CS 521: Linear Programming **Rutgers: Fall 2022**

Lecture 7

October 21, 2022

Instructor: Sepehr Assadi *Scribe: Xinyue Feng, Songhua He, Shuning Jin, Zejun Xie*

Convexity in Linear Programming

1 Basics of Convexity

In this lecture, we introduce notions and applications of convexity in linear programming (or alternatively, study linear programming from a geometric point of view). We start with basic definitions, and then consider several canonical convex sets, and finally see a high level connection between convexity and linear programming.

1.1 Basic Definitions

Convex set. We start with the definition of the convex set.

Definition 1 (Convex Set)**.** A set $S \subseteq \mathbb{R}^n$ is convex if $\forall x, y \in S, t \in [0, 1]$, we have $tx + (1 - t)y \in S$.

Geometrically, given two points x and y, we draw a line segment xy between them. We can obtain all the points on xy by changing the value of $t \in [0, 1]$ in $tx + (1 - t)y$. Thus, a set is convex if the segment joining any two of its points is contained in the set. Figure 1 gives some examples.

There are also other ways to define a convex set, e.g., a convex set is a subset that intersects every line into a single line segment (possibly empty).

Claim 2. *The intersection of any collection of convex sets is also convex.*

Proof. Let $\mathcal{C}$ be any collection of convex sets. If $x, y \in \cap\mathcal{C}$, $x, y \in C$ for each $C \in \mathcal{C}$. Therefore, $\forall t \in [0, 1], tx + (1 - t)y \in C$ for each $C \in \mathcal{C}$, which means $tx + (1 - t)y \in \cap\mathcal{C}$ and $\mathcal{C}$ is a convex set. $\square$

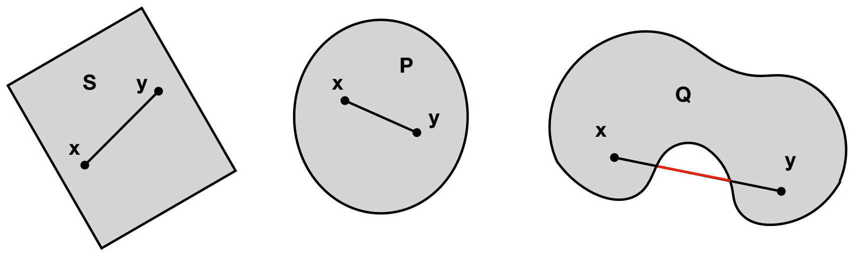

Figure 1: Sets S and P are convex, but set Q is not, because segment xy is not contained in Q

Convex function. Convex functions and convex sets are closely related and connected by the definition of epigraphs (Figure 2). An **epigraph** or **supergraph** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set

$$\mathrm{epi}(f) = \{(x, y) \in \mathbb{R}^{n+1} \mid f(x) \leq y\} \quad \text{with } x \in \mathbb{R}^n, y \in \mathbb{R}.$$

Then we can define that a function f is convex if and only if its epigraph is a convex set.

Equivalently, f is convex if and only if the line segment between any two points on the graph of the function lies above the graph between the two points. Thus,

Definition 3 (Convex Function)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real function,

(1) f is convex if and only if its epigraph $\mathrm{epi}(f) = \{(x, y) \in \mathbb{R}^{n+1} \mid f(x) \leq y\}$ is a convex set.

(2) f is convex if and only if $\forall x, y \in \mathbb{R}^n, f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$.

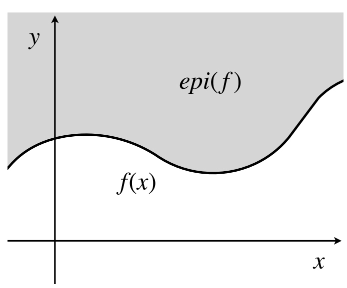

Figure 2: The epigraph of a function f is all points lying on or above its graph.

Convex combination. Recall that a linear combination is the weighted average of points: $\sum_i \lambda_i x_i$, with constants $\lambda_i \in \mathbb{R}$. A convex combination is similar, but with extra constraints on constants λ_i.

Definition 4 (Convex Combination)**.** Convex combination for a set of points $\{x_1, \ldots, x_m \in \mathbb{R}^n\}$ is

$$\sum_{i=1}^{m} \lambda_i x_i \ \text{ s.t. } \sum_i \lambda_i = 1,\ \lambda_i \geq 0.$$

Therefore, convex combination is a special type of linear combination: every convex combination is a linear combination, but the opposite is not true. We can also interpret the convex combination from another point of view: It is an expectation of any probability distribution over points $x_1, ..., x_m$.

Convex combinations of two points x_1 and x_2 fill exactly the segment x_1x_2. It is also easy to see that all convex combinations of three points x_1, x_2, x_3 fill exactly the triangle $x_1x_2x_3$; see Figure 3 for an illustration.

