CO452 Integer Programming

Riley Jackson

January 31, 2021

Contents

Abstract

These are my personal notes for Professor Ricardo Fukasawa's Winter 2021 offering of CO452, Integer Programming. In this course we study properties and methods that are useful when solving integer programs.

@IMPORTANT: These are notes for an introductory graduate level course on Integer Programming and are not intended to be used as a reference on the field. I know that many of these results can be generalized, and I am aware that my transcription may introduce some errors. Furthermore, I skipped proofs that I found boring or non-insightful.

An **integer program** in the context of this course is an optimization problem of the form:

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax \leq b \\ & x_j \in \mathbb{Z} \quad \text{for} \quad j \in I \end{aligned} \tag{IP}$$

where $c \in \mathbb{Q}^n, A \in \mathbb{Q}^{m \times n}, b \in \mathbb{Q}^m, \varnothing \subsetneq I \subseteq [n]$. We define

$$P = \{x : Ax \leq b\}$$
$$P_I = \{x : Ax \leq b \text{ and } \forall_{j \in I} x_j \in \mathbb{Z}\}$$

When $I = [n]$ we say that IP is a **pure integer program**, otherwise it's a **mixed integer program**. If we have $x \in \{0,1\}^n$ then we say IP is a **binary integer program**.

1 Polyhedral Theory

We begin with some basic geometric definitions:

Definition

- A **halfspace** is a set of the form

$$\left\{x \in \mathbb{R}^n : \alpha^\top x \leq \beta\right\}$$

  for $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$.

- A **hyperplane** is a set of the form

$$\left\{x \in \mathbb{R}^n : \alpha^\top x = \beta\right\}$$

  for $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$.

- A **polyhedron** is a set of the form

$$\{x \in \mathbb{R}^n : Ax \leq b\}$$

  for $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. If A, b are rational then we say say this is a **rational polyhedron**.

Notice that a polyhedron is an intersection of finitely many half spaces.

Definition Let $x^1, x^2 \in \mathbb{R}^n$, then the **line segment** between them is denoted by

$$\left[x^1, x^2\right] \coloneqq \left\{x \in \mathbb{R}^n : x = \lambda x^1 + (1-\lambda)\, x^2, \lambda \in [0,1]\right\}$$

We now introduce (arguably) the most important type of set in optimization:

Definition A set $C \subseteq \mathbb{R}^n$ is said to be **convex** if

$$\forall_{x^1, x^2 \in C} \left[x^1, x^2\right] \subseteq C$$

So convex sets are those which contain all of their line segments. The following is an immediate consequence of convexity:

Proposition Let $\{C_\alpha\}_{\alpha \in \mathcal{A}}$ be a collection of convex sets, then $\bigcap_{\alpha \in \mathcal{A}} C_\alpha$ is convex.

Unfortunately, most sets aren't convex; this prompts the following definition:

Definition Given $M \subseteq \mathbb{R}^n$, we say that the **convex hull of** M, denoted $\operatorname{conv}(M)$, to be the inclusion wise minimum convex set containing M.

The previous proposition ensures that there is an inclusion wise minimum, and tells us that it is the intersection of every convex set containing M.

Definition We say that $x \in \mathbb{R}^n$ is a convex combination of $x^1, \ldots, x^1$ if there is some $\lambda \in \mathbb{R}^q_+$ such that

$$x = \sum_{i=1}^{q} \lambda_i x^i \quad \text{and} \quad \mathbb{1}^\top \lambda = 1$$

It seems that this notion is somewhat useless for Integer Programming since the feasible region will never be convex if there are atleast 2 feasible solutions. The following is a beautiful theoretical result:

Theorem

$$\min\left\{c^\top x : x \in P_I\right\} = \left\{c^\top x : x \in \operatorname{conv}(P_I)\right\}$$

Proof. It is clear that $P_I = \varnothing \iff \operatorname{conv}(P_I) = \varnothing$, so we have WLOG that the problem is feasible. Notice that, since $P_I \subseteq \operatorname{conv}(P_I)$, we clearly have LHS $\geq$ RHS. Suppose now that there is some $\bar{x} \in \operatorname{conv}(P_I)$ such that $c^\top \bar{x} < \min\left\{c^\top x : x \in P_I\right\}$. By the above, we can write $\bar{x} = \sum_{i=1}^{q} \lambda_i x^i$ where $\left\{x^i\right\}_{i=1}^{q} \subseteq P_I$ and $\lambda \in \mathbb{R}^q_+$ with $\mathbb{1}^\top \lambda = 1$. But then we have

$$c^\top \bar{x} = \sum_{i=1}^{q} \lambda_i c^\top x^i > \sum_{i=1}^{q} \lambda_i c^\top \bar{x} = c^\top \bar{x}$$

which is a contradiction. It follows that RHS $\geq$ LHS and the result is proven. □

Lemma $C \subseteq \mathbb{R}^n$ is convex if and only if all finite convex combinations of points of C are in C.

