an index sequence $T \subseteq T_2$ such that

$$\lim_{t\in T} \lambda_i^t \left(u^i\right)^t = \hat{\lambda}_i \hat{u}^i,$$

where $\lim_{t\in T} \left(u^i\right)^t = \hat{u}^i \in K^i$ because K^i is closed.

In the remainder of the proof we show that for those $i \in I$ such that $\hat{\lambda}_i = 0$ we have that

$$\hat{v}^i \equiv \lim_{t\in T} \lambda_i^t \left(u^i\right)^t \in K^i_\infty.$$

Since K^i is nonempty, we choose an arbitrary $\hat{u}^i \in K^i$. We have to prove that $\hat{u}^i + \alpha\hat{v}^i \in K^i$, for every $\alpha \ge 0$. Since $\lim_{t\in T} \lambda_i^t = \hat{\lambda}_i = 0$, we have that, for all but a finite number of t's in T,

$$0 \le \alpha\lambda_i^t \le 1 \quad \text{and} \quad \left(1 - \alpha\lambda_i^t\right)\hat{u}^i + \alpha\lambda_i^t \left(u^i\right)^t \in K^i.$$

Clearly,

$$\lim_{t\in T} \left(1 - \alpha\lambda_i^t\right)\hat{u}^i + \alpha\lambda_i^t \left(u^i\right)^t = \hat{u}^i + \alpha\hat{v}^i,$$

and so, $\hat{u}^i + \alpha\hat{v}^i \in K^i$ because K^i is closed. Thus, one inclusion is proved.

In order to prove the "$\supseteq$" part, let $x \in \operatorname{conv}\left(\cup_{i\in I} K^i\right) + \sum_{i\in I} K^i_\infty$. After some algebra, x can be written as

$$x = \sum_{i\in I} \lambda_i x^i + \sum_{i\in I} d^i, \tag{5}$$

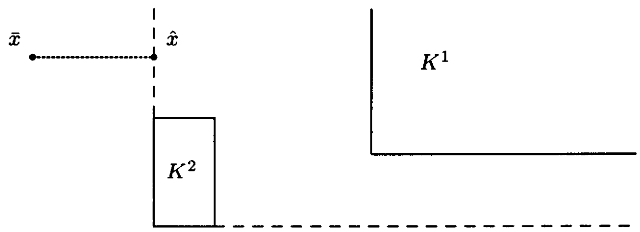

Fig. 1. The optimal solution over P is not in K

where $x^i \in K^i$, $d^i \in K^i_\infty$ and $\lambda = (\lambda_i) \in \Delta_I$. Now, define $I_+ = \{i \in I : \lambda_i > 0\}$ and $I_0 = \{i \in I : \lambda_i = 0\}$ and pick an index $j \in I_+$, a nonempty set because $\sum_{i\in I} \lambda_i = 1$, so that

$$x_\epsilon = \sum_{i\in I_+\setminus\{j\}} \left(\lambda_i x^i + d^i\right) + \Big(\lambda_j - \sum_{i\in I_0} \epsilon_i\Big) x^j + d^j + \sum_{i\in I_0} \left(\epsilon_i x^i + d^i\right),$$

where $\epsilon = (\epsilon_i)$ is such that $0 < \epsilon_i < \lambda_j / \mid I_0 \mid$, for every $i \in I_0$. For every ϵ we have that $x_\epsilon \in \text{conv}\,(\cup_{i\in I} K^i)$. It is now easy to construct a sequence of points in $\text{conv}\,(\cup_{i\in I} K^i)$ convergent to x. Hence, $x \in P$ and the mutual inclusion is proved.

Note that, from Proposition 1, the Minkowski sum of all the recession cones associated with nonempty sets K^i is enough to "close" the set $\text{conv}\,(\cup_{i=1}^p K^i)$. See Fig. 1 for an example where the set $\text{conv}\,(K^1 \cup K^2)$ does not contain the dashed lines and, therefore, it is not closed.

Proposition 1 provides an algebraic characterization of P as the projection into the space of the x variables of some higher-dimensional set. In fact, $x \in P$ if and only if there exist vectors (λ_i, z^i, d^i), for every $i \in I$, such that the following nonlinear system holds

$$x = \sum_{i\in I} \lambda_i z^i + \sum_{i\in I} d^i \tag{6}$$

$$G^i(z^i) \leq 0, \quad {G^i}'_\infty(d^i) \leq 0, \quad i \in I \tag{7}$$

$$\sum_{i\in I} \lambda_i = 1, \quad \lambda_i \geq 0, \quad i \in I, \tag{8}$$

where ${G^i}'_\infty$ is the recession, or asymptotic, function of G^i (see the appendix).

There are two sources of difficulty with the characterization (6)–(8). The first one comes from the fact that the functions ${G^i}'_\infty$ do not admit a closed-form expression in general. The other arises from the presence of the nonlinear equality constraint (6) which complicates the optimization stage as our next result shows.

Proposition 2. *The higher-dimensional set $\mathcal{P}$ defined by (6)–(8) is convex if and only if every set K^i, for $i \in I$, is a singleton.*

Proof. Let $\tilde{X} = (\tilde{x}, \tilde{\lambda}, \tilde{z}, \tilde{d})$ and $\hat{X} = (\hat{x}, \hat{\lambda}, \hat{z}, \hat{d})$ be any two points in $\mathcal{P}$ and consider an arbitrary convex combination $\delta\tilde{X} + (1-\delta)\hat{X}$, where $\delta \in [0, 1]$. Since $\sum_{i\in I} \delta\tilde{\lambda}_i +$

$(1-\delta)\hat{\lambda}_i = 1$, $\delta\tilde{z}^i + (1-\delta)\hat{z}^i \in K^i$ and $\delta\tilde{d}^i + (1-\delta)\hat{d}^i \in K^i_\infty$ always, then $\mathcal{P}$ is convex if and only if

$$\delta\tilde{x}+(1-\delta)\hat{x}-\sum_{i\in I}\left(\delta\tilde{\lambda}_i+(1-\delta)\hat{\lambda}_i\right)\left(\delta\tilde{z}^i+(1-\delta)\hat{z}^i\right)-\sum_{i\in I}\left(\delta\tilde{d}^i+(1-\delta)\hat{d}^i\right)=0, \quad (9)$$

for any $\delta \in [0, 1]$ and any two feasible points in $\mathcal{P}$. Simplifying the left-hand-side of (9) we obtain an equivalent expression as

$$\delta(1-\delta)\sum_{i\in I}\left(\hat{\lambda}_i - \tilde{\lambda}_i\right)\left(\tilde{z}^i - \hat{z}^i\right) = 0.$$

Since we may select $\hat{\lambda}_i$ and $\tilde{\lambda}_i$ distinctly, the desired result follows.

