Optimality certificates for convex minimization and Helly numbers

Amitabh Basu Michele Conforti Gérard Cornuéjols
Robert Weismantel Stefan Weltge

May 10, 2017

Abstract

We consider the problem of minimizing a convex function over a subset of $\mathbb{R}^n$ that is not necessarily convex (minimization of a convex function over the integer points in a polytope is a special case). We define a family of duals for this problem and show that, under some natural conditions, strong duality holds for a dual problem in this family that is more restrictive than previously considered duals.

1 Introduction

Insights obtained through duality theory have spawned efficient optimization algorithms (combinatorial and numerical) which simultaneously work on a pair of primal and dual problems. Striking examples are Edmonds' seminal work in combinatorial optimization, and interior-point algorithms for numerical/continuous optimization.

Compared to duality theory for continuous optimization, duality theory for mixed-integer optimization is still underdeveloped. Although the linear case has been extensively studied, see, e.g., [4, 5, 12, 13], nonlinear integer optimization duality was essentially unexplored until recently. An important step was taken by Morán et al. for conic mixed-integer problems [11], followed up by Baes et al. [2] who presented a duality theory for general convex mixed-integer problems. The approach taken by Moran et al. was essentially algebraic, drawing on the theory of subadditive functions. Baes et al. took a more geometric viewpoint and developed a duality theory based on lattice-free polyhedra. We follow the latter approach.

Given $S \subseteq \mathbb{R}^n$ and a convex function $f : \mathbb{R}^n \to \mathbb{R}$, we consider the problem

$$\inf_{s \in S} f(s). \tag{1}$$

We describe a geometric dual object that can be used to certify optimality of a solution to (1). For simplicity, let us consider the case when there exists an $x_0 \in \mathbb{R}^n$ such that $f(x_0) \leq f(s)$ for all $s \in S$. We say that a closed set C is an *S-free neighborhood of x_0* if $x_0 \in \text{int}(C)$ and $\text{int}(C) \cap S = \emptyset$. Using the convexity of f, it follows that for any $\bar{s} \in S$ and any C that is an S-free neighborhood of x_0, the following holds:

$$f(\bar{s}) \geq \inf_{z \in \text{bd}(C)} f(z) =: L(C), \tag{2}$$

where $\text{bd}(C)$ denotes the boundary of C (to see this, consider the line segment connecting $\bar{s}$ and x_0 and a point at which this line segment intersects $\text{bd}(C)$). Thus, an S-free neighborhood of x_0 can be interpreted as a "dual object" that provides a *lower bound* of the type (2). As a consequence, the following is true.

Proposition 1 (Weak duality). *If there exist $\bar{s} \in S$ and $C \subseteq \mathbb{R}^n$ that is an S-free neighborhood of x_0, such that equality holds in* (2)*, then $\bar{s}$ is an optimal solution to* (1)*.*

2 The dual problem

This motivates the definition of a dual optimization problem to (1). For any family $\mathcal{F}$ of S-free neighborhoods of x_0, define the $\mathcal{F}$-dual of (1) as

$$\sup_{C \in \mathcal{F}} L(C). \tag{3}$$

Assuming very mild conditions on S and f (e.g., when S is a closed subset of $\mathbb{R}^n$ disjoint from $\arg\inf_{x \in \mathbb{R}^n} f(x)$), it is straightforward to show that if $\mathcal{F}$ is the family of *all* S-free neighborhoods of x_0, then strong duality holds, i.e., there exists $\bar{s} \in S$ and $C \in \mathcal{F}$ such that the condition in Proposition 1 holds. However, the entire family of S-free neighborhoods is too unstructured to be useful as a dual problem. Moreover, the inner optimization problem (2) of minimizing on the boundary of C can be very hard if C has no structure other than being S-free. Thus, we would like to *identify subfamilies $\mathcal{F}$ of S-free neighborhoods that still maintain strong duality, while at the same time, are much easier to work with inside a primal-dual framework.* We list below three subclasses that we expect to be useful in this line of research. First, we need the concept of a *gradient polyhedron*:

Definition 2. *Given a set of points $z_1, \ldots, z_k \in \mathbb{R}^n$,*

$$Q := \{x \in \mathbb{R}^n : \langle a_i, x - z_i \rangle \leq 0, \quad i = 1, \ldots, k\}$$

is said to be a gradient polyhedron *of $z_1, \ldots, z_k$ if for every $i = 1, \ldots, k$, $a_i \in \partial f(z_i)$, i.e., a_i is a subgradient of f at z_i.*

Remark 3. *For every gradient polyhedron* Q *of points* $z_1, \dots, z_k$ *we have* $L(C) = \inf_{z \in \mathrm{bd}(C)} = \min_{i \in [k]} f(z_i)$.