Proof. The backwards direction is obvious since a line segment is a convex combination of 2 elements. For the forwards direction we proceed by induction on the number of elements in the convex combinations. The base case is trivial so suppose that $\bar{x} = \sum_{i=1}^{k} \lambda_i x^i$ is a convex combination of k elements with $k > 1$, then we have

$$\bar{x} = \sum_{i=1}^{k-1} \lambda_i x^i + \lambda_k x^k = (1-\lambda_k) \sum_{i=1}^{k-1} \frac{\lambda_i}{1-\lambda_k} x^i + \lambda_k x^k$$

which is a convex combination of x^k and $\sum_{i=1}^{k-1} \lambda_i x^i$ (which is in C by the induction hypothesis). □

The following is a useful characterization of convex hulls:

Theorem For any $M \subseteq \mathbb{R}^n$, we have that $\operatorname{conv}(M)$ is the set of all finite convex combinations of points in M

Proof. Let C be the set of all finite convex combinations of points in M, then C is clearly a convex set containing M and we get that $\operatorname{conv}(M) \subseteq C$. But if $\bar{x} \in C$ then $\bar{x}$ is a convex combination of points in $M \subseteq \operatorname{conv}(M)$, hence it's in $\operatorname{conv}(M)$ by the previous lemma. □

Continuing with basic geometry, we have

Definition Given $x^1, \ldots, x^k \in \mathbb{R}^n$, we say that $\bar{x}$ is a **linear combination** of $x^1, \ldots, x^k$ if there is some $\lambda \in \mathbb{R}^k$ such that

$$\bar{x} = \sum_{i=1}^{k} \lambda_i x^i$$

Definition $x^1, \ldots, x^k \in \mathbb{R}^n$ are **linearly independent** if $\lambda = \mathbb{0}$ is the unique solution to

$$\sum_{i=1}^{k} \lambda_i x^i = \mathbb{0}$$

Definition $\varnothing \subsetneq \mathcal{L} \subseteq \mathbb{R}^n$ is a **linear space** if it's closed under taking linear combinations.

Note Linear spaces are exactly those that can be represented in the form $\{x \in \mathbb{R}^n : Dx = \mathbb{0}\}$ for some $D \in \mathbb{R}^{m \times n}$.

Definition A **basis** of a linear space is an inclusion wise maximal set of linearly independent vectors.

Definition The size of a basis is the **dimension** of the linear space, denoted $\dim(\mathcal{L})$.

Note It's a basic theorem of linear algebra that the above definition makes sense. Furthermore, the rank-nullity theorem gives that $\dim(\mathcal{L}) = n - \operatorname{rank}(D)$ if $\mathcal{L} = \{x : Dx = \mathbb{0}\}$.

Definition Given $x^1, \ldots, x^k \in \mathbb{R}^n$, we say that $\bar{x}$ is an **affine combination** of $x^1, \ldots, x^k$ if there is some $\lambda \in \mathbb{R}^k$ such that

$$\bar{x} = \sum_{i=1}^{k} \lambda_i x^i \quad \text{and} \quad \mathbb{1}^\top \lambda = 1$$

Definition $\varnothing \subseteq \mathcal{A} \subseteq \mathbb{R}^n$ is an **affine space** if it's closed under affine combinations.

We would like to generalize our notion of independence and dimension from linear spaces to affine spaces, however blindly applying the current definitions will not match our intuition

Example Consider $\mathbb{R}^2$ with the standard basis $\{e_1, e_2\}$, clearly the line $[e_1, e_2]$ is "one-dimensional", but it contains 2 linearly independent points. Furthermore, e_1 and e_2 should be "affinely dependant" since they lie in a "one-dimensional" affine space, yet they are linearly independent.