Though, in the higher-dimensional space, the representation (6)–(8) defines a non-convex set, there is an alternative representation that defines a convex set. Both representations are equivalent in the sense that they project the same set in the space of the x variables. When the G^i's are linear mappings the required algebraic manipulation was referred to as *convexification* by Balas [2]. Theorem 1 below extends this argument to the nonlinear setting. We note, however, that Jeroslow [9, Example 4.1] had already found the same result in a more particular setting, namely that $P \subseteq \mathbb{R}^n_+$ and bounded. Our generality is mainly due to the connection between the convexification argument and the convex analysis concept of perspective function. In fact, we have already seen that $x \in P$ if and only if

$$x = \sum_{i\in I:\,\lambda_i>0}\lambda_i z^i + \sum_{i\in I:\,\lambda_i=0} d^i,$$

for some suitable choice of the remaining variables. This is equivalent to saying that $x = \sum_{i\in I} x^i$, where $\lambda_i G^i(x^i/\lambda_i) \le 0$, for every $i \in I\colon \lambda_i > 0$ and ${G^i}'_\infty(x^i) \le 0$, for every $i \in I\colon \lambda_i = 0$. The perspective mapping captures these two different expressions into a single one because $\left(\mathrm{cl}\ \tilde{G}^i\right)\left(\lambda_i, x^i\right) = \lambda_i G^i(x^i/\lambda_i)$ whenever $\lambda_i > 0$ and $\left(\mathrm{cl}\ \tilde{G}^i\right)\left(0, x^i\right) = {G^i}'_\infty(x^i)$.

Theorem 1. *Let P be given by (2) and $I \equiv \{i\colon K^i \neq \emptyset\}$. If the set K is bounded below or above then $x \in P$ if and only if there exist vectors (λ_i, x^i), for every $i \in I$, such that the following nonlinear system is feasible*

$$x = \sum_{i\in I} x^i \quad (10)$$

$$\left(cl\ \tilde{G}^i\right)\left(\lambda_i, x^i\right) \le 0, \quad i \in I, \quad (11)$$

$$\sum_{i\in I}\lambda_i = 1, \quad \lambda_i \ge 0, \quad i \in I, \quad (12)$$

where, generically, $\left(cl\ \tilde{G}\right)(\lambda, x)$ denotes the closure of the perspective mapping of G at (λ, x).

Proof. From basic results recalled in the appendix, we have that, for every $i \in I$,

$$\left(\mathrm{cl}\, \tilde{G}^i\right)\left(\lambda_i, x^i\right) \leq 0 \iff \begin{cases} G^i(x^i/\lambda_i) \leq 0 & \text{if } \lambda_i > 0 \\ {G^i}'_\infty(x^i) \leq 0 & \text{if } \lambda_i = 0. \end{cases}$$

Thus, (10)–(12) characterizes all the points x that are the sum of a convex combination of points in every set K^i and directions of the same sets K^i, for $i \in I$. From Proposition 1, this is equivalent to saying that $x \in P$.

Now, we may explain why the boundedness assumption on K is not a requirement for Proposition 1 and Theorem 1 to hold. In general, we have that

$$\mathrm{conv}\left(\bigcup_{i \in I} K^i\right) \subseteq \mathrm{conv}\left(\bigcup_{i \in I} K^i\right) + \sum_{i \in I} K^i_\infty \subseteq \mathrm{cl}\,\mathrm{conv}\left(\bigcup_{i \in I} K^i\right) = P.$$

But, characterization (10)–(12), which defines the set in the middle regardless of the boundedness assumption, defines a closed set because it is the projection into the space of the x variables of a closed higher-dimensional set. Thus, the second inclusion is actually an equality and we may conclude that (10)–(12) defines P regardless of the boundedness assumption.

The convexification argument of Balas when $G^i(x) = b^i - A^i x$ becomes a natural corollary of Theorem 1. In fact, it is easy to check that

$$\left(\mathrm{cl}\, \tilde{G}^i\right)\left(\lambda_i, x^i\right) = \begin{cases} \lambda_i b^i - A^i v^i & \text{if } \lambda_i \geq 0, \\ +\infty & \text{if } \lambda_i < 0, \end{cases}$$

so that that (10)–(12) and Balas' characterization of P, see [2], are essentially the same.

Characterization (10)–(12) of P requires the knowledge of which sets K^i are nonempty. If the set $\{1, \ldots, p\}$ is used instead of I then it may define a larger convex set $\hat{P}$ than the intended P. We recall that the nonlinear system $\left(\mathrm{cl}\, \tilde{G}^i\right)\left(0, x^i\right) \leq 0$ admits the trivial solution $x^i = 0$ even if the set K^i is empty. A necessary and sufficient condition for $\hat{P} = P$ is that

$$\sum_{i=1}^{p} \left\{d \in \mathbb{R}^n : \left(\mathrm{cl}\, \tilde{G}^i\right)(0, d) \leq 0\right\} = \sum_{i \in I} \left\{d \in \mathbb{R}^n : \left(\mathrm{cl}\, \tilde{G}^i\right)(0, d) \leq 0\right\}, \tag{13}$$

which depends upon the chosen algebraic representation of every set K^i. This condition holds trivially in, at least, two interesting situations. When P is the convex hull of the feasible region of a mixed integer convex program, each set K^i corresponds to a particular assignment of the integer variables. In this case, the set $\{d \in \mathbb{R}^n : \left(\mathrm{cl}\, \tilde{G}^i\right)(0, d) \leq 0\}$ is independent of i and, thus, condition (13) follows. When P is bounded and every set K^i (or at least the empty ones) contains among its algebraic representation a subset of constraints defining a nonempty polytope, condition (13) also holds because $\{d \in \mathbb{R}^n : \left(\mathrm{cl}\, \tilde{G}^i\right)(0, d) \leq 0\} = \{0\}$. These two special cases are natural extensions of results presented in [2].

Therefore, the problem of finding the minimum of some convex function f over the set P defined by (2) can be algebraically formulated as

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & (10)\text{–}(12) \end{array} \tag{14}$$

This program is not amenable by standard convex programming algorithms because at points of the form $(0, x^i)$, the mapping cl $\tilde{G}^i$ may not have a closed-form expression and, in general, it is not differentiable even if the G^i's are continuously differentiable everywhere. If we knew in advance the existence of an optimal solution where the components λ_i, $i \in I$, would all be positive then this difficulty could be overcome by simply imposing a sufficiently small positive lower bound on the variables λ_i, $i \in I$. In the following sections we will explain a procedure for solving Program (14) in general. We start by elaborating on necessary and sufficient conditions for optimality in the next section.

3. Optimality conditions

In this section we focus on deriving optimality conditions for

$$\min_{x \in P} f(x), \tag{15}$$

where P is defined by (2) and f is a closed convex function. Additionally, we assume that f and every G^i are continuous in an open set containing P.