We consider the following families.

- The family $\mathcal{F}_{\max}$ of maximal convex S-free neighborhoods of x_0, i.e., those S-free neighborhoods that are convex, and are not strictly contained in a larger convex S-free neighborhood.

- The family $\mathcal{F}_{\partial}$ of convex S-free neighborhoods of x_0 that are also gradient polyhedra for some finite set of points in $\mathbb{R}^n$.

- The family $\mathcal{F}_{\partial,S} \subseteq \mathcal{F}_{\partial}$ of convex S-free neighborhoods of x_0 that are also gradient polyhedra for some finite set of points in S.

We propose the above families so as to leverage a recent surge of activity analyzing their structure; the surveys [3] and Chapter 6 of [6] provide good overviews and references for this whole line of work. This well-developed theory provides powerful mathematical tools to work with these families. As an example, this prior work shows that for most sets S that occur in practice (which includes the integer and mixed-integer cases), the family $\mathcal{F}_{\max}$ only contains polyhedra. This is good from two perspectives:

- polyhedra are easier to represent and compute with than general S-free neighborhoods,

- the inner optimization problem (2) of computing $L(C)$ becomes the problem of solving finitely many continuous convex optimization problems, corresponding to the facets of C.

guarantees that the cost of these inner optimization problems will not be arbitrarily high.

Of course, the first question to settle is whether these three families actually enjoy strong duality, i.e., do we have strong duality between (1) and the $\mathcal{F}_{\max}$-dual, $\mathcal{F}_{\partial}$-dual and $\mathcal{F}_{\partial,S}$-dual? It turns out that the main result in [2] shows that for the mixed-integer case, i.e., $S = C \cap (\mathbb{Z}^{n_1} \times \mathbb{R}^{n_2})$ for some convex set C, the $\mathcal{F}_{\partial}$-dual enjoys strong duality under conditions of the Slater type from continuous optimization. It is not hard to strengthen their result to also show that the $\mathcal{F}_{\max} \cap \mathcal{F}_{\partial}$-dual is a strong dual, under some additional assumptions.

In this paper, we give conditions on S and f such that strong duality holds for the dual problem (3) associated with $\mathcal{F}_{\max} \cap \mathcal{F}_{\partial,S}$. Below we give an explanation as to why this family is very desirable. If these conditions on S and f are met, our result is stronger than Baes et al. [2]. For example, when S is the set of integer points in a compact convex set and f is any convex function, our certificate is a stronger one. However, our conditions on S and f do not cover certain mixed-integer problems;

whereas, the certificate from Baes et al. still exists in these settings. Nevertheless, it can be shown that in such situations, a strong certificate like ours does not necessarily exist.

3 Strong optimality certificates

Definition 4. *A* strong optimality certificate of size k *for* (1) *is a set of points* $z_1, \ldots, z_k \in S$ *together with subgradients* $a_i \in \partial f(z_i)$ *such that*

$$Q := \{x \in \mathbb{R}^n : \langle a_i, x - z_i \rangle \leq 0,\ i = 1, \ldots, k\} \text{ is } S\text{-free}, \tag{4}$$

$$\langle a_i, z_j - z_i \rangle < 0 \text{ for all } i \neq j. \tag{5}$$

Remark 5. *If a strong optimality certificate exists, then the infimum of f over S is attained and we have $\min_{s \in S} f(s) = \min_{i \in [k]} f(z_i)$. In other words, given a strong optimality certificate, we can compute* (1) *by simply evaluating $f(z_1), \ldots, f(z_k)$.*

Indeed, recall that $a \in \partial f(z)$ means that $f(x) \geq f(z) + \langle a, x - z \rangle$ holds for all $x \in \mathbb{R}^n$. Since Q is S-free, for every $s \in S$ there is some $i \in [k]$ such that $\langle a_i, s - z_i \rangle \geq 0$ and hence $f(s) \geq f(z_i)$.

In order to verify that $z_1, \ldots, z_k$ together with $a_1, \ldots, a_k$ form a strong optimality certificate, one has to check whether the polyhedron Q is S-free. Deciding whether a *general* polyhedron is S-free might be a difficult task. However, Property (5) ensures that Q is *maximal* S-free, i.e., Q is not properly contained in any other S-free closed convex set: Indeed, Property (5) implies that Q is a full-dimensional polyhedron and that $\{x \in Q : \langle a_i, x \rangle = 0\}$ is a facet of Q containing $z_i \in S$ in its relative interior for every $i \in [k]$. Since every closed convex set C that properly contains Q contains the relative interior of at least one facet of Q in its interior, C cannot be S-free.