The problem essentially boils down to the fact that affine spaces may not contain $\mathbb{0}$, and luckily all problems go away if we first translate the affine space to contain $\mathbb{0}$. More precisely, we have:

Definition $x^1, \ldots, x^k \in \mathbb{R}^n$ are **affinely independent** if $\lambda = \mathbb{0}$ is the only solution to

$$\sum_{i=1}^{k} \lambda_i x^i = \mathbb{0} \quad \text{and} \quad \sum_{i=1}^{k} \lambda_i = 0$$

Definition An **affine basis** is an inclusion wise maximal set of affinely independent vectors.

Definition The **dimension** of an affine space is the cardinality of an **affine basis**.

The following proposition essentially proves that the above definition is what we want it to be.

Proposition The following are equivalent:

(i) $\{x^0, \ldots, x^k\}$ are affinely independent.

(ii) $\{x^1 - x^0, \ldots, x^k - x^0\}$ are linearly independent.

(iii) No point in $\{x^0, \ldots, x^k\}$ is an affine combination of the others.

Similar to linear spaces and convex spaces, we have the following descriptions of affine spaces:

Proposition The following are equivalent:

(i) $\mathcal{A}$ is an affine space.

(ii) $\forall_{x,y\in\mathcal{A}}\forall_{\lambda\in\mathbb{R}}\lambda x+(1-\lambda)\,y\in\mathcal{A}$.

(iii) $\mathcal{A}=\{x\in\mathbb{R}^n : Dx=d\}$ for some $D\in\mathbb{R}^{m\times n}$ and $d\in\mathbb{R}^m$.

(iv) There is some linear space $\mathcal{L}$ such that for any $x\in\mathcal{A}$ we have $\mathcal{A}=\{x\}+\mathcal{L}$.

Also, as we would expect, $\dim(\mathcal{A})=\dim(\mathcal{L})=n-\operatorname{rank}(D)$.

Also similar to convex hull and linear span, we have

Definition Given $M\subseteq\mathbb{R}^n$ we define the **affine hull**, denoted $\operatorname{aff}(M)$, to be the inclusion wise minimal affine space containing M.

and we have

Proposition Let $M\subseteq\mathbb{R}^n$, then we have:

(i) $\operatorname{aff}(M)$ is the set of all finite affine combinations of points in M.

(ii) $\operatorname{aff}(M)=\bigcap_{\substack{M\subseteq\mathcal{A}\\ \mathcal{A}\text{ affine}}}\mathcal{A}$.

Finally, it turns out that an ever more general definition of dimension will be useful, hence we say (for the rest of these notes) that

Definition Let $S\subseteq\mathbb{R}^n$, then the **dimension** of S, denoted $\dim(S)$, is the maximum number of affinely independent points in S minus 1.

It is clear, and will be useful, that $\dim(S)=\dim(\operatorname{aff}(S))$.

We will now develop methods that take advantage of additional structure that facilitate finding the dimension of polyhedra. So let $P=\{x\in\mathbb{R}^n : Ax\le b\}$ and write

$$A=\begin{pmatrix}— & a_1^\top & —\\ & \vdots & \\ — & a_m^\top & —\end{pmatrix}\qquad b=\begin{pmatrix}b_1\\ \vdots\\ b_m\end{pmatrix}$$

so we have that $x\in P$ if and only $a_i^\top x\le b_i$ for all $1\le i\le m$.

Definition We say that $a_i^\top x\le b_i$ is an **implicit equality** if $x\in P\implies a_i^\top x=b_i$.

Example The polyhedron

$$P=\left\{x\in\mathbb{R} : \begin{pmatrix}1\\-1\end{pmatrix}x\le\begin{pmatrix}0\\0\end{pmatrix}\right\}$$

has 2 inequality constraints, but clearly $P=\{0\}$, hence they are both implicit equality constraints.

Note We can find implicit equalities in poly time by solving for each $1\le i\le m$ the linear program

$$\begin{aligned}\min\quad & a_i^\top x\\ \text{s.t.}\quad & Ax\le b\end{aligned}$$

If the optimal value is b_i, the i-th constraint is an implicit equality.

Let M^E denote the indices of the implicit equality constraints of P and let $M^L=[m]\setminus M^E$. Let A^E,b^E,A^L,b^L denote the submatrices induced by rows of M^E,M^L respectively, then we have

$$x\in P\implies A^Ex=b^E\qquad\text{and}\qquad\forall_{i\in M^L}\exists_{x_i\in P}a_i^\top x<b_i$$

Notice that setting $\bar{x}=\frac{1}{|M^L|}\sum_{i\in M^L}x_i$ give $A^L\bar{x}<b^L$ ($\bar{x}\in P$ by convexity).