First, observe that if f is linear then it is equivalent to solving Program (15) over P or over K. In fact, let $\hat{x} \in P$ be defined by

$$\hat{x} = \sum_{i \in I:\, \hat{\lambda}_i > 0} \hat{\lambda}_i \hat{z}^i + \sum_{i \in I:\, \hat{\lambda}_i = 0} \hat{d}^i, \tag{16}$$

where $I = \{i : K^i \neq \emptyset\}$ and $\hat{z}^i \in K^i$, for every $i \in I : \hat{\lambda}_i > 0$, $\hat{d}^i \in K^i_\infty$, for every $i \in I : \hat{\lambda}_i = 0$, $(\hat{\lambda}_i) \in \Delta_I$. Assume that $\hat{x}$ is optimal for Program (15) and $f(x) = cx$. Since $K^i \subseteq P$, for every $i \in I$, then

$$\begin{aligned} \min_{z^i \in K^i} cz^i &= c\hat{z}^i = c\hat{x},\ i \in I : \hat{\lambda}_i > 0, \\ \min_{z^i \in K^i} cz^i &\geq c\hat{x},\ i \in I : \hat{\lambda}_i = 0. \end{aligned} \tag{17}$$

The reciprocal is also true, i.e., conditions (17) are also sufficient for optimality of $\hat{x}$ defined by (16). Therefore, since at least one $\hat{\lambda}_i$ is positive then it is equivalent to solve Program (15) over K or over P, in the sense that the optimal value is the same and at least one optimal solution belongs to both sets. Furthermore, the set of optimal solutions $\hat{S}$ can be fully characterized simply by knowing one of the optimal solutions. In fact, the set of optimal solutions $\hat{S}$ is given by

$$\hat{S} = \text{cl conv}\left(\bigcup_{i \in I} \left\{ z^i \in K^i : cz^i = c\hat{x} \right\} \right).$$

When f is nonlinear then the minimum value of f over P is a lower bound to that over K. For example, Fig. 1 illustrates a point $\hat{x}$ that uniquely minimizes the l_2 distance function $\|x - \bar{x}\|$ over $P = \text{cl conv}\,(K^1 \cup K^2)$, but that it is not in $K^1 \cup K^2$.

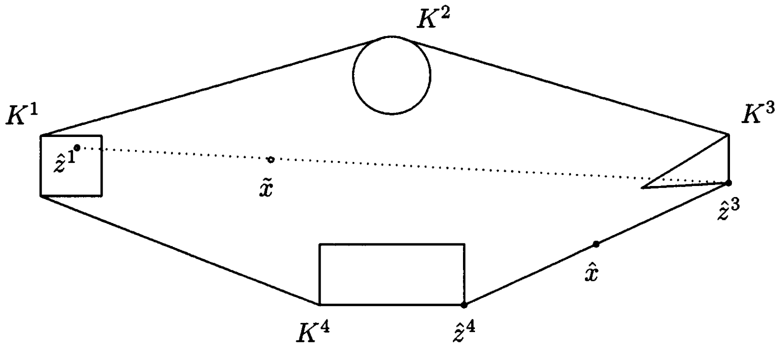

Fig. 2. Geometric interpretation of optimality

A point $\hat{x} \in P$ defined by (16) is optimal for Program (15) if and only if there exists a subgradient $\hat{\xi} \in \partial f(\hat{x})$ such that

$$\hat{\xi}(x - \hat{x}) \geq 0, \tag{18}$$

for every $x \in P$, see [7, Theorem VI.1.1.1] for example. Thus, by using (17), this is equivalent to saying that $\hat{x} \in P$ defined by (16) is optimal for Program (15) if and only if there exists a subgradient $\hat{\xi} \in \partial f(\hat{x})$ such that for $\hat{\alpha} = -\hat{\xi}$,

$$\begin{aligned} h^i(\hat{\alpha}) &= \hat{\alpha}\hat{z}^i = \hat{\alpha}\hat{x},\ i \in I : \hat{\lambda}_i > 0, \\ h^i(\hat{\alpha}) &\leq \hat{\alpha}\hat{x},\ i \in I : \hat{\lambda}_i = 0, \end{aligned} \tag{19}$$

where $h^i(\alpha) \equiv \sup_{z^i \in K^i} \alpha z^i$. Furthermore, extending an earlier result of Mangasarian [11], Burke and Ferris [4] showed that the set of subgradients that satisfy (18) is the same for any point $\hat{x}$ in the set of optimal solutions $\hat{S}$. This implies, for example, that if f is differentiable in an open set containing P then

$$\hat{S} \subseteq \text{cl conv}\left(\bigcup_{i \in I}\left\{z^i \in K^i : \hat{\xi}z^i = \hat{\xi}\hat{x}\right\}\right), \tag{20}$$

where $\hat{\xi} = \nabla f(\hat{x})$. The inclusion may be strict as the two-dimensional example of Fig. 2 shows. Assume that the interior point $\tilde{x}$ is the unique optimal solution so that $\hat{\xi} = 0$. Then, the set on the right-hand-side of (20) is $P = \text{conv}\,(\cup_{i=1}^4 K^i)$, i.e., everything. In the same figure, suppose that $\hat{x}$ is optimal and f is differentiable, the point $\hat{x}$ can only be expressed as a convex combination of the points $\hat{z}^3$ and $\hat{z}^4$. Thus, $\hat{S}$ lies in the segment $\delta\hat{z}^3 + (1 - \delta)\hat{z}^4$, for $\delta \in [0, 1]$.

In order to check whether a given point $\hat{x} \in P$ is optimal we need to solve $|\ I\ |$ optimization problems if f is differentiable, according to (19). If we want to avoid solving that many optimization problems then we need more explicit optimality conditions involving Lagrange multipliers. Proposition 3 below shows conditions that are sufficient for optimality of Program (15). The conditions that we propose involve the following set which is related to the normal cone to P at some point $\hat{x}$ defined by (16),

$$N'_G\big(\hat{x}, (\hat{\lambda}_i, \hat{z}^i, \hat{d}^i)_{i\in I}\big) \equiv \left\{\alpha \in \mathbb{R}^n : \begin{array}{lll} \alpha \in N'_{G^i}(\hat{z}^i), & \alpha\hat{z}^i \le \alpha\hat{x}, & i \in I : \hat{\lambda}_i > 0, \\ \alpha z^i \le \alpha\hat{x}, & \text{for any } z^i \in K^i, & i \in I : \hat{\lambda}_i = 0, \end{array}\right\}.$$

where, following the notation in [7], the set $N'_{G^i}(\hat{z}^i)$ is related to the normal cone to K^i at $\hat{z}^i$ and defined by

$$N'_{G^i}(\hat{z}^i) = \left\{\alpha \in \mathbb{R}^n : \alpha = u^i S^i, S^i \in \partial G^i(\hat{z}^i), u^i G^i(\hat{z}^i) = 0, u^i \ge 0\right\}.$$

Later, we will show that the sufficient condition for optimality stated in Proposition 3 is also necessary under a constraint qualification. We remark that the sets N'_{G^i} and N'_G above are defined differently. While in one case, G^i identifies the specific mapping that algebraically characterizes the set K^i, in the other case the "G" is a generic symbol.

Proposition 3. *Let $\hat{x} \in P$ be defined by (16). If*

$$0 \in \partial f(\hat{x}) + N'_G\big(\hat{x}, (\hat{\lambda}_i, \hat{z}^i, \hat{d}^i)_{i\in I}\big), \tag{21}$$

then $\hat{x}$ is optimal for Program (15).

Proof. Let $\xi \in \partial f(\hat{x})$ be such that $\alpha = -\xi \in N'_G(\hat{x}, (\hat{\lambda}_i, \hat{z}^i, \hat{d}^i)_{i\in I})$. Since $N'_{G^i}(\hat{z}^i) \subseteq N_{K^i}(\hat{z}^i) \equiv \{\alpha : h^i(\alpha) = \alpha\hat{z}^i\}$, see [7, Lemma VII.2.1.3, page 305], then we have that $\alpha x \le \alpha\hat{x}$, for every $x \in K$. Thus, by continuity and from the definition of α, we conclude that $\xi(x - \hat{x}) \ge 0$, for every $x \in P$, which shows the optimality of $\hat{x}$.

