Closing the Duality Gap on Integer Linear Programming

Scott Sirri
Washington University in St. Louis
scottsirri@wustl.edu

2024 60th Annual Allerton Conference on Communication, Control, and Computing (Allerton) | 979-8-3315-4103-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ALLERTON63246.2024.10735274

***Abstract*—Dual formulations for optimization problems provide a new way to view the same problem, revealing its structure and bounding optimal solutions. A strong dual formulation constructs a problem that can be solved in place of or alongside the primal and converges to the same optimal objective function value. In this paper we examine the general dual framework for nonlinear programs suggested by Everett and developed by Gould. This is leveraged to provide a novel strong dual formulation for general and 0-1 integer linear programming with qualitatively different structure from other known strong dual formulations. In particular, the dual problem optimizes three variables subject to a bounded, exponential number of constraints. The zero duality gap and small domain may prove useful in developing new techniques for solving integer linear programs.**

***Index Terms*—Duality, Lagrangian multipliers, Integer linear programming, Nonconvex optimization**

I. Introduction

This section introduces the topic under consideration and provides context for the state of the art.

Constrained optimization problems typically involve the maximization or minimization of some *objective function* subject to a set of *equality* or *inequality constraints* on the variables. *Nonconvex optimization* problems are notorious for their intractability. Significant resources are expended in developing sufficiently efficient techniques for solving them.

One common approach is to leverage *duality theory*. This involves the formulation of a related *dual problem*, whose solution tells you something about the solution to the original, or *primal*, problem. In some scenarios, it is faster to solve the primal and dual problems in tandem than to solve either separately. In other scenarios, the dual problem is significantly easier to solve than the primal. The difference between the optimal objective function values of the pair is called the *duality gap*. For many formulations, the only guarantee that can be provided is that of *weak duality*, i.e., that the optimal objective function values of the primal and dual provide upper or lower bounds on each other. In some cases, however, the dual formulation guarantees *strong duality*, in which the optimal objective function values of the twin problems are equal. This is a highly desirable property for the flexibility it enables in how one may go about solving the problems. This applies particularly to nonconvex optimization on account of how inherently difficult the primal problems may be, and so this is an active research area spanning a range of disciplines [7], [8], [11].

One novel formulation, proposed by [4] and extended by [5], introduces a generalized form of Lagrange multipliers. Lagrangian duality is a canonical approach to a variety of problems, constructing a dual problem with scalar variables. These two papers, along with a number of other contemporary works [1], [3], [12], all of which are cogently surveyed in [14], conceive of the set of Lagrange multipliers as defining a linear function and then replace this with a more general function. The Lagrangian duality problem then becomes an optimization problem over a set of functions. The domain, i.e., the set of functions that is selected, determines the tractability of the dual problem as well as whether strong duality holds. Classes of functions which are sufficient for strong duality are given by [14] for integer programming, integer linear programming, and nonlinear programming under certain conditions. However, these are for the most part of only theoretical interest on account of the high- or infinite-dimensional domains they optimize over.

In particular, a *superadditive dual* is introduced for integer linear programming optimizing over the set of superadditive nondecreasing functions [6]. It's proved that there exists a superadditive, nondecreasing function solving it which trivially guarantees strong duality. Therefore, if the domain is the set of superadditive, nondecreasing functions, then it must contain an optimal solution. This particular function is, however, much more difficult to find than it is to just solve the primal. One alternate approach is to construct a solution in stages corresponding to the Chvátal ranks of a polytope, essentially encoding the operation of a branch-and-cut algorithm in a single function [6]. There is additionally a *branch and bound dual* and an *inference dual* which provide strong dual formulations for integer linear programming. However, these are in practice structured and solved according to branch-and-bound techniques, meaning that they are unlikely to yield lower runtimes and do not seem to lend new insight into the composition of the problem [6]. Our understanding is that, as concisely put by Bertsimas and Tsitsiklis, "unlike linear programming, integer [linear] programming does not have a strong duality theory" [2].

A. Problem Statement

Integer linear programming is the problem of maximizing a linear objective function subject to a set of integer equality or inequality constraints and with the variables' domain constrained to a subset of the integers. This problem is nonconvex on account of its discrete domain, and its decision variant is NP-complete. A strong dual formulation for integer linear programming generates a related program with the same optimal objective function value.

B. Main Results

This paper presents a novel strong dual formulation for integer linear programming with a finite-dimensional domain.

C. Contributions

This paper applies the generalized dual framework introduced by [4], extended by [5], and surveyed by [14] to develop this novel dual formulation.

II. Background

This section provides background knowledge useful for understanding the results of this paper.

Linear programming is the problem of optimizing a linear objective function subject to linear equality or inequality constraints on variables usually understood to be spanning the real numbers. This problem is well-understood and enjoys a wealth of efficient solutions. In particular, for a linear program $\max\{cx : Ax \leq b, x \in \mathbb{R}^n\}$, there exists a dual problem $\min\{ub : uA = c, u \geq 0, u \in \mathbb{R}^n\}$ where, for two vectors u, v, the expression $u \leq v$ denotes that $u_i \leq v_i$ for every entry. The expression $u < v$ is defined analogously.

Definition II.1. Consider a maximization (minimization) problem P which is dual to a minimization (maximization) problem D. *Weak duality* holds for this system if it is guaranteed that the optimal objective function value for P is less (greater) than or equal to that of D; *strong duality* holds if they are guaranteed to be equal.

Without loss of generality, it will be assumed in the rest of this paper that primal problems are maximization instances and dual problems are minimization instances; the same results hold if these are swapped.

It is well known that this dual formulation for linear programming exhibits strong duality. Therefore, their optimal objective function values are equal if both of the problems are feasible and bounded.

The problem is drastically changed in a qualitative way if we add the constraint that solution vectors must be integral, i.e., $\max\{cx : Ax \leq b, x \in \mathbb{Z}^n\}$. This is because linear programming is a convex problem, meaning that any local optimum is also a global optimum. The integrality requirement turns this into a nonconvex problem and eliminates that assurance. An optimal solution to a linear programming problem, if it exists, will be at a vertex of the convex polytope representing the feasible solutions. But the integrality requirement may take this optimum and displace it to an integer vector which is difficult to find. A variety of approaches have been proposed to solve such problems, but they are little better than an exhaustive search in the worst-case scenario.