Theorem Let $P\neq\varnothing$ be as above, then

(i) $\operatorname{aff}(P)=\left\{x : A^Ex\le b^E\right\}=\left\{x\in\mathbb{R}^n : A^Ex=b^E\right\}$

(ii) $\dim(P) = n - \text{rank}\left(A^E\right)$

Proof. Clearly we have $P \subseteq \left\{x \in \mathbb{R}^n : A^E x = b^E\right\}$, so since the RHS is affine we get (by taking affine hulls)

$$\text{aff}(P) \subseteq \left\{x \in \mathbb{R}^n : A^E x = b^E\right\} \subseteq \left\{x \in \mathbb{R}^n : A^E x \leq b^E\right\}$$

Now let $\hat{x} \in P \setminus \left\{x \in \mathbb{R}^n : A^E x \leq b^E\right\}$, if $\hat{x}$ doesn't exist then we have

$$\left\{x \in \mathbb{R}^n : A^E x \leq b^E\right\} \subseteq P \subseteq \text{aff}(P)$$

Let $\bar{x} \in P$ be such that $A^L \bar{x} < b^L$ and consider $x(\varepsilon) = \bar{x} + \varepsilon(\hat{x} - \bar{x})$, clearly we can take ε small enough so that $x(\varepsilon) \in P$, but then $\hat{x}$ is an affine combination of $\bar{x}$ and $x(\varepsilon)$, hence $\hat{x} \in \text{aff}(P)$. It follows immediately that $\left\{x \in \mathbb{R}^n : A^E x \leq b^E\right\} \subseteq \text{aff}(P)$. $\square$

Example Consider the knapsack polyhedron

$$K = \text{conv}\left(\left\{x \in \mathbb{R}^n : \sum_{i=1}^{n} a_i x_i \leq b\right\}\right)$$

where $0 \leq a_1, \ldots, a_n \leq b$.

We will show how to compute the dimension in 2 ways.

(i) First, we note that $0, e_1, \ldots, e_n \in K$, hence we have found $n+1$ affinely independent points and the dimension is $\geq n$. Since this is maximal it must be exactly equal; that is, $\dim(K) = n$.

(ii) Second, suppose that K has an implicit equality $\alpha^\top x = \beta$. Since $0 \in K$ we must have $\beta = 0$, and since $e_i \in K$ we get that $0 = \alpha^\top e_i = \alpha_i \implies \alpha = 0$. That is, the only implicit equality constraints are degenerate, hence $\text{rank}\left(A^E\right) = 0$, and we get that $\dim(K) = n$.

We now pivot to talking about representations of polyhedra. A tool that will be useful is the following:

Definition Given $P = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^p : Ax + Gy \leq b\}$ we say the **projection of P w.r.t x** is

$$\text{proj}_x(P) = \{x \in \mathbb{R}^n : \exists_{y \in \mathbb{R}^p} Ax + Gy \leq b\}$$

Projections of polyhedra can be computed using Fourier-Motzkin elimination, and the algorithm gives the following:

(i) $\text{proj}_x(P)$ is also a polyhedron

(ii) If P is rational then so is $\text{proj}_x(P)$

(iii) If $b = 0$ then $\text{proj}_x(P)$ can be written as $\{x \in \mathbb{R}^n : Dx \leq \mathbb{0}\}$

(iv) $P = \varnothing \iff \text{proj}_x(P) = \varnothing$

Definition Let $x^1, \ldots, x^k \in \mathbb{R}^n$, we say that $\bar{x}$ is a **conic combination** of $x^1, \ldots, x^k$ if $\exists_{\lambda_1, \ldots, \lambda_k \geq 0}$ such that $\bar{x} = \sum_{i=1}^{k} \lambda_i x^i$

Definition A set $C \subseteq \mathbb{R}^n$ is a **cone** if $\mathbb{0} \in C$ and $\forall_{x \in C} \forall_{\lambda \geq 0} \lambda x \in C$

Note Some authors say cone to mean convex cone, we will always specify. It's easy to see that a cone is convex if and only if it's closed under conic combinations.