We would expect that checking whether a given point $\hat{x} \in P$ defined by (16) is optimal for Program (14) to be an easy linear feasibility problem as it usually occurs with standard differentiable convex programs. This is not the case with the condition (21) for, suppose that all the functions involved are continuously differentiable just to make subgradients uniquely determined. Since $N'_{G^i}(\hat{z}^i) \subseteq N_{K^i}(\hat{z}^i)$, for every $i \in I : I \cap \{i : \hat{\lambda}_i > 0\}$, see [7, Lemma VI.2.1.3, page 305], condition (21) requires that, not only we solve $|\ I \cap \{i : \hat{\lambda}_i > 0\}\ |$ linear feasibility problems in the variables u^i, but also we have to make sure that, for $\hat{\alpha} = -\nabla f(\hat{x})$,

$$h^i(\hat{\alpha}) = \begin{array}{l} \max\ \hat{\alpha} z^i \\ \text{s.t.}\quad z^i \in K^i \end{array} \le \hat{\alpha}\hat{x},$$

for every $i \in I : \hat{\lambda}_i = 0$. Thus, we also have to solve $|\ I \cap \{i : \hat{\lambda}_i = 0\}\ |$ optimization problems or, at least, prove that the optimal value of all of them is bounded by $\hat{\alpha}\hat{x}$.

A constraint qualification is needed to guarantee the existence of the multipliers involved in condition (21) at an optimal solution of Program (15), as with any standard convex program. One possible constraint qualification simply imposes that the cone $N'_G(\hat{x}, (\hat{\lambda}_i, \hat{z}^i, \hat{d}^i)_{i\in I})$ coincides with the normal cone to P at $\hat{x} \in P$, $N_P(\hat{x})$, defined by

$$\begin{aligned} N_P(\hat{x}) &\equiv \left\{\alpha \in \mathbb{R}^n : \alpha(x - \hat{x}) \le 0,\ \text{for every } x \in P\right\} \\ &= \left\{\alpha \in \mathbb{R}^n : h(\alpha) = \alpha\hat{x}\right\}, \end{aligned}$$

where $h(\alpha) \equiv \max_{x \in P} \alpha x$. Note that $N'_G(\hat{x}, (\hat{\lambda}_i, \hat{z}^i, \hat{d}^i)_{i \in I}) \subseteq N_P(\hat{x})$ holds at any $\hat{x} \in P$ defined by (16). Our constraint qualification implies that the reciprocal inclusion also holds for any $\hat{x} \in P$.

Definition 1. *We say that the basic constraint qualification holds for K defined by (2) if*

$$N'_{G^i}(z^i) = N_{K^i}(z^i),$$

for every $z^i \in K^i$ and every $i \in I$.

When $p = 1$ this definition coincides with the basic constraint qualification introduced in [7] where the concept is also related to other constraint qualifications in constrained optimization.

Assuming that the basic constraint qualification holds, let $\alpha \in N_P(\hat{x})$. In particular, $\alpha z^i \leq \alpha \hat{x}$, for every $z^i \in K^i$ and for every $i \in I$. Since

$$\alpha \hat{x} = \sum_{i \in I:\, \hat{\lambda}_i > 0} \hat{\lambda}_i \alpha \hat{z}^i + \sum_{i \in I:\, \hat{\lambda}_i = 0} \alpha \hat{d}^i$$

and $\alpha \hat{d}^i \leq 0$, for every $i \in I: \hat{\lambda}_i = 0$, or otherwise we wouldn't have a finite $h^i(\alpha)$, then we must have $\alpha \hat{z}^i = \alpha \hat{x}$, for every $i \in I: \hat{\lambda}_i > 0$. Thus, $\alpha \in N_{K^i}(\hat{z}^i) = N'_{G^i}(\hat{z}^i)$, for every $i \in I: \hat{\lambda}_i > 0$, which implies that

$$N'_G(\hat{x}, (\hat{\lambda}_i, \hat{z}^i, \hat{d}^i)_{i \in I}) = N_P(\hat{x}),$$

for any $\hat{x} \in P$ defined by (16). Now, since optimality of Program (15) is equivalent to

$$0 \in \partial f(\hat{x}) + N_P(\hat{x})$$

then the basic constraint qualification is enough to guarantee (21) at optimality.

If every function G^i is affine, i.e., $G^i(z^i) = b_i - A^i z^i$ then, since the basic constraint qualification holds, see [7, Proposition VI.2.2.2], the existence of the multipliers at an optimal solution of Program (15) is guaranteed. If some function G^i is nonlinear then the basic constraint qualification is hard to check from its definition. However, a well known sufficient condition for $N_{G^i}(z^i) = N_{K^i}(z^i)$ to hold for any $z^i \in K^i$ is that the weak Slater condition is satisfied. The weak Slater condition is said to hold for $K^i = \{z^i: G^i(z^i) \leq 0\}$ if there exists a vector $\tilde{z}^i \in K^i$ such that $G^i_k(\tilde{z}^i) < 0$, for every $k \in J_i$, where J_i denotes the index sets corresponding to the nonlinear functions. So, if the weak Slater condition holds for every set K^i then the existence of multipliers at an optimal solution of Program (15) is guaranteed.

4. A primal procedure

The optimal solution of Program (15) was characterized in the previous section. In this section we propose a primal procedure for solving Program (14) by solving a sequence of convex programs defined by one of the following problems:

$$\begin{array}{ll} \min & f(x) - \mu \sum_{i\in I} \ln \lambda_i \\ & x = \sum_{i\in I} x^i \\ \text{s.t.} & \left(\text{cl}\, \tilde{G}^i\right)\left(\lambda_i, x^i\right) \le 0, \quad i \in I \\ & \sum_{i\in I} \lambda_i = 1, \\ & \lambda_i \ge 0, \quad i \in I, \end{array} \qquad (22)$$

$$\begin{array}{ll} \min & f(x) \\ & x = \sum_{i\in I} x^i \\ \text{s.t.} & \left(\text{cl}\, \tilde{G}^i\right)\left(\lambda_i, x^i\right) \le 0, \quad i \in I \\ & \sum_{i\in I} \lambda_i = 1, \\ & \lambda_i \ge \mu, \quad i \in I. \end{array} \qquad (23)$$

When $\mu > 0$ and fixed, both of these programs are convex and have closed level sets. Moreover, since every set K^i is nonempty then every $(x, (\lambda_i, x^i)_{i\in I})$ of any level set is such that $\lambda_i > 0$, for every $i \in I$, meaning that if we use an algorithm that keeps iterates in the same level set then no points of the form $(0, x^i)$ will be generated. Furthermore, at any of those points

$$\left(\text{cl}\, \tilde{G}^i\right)\left(\lambda_i, x^i\right) = \lambda_i G^i(x^i/\lambda_i),$$

so that we may exploit the differentiability properties of the original program. Theorem 2 below states that when μ goes to zero, any accumulation point of the sequence defined by the optimal points for Program (22), or Program (23), is optimal for Program (14). The theorem applies to a particular convergent subsequence. Existence can be guaranteed under appropriate compactness assumptions.