Now consider the constrained optimization problem

$$\text{(P)}\quad \begin{aligned} \max_x \quad & f(x), \\ & g(x) \leq b, \\ & x \in X, \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $b \in \mathbb{R}^m$, and $X \subseteq \mathbb{R}^n$. A common tool for such problems is *Lagrangian duality*. We define this problem's *Lagrangian function* $L(x, \lambda) = f(x) + \lambda^T(b - g(x))$, its *dual function* $J(\lambda) = \max_x L(x, \lambda)$, and construct its Lagrangian dual, which is simply

$$\text{(D)}\quad \min_{\lambda \geq 0} \; J(\lambda).$$

Note that for all feasible (x, λ), it must hold that $f(x) \leq J(\lambda)$. In particular, $\max_{x \in X} f(x) \leq \min_{\lambda \geq 0} J(\lambda)$ if $g(x) \leq b$. Therefore, weak duality holds for problems P and D.

In [10], the authors reformulated this pair of problems as $\max_{x \in X} \min_{\lambda \geq 0} L(x, \lambda)$ and $\min_{\lambda \geq 0} \max_{x \in X} L(x, \lambda)$, respectively, and interpreted them as a pair of related two-player, zero-sum games. Player A chooses x and wants to maximize their objective function, while player B chooses λ and wants to minimize their own objective function.

In the primal, player A goes first and chooses an x to maximize $\min_{\lambda \geq 0} L(x, \lambda)$. Player A has to be careful to not choose an x that will result in $b - g(x)$ having a negative entry, otherwise then player B can choose a λ vector with an arbitrarily large positive value in that entry to make the objective function an arbitrarily large negative number. Therefore, it's in the interests of player A to choose an x that will make $f(x)$ as large as possible without making any entry of $b - g(x)$ negative, i.e., $\max_{x \in X} f(x)$ subject to $g(x) \leq b$. That's why $\max\{f(x) : g(x) \leq b, x \in X\} = \max_{x \in X} \min_{\lambda \geq 0} L(x, \lambda)$.

The dual problem, on the other hand, has the order of play reversed. Player B goes first and chooses a λ to minimize $\max_{x \in X} L(x, \lambda)$. It would be unwise for Player B to make all the λ_i equal zero because then Player A would be unconstrained and might make $f(x)$ arbitrarily large. However, Player B has to be thoughtful in their choice because Player A may also be able to make certain entries of $b - g(x)$ large by leaving slack in the primal constraint. In that case, if any λ_i are carelessly chosen, then Player A could make the objective function large.

We've only stated the weak duality of this system because strong duality is not guaranteed to hold in general. And yet, the only difference between $\max_{x \in X} \min_{\lambda \geq 0} L(x, \lambda)$ and $\min_{\lambda \geq 0} \max_{x \in X} L(x, \lambda)$ is the order in which Players A and B make their moves. Whichever player makes the first move is at a disadvantage because their opponent can react optimally. Intuitively, if we were to enlarge the space of strategies which the players could choose from, then that might give the first player a sufficient advantage to close this duality gap.

A generalized Lagrangian dual, called the $\mathcal{F}$-dual of problem P, is given in [14] of the form

$$\text{(D}_{\text{G}}\text{)}\quad \begin{aligned} \min_F \quad & F(b), \\ & F(g(x)) \geq f(x) \;\; \forall x \in X, \\ & F \in \mathcal{F} \subseteq \mathcal{F}_+^m, \end{aligned}$$

where $\mathcal{F}_+^m$ is the set of nondecreasing functions $F : \mathbb{R}^m \to \mathbb{R}$, meaning $u \leq v \implies F(u) \leq F(v)$. If x, F are feasible for their respective problems, then $f(x) \leq F(g(x)) \leq F(b)$. The first inequality holds by dual feasibility, and the second inequality holds by primal feasibility coupled with F being nondecreasing. In particular, these hold for the optimal (x^*, F^*) and so weak duality holds between problems P and D_G.

We denote the *value function* of problem P by $\phi : \mathbb{R}^m \to \overline{\mathbb{R}}$ and define it, as in [14], $\phi(d) = \sup_x\{f(x) : g(x) \leq d, x \in X\}$. Note that if there is no x such that $g(x) \leq d$, then $\phi(d) = -\infty$. This function gives the optimal objective function value for problem P subject to a specific loosening or tightening of its constraints. It's not hard to see that $u \leq v \implies \phi(u) \leq \phi(v)$ because a loosening of P's constraints can't decrease its optimal objective function value, so $\phi \in \mathcal{F}_+^m$. Therefore if $\mathcal{F} = \mathcal{F}_+^m$, then $\phi \in \mathcal{F}$ and strong duality trivially holds between problems P and D_G. However, ϕ is of course even more general and difficult to find than the optimal solution to problem P. Therefore for a given problem, it is advantageous to identify a small subset of $\mathcal{F}_+^m$, not containing ϕ but still sufficient to guarantee strong duality.

We now consider a formulation equivalent to problem D_G, proposed by [5], which lends some geometric insight:

$$(\mathrm{D}'_\mathrm{G}) \qquad \begin{aligned} \min_F \quad & F(b), \\ & F(d) \geq \phi(d) \ \ \forall d \in \mathbb{R}^m, \\ & F \in \mathcal{F} \subseteq \mathcal{F}_+^m. \end{aligned}$$

Lemma 1 (Lemma 2.8 from [14]). *Problems D_G and D'_G are equivalent.*

Proof.

$$\begin{aligned} &\{F : F(g(x)) \geq f(x) \ \ \forall x \in X\} \\ =&\{F : F(d) \geq f(x) \ \ \forall[(d, x) : d \geq g(x), x \in X]\} \\ =&\{F : F(d) \geq \phi(d) \ \ \forall d \in \mathbb{R}^m\} \end{aligned}$$

□

The formulation of problem D'_G provides an interesting new geometric perspective on the problem. If we can characterize the shape of ϕ, then we only need $\mathcal{F}$ to be a class of nondecreasing functions, at least one of which upper-bounds ϕ everywhere and equals it at b. That will be sufficient to guarantee strong duality.

This paper will largely focus on the problem of 0-1 integer linear programming, whose primal and corresponding pair of duals, formulated as above, are given below:

$$(\mathrm{P}_{\text{0-1 ILP}}) \qquad \begin{aligned} \max_x \quad & c^T x, \\ & Ax \leq b, \\ & x \in \{0,1\}^n, \end{aligned}$$

$$(\mathrm{D}_{\text{0-1 ILP}}) \qquad \begin{aligned} \min_h \quad & h(b), \\ & h(Ax) \geq c^T x \ \ \forall x \in \{0,1\}^n, \\ & h \in \mathcal{F} \subseteq \mathcal{F}_+^m, \end{aligned}$$

$$(\mathrm{D}'_{\text{0-1 ILP}}) \qquad \begin{aligned} \min_h \quad & h(b), \\ & h(d) \geq \phi(d) \ \ \forall d \in \mathbb{R}^m, \\ & h \in \mathcal{F} \subseteq \mathcal{F}_+^m \end{aligned}$$

for $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$. The resulting theorem will then be extended to general integer linear programming.