For particular sets S, the properties of S-free sets that are maximal have been extensively studied and are much better understood than general S-free sets. For instance, if $S = (\mathbb{R}^d \times \mathbb{Z}^n) \cap C$ where C is a closed convex subset of $\mathbb{R}^{n+d}$, maximal S-free sets are polyhedra with at most 2^n facets [10]. In particular, if $S = \mathbb{Z}^2$ the characterizations in [8, 9] yield a very simple algorithm to detect whether a polyhedron is maximal $\mathbb{Z}^2$-free.

In order to state our main result, we need the notion of the *Helly number* $h(S)$ of the set S, which is the largest number m such that there exist convex sets $C_1, \ldots, C_m \subseteq \mathbb{R}^n$ satisfying

$$\bigcap_{i \in [m]} C_i \cap S = \emptyset \quad \text{and} \quad \bigcap_{i \in [m] \setminus \{j\}} C_i \cap S \neq \emptyset \text{ for every } j \in [m]. \tag{6}$$

Theorem 6. *Let $S \subseteq \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function such that* *(i)* $\mathbb{O} \notin \partial f(s)$ *for all* $s \in S$*,*

(ii) $h(S)$ *is finite, and*

(iii) for every polyhedron $P \subseteq \mathbb{R}^n$ *with* $\operatorname{int}(P) \cap S \neq \emptyset$ *there exists an* $s^\star \in \operatorname{int}(P) \cap S$ *with* $f(s^\star) = \inf_{s \in \operatorname{int}(P) \cap S} f(s)$*.*

Then there exists a strong optimality certificate of size at most $h(S)$*.*

Let us comment on the assumptions in Theorem 6. First, if $\mathbb{O} \in \partial f(s^\star)$ for some $s^\star \in S$, then $s^\star$ is an optimal solution to (1) as well as $s^\star \in \arg\inf_{x \in \mathbb{R}^n} f(x)$. An easy certificate of optimality in this case is the subgradient $\mathbb{O}$.

Second, a quite general situation in which (ii) is always satisfied is the case $S = (\mathbb{R}^d \times \mathbb{Z}^n) \cap C$ where $C \subseteq \mathbb{R}^{d+n}$ is a closed convex set. In this situation, one has $h(S) \leq 2^n(d+1)$. The characterization of closed sets S for which $h(S)$ is finite has received a lot of attention, see, e.g., [1].

Third, note that (iii) implies that the minimum in (1) actually exists. As an example, (iii) is fulfilled whenever S is discrete (every bounded subset of S is finite) and the set $\{x \in \mathbb{R}^n : f(x) \leq \alpha\}$ is bounded and non-empty for some $\alpha \in \mathbb{R}$ (implying that the set is actually bounded for every $\alpha \in \mathbb{R}$). This latter condition is satisfied, e.g., when f is strictly convex and has a minimizer. Another situation where (iii) is satisfied is when S is a finite set, e.g., $S = C \cap \mathbb{Z}^n$ where C is a compact convex set.

Also, if conditions (i) and (ii) hold, but (iii) does not hold, a strong optimality certificate may not exist. For example, consider $S = \{x \in \mathbb{Z}^2 : \sqrt{2}x_1 - x_2 \geq 0, \;\; x_1 \geq \frac{1}{2}, \;\; x_2 \geq 0\}$ and $f(x) = \sqrt{2}x_1 - x_2$. In this case, no strong optimality certificate can exist, as the infimum of f over S is 0, but it is not attained by any point in S.

Finally, we remark that Theorem 6 yields a situation in which the following strong duality (see Proposition 1) holds.

Corollary 7 (Strong duality)**.** *If the conditions in Theorem 6 are met, then there exist* $\bar{s} \in S$ *,* $x_0 \in \mathbb{R}^n$ *and* $C \subseteq \mathbb{R}^n$ *such that* C *is an* S*-free neighborhood of* x_0 *and* (2) *holds with equality.*

Proof. Note that conditions (i) and (iii) imply that there exists an $x_0 \in \mathbb{R}^n$ with $f(x_0) < f(s)$ for all $s \in S$. By Theorem 6, there exists a strong optimality certificate defined by points $z_1, \ldots, z_k \in S$. Consider the S-free convex set Q defined in (4). The inequalities in (5) imply that Q is an S-free neighborhood of x_0. Furthermore, by Remarks 3 and 5 we also know that

$$L(Q) = \inf_{z \in \operatorname{bd}(Q)} f(z) = \min_{i \in [k]} f(z_i) = \min_{s \in S} f(s)$$

holds, which yields the claim. □

Note that the above proof actually shows that x_0 in Corollary 7 can be chosen as any point that satisfies $f(x_0) < f(s)$ for all $s \in S$.