Theorem 2. *Let $\{x^t\}$ be a convergent subsequence of optimal points for Program (22), or Program (23), for some sequence $\{\mu_t\}$ of positive numbers converging to zero. Then, $\hat{x} = \lim x^t$ is optimal for Program (14).*

Proof. Consider Program (22) first and let $(\xi^t, (-\mu_t/\lambda_i^t, 0)_{i\in I})$ be a subgradient of the objective function at the optimal point of the t-th problem that satisfies

$$\xi^t(x - x^t) - \sum_{i\in I} \frac{\mu_t}{\lambda_i^t}\left(\lambda_i - \lambda_i^t\right) \ge 0, \qquad (24)$$

for every $(x, (\lambda_i, x^i)_{i\in I}) \in \mathcal{P}$, the feasible region of Program (22).

Since $\lim \partial f(x^t) \subseteq \partial f(\hat{x})$, which is a compact set because $\hat{x} \in P$ and f is continuous at $\hat{x}$, there exists an accumulation point $\hat{\xi} \in \partial f(\hat{x})$ of the sequence $\{\xi^t\}$. We assume without loss of generality that $\{\xi^t\}$ converges to $\hat{\xi}$.

Now, we show that the sequence $\{\mu_t/\lambda_i^t\}$ is bounded, for every $i \in I$. By contradiction, suppose that there exists a subsequence $\{\mu_t/\lambda_i^t\}_{t\in T}$ such that $\lim_{t\in T} \mu_t/\lambda_i^t = +\infty$. In particular, we must have $\lim_{t\in T} \lambda_i^t = 0$ because $\lim_{t\in T} \mu_t = 0$. Pick any vector $(\tilde{x}, (\tilde{\lambda}_i, \tilde{x}^i)_{i\in I}) \in \mathcal{P}$ such that $\tilde{\lambda}_i > 0$. Then, from (24), we have that

$$\xi^t(\tilde{x} - x^t) - \sum_{i\in I} \frac{\mu_t}{\lambda_i^t}(\tilde{\lambda}_i - \lambda_i^t) \ge 0, \qquad (25)$$

for every $t \in T$. But, we have reached a contradiction because as $t \in T$ approaches $+\infty$ the left-hand-side of (25) approaches $-\infty$. Thus, we conclude that $\{\mu_t/\lambda_i^t\}$ is bounded, for every $i \in I$.

Since both sequences $\{\mu_t/\lambda_i^t\}$ and $\{\lambda_i^t\}$ are bounded, let δ_i and $\hat{\lambda}_i$ be two accumulation points of these sequences. Note that $\delta_i > 0$ implies that $\hat{\lambda}_i = 0$ because $\lim_t \mu_t = 0$. Taking limits in (24) over these convergent subsequences, we conclude that

$$\hat{\xi}(x - \hat{x}) - \sum_{i\in I} \delta_i \lambda_i \ge 0,$$

for every $(x, (\lambda_i, x^i)_{i\in I}) \in \mathcal{P}$, which in particular implies that $\hat{\hat{\xi}}(x - \hat{x}) \geq 0$, for every $x \in P$. Thus, $\hat{x}$ is optimal for Program (14).

Now, consider Program (23) and let $(\xi^t, (0, 0)_{i\in I})$ be the subgradient of the objective function at the optimal point of the t-th problem that satisfies

$$\xi^t(x - x^t) \geq 0, \tag{26}$$

for every $(x, (\lambda_i, x^i)_{i\in I}) \in \mathcal{P}_{\mu_t}$, the feasible region of Program (23). Since $\lim \partial f(x^t) \subseteq \partial f(\hat{x})$, which is a compact set because $\hat{x} \in P$ and f is continuous at $\hat{x}$, there exists an accumulation point $\hat{\hat{\xi}} \in \partial f(\hat{x})$ of the sequence $\{\xi^t\}$. We assume without loss of generality that $\{\xi^t\}$ converges to $\hat{\hat{\xi}}$. Since $\lim \mathcal{P}_{\mu_t} = \mathcal{P}$ then taking limits (26) we conclude that $\hat{\hat{\xi}}(x - \hat{x}) \geq 0$, for every $x \in P$. Thus, $\hat{x}$ is optimal for Program (14).

Theorem 2 shows that we may solve Program (14) by solving an infinite sequence of standard convex programs. Since it is not possible to solve such a large number of problems then the next question is how to verify whether a given point $\hat{x}$ is optimal. The necessary and sufficient condition (18) provided a possible answer because $\hat{x}$ is optimal if and only if for $\hat{\alpha} = -\hat{\hat{\xi}}$

$$h^i(\hat{\alpha}) \equiv \begin{array}{l} \max\ \hat{\alpha} z^i \\ \text{s.t.}\quad z^i \in K^i \end{array} \leq \hat{\alpha}\hat{x},$$

for every $i \in I$, which amounts to solving $\mid I \mid$ optimization problems if f is differentiable.

We now provide a slightly different version of Theorem 2 by showing that if $\hat{x}$ is an accumulation point of a sequence of KKT points for Program (22) then $\hat{x}$ satisfies condition (21) for a suitable choice of the multipliers. A suitable analysis can be carried out for Program (23). For Program (22), the KKT conditions are satisfied at $(x, (\lambda_i, x^i)_{i\in I}) \in \mathcal{P}$ if there is a multiplier row vector $(\alpha, \delta, (u^i)_{i\in I})$ and subgradients $\xi \in \partial f(x)$, $S^i \in \partial G^i(x^i/\lambda_i)$ that satisfy

$$\left.\begin{array}{lllcl} \xi & +\alpha & & = & 0, \\ & -\alpha x^i/\lambda_i & +\delta & = & \mu/\lambda_i \\ & -\alpha & \quad +u^i S^i & = & 0, \end{array}\right\}, i \in I, u^i \geq 0, i \in I \tag{27}$$

and the complementarity equations

$$u^i G^i(x^i/\lambda_i) = 0, i \in I. \tag{28}$$

For Program (23), the KKT conditions are satisfied at $(x, (\lambda_i, x^i)_{i\in I}) \in \mathcal{P}_\mu$ if there is a multiplier row vector $(\alpha, \delta, (\gamma_i, u^i)_{i\in I})$ and subgradients $\xi \in \partial f(x)$, $S^i \in \partial G^i(x^i/\lambda_i)$ that satisfy

$$\begin{array}{l} \left.\begin{array}{lllll} \xi & +\alpha & & & = 0, \\ & -\alpha x^i/\lambda_i & +\delta & -\gamma_i & = 0 \\ & -\alpha & \quad +u^i S^i & & = 0, \end{array}\right\}, i \in I, \\ u^i, \gamma_i \geq 0, i \in I \end{array} \tag{29}$$

and the complementarity equations

$$\left.\begin{array}{ll} u^i G^i(x^i/\lambda_i) & = 0, \\ \gamma_i(\lambda_i - \mu) & = 0, \end{array}\right\}, \; i \in I, \tag{30}$$

Note that, since the KKT conditions are sufficient for optimality then, as explained before, $\lambda_i > 0$, for every $i \in I$. We also note that the derivation of the KKT conditions required the knowledge of the subdifferential of cl $\tilde{G}^i$ in terms of the subdifferential of G^i. This is explained in the appendix.

Theorem 3 below states that when μ goes to zero, any accumulation point of the sequence defined by the KKT points for Program (22) satisfies (21). An analogous result holds for Program (23).