Additionally, a brief review of simplices and barycentric coordinates is in order. An *n-simplex* is a generalization of the triangle or tetrahedron to n dimensions. It is the n-dimensional convex hull of its exactly $n+1$ vertices, and it is the polytope with the fewest vertices that requires n dimensions. A simplex is a point for $n = 0$, a line segment for $n = 1$, a triangle for $n = 2$, a tetrahedron for $n = 3$, a 5-cell for $n = 4$, et cetera. A simplex has an *orthogonal corner* if it has some vertex with all incident edges pairwise orthogonal, and a simplex can only have at most one orthogonal corner. In this paper, we refer to the $(n-1)$-face "opposite" a simplex's orthogonal corner as its *base*; an n-simplex's base is defined by the set of n vertices connected to but not containing the orthogonal corner.

The barycentric coordinate system is a useful tool for uniquely describing every point inside a simplex. The coordinates for a point u relative to an n-simplex are, intuitively, the values of the $n+1$ masses λ_i you would place at the vertices of the simplex if you wanted u to be their center of mass. At least one of the masses λ_i is negative if and only if u is outside the simplex. It is usually required that the coordinates are normalized, i.e., $\sum_{i=1}^{n+1} \lambda_i = 1$. That way, for a given n-simplex with its $n+1$ vertices denoted s^i, you can write $u = \sum_{i=0}^{n} \lambda_i s^i$. If there are more than $n+1$ points s^i being used in the barycentric expression of a point u relative to an n-simplex, then u is not uniquely determined by a single specific set $\{\lambda_i\}$. The authors choose to call an n-simplex *open* if it does not contain any point which is an affine combination of fewer than $n+1$ of its vertices (under the constraint that $\lambda_i \in [0, 1]$). Therefore, every point in an open simplex has $\lambda_i > 0$ for every coordinate i. Finally, we denote by $[m]$ the set $\{i : 1 \leq i \leq m, i \in \mathbb{Z}\}$.

III. A Sufficient Class for 0-1 ILP

This section explains and proves the sufficiency of a class of functions which, serving as the domain of a dual problem for 0-1 integer linear programming (problems $D_{\text{0-1 ILP}}$ and $D'_{\text{0-1 ILP}}$), guarantees strong duality. On a first reading of this paper, the intuition provided by section III-A may be a more productive place to start before returning here.

These functions will be called the class $\mathcal{H}$ of piecewise linear *hatch functions* for the shape they take in low-dimensional instances of 0-1 ILP. We denote by $\mathbf{1}_m$ the $\mathbb{R}^m$ vector containing 1 in every entry, and we denote by ϵ (as will be more thoroughly explained later) the radius of a small open neighborhood around b, which our result requires such that $\phi(b) = \phi(d)$ for all d in the neighborhood. All functions in the set $\mathcal{H}$ of hatch functions are of the form $h : \mathbb{R}^m \to \mathbb{R}$, have an associated open simplex $S_h \subseteq \mathbb{R}^m$, and are defined piecewise as follows:

$$h(d) = \begin{cases} \alpha \mathbf{1}_m^T (b - \frac{l}{\sqrt{m}} \mathbf{1}_m) + \beta & \text{if } d \in S_h \\ \alpha \mathbf{1}_m^T d + \beta & \text{otherwise} \end{cases} \tag{1}$$

where $\alpha, \beta, l \in \mathbb{R}$ and $b \in \mathbb{R}^m$ are parameters of h satisfying $\alpha > 0, l \geq 0$ (when the primal is a maximization problem). Note that b will be fixed by the problem you're seeking a hatch function for.

The simplex S_h corresponding to $h \in \mathcal{H}$ has exactly $m+1$ vertices at the points

$$\begin{aligned} s^0 &= b + \frac{\epsilon}{\sqrt{m}} \mathbf{1}_m \in \mathbb{R}^m, \\ s^i &= s^0 - (l+\epsilon)\sqrt{m} e_i \in \mathbb{R}^m \text{ for } i \in [m], e_i \in \mathbb{R}^m \text{ standard} \\ &\qquad \text{basis vector.} \end{aligned} \tag{2}$$

Observe that $w < s^0$ for every $w \in S_h$. This follows because, for every $w \in S_h$, the normalized barycentric representation of w with coordinates $\lambda_0 \in \mathbb{R}, \lambda \in \mathbb{R}^m$ is given by

$$\begin{aligned} w &= \sum_{i=0}^{m} \lambda_i s^i = \lambda_0 s^0 + \sum_{i=1}^{m} \lambda_i (s^0 - (l+\epsilon)\sqrt{m} e_i) \\ &= s^0 - (l+\epsilon)\sqrt{m} \sum_{i=1}^{m} \lambda_i e_i, \\ w_i &= (b_i + \frac{\epsilon}{\sqrt{m}}) - \lambda_i (l+\epsilon)\sqrt{m} < b_i + \frac{\epsilon}{\sqrt{m}} = s_i^0 \end{aligned} \tag{3}$$

because S_h is open, so $\lambda_i > 0$. Additionally, note that for each $i > 0$, it follows that $h(s^i) = \alpha \mathbf{1}_m^T (b - \frac{l}{\sqrt{m}} \mathbf{1}_m) + \beta$, which can be verified by plugging in their definition to the hyperplane equation.

A. Intuition

To find an appropriate dual domain that guarantees strong duality, we need a set of nondecreasing functions of the form $h : \mathbb{R}^m \to \mathbb{R}$ that can upper-bound ϕ and fulfill $h(b) = \phi(b)$. We first must intuitively characterize the shape of ϕ. The below description is informal to give an idea of why hatch functions are appropriate.

For a 0-1 ILP, the first things to recognize are that ϕ exists at every point $d \in \mathbb{R}^m$, is discontinuous, its gradient is 0 when it exists, and it's nondecreasing. This already gives us quite a bit of information about how it will look. Its graph will consist of a set of level *platforms*, each corresponding to an optimal objective function value for instances of problem $P_{\text{0-1 ILP}}$ with constraint vectors near each other. The platforms get monotonically higher the further they are from the origin.

A function h that we choose for problem $D'_{\text{0-1 ILP}}$ has to not only upper-bound ϕ everywhere, but it may need to be able to coincide with points in the interior of platforms without ever decreasing on the way there from the origin. If h ever exceeds $\phi(b)$ at some point $d \leq b$, then it cannot decrease back down to $\phi(b)$ before its surface passes over b. This is restrictive because the gradient on the platforms is 0 where it exists, so this suggests that h may be a piecewise function with at least one of those pieces a level surface with a constant gradient of 0 that can coincide with platforms.