Definition Sets of the form $C = \{x \in \mathbb{R}^n : Ax \leq \mathbb{0}\}$ are called **polyhedral cones**

Definition Given $r^1, \ldots, r^n \in \mathbb{R}^k$, the **cone generated by** $r^1, \ldots, r^k$, denoted $\text{cone}\left(r^1, \ldots, r^k\right)$, is defined to be

$$\text{cone}\left(r^1, \ldots, r^k\right) := \left\{x \in \mathbb{R}^n : \exists_{\mu \geq \mathbb{0}} x = \sum_{i=1}^{k} \mu_i r^i\right\}$$

Note (i) We say that $\text{cone}\left(r^1, \ldots, r^k\right)$ is a **finitely generated cone** and that $r^1, \ldots, r^k$ are its **generators**.

(ii) If

$$R = \begin{pmatrix} | & & | \\ r^1 & \cdots & r^k \\ | & & | \end{pmatrix}$$

then we say $\text{cone}\,(R) = \text{cone}\,\left(r^1, \ldots, r^k\right)$

(iii) We say $\text{cone}\,(\varnothing) = \{\mathbb{0}\}$

Let C be the (closed) convex cone given by $C = \{x \in \mathbb{R}^n : \langle \alpha_j, x\rangle \leq 0 : j \in J\}$

Definition The **polar cone**, denoted C°, is defined to be

$$C^\circ \coloneqq \{y \in \mathbb{R}^n : \forall_{x\in C} \langle x, y\rangle \leq 0\}$$

Notice that C° is also a (closed) convex cone.

Proposition $\left(C^\circ\right)^\circ = C$

Proof. Let $x \in C$, then $\langle x, y\rangle \leq 0$ for all $y \in C^\circ$, so $x \in \left(C^\circ\right)^\circ$. Now suppose that $x \in \left(C^\circ\right)^\circ$, notice that $\alpha_j \in C^\circ$ for all $j \in J$, hence $\langle \alpha_j, x\rangle \leq 0$ for all $j \in J$, so $x \in C$. □

The above is an example of a tradeoff between generality and simplicity (which will not be illustrated again), the general result is:

Theorem Let C be a cone, then $\left(C^\circ\right)^\circ = \text{cl}\,(\text{conv}\,(C))$

Note For those who have seen **dual cones**, we trivially have $C^* = -C^\circ$.

Theorem Minkowski-Weyl C is a polyhedral cone if and only if it's finitely generated.

Proof. Suppose first that $C = \text{cone}\,\left(r^1, \ldots, r^k\right)$ and define

$$\bar{C} \coloneqq \left\{(x, \mu) \in \mathbb{R}^n \times \mathbb{R}^k : x - \sum_{i=1}^{k} \mu_i r^i = 0 \text{ and } \mu \geq \mathbb{0}\right\}$$

Clearly $\bar{C}$ is a polyedral cone and $C = \text{proj}_x\left(\bar{C}\right)$, hence C is a polyhedral cone as well.

Suppose now that

$$C = \{x \in \mathbb{R}^n : Ax \leq b\} \qquad \text{where} \qquad A = \begin{pmatrix} — & a_1^\top & — \\ & \vdots & \\ — & a_k^\top & — \end{pmatrix}$$

Claim $C^\circ = \overbrace{\left\{y \in \mathbb{R}^n : \exists_{\mu\geq\mathbb{0}} A^\top \mu = y\right\}}^{C'}$

Proof. Let $y \in C'$, then $\langle y, x\rangle = y^\top x = \left(\mu^\top A\right) x = \sum_{i=1}^{k} \mu_i a_i^\top x \leq 0$ so $y \in C^\circ$. Suppose now that $y \notin C'$, then

$$\left\{\mu \in \mathbb{R}^n : A^\top \mu = y \text{ and } \mu \geq \mathbb{0}\right\} = \varnothing$$

By Farkas' lemma, it follows that there is some $x \in \mathbb{R}^n$ such that $x^\top y > 0$ and $x^\top A^\top = \mathbb{0}$. That is, there is some $x \in \mathbb{R}^n$ such that $Ax \leq \mathbb{0}$ and $\langle x, y\rangle > 0$, meaning that $y \notin C^\circ$. By double inclusion we conclude that $C^\circ = C'$. □

It follows immediately that $C = \left(C^\circ\right)^\circ$ is finitely generated. □

Note This proof can be extended to show that C is a rational polyhedral cone if and only if C can be genreated by finitely many rational vectors.