Theorem 3. *Let $\{x^t = \sum_{i\in I}(x^i)^t\}_{t\in T}$ be a subsequence of KKT points for Program (22) for some sequence $\{\mu_t\}$ of positive numbers converging to zero such that*

$$\left.\begin{array}{lll} & \lim_{t\in T} \lambda^t = \hat{\lambda} \in \Delta_I, & \\ \lim_{t\in T}(x^i)^t/\lambda_i^t = \hat{z}^i \in K^i, & \lim_{t\in T}(S^i)^t = \hat{S}^i \in \partial G^i(\hat{z}^i), & i \in I : \hat{\lambda}_i > 0, \\ \lim_{t\in T}(x^i)^t = \hat{d}^i \in K_\infty^i, & & i \in I : \hat{\lambda}_i = 0, \\ \lim_{t\in T} x^t = \hat{x} \in P, & \lim_{t\in T} \xi^t = \hat{\xi} \in \partial f(\hat{x}), & \end{array}\right\} \tag{31}$$

If $\hat{u}^i$, $i \in I : \hat{\lambda}_i > 0$, are accumulation points of the sequence of multipliers $\{(u^i)^t\}_{t\in T}$, $i \in I : \hat{\lambda}_i > 0$, then

$$\hat{x} = \lim_{t\in T} x^t = \sum_{i\in I : \hat{\lambda}_i > 0} \hat{\lambda}_i \hat{z}^i + \sum_{i\in I : \hat{\lambda}_i = 0} \hat{d}^i \in P$$

satisfies (21) with those values of the multipliers.

Proof. As we saw in the proof of Theorem 2, $\hat{\xi}\left(x - \hat{x}\right) \geq 0$, for every $x \in P$. In particular, since $K^i \subseteq P$, we conclude that $h^i(\hat{\alpha}) \leq \hat{\alpha}\hat{x}$, for every $i \in I$. Now, consider only those $i \in I : \hat{\lambda}_i > 0$. From the second set of equations in (27),

$$-\alpha^t \frac{(x^i)^t}{\lambda_i^t} + \delta_t = \frac{\mu_t}{\lambda_i^t}. \tag{32}$$

Taking limits in (32) over T, we conclude that the sequence $\{\delta_t\}_{t\in T_1}$ converges to $\hat{\delta} = \hat{\alpha}\hat{z}^i$. Moreover, from the third set of equations (27),

$$-\alpha^t + \left(u^i\right)^t \left(S^i\right)^t = 0. \tag{33}$$

Taking limits in (33) over T, we conclude that $\hat{\alpha} + \hat{u}^i \hat{S}^i = 0$. The complementarity $\hat{u}^i G^i(\hat{z}^i) = 0$ follows from the continuity of G^i.

The existence of KKT points for Program (22), or Program (23), is assumed in Theorem 3. But, if the basic constraint qualification holds then the existence of KKT points for Program (22) is guaranteed, regardless of the value of μ, because the constraint qualification is independent of the objective function.

Most importantly, Theorem 3 shows that if we are solving Program (15) through solving Program (22), for some sequence of positive numbers $\{\mu_t\}$ converging to zero then, the corresponding multipliers $(u^i)^t$ at the t-th optimal solution and the subgradient ξ^t define approximated values with which we can verify condition (21). We remark that checking whether $\xi \in \partial f(\hat{x})$ is easy in many cases of interest. An especially important example is $f(x) = \|x - \bar{x}\|$ where

$$\partial f(x) = \left\{\xi \in \mathbb{R}^n : \xi(x - \bar{x}) = \|x - \bar{x}\|, \|\xi\|_* = 1\right\}.$$

There is a limitation in Theorem 3 in that it does not show any value in the accumulation points $\hat{u}^i, i \in I : \hat{\lambda}_i = 0$ of the sequence of multipliers $\{(u^i)^t\}_{t\in T}, i \in I : \hat{\lambda}_i = 0$. However, we think that these multipliers $\hat{u}^i, i \in I : \hat{\lambda}_i = 0$ may considerably speed-up the verification that $h^i(\alpha) \leq \alpha\hat{x}, i \in I : \hat{\lambda}_i = 0$ in an effort to establish the optimality of $\hat{x}$.

Theorem 3 required the existence of accumulation points of the sequence of multipliers. A stronger version of the Slater condition guarantees the existence *a priori*. We say that the strong Slater condition holds for K defined by (2) if the strong Slater condition holds for every set K^i, for $i \in I$, see [7, Def VI.2.3.1, page 311] for example. We recall that the strong Slater condition holds for a convex set $K^i = \{z^i : G^i(z^i) \leq 0\}$ if there exists a feasible point where the only binding constraints are linear and its normals are linearly independent. Theorem 4 formalizes the usefulness of this concept by showing that the multipliers sequence remains in a compact set, which guarantees the existence of accumulation points.

Theorem 4. *Assume that the strong Slater condition holds for the set K defined by (2) and that (31) holds. Then, the multipliers sequences $\{(u^i)^t\}_{t\in T}, i \in I : \hat{\lambda}_i > 0$, associated with Program (22), for some sequence of parameter values converging to zero, are bounded.*

Proof. Consider only those indices $i \in I : \hat{\lambda}_i > 0$. Equation (33) is equivalent to

$$\sum_{j=1}^{m_i} \left(u^i_j\right)^t \left(S^i_j\right)^t = \alpha^t, \tag{34}$$

where $(S^i_j)^t$ denotes the jth row vector of the matrix $(S^i)^t$.

Let $\tilde{z}^i$ be the point in K^i where the strong Slater condition is satisfied, J be the set of the indices corresponding to the non-binding constraints at $\tilde{z}^i$ and $\tilde{J}$ be the set of the indices corresponding to the binding ones. By the definition of subgradient, we have that

$$0 > G^i_j\left(\tilde{z}^i\right) \geq G^i_j\left((x^i)^t/\lambda^t_i\right) + \left(S^i_j\right)^t\left(\tilde{z}^i - (x^i)^t/\lambda^t_i\right),$$

for every $j \in J$. Since $G^i_j(\tilde{z}^i) = a^j \tilde{z}^i - b_j = 0$, for every $j \in \tilde{J}$, and it holds that $(u^i_j)^t G^i_j((x^i)^t/\lambda^t_i) = 0$, we have that

$$\sum_{j\in J} \left(u^i_j\right)^t G^i_j\left(\tilde{z}^i\right) \geq \sum_{j=1}^{m_i} \left(u^i_j\right)^t \left(S^i_j\right)^t \left(\tilde{z}^i - (x^i)^t/\lambda^t_i\right) \tag{35}$$

$$= \alpha^t \left(\tilde{z}^i - (x^i)^t/\lambda^t_i\right). \tag{36}$$

The left-hand-side of (35) is always non-positive because $(u^i_j)^t \geq 0$. The right-hand-side converges to $(\tilde{z}^i - \hat{z}^i)$ over T. Since $G^i_j(\tilde{z}^i) < 0$, for every $j \in J$, we conclude that every sequence $\{(u^i_j)^t\}_{t\in T}$, with $j \in J$, is bounded.