Consider the example 0-1 ILP

$$\begin{aligned} \max_{x} \quad & 2x_1 + 4x_2 + 6x_3 + 8x_4, \\ & 8x_1 + 4x_2 + 6x_3 + 10x_4 \leq 11, \\ & x_i \in \{0,1\}. \end{aligned} \tag{4}$$

The functions h and ϕ both have domain $\mathbb{R}$, so they are depicted as a two-dimensional graph in figure 1. We will take $b = 11$ to be the problem that we're interested in, and that value of b is depicted as a purple dot on the graph among all the other possible values for b.

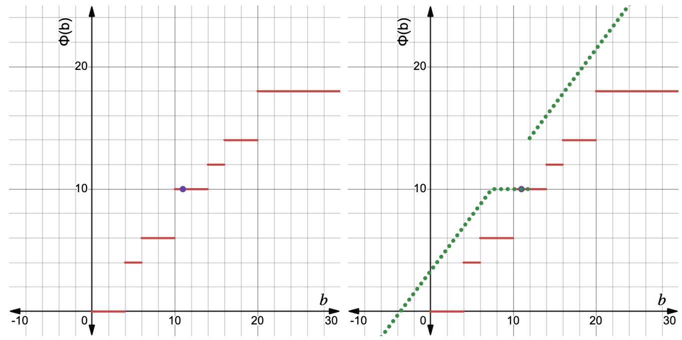

Fig. 1. The value function (red solid, left) and hatch function (green dotted, right) for problem 4 with $m = 1$ and $b = 11$ (purple dot)

The shape of ϕ agrees with our inferences, and so the first step is to choose a function that can upper-bound it everywhere. It is preferable to select a function that requires few parameters to describe it, so we choose a linear function with positive slope. It's easy to find a function satisfying this upper-bound property just by looking at the graph, but now we need to replace a segment of it so that $h(b) = \phi(b)$ without violating the nondecreasing requirement. However, we've already described exactly what we need that to look like: a level, flat surface with a constant gradient of 0 that can rest on top of platforms. So we simply change h to have slope 0 once it reaches $h(d) = \phi(b)$ for some $d \leq b$, it holds constant until it reaches b, and then snaps back up to its linear function. This can be informally analogized to unlocking a hatch that's built into the linear function of h and allowing it to swing down, coming to rest on top of the platform: hence the name of the "hatch functions". It clearly is still a nondecreasing function in figure 1, and therefore we've characterized a sufficient class of functions for $m = 1$.

Now we need to extend this intuition to $m = 2$. Consider the example 0-1 ILP

$$\begin{aligned} \max_{x,y} \quad & x + 2y, \\ & 2x + 4y \leq 5, \\ & 3x + y \leq 3, \\ & x, y \in \{0,1\}. \end{aligned} \tag{5}$$

The functions h and ϕ now have domains of $\mathbb{R}^2$, so we need a three-dimensional graph in figure 2. The shape of ϕ still agrees with our inferences and there's nothing surprising there, but we now need to extend our definition of h such that it remains nondecreasing in higher-dimensional spaces. We intuitively generalize our line with positive slope to a plane with gradient $(1,1)$. We can guess that the hatch will once again be some flat, level surface with gradient 0 hanging off the underside of the linear function, and now must choose the shape of the hatch. We can't allow there to be any pair of points $u \leq v$ such that v is on the hatch and u is a point neighboring the hatch at a height greater than that of v. Intuitively, we can't allow an ant walking around on h, constrained to walk only along the $+x$ or $+y$ directions, to ever fall onto the hatch from a neighboring point of greater height. This is simply resolved by making the hatch a right triangle with its base perpendicular to the gradient, while the other two are each parallel to a respective axis that b varies along. The triangle is rotated so that the only way to fall onto the hatch from a neighboring point is for the ant to walk in the direction of some vector in $\mathbb{R}^2$ that has at least one negative component. The hatch function is now nondecreasing at every point and the location of the hatch can be adjusted so it rests on the platform of b, covering the point corresponding to b itself it with just its "tip", or orthogonal corner.

Generalizations to higher dimensions reveal that the hatch, using an approach analogous to this, is always a simplex with an orthogonal corner, with each orthogonal edge parallel to an axis in $\mathbb{R}^m$ and the base orthogonal to $\mathbf{1}_m$. This is a generalization of our observation that two of the triangle's edges must each be parallel to an axis so as to maintain the nondecreasing property. The simplex is embedded in a hyperplane and the function's value on the simplex is always equal to that of the hyperplane along its interface with the simplex (where the interface is a point for $m = 1$, a line segment for $m = 2$, a face for $m = 3$, et cetera).

In the rest of the paper, the variables l and ϵ are frequently used as parameters for the hatch function and problem under consideration, respectively. To give them a geometric meaning, l is the distance from the base of the simplex to b, and ϵ is the radius of the open neighborhood around b which is level on ϕ. That's why the length of the hatch, from base to orthogonal corner, is $l+\epsilon$. However, the math is expressed in terms of rotated vectors whose Euclidean lengths therefore depend on their dimension. As a result, the displacement from b to the base of the simplex is always expressed as $-\frac{l}{\sqrt{m}}\mathbf{1}_m$ and the displacement from b to the orthogonal corner of the simplex is $+\frac{\epsilon}{\sqrt{m}}\mathbf{1}_m$. And the value $(l+\epsilon)\sqrt{m}$ in eq. (2) is just the length for each edge in the set of orthogonal edges for a simplex with an orthogonal corner and height $l+\epsilon$.

In the bottom picture of figure 2, the black line from the hypotenuse of the triangle to the red dot at b represents the length l, and the circle around the red dot represents the neighborhood of radius ϵ.

B. Functions in $\mathcal{H}$ are nondecreasing

In this section, we will prove that every function in $\mathcal{H}$ is nondecreasing. For convenience of notation below, we will also define the function $h' : \mathbb{R}^m \to \mathbb{R}^{m+1}$ as $h'(d) = \begin{pmatrix} d \\ h(d) \end{pmatrix}$. To clarify, h' is not a member of $\mathcal{H}$.

Theorem 1. *Let $h \in \mathcal{H}$ and $u \leq v$ for any $u, v \in \mathbb{R}^m$. Then h is nondecreasing, i.e., $h(u) \leq h(v)$.*

Proof. We will consider the four exhaustive cases below and prove $h(u) \leq h(v)$ in each:

a) Case 1: If u, v are both inside S_h, then $h(u) = h(v)$ and so the nondecreasing property is fulfilled.

b) Case 2: If u, v are both outside S_h, then $u \leq v$ means that $u_i \leq v_i$ for all $i \leq m$. As a result, $h(u) = \alpha \mathbf{1}_m^T u + \beta = \alpha \sum_i u_i + \beta \leq \alpha \sum_i v_i + \beta = h(v)$ and so the nondecreasing property is fulfilled.