4 Proof of Theorem 6

We make use of the following simple observation (see [7, Thm. 3] for a slightly stronger statement). Let $\operatorname{conv}(\cdot)$ denote the convex hull and $\operatorname{vert}(P)$ denote the set of vertices of a polyhedron P.

Lemma 8. *Let $S \subseteq \mathbb{R}^n$ and $V \subseteq S$ finite such that $V = \operatorname{conv}(V) \cap S = \operatorname{vert}(\operatorname{conv}(V))$. Then we have $|V| \leq h(S)$.*

Proof. Let $V = \{v_1, \dots, v_m\}$ and for every $i \in [m]$ let $C_i := \operatorname{conv}(V \setminus \{v_i\})$. Since $V = \operatorname{conv}(V) \cap S = \operatorname{vert}(\operatorname{conv}(V))$, we have $C_i \cap S = V \setminus \{v_i\}$ for every $i \in [m]$. Thus, $C_1, \dots, C_m$ satisfy (6) and hence $m \leq h(S)$. □

We are ready to prove Theorem 6. Let us consider the following algorithm (in fact, we will see that this is indeed a finite algorithm):

$Q_0 \leftarrow \mathbb{R}^n$, $k \leftarrow 1$

while $\operatorname{int}(Q_{k-1}) \cap S \neq \emptyset$:

$$t_k \leftarrow \min\{f(s) : s \in \operatorname{int}(Q_{k-1}) \cap S\} \tag{7}$$

$C_k \leftarrow \{x \in \mathbb{R}^n : f(x) \leq t_k\}$

$$z_k \leftarrow \text{any } s \in \operatorname{int}(Q_{k-1}) \cap S \text{ with } f(s) = t_k \text{ such that } \dim(F_{C_k}(s)) \text{ is largest possible} \tag{8}$$

$a_k \leftarrow$ any point in $\operatorname{relint}(\partial f(z_k))$

$Q_k \leftarrow \{x \in Q_{k-1} : \langle a_k, x - z_k \rangle \leq 0\}$

$k \leftarrow k + 1$

In the above, $\operatorname{relint}(\cdot)$ denotes the relative interior and $\dim(\cdot)$ the affine dimension. For a closed convex set $C \subseteq \mathbb{R}^n$ and a point $p \in C$ we denote by $F_C(p)$ the smallest face of C that contains p.

Remark that iteration k of the algorithm can always be executed, as the set Q_k is a polyhedron and hence by the assumption in (iii) the minimum in (7) always exists. Furthermore, since $a_k \in \operatorname{relint}(\partial f(z_k))$ we have

$$F_k := F_{C_k}(z_k) = \{x \in C_k : \langle a_k, x - z_k \rangle = 0\} \tag{9}$$

Claim 1: For every k we have that $\langle a_i, z_j - z_i \rangle < 0$ holds for all $i, j \leq k$ with $i \neq j$.

Let $k \geq 2$ and assume that the claim is satisfied for all $i, j \leq k - 1$, $i \neq j$. Since $z_k \in \operatorname{int}(Q_{k-1})$ and $a_i \neq \mathbb{O}$ by assumption (i), we have that $\langle a_i, z_k - z_i \rangle < 0$ for every $i < k$.

It remains to show that $\langle a_k, z_i - z_k \rangle < 0$ for every $i < k$. Since $a_k \in \partial f(z_k)$, we have that $\langle a_k, z_i - z_k \rangle \leq f(z_i) - f(z_k)$ and for $i < k$ by (7) we have $f(z_i) \leq f(z_k)$. Therefore

$\langle a_k, z_i - z_k \rangle \leq 0$ and if $\langle a_k, z_i - z_k \rangle = 0$, then $f(z_i) = f(z_k)$. Assume this is the case. Since $\langle a_i, z_k - z_i \rangle < 0$ we have $z_k \notin F_i$ and in particular

$$F_i \neq F_k. \tag{10}$$

By (9) this means that $z_i \in F_k$ holds. Since F_i is the smallest face of $C_i = C_k$ that contains z_i, this implies $F_i \subseteq F_k$. By (8), we have that $\dim(F_i) \geq \dim(F_k)$ and thus $F_i = F_k$, a contradiction to (10).