From (34), we have that

$$\sum_{j\in\tilde{J}} \left(u^i_j\right)^t a^j = \alpha^t - \sum_{j\in J} \left(u^i_j\right)^t \left(S^i_j\right)^t, \tag{37}$$

for every $t \in T$. The right-hand-side of (37) is bounded because $\{(u^i_j)^t\}_{t\in T_1}$, for $j \in J$, is bounded and everything else converges over T_1. Suppose that, for some $j \in \tilde{J}$, the sequence $\{(u^i_j)^t\}_{t\in T_1}$ is unbounded. This would imply that the right-hand-side in (37) would not be bounded, because the set $\{a^j\}_{j\in\tilde{J}}$ is linearly independent. By contradiction, we conclude that every sequence $\{(u^i_j)^t\}_{t\in T_1}$, for every $j \in \tilde{J}$, is also bounded over T_1.

5. Applications and conclusions

In most practical applications of disjunctive convex programming the set P is defined in (2) by a large number p of individual sets, frequently exponential in the number of variables. In these circumstances, the index set I is either unknown or too large. Therefore, our procedure should be applied within a framework that takes care of the dimensionality issue in the form of a cutting-plane algorithm, a branch-and-bound algorithm, or a combination of these.

Basically, the idea behind such a global algorithm is to solve the original problem where p is too large by solving a sequence of problems that are defined by a small number of individual sets. In this setting we envisage two circumstances where our algorithm may be applied. One occurs when looking for a separating hyperplane between a given point $\bar{x}$ and a convex set P defined by (2). For example, in the context of convex programming with integer variables, the point $\bar{x}$ may be the optimal solution of some nonlinear programming relaxation $\bar{P}$ such that some component $\bar{x}_j$ is fractional and it should be either 0 or 1. Then, it might be the case that $\bar{x}$ does not belong to $P \equiv \text{cl conv}\,(K^0 \cup K^1)$, where $K^0 = \bar{P} \cap \{x : x_j = 0\}$ and $K^1 = \bar{P} \cap \{x : x_j = 1\}$. Then, as implied by Stubbs and Mehrotra [13], a separating hyperplane between $\bar{x}$ and P may be found by using the following Fenchel duality result

$$\begin{array}{ll}\max & \alpha\bar{x} - h(\alpha) \\ \text{s.t.} & \|\alpha\|_* = 1,\end{array} = \begin{array}{ll}\min & f(x) \equiv \|x - \bar{x}\| \\ \text{s.t.} & x \in P\end{array} \tag{38}$$

where $h(\alpha) \equiv \max\{\alpha x : x \in P\}$. If $\hat{x}$ is an optimal solution in (38), i.e., $\hat{x}$ is the projection of $\bar{x}$ into the set P and $\hat{\xi} \in \partial f(\hat{x})$ satisfies (18) then the vector $\hat{\alpha} = -\hat{\xi}$ is an optimal dual solution in (38) defining the following valid inequality for P,

$$\hat{\alpha} x \leq h(\hat{\alpha}) \qquad \left(= \hat{\alpha}\hat{x}\right),$$

which is in the all-linear case referred to as the *deepest cut* because it is the valid inequality for P that cuts-off $\bar{x}$ by the largest amount. The value of $\hat{\alpha}$ is available at optimality when the procedure described in the previous sections is used.

Another possible setting occurs when at a given node of the branch-and-bound tree we want a lower bound on the optimal value of the nonlinear programming relaxations associated with all the nodes emanating from the current node. Instead of solving all those problems individually we may consider them as defining the individual sets of some set P. Thus, our procedure would require solving only one convex program though with a larger number of variables.

Finally, following a remark of Claude Lemaréchal we have realized that another type of solution procedure may be envisaged that uses an idea akin to the Bundle method, or Frank–Wolfe's algorithm. This is a primal-dual procedure which is particularly interesting in the generation of cutting-planes for convex programming with integer variables using (38) for two main reasons: one, the procedure can be implemented as a decomposition algorithm and, two, the procedure may be terminated before optimality is achieved while still guaranteeing a cut. This feature is not found in the primal approach described in this paper, in which the cut is only found assimptotically. However, this procedure lacks the usage of second-order objective function information that our approach can use in the same way an interior-point code does. We are in the process of testing both approaches in the context of solving mixed-integer nonlinear programs through a cutting-plane algorithm.

Appendix. Recession and perspective functions

We recall and expand some basic results stated in [7, pages 178–183] for recession functions and [7, pages 160–162] for perspective functions. We consider only the relevant results to the context of this paper. The interested reader may find in [7,8] a broad treatment of this topic.

Given a closed convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the recession function $f'_\infty : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$f'_\infty(d) \equiv \lim_{t \to \infty} \frac{f(x^0 + td) - f(x^0)}{t}, \tag{39}$$

where x^0 is an arbitrary point of $\operatorname{dom}(f)$. If $F : \mathbb{R}^n \to (\mathbb{R} \cup \{+\infty\})^m$ is a vector mapping whose components are closed convex functions then F'_∞ is a vector mapping whose components are the respective recession functions. Recession functions reflect the behavior of f at ∞ along a direction.

It is known that the recession function of a closed convex function is also closed and convex [7, Proposition VI.3.2.2]. Our interest in this function is that if $P =$

$\{x \in \mathbb{R}^n : G(x) \le 0\}$ then $P_\infty = \left\{d \in \mathbb{R}^n : G'_\infty(d) \le 0\right\}$, see [7, Proposition 3.2.5]. It is easy to check that if $G(x) = b - Ax$ then $G'_\infty(d) = -Ad$, but, in general, G'_∞ does not have a closed-form expression.

Given a closed convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the perspective function $\tilde{f} : \mathbb{R}^{n+1} \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\tilde{f}\,(\lambda, x) \equiv \begin{cases} \lambda\, f(v/\lambda) & \text{if } \lambda > 0, \\ +\infty & \text{if } \lambda \le 0. \end{cases} \tag{40}$$

If $F : \mathbb{R}^n \to (\mathbb{R} \cup \{+\infty\})^m$ is a vector mapping whose components are closed convex functions then $\tilde{F}$ is is a vector mapping whose components are the respective perspective functions.

It is known that the perspective function of a closed convex function is convex, see [7, Proposition VI.2.2.1], but it need not be closed. The closure of $\tilde{f}$ is shown in [7, Proposition VI.2.2.2] to be defined by,

$$\left(\text{cl }\tilde{f}\right)(\lambda, x) = \begin{cases} \lambda\, f(x/\lambda) & \text{if } \lambda > 0, \\ \lim_{\lambda \to 0^+} \lambda\, f(\tilde{x} - x + x/\lambda) & \text{if } \lambda = 0, \\ +\infty & \text{if } \lambda < 0, \end{cases} \tag{41}$$

where $\tilde{x}$ is an arbitrary point of ri dom (f). Our interest in this function is that it is employed in deriving a convex algebraic characterization of the set P defined by (2).

Proposition 4 below provides a characterization of the subdifferential of (cl $\tilde{f}$) in terms of subdifferential of f at all the points of interest to the context of this paper.