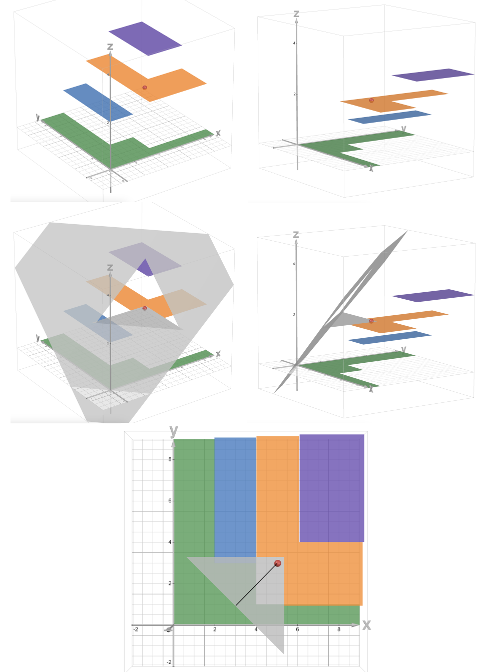

Fig. 2. The value function (color coded by the value of $\phi(d)$, top two), hatch function (grey, middle two), and isolated hatch (grey, bottom) for problem 5 with $m = 2$ and $b = (5, 3)$ (red dot)

c) Case 3: If u is inside S_h and v is not, then we first must prove that $h'(u)$ and $h'(v)$ are both in the positive open half-space (in the direction that the normal points) of the hyperplane $K \subseteq \mathbb{R}^{m+1}$ defined by $N \cdot (d - h'(b - \frac{l}{\sqrt{m}}\mathbf{1}_m)) = 0$, where $N \in \mathbb{R}^{m+1}$ is the normal vector of K with $N_i = 1$ for $i < m+1$ and $N_{m+1} = m\alpha$. Note that N is orthogonal to the normal of the hyperplane $d_{m+1} = \alpha \mathbf{1}_m^T d + \beta$ in h, so the two hyperplanes are orthogonal. Some algebra shows that the subset of $\mathbb{R}^m$ corresponding to their intersection is given by:

$$\begin{aligned} \alpha \mathbf{1}_m^T d + \beta &= \alpha \mathbf{1}_m^T (b - \frac{l}{\sqrt{m}}\mathbf{1}_m) + \beta \\ \sum_{i=1}^m d_i &= \sum_{i=1}^m b_i - \sqrt{m} l. \end{aligned} \tag{6}$$

To understand why $h'(u)$ and $h'(v)$ are on the positive side of K, we first note that $h'(s^i) \in K$ for $i > 0$. This may be verified by plugging in to the equation for K. Additionally, $h'(s^0)$ is in the positive *open* half-space of K because $h'(b - \frac{l}{\sqrt{m}}\mathbf{1}_m) \in K$ by construction and the angle that the free vector $h'(s^0) - h'(b - \frac{l}{\sqrt{m}}\mathbf{1}_m)$ makes with N is acute, as seen by evaluating their dot product:

$$\begin{aligned} &N \cdot (h'(s^0) - h'(b - \frac{l}{\sqrt{m}}\mathbf{1}_m)) \\ =&\mathbf{1}_m \cdot (\frac{\epsilon + l}{\sqrt{m}}\mathbf{1}_m) + m\alpha(m\alpha\frac{\epsilon + l}{\sqrt{m}}) \\ =&(1 + m\alpha^2)\sqrt{m}(\epsilon + l) > 0. \end{aligned} \tag{7}$$

Therefore, the projection of S_h onto the surface of h is contained entirely in the positive open half-space of K because m of its vertices are in K, the last vertex is in the positive open half-space, and S_h is a convex open polytope. Now recall that u is inside S_h, therefore $h'(u)$ is in the positive open half-space of K.

Now suppose for contradiction that $h'(v)$ is in the negative closed half-space of K. Then because $u \leq v$ by their definition, there must exist some $w \in \mathbb{R}^m$ such that $u \leq w \leq v$ and $h'(w)$ is in the negative *open* half-space of K. Consider some $p \in K$. We know $N \cdot (u - p) > 0$ and $N \cdot (w - p) < 0$, and therefore $N \cdot (u - w) = N \cdot (u - p) + N \cdot (p - w) > 0$. And yet, $u_i - w_i \leq 0$ and every entry of N is positive, so it must be that $N \cdot (u - w) \leq 0$. Hence we derive our contradiction, and so $h'(v)$ is in the positive open half-space of K.

We denote $h(d)$ by d_{m+1} for clarity. Now, by setting their equations equal to each other, we can see that the intersection of the graph of $d_{m+1} = \alpha \mathbf{1}_m^T d + \beta$ with K defines a level curve (which makes sense because K is orthogonal to the former's constant gradient):

$$
\begin{aligned}
\frac{-1}{\alpha} d_{m+1} + \sum_{i=1}^{m} d_i + \frac{\beta}{\alpha} =& N \cdot d - N \cdot h'(b - \frac{l}{\sqrt{m}} \mathbf{1}_m) \\
\frac{-1}{\alpha} d_{m+1} + \frac{\beta}{\alpha} =& m\alpha d_{m+1} - \sum_{i=1}^{m} (b_i - \frac{l}{\sqrt{m}}) \\
& - m\alpha(\alpha \sum_{i=1}^{m} (b_i - \frac{l}{\sqrt{m}}) + \beta) \\
\frac{m\alpha^2 + 1}{\alpha} d_{m+1} =& (1 + m\alpha^2) \sum_{i=1}^{m} (b_i - \frac{l}{\sqrt{m}}) \\
& + \beta \frac{m\alpha^2 + 1}{\alpha} \\
d_{m+1} =& \alpha \sum_{i=1}^{m} (b_i - \frac{l}{\sqrt{m}} \mathbf{1}_m) + \beta \\
d_{m+1} =& h(b - \frac{l}{\sqrt{m}} \mathbf{1}_m).
\end{aligned} \tag{8}
$$

The hyperplanes of K and h contain the point $h'(b - \frac{l}{\sqrt{m}} \mathbf{1}_m)$ by definition of the hyperplane and by definition of the point itself, respectively, and so the fact that their level curve intersection has its value makes sense. As a result, the level curve tells us that $h(b - \frac{l}{\sqrt{m}} \mathbf{1}_m) \leq h(w)$ for every $h'(w)$ in the positive closed half-space, particularly $h(u) = h(b - \frac{l}{\sqrt{m}} \mathbf{1}) \leq h(v)$, and so the nondecreasing property is fulfilled.

d) Case 4: If v is inside S_h and u is not, then there are two subcases we split this into.