Claim 2: For every k we have that $V := \{z_1, \ldots, z_k\}$ satisfies $V = \operatorname{conv}(V) \cap S = \operatorname{vert}(\operatorname{conv}(V))$.

It is easy to see that Claim 1 implies $V = \operatorname{vert}(\operatorname{conv}(V))$. For the sake of contradiction, assume there exists some $s \in (\operatorname{conv}(V) \setminus V) \cap S$. By Claim 1, we have $s \in \operatorname{int}(Q_k)$. Therefore by (7) we have $f(s) \geq t_k$. Since f is convex and $s \in \operatorname{conv}(V)$, this implies $f(s) = t_k$. Let $a \in \operatorname{relint}(\partial f(s))$ and consider $F := F_{C_k}(s) = \{x \in C_k : \langle a, x - s \rangle = 0\}$. Since $V \subseteq C_k$, we have that $z_i \in F$ for at least one $i \in [k]$. Due to $\langle a, z_i - s \rangle \leq f(z_i) - f(s)$ we must have $f(z_i) = t_k$ and hence $F_i \subseteq F$. By (8), we further have $\dim(F_i) \geq \dim(F)$, which shows $F_i = F$. However, by Claim 1 we have $z_j \notin F_i$ for all $j \neq i$ and hence $s \notin F_i$, a contradiction since $s \in F$.

Claim 3: The algorithm stops after at most $h(S)$ iterations and $Q := Q_k$ is S-free.

Note that the set $V := \{z_1, \ldots, z_k\}$ becomes larger in every iteration. By Claim 2 and Lemma 8 we must have $k \leq h(S)$ and hence the algorithm stops after at most $h(S)$ iterations. Since the algorithm stops if and only if Q_k is S-free, this proves the claim. □

Acknowledgements

This work was performed during a stay of the first three authors at ETH Zürich. Their visit had been funded by the Research Institute of Mathematics (FIM). Furthermore, part of this work was supported by NSF grant 1560828 and ONR grant 00014-15-12082. We are grateful for this support.

We would also like to thank the anonymous referee for her/his valuable comments and suggestions to improve the quality of the paper.

References

[1] Gennadiy Averkov. On maximal S-free sets and the Helly number for the family of S-convex sets. *SIAM Journal on Discrete Mathematics*, 27(3):1610–1624, 2013.

[2] Michel Baes, Timm Oertel, and Robert Weismantel. Duality for mixed-integer convex minimization. *Mathematical Programming*, 158(1):547–564, 2016.

[3] Amitabh Basu, Michele Conforti, and Marco Di Summa. A geometric approach to cut-generating functions. *Mathematical Programming*, 151(1):153–189, 2015.

[4] Charles E Blair and Robert G Jeroslow. Constructive characterizations of the value-function of a mixed-integer program I. *Discrete Applied Mathematics*, 9(3):217–233, 1984.

[5] Charles E Blair and Robert G Jeroslow. Constructive characterizations of the value function of a mixed-integer program II. *Discrete Applied Mathematics*, 10(3):227–240, 1985.

[6] Michele Conforti, Gérard Cornuéjols, and Giacomo Zambelli. *Integer programming*, volume 271. Springer, 2014.

[7] Michele Conforti and Marco Di Summa. Maximal \$s\$-free convex sets and the helly number. *SIAM Journal on Discrete Mathematics*, 30(4):2206–2216, 2016.

[8] Santanu S. Dey and Laurence A. Wolsey. Two row mixed-integer cuts via lifting. *Mathematical Programming*, 124:143–174, 2010.

[9] Cor A J Hurkens. Blowing up convex sets in the plane. *Linear Algebra and its Applications*, 134:121–128, 1990.

[10] Diego A Morán R and Santanu S Dey. On maximal S-free convex sets. *SIAM Journal on Discrete Mathematics*, 25(1):379–393, 2011.

[11] Diego A Morán R, Santanu S Dey, and Juan Pablo Vielma. A strong dual for conic mixed-integer programs. *SIAM Journal on Optimization*, 22(3):1136–1150, 2012.

[12] Laurence A Wolsey. The b-hull of an integer program. *Discrete Applied Mathematics*, 3(3):193–201, 1981.

[13] Laurence A Wolsey. Integer programming duality: Price functions and sensitivity analysis. *Mathematical Programming*, 20(1):173–195, 1981.