Proposition 4. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a closed convex function and $(\hat{\lambda}, \hat{x})$ be such that $\hat{\lambda} > 0$. Then, if we define $\hat{z} = \hat{x}/\hat{\lambda}$,*

1. *$\partial\left(cl\ \tilde{f}\right)\left(\hat{\lambda}, \hat{x}\right)$ is empty if and only if $\partial f(\hat{z})$ is empty.*
2. *If $\partial f(\hat{z})$ is nonempty then*

$$\partial\left(cl\ \tilde{f}\right)\left(\hat{\lambda}, \hat{x}\right) = \left\{\left(f(\hat{z}) - \xi\hat{z}, \xi\right) : \xi \in \partial f(\hat{z})\right\}. \tag{42}$$

Proof. Assume that $(\delta, \xi) \in \partial\left(\text{cl }\tilde{f}\right)\left(\hat{\lambda}, \hat{x}\right)$, i.e.,

$$\left(\text{cl }\tilde{f}\right)(\lambda, x) \ge \hat{\lambda}\, f(\hat{x}/\hat{\lambda}) + (\delta, \xi) \begin{bmatrix} \lambda - \hat{\lambda} \\ x - \hat{x} \end{bmatrix}, \tag{43}$$

for any $(\lambda, x) \in \mathbb{R}^{n+1}$. In particular, when $\lambda = \hat{\lambda}$, $\hat{\lambda}\, f(x/\hat{\lambda}) \ge \hat{\lambda}\, f(\hat{z}) + \xi\left(x - \hat{x}\right)$, for every $x \in \mathbb{R}^n$. But, since any $z \in \mathbb{R}^n$ can be written as $z = x/\hat{\lambda}$ then

$$f(z) \ge f(\hat{z}) + \xi\left(z - \hat{z}\right),$$

for every $z \in \mathbb{R}^n$, which shows that that $\xi \in \partial f(\hat{z})$. Moreover, since (43) can be equivalently written as

$$\lambda\left[f(x/\lambda) - \xi\frac{x}{\lambda} - \delta\right] \ge \hat{\lambda}\left[f(\hat{x}/\hat{\lambda}) - \xi\frac{\hat{x}}{\hat{\lambda}} - \delta\right].$$

for any $(\lambda, x) \in \mathbb{R}^{n+1}$ such that $\lambda > 0$, we can easily conclude that $\delta = f(\hat{z}) - \xi\hat{z}$ by suitably choosing values for λ keeping $x/\lambda = \hat{x}/\hat{\lambda}$. Thus, we have proved that $\delta = f(\hat{z}) - \xi\hat{u}$ and $\xi \in \partial f(\hat{z})$.

Now, let $\xi \in \partial f(\hat{z})$, which in particular implies that $\hat{z} \in \text{dom}\,(f)$. We need to prove that

$$\left(\text{cl}\ \tilde{f}\right)(\lambda, x) \geq \hat{\lambda}\, f(\hat{x}/\hat{\lambda}) + \left(f(\hat{z}) - \xi\hat{z}, \xi\right)\begin{bmatrix} \lambda - \hat{\lambda} \\ x - \hat{x} \end{bmatrix}, \tag{44}$$

for every $(\lambda, x) \in \mathbb{R}^{n+1}$. This is trivially true when $\lambda < 0$. It also holds when $\lambda > 0$ because

$$\begin{aligned}\left(\text{cl}\ \tilde{f}\right)(\lambda, x) &\geq \lambda\left[f(\hat{z}) + \xi\left(\frac{x}{\lambda} - \frac{\hat{x}}{\hat{\lambda}}\right)\right] \\ &= \hat{\lambda}\, f(\hat{z}) + \left(f(\hat{z}) - \xi\hat{z}, \xi\right)\begin{bmatrix} \lambda - \hat{\lambda} \\ x - \hat{x} \end{bmatrix}.\end{aligned}$$

In order to prove (44) when $\lambda = 0$ note that, for any t, we have that, for every $x \in \mathbb{R}^n$,

$$\begin{aligned} &\frac{f(\hat{z} + tx) - f(\hat{z})}{t}) \geq \xi x \\ \Longrightarrow\quad & f'_\infty(x) \geq \xi x, \\ \Longleftrightarrow\quad & f'_\infty(x) \geq \hat{\lambda}\, f(\hat{z}) + \left(f(\hat{z}) - \xi\hat{z}, \xi\right)\begin{bmatrix} 0 - \hat{\lambda} \\ x - \hat{x} \end{bmatrix},\end{aligned}$$

from where the desired result follows because $\left(\text{cl}\ \tilde{f}\right)(0, x) = f'_\infty(x)$, see [7, Remark IV.2.2.3]. We have proved both statements.

The characterization of the subdifferential of $(\text{cl}\ \tilde{f})$ at points of the form $(0, x)$ is not so simple or informative. It was shown to us by Claude Lemaréchal ([10]) that if $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a closed convex function then,

$$\partial\left(\text{cl}\ \tilde{f}\right)(0, 0) = \left\{(\delta, \xi) \in \mathbb{R}^{n+1} : \delta + f^*(\xi) \leq 0\right\},$$

where $f^*(\xi) \equiv \sup_{x \in \mathbb{R}^n} \xi x - f(x)$ is the conjugate function of f at ξ. Since

$$\bigcup_{z \in \text{dom}(f)} \{(f(z) - \xi z, \xi) : \xi \in \partial f(z)\} \subseteq \left\{(\delta, \xi) : \delta + f^*(\xi) \leq 0\right\},$$

we conclude that the set $\partial\left(\text{cl}\ \tilde{f}\right)(0, 0)$ is too large to be of value.

References

1. Balas, E. (1975): Disjunctive programming: cutting planes from logical conditions. In: Nonlinear Programming **2**, pp. 279–312
2. Balas, E. (1985): Disjunctive programming and a hierarchy of relaxations for discrete optimization problems. SIAM J. Algebraic Discrete Methods **6**, 466–486
3. Bienstock, D. (1996): Computational study of a family of mixed-integer quadratic programming problems. Math. Program. **74**, 121–140

4. Burke, J., Ferris, M. (1991): Characterization of solution sets of convex programs. Oper. Res. Lett. **10**, 57–60
5. Butera, G. (1997): The solution of a class of limited diversification portfolio selection problems. PhD thesis. Department of Computational and Applied Mathematics, Rice University, Houston, Texas, May 1997
6. Duran, M., Grossmann, I. (1986): An outer-approximation algorithm for a class of mixed-integer nonlinear programs. Math. Program. **36**, 307–339
7. Hiriart-Urruty, J., Lemaréchal, C. (1993): Convex Analysis and Minimization Algorithms, Vol. 1. Springer-Verlag, Berlin Heidelberg New York
8. Hiriart-Urruty, J., Lemaréchal, C. (1993): Convex Analysis and Minimization Algorithms, Vol. 2. Springer-Verlag, Berlin Heidelberg New York
9. Jeroslow, R. (1977): Representability in mixed integer programming, I: Characterization results. Discrete Appl. Math. **17**, 223–243
10. Lemaréchal, C. (1997): Geometric characterization of the subdifferential of a perspective function at points $(0, x)$. Private Communication, May 1997
11. Mangasarian, O. (1988): A simple characterization of solution sets of convex programs. Oper. Res. Lett. **7**, 21–26
12. Sherali, H., Shetty, C. (1980): Optimization with disjunctive constraints. In: Beckman, M., Kunzi, H., eds., Lect. Notes Econ. Math. Systems **181**. Springer-Verlag, Berlin Heidelberg New York
13. Stubbs, R., Mehrotra, S. (1996): A branch-and-cut method for 0-1 mixed convex programming. Technical Report 96-01, Northwestern University, Evanston, Illinois, Jan 1996