If $h'(u)$ is in the closed negative half-space of K, then $h(u) \leq h(v)$ by a similar argument as above in case 3. The structure of the proof is that $h(w) = \alpha \mathbf{1}_m^T w + \beta \leq \alpha \mathbf{1}_m^T (b - \frac{l}{\sqrt{m}} \mathbf{1}_m) + \beta = h(v)$ for all $w \leq v$, particularly $w = u$, and so the nondecreasing property is fulfilled.

Finally, it suffices to show it is not possible for $h'(u)$ to be in the positive open half-space of K (as defined in case 3) under the above conditions, namely, that $u \leq v$, $u \notin S_h$, and $v \in S_h$. We've proven that $w < s^0$ for all $w \in S_h$, therefore $u \leq v < s^0$.

An argument based on the normalized barycentric coordinates of u can derive a contradiction under the assumption that u is in the positive open half-space of K. Let $u = \sum_{i=0}^{m} \lambda_i s^i = s^0 - \sum_{i=1}^{m} \lambda_i (l + \epsilon) \sqrt{m} e_i$ for $\sum_i \lambda_i = 1$, so $u_i = b_i + \frac{\epsilon}{\sqrt{m}} - \lambda_i (l + \epsilon) \sqrt{m}$. We will now prove that if $h'(u)$ is in the positive *open* half-space of K, or $\sum_{i=1}^{m} u_i > \sum_{i=1}^{m} b_i - \sqrt{m} l$ as some algebra shows, then $\lambda_i > 0$ for all i. We can see that $\lambda_0 > 0$ by plugging its entry-wise definition into the preceding inequality:

$$
\begin{aligned}
\sum_{i=1}^{m} u_i &> \sum_{i=1}^{m} b_i - \sqrt{m} l \\
\sum_{i=1}^{m} (b_i + \frac{\epsilon}{\sqrt{m}} - (l + \epsilon) \sqrt{m} \lambda_i) &> \sum_{i=1}^{m} b_i - \sqrt{m} l \\
\sum_{i=1}^{m} b_i + \sqrt{m} \epsilon - (l + \epsilon) \sqrt{m} (1 - \lambda_0) &> \sum_{i=1}^{m} b_i - \sqrt{m} l \\
(l + \epsilon) \sqrt{m} \lambda_0 > 0 &\implies \lambda_0 > 0.
\end{aligned} \tag{9}
$$

And to prove $\lambda_i > 0$ for $i > 0$, we recall that $u < s^0$:

$$
\begin{aligned}
u < s^0 &\iff u_i < s_i^0 \ \forall i \\
b_i + \frac{\epsilon}{\sqrt{m}} - (l + \epsilon) \sqrt{m} \lambda_i &< b_i + \frac{\epsilon}{\sqrt{m}} \\
-(l + \epsilon) \sqrt{m} \lambda_i < 0 &\implies \lambda_i > 0.
\end{aligned} \tag{10}
$$

On the other hand, it's a basic property of barycentric coordinates that $u \notin S_h$ implies $\lambda_i < 0$ for some i. Thus the contradiction arises, and so it's not possible for $h'(u)$ to be in the closed negative half-space of K under the provided assumptions, so there's no need to worry about proving $h(u) \leq h(v)$ under those conditions.

That concludes the fourth and last case. All other cases have been exhausted, and so the proof that h is nondecreasing is complete. ☐

C. $\mathcal{H}$ guarantees strong duality

This section proves that $\mathcal{F} = \mathcal{H}$ guarantees strong duality for problem $\mathrm{P}_{\text{0-1 ILP}}$ with both problems $\mathrm{D}_{\text{0-1 ILP}}$ and $\mathrm{D}'_{\text{0-1 ILP}}$.

Lemma 2. *Consider the value function $\phi : \mathbb{R}^m \to \mathbb{R}$ corresponding to feasible problem $\mathrm{P}_{\text{0-1 ILP}}$ with a bounded solution. There exists a function the form $h(d) = \alpha \mathbf{1}_m^T d + \beta$ with parameters $\alpha, \beta \in \mathbb{R}$ and $\alpha > 0$ that upper-bounds ϕ, i.e., $h(d) \geq \phi(d)$ for all $d \in \mathbb{R}^m$.*

Proof. For constraint vectors d that give a feasible primal, we can express $\phi(d)$ as $\sum_{i \in I_d} c_i$ where $I_d \subseteq [n]$ contains the indices of the 1-entries in an optimal solution x^* to problem $\mathrm{P}_{\text{0-1 ILP}}$ with constraint vector d, i.e., $x_i^* = 1 \iff i \in I_d$. As a result, there's only a finite number of values that ϕ can take because there are only a finite number of possible settings of I (subsets of $[n]$). In particular, there is a finite upper bound on ϕ of $\sum_{c_i > 0} c_i$.

Additionally, any point $t \in \mathbb{R}^m$ with some entry $t_j < \min_{I \subseteq [n]} \sum_{i \in I} A_{j,i}$ has $\phi(t) = -\infty$ because no 0-1 combination of the columns of A can satisfy problem $\mathrm{P}_{\text{0-1 ILP}}$ with constraint vector t: it is infeasible. Note that the subset of all such points in $\mathbb{R}^m$ is unbounded; for example, all $t' \in \mathbb{R}^m$ with $t'_j < t_j$ as defined above will also have $\phi(t') = -\infty$.

Now define $t^* \in \mathbb{R}^m$ by $t^*_j = \min_{I \subseteq [n]} \sum_{i \in I} A_{j,i}$ for each $j \in [m]$, so every entry t^*_j is an inclusive lower bound on $(Ax)_j$ in problem $P_{\text{0-1 ILP}}$. It then must be that $\phi(d) = -\infty$ for all d satisfying $\mathbf{1}_m^T(d - t^*) < 0$ because this is equivalent to saying that $\sum_{i=1}^m (d_i - t^*_i) < 0$, in which case some entry satisfies $d_i < t^*_i$. Therefore, we define $h(d) = \sum_{i=1}^m (d_i - (t^*_i - \delta))$ with $\delta > 0$ so that its hyperplane intersects the level curve $d_{m+1} = 0$ along the line $\sum_{i=1}^m (d_i - (t^*_i - \delta)) = 0$. It is possible in general that ϕ has a finite value at t^*, so the inclusion of δ shifts the line slightly and ensures that $\phi(d) = -\infty$ at all points satisfying the equation of this line. As a result, $h(d) = 0 > \phi(d) = -\infty$ for all d satisfying $\sum_{i=1}^m (d_i - (t^*_i + \delta)) = 0$, and $h(d) \in (-\infty, 0)$ for every finite d satisfying $\sum_{i=1}^m (d_i - (t^*_i + \delta)) < 0$, while $\phi(d) = -\infty$ for all such points. Therefore, h upper-bounds ϕ for all constraint vectors d satisfying $\sum_{i=1}^m (d_i - (t^*_i + \delta)) \leq 0$.

With this in mind, consider any function $h(d) = \alpha \sum_{i=1}^m (d_i - (t^*_i - \delta))$ with $\alpha, \delta > 0$. Clearly this still upper-bounds the same set of points described above. That means if h does not upper-bound ϕ at every point, then the points at which $h(d) < \phi(d)$ must satisfy $\sum_{i=1}^m (d_i - (t^*_i - \delta)) > 0$. In this case, the function $g(d) = \alpha \sum_{i=1}^m (d_i - (t^*_i - \delta)) + \beta$ where $\beta = \sum_{c_i > 0} c_i$, is still of the same form described in the lemma statement and does upper-bound ϕ at every point. This follows for all d satisfying $\sum_{i=1}^m (d_i - (t^*_i + \delta)) > 0$ because $h(d) \geq 0$ and $\phi(d) \leq \sum_{c_i > 0} c_i$ on this set, therefore $h(d) - \phi(d) \geq 0 - \sum_{c_i > 0} c_i$ and so $g(d) - \phi(d) \geq 0$ on this set. For all d satisfying $\sum_{i=1}^m (d_i - (t^*_i + \delta)) \leq 0$, $g(d) \in (-\infty, 0]$ clearly still upper-bounds $\phi(d)$ and so g upper-bounds ϕ everywhere. □

Theorem 2. *Let the domain $\mathcal{F}$ of problem $D'_{\text{0-1 ILP}}$ be $\mathcal{H}$. Assume there is some open neighborhood around b of radius $\epsilon > 0$ such that $\phi(d) = \phi(b)$ for every d in that neighborhood. Then strong duality holds with problem $P_{\text{0-1 ILP}}$ when feasible with a bounded solution, i.e., there exists some $h^* \in \mathcal{H}$ such that $h^*(d) \geq \phi(d)$ for all $d \in \mathbb{R}^m$ and $h^*(b) = \phi(b)$.*

Proof. Once again, for convenience of notation we will define the function $h' : \mathbb{R}^m \to \mathbb{R}^{m+1}$ as $h'(d) = \begin{pmatrix} d \\ h(d) \end{pmatrix}$.

We know from lemma 2 that there must be some function $g(d) = \alpha \mathbf{1}_m^T d + \beta$ such that $g(d) \geq \phi(d)$ for all $d \in \mathbb{R}^m$. We also know ϕ is a nondecreasing function because any loosening of the constraints in problem $P_{\text{0-1 ILP}}$ (replacing b with $b' \geq b$) will only maintain or increase its optimal value. As a result, $\phi(b - r) \leq \phi(b)$ for every $r \in \mathbb{R}^m$, $r \geq 0$. The same thing holds for every vector inside the level open neighborhood of b by definition, in particular, $\phi(b) = \phi(b + (\frac{\epsilon}{\sqrt{m}} - \delta)\mathbf{1}_m)$ for sufficiently small $\delta \in (0, \frac{\epsilon}{\sqrt{m}})$. Therefore, $\phi(d) \leq \phi(b)$ for every d in the open simplex defined by the vertices of eq. (2) given the unique value for l which satisfies $\alpha \mathbf{1}_m^T(b - \frac{l}{\sqrt{m}}\mathbf{1}_m) + \beta = \phi(b)$ using the same α, β from g. Such a value for l must exist because we know $\alpha \mathbf{1}_m^T b + \beta \geq \phi(b)$, so there must be some $l \geq 0$ such that $\alpha \mathbf{1}_m^T b - \sqrt{m} l + \beta = \alpha \mathbf{1}_m^T(b - \frac{l}{\sqrt{m}}\mathbf{1}_m) + \beta = \phi(b)$.

As a result, if we change g just by lowering its values in that open simplex to be identical to $\alpha \mathbf{1}_m^T(b - \frac{l}{\sqrt{m}}\mathbf{1}_m) + \beta$, then it remains an upper bound on ϕ at every point. It is also now of the form defined in eq. (1), so it is nondecreasing by theorem 1. Finally, $g(b)$ is now equal to $\phi(b)$. Therefore, $\mathcal{H}$ meets the criteria for strong duality according to definition II.1 and problems $D_{\text{0-1 ILP}}$ and $D'_{\text{0-1 ILP}}$. □

IV. Extending Hatch Functions to General ILP

This section extends the results of section III to general integer linear programs in a straightforward way. Consider a general integer linear program and its dual of the form

$$(\mathrm{P_{ILP}}) \quad \begin{aligned} \max_x \quad & c^T x, \\ & Ax \leq b, \\ & x \in \mathbb{Z}^n, \end{aligned}$$

$$(\mathrm{D'_{ILP}}) \quad \begin{aligned} \min_h \quad & h(b), \\ & h(d) \geq \phi(d) \quad \forall d \in \mathbb{R}^m, \\ & h \in \mathcal{F} \subseteq \mathcal{F}_+^m \end{aligned}$$

with $A \in \mathbb{R}^{m \times n}$ and $\phi : \mathbb{R}^m \to \mathbb{R}$ the value function of problem P_{ILP}.

Theorem 3. *Taking the primal problem to be problem P_{ILP}, let the domain $\mathcal{F}$ of problem D'_{ILP} be $\mathcal{H}$. Assume there is some open neighborhood around b of radius $\epsilon > 0$ such that $\phi(d) = \phi(b)$ for every d in that neighborhood. Then strong duality holds with problem P_{ILP} when feasible with a bounded solution, i.e., there exists some $h^* \in \mathcal{H}$ such that $h^*(d) \geq \phi(d)$ for all $d \in \mathbb{R}^m$ and $h^*(b) = \phi(b)$.*

Proof. The argument is identical to that of theorem 2 except for the fact that now, there is no finite upper or lower bound on ϕ across all of $\mathbb{R}^m$. However, note that if we take A, b, c all to be constants fixed by problem P_{ILP}, then for $k \in \mathbb{R}$, $\phi(b + ke_i) = O(k)$ for each ith unit basis vector e_i. In other words, ϕ will not increase superlinearly along any axis in $\mathbb{R}^m$. Suppose for contradiction that it did, i.e., $\phi(b + ke_i) = \omega(k)$. That would imply that $c^T x(k) = \omega(k)$ where $x(k) \in \mathbb{Z}^n$ is the optimal solution such that $Ax(k) \leq b + ke_i$. All entries in c are treated as constants here, so the superlinear growth comes from $\sum_{j=1}^n x(k)_j = \omega(k)$. However, in the ith row of that matrix inequality, $\sum_{j=1}^n A_{i,j} x(k)_j \leq b_i + k$, the right-hand side only grows linearly in k and the left-hand side is $\Omega(\sum_{j=1}^n x(k)_j)$, so the constraint would be violated eventually. Hence, $\phi(b + ke_i) = O(k)$.

The linear growth therefore also holds for linear combinations of the constraint vector axes, so there exists some linear function $g(d) = \sum_{i=1}^m \alpha_i d_i + \beta$ which upper-bounds $\phi(b + k\mathbf{1}_m)$ as a function of $k \in \mathbb{R}$. The same function will upper-bound $\phi(b + ke_i)$ for each e_i because problem P_{ILP} with constraint vector $b + ke_i$ will have no greater an objective function value than that with $b + k\mathbf{1}_m$. Now define $h(d) = \alpha \sum_{i=1}^m d_i + \beta$ for $\alpha = \max\{\alpha_i\}$ and it will still upper-bound ϕ. Insert the hatch in the same way as for the 0-1 problem with the same justification, and this is a feasible solution to problem D'_{ILP} with value at b equal to $\phi(b)$.

Therefore, there exists some $h^* \in \mathcal{H}$ that upper-bounds ϕ for all d such that $h^*(b) = \phi(b)$. $\square$

V. Discussion and Further Work

This section provides some discussion of the application and structure of these results.

The first thing to clarify is that, recalling lemma 1, although problem $D'_{\text{0-1 ILP}}$ has an unbounded number of constraints and requires knowledge of ϕ, it is equivalent to problem $D_{\text{0-1 ILP}}$. Problem $D'_{\text{0-1 ILP}}$ helps prove the result, but problem $D_{\text{0-1 ILP}}$ appears more readily solvable by conventional techniques. The analogous observation holds for general ILPs.

Next, we would like to highlight the Sherali-Adams [13] hierarchy's contributions in this area. This is a so-called "lift and project" method which takes a 0-1 ILP and produces a hierarchy of increasingly tight relaxations of the integer convex hull. This may be leveraged to construct a strong dual problem corresponding to the proof system encoding this algorithm's logic. However, to our knowledge, this is only applicable to 0-1 ILP problems. In addition, the constant-size domain of our formulation makes it qualitatively distinct from the Sherali-Adams formulation.

The next thing to note is the exponential number of constraints generated by this dual formulation. Though this is restrictive, a well-known result by Lenstra [9] demonstrates that for a constant number of variables, there is an algorithm for integer linear programming with runtime polynomial in the instance size. Therefore, the exponential number of constraints is unsurprising; if the dual instance had size polynomial in that of the primal, then Lenstra's Algorithm could be applied to solve the primal instance via the dual in polynomial time.

It's also interesting to observe that a certain subset of constraints are in some sense responsible for the hardness of this dual formulation. For any positive setting of α, it's trivial to set β such that $(\alpha \mathbf{1}_m^T A - c)x + \beta \geq 0$ for all $x \in \{0,1\}^n$. Therefore, all dual constraints associated with x which $h \circ A$ maps onto the hyperplane part of h are satisfied. The complexity in solving the dual is therefore rooted in constraints falling on the hatch part of the hatch function, rather than the hyperplane. If l is initially set to 0 with the above settings of α and β, then all constraints are satisfied but this is unlikely to yield the optimal solution. Next, l can be gradually increased, expanding the simplex to start absorbing additional constraints while also reducing the objective function value towards the optimal. Any constraints violated by the increase of l will be associated with x which have $Ax \in S_h$. The identification and checking of these constraints is the root of the complexity. The procedure detailed above suggests itself as one natural approach for solving this dual. In particular, problem $D_{\text{0-1 ILP}}$ can be rephrased as a problem optimizing over just l and the set of constraints pared down to only those associated with x for which $Ax \leq b$.

Empirical analysis of Lenstra's algorithm for a fixed number of variables [9] applied to this formulation may also prove an interesting further avenue of exploration. The NP-completeness of integer linear programming will limit any application to sufficiently small instances, here owing to the exponential number of constraints. However, it may be possible to beat the naïve approach by a polynomial factor using techniques for pruning the space of constraints to be checked at each step. There is also potential for development of an approximation algorithm, only checking a small subset of the constraints and therefore surpassing the optimal solution by at most some bounded amount.

In addition, it is possible that a similar approach may be applied to obtain a strong dual formulation for mixed integer linear programming. We have not found a comprehensive duality theory for MILPs, and so that may prove a useful generalization of our result. The value function would have "tilted" platforms rather than level ones, as would the hatch function. It may prove straightforward to obtain such a formulation.

VI. Acknowledgments

This research was funded in part by the US National Science Foundation (Grant No CNS-2141256).

We give our heartfelt thanks to Dr. Laurence A. Wolsey for answering our email containing questions about his paper [14], which served as the indispensable foundation of this project, over forty years after its publication.

Our gratitude goes to Dr. Brendan Juba for his advice on the context, framing, and presentation of our result. We also thank Alex Nelson for his invaluable feedback on early drafts of this paper.

References

[1] Erik Balder. An extension of duality-stability relations to nonconvex optimization problems. *SIAM Journal on Control and Optimization*, 15:329–343, 1977.

[2] Dimitris Bertsimas and John Tsitsiklis. *Introduction to Linear Optimization*, page 495. Athena Scientific, 1st edition, 1997.

[3] Szymon Dolecki and Stanislaw Kurcyyusz. On phi-convexity in extremal problems. *SIAM Journal on Control and Optimization*, 16(2):277, 1978.

[4] Hugh Everett. Generalized lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research*, 11(3):399–417, 1963.

[5] F. J. Gould. Extensions of lagrange multipliers in nonlinear programming. *SIAM Journal on Applied Mathematics*, 17(6):1280–1297, 1969.

[6] J. N. Hooker. *Integer programming dualityInteger Programming Duality*, pages 1657–1667. Springer US, Boston, MA, 2009.

[7] Cédric Josz and Didier Henrion. Strong duality in lasserre's hierarchy for polynomial optimization, 2014.

[8] Javad Lavaei and Steven H Low. Zero duality gap in optimal power flow problem. *IEEE Transactions on Power systems*, 27(1):92–107, 2011.

[9] H. W. Lenstra. Integer programming with a fixed number of variables. *Mathematics of Operations Research*, 8(4):538–548, 1983.

[10] Panos Parpas and Berç Rustem. *Duality Gaps in Nonconvex Optimization*, pages 802–805. Springer US, Boston, MA, 2009.

[11] Teemu Pennanen and Ari-Pekka Perkkiö. Stochastic programs without duality gaps. *Mathematical Programming*, 136(1):91–110, 2012.

[12] R. Rockafeller. Augmented lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12(2), 1974.

[13] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990.

[14] Jorgen Tind and Laurence A. Wolsey. An elementary survey of general duality theory in mathematical programming. *Mathematical Programming*, 21:241–261, 1981.