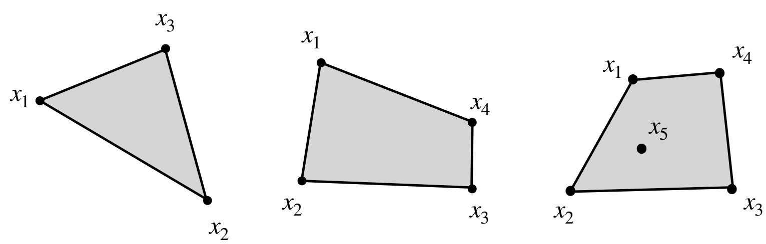

Figure 3: Convex hull of sets containing different numbers of points.

Convex hull. Based on the definition of convex combination, we can define a convex hull as below.

Definition 5 (Convex Hull)**.** The convex hull for a set of points $S := \{x_1, \ldots, x_m \in \mathbb{R}^n\}$ is

$$\text{ConvexHull}(S) := \{y \in \mathbb{R}^n \mid y \text{ is a convex combination of } x_1, \ldots, x_m\}.$$

The convex hull can also be described equivalently in Lemma 6 as the intersection of all convex sets that contains the original points.

Lemma 6. *The convex hull of points $S := \{x_1, ..., x_m\}$ is convex and is a subset of any other convex set that contains S (informally speaking,* ConvexHull(S) *is the "smallest" convex set containing all of S).*

Proof. We shall prove the convex hull is a convex set and also the smallest. Let $C := \text{ConvexHull}(S)$.

C is a convex set. Any two points $y, z \in C$ can be represented by convex combinations of $\{x_1, ..., x_m\}$,

$$y = \lambda_1 x_1 + \cdots + \lambda_m x_m, \qquad \sum_{i=1}^{m} \lambda_i = 1, \quad \lambda_i \geq 0$$

$$z = \lambda'_1 x_1 + \cdots + \lambda'_m x_m, \qquad \sum_{i=1}^{m} \lambda'_i = 1, \quad \lambda'_i \geq 0.$$

Thus, for any $t \in [0, 1]$, we have,

$$ty + (1-t)z = \sum_{i=1}^{m} (t\lambda_i + (1-t)\lambda'_i) \cdot x_i.$$

As $t, 1-t, \lambda_i$, and λ'_i are all greater than 0, and

$$\sum_{i=1}^{m} [t\lambda_i + (1-t)\lambda'_i] = t \sum_{i=1}^{m} \lambda_i + (1-t) \sum_{i=1}^{m} \lambda'_i = t + (1-t) = 1,$$

we have that $ty + (1-t)z$ is also a convex combination of points in S and thus belongs to C. Therefore, C is convex based on the definition of convex set.

Any convex set $\tilde{C}$ containing $x_1, ..., x_m$, also contains C. C is composed of all convex combinations $y = \sum_{i=1}^m \lambda_i x_i$ of S. We prove that y belongs to $\tilde{C}$ as well. If $m = 1$, it is obviously that $y \in \tilde{C}$. If $m = 2$, $y = \lambda x_1 + (1-\lambda)x_2$, and thus since $\tilde{C}$ is a convex set, $x_1, x_2 \in \tilde{C}$, and $\lambda \in [0,1]$, then $y \in \tilde{C}$.

Now we shall prove the following proposition by induction: for all $m \geqslant 3$, y can be rewritten as

$$y = \sum_{i=1}^{m-1} \lambda_i z_{m-2} + \left(1 - \sum_{i=1}^{m-1} \lambda_i\right) x_m \in \tilde{C},$$

where

$$z_{m-2} = \frac{\sum_{i=1}^{m-2} \lambda_i}{\sum_{i=1}^{m-1} \lambda_i} \cdot z_{m-3} + \frac{\lambda_{m-1}}{\sum_{i=1}^{m-1} \lambda_i} \cdot x_{m-1} \in \tilde{C}, z_0 = x_1.$$

When $m = 3$, $y = \lambda_1 x_1 + \lambda_2 x_2 + (1-\lambda_1-\lambda_2)x_3$. Here we define $z_1 = \frac{\lambda_1 x_1}{\lambda_1+\lambda_2} + \frac{\lambda_2 x_2}{\lambda_1+\lambda_2}$. Since $\frac{\lambda_1}{\lambda_1+\lambda_2} + \frac{\lambda_2}{\lambda_1+\lambda_2} = 1$ and all weights are greater than 0, we have $z_1 \in \tilde{C}$. Then we can rewrite y as $y = (\lambda_1+\lambda_2)z_1 + (1-\lambda_1-\lambda_2)x_3$. As z_1 and x_3 all belong to $\tilde{C}$, $y \in \tilde{C}$. The proposition holds.

Assume that the proposition holds when $m = k$, i.e.,

$$y = \sum_{i=1}^{k-1} \lambda_i z_{k-2} + \left(1 - \sum_{i=1}^{k-1} \lambda_i\right) x_k.$$

where

$$z_{k-2} = \frac{\sum_{i=1}^{k-2} \lambda_i}{\sum_{i=1}^{k-1} \lambda_i} z_{k-3} + \frac{\lambda_{k-1}}{\sum_{i=1}^{k-1} \lambda_i} x_{k-1} \in \tilde{C}.$$

Then when $m = k+1$,

$$y = \sum_{i=1}^{k-1} \lambda_i z_{k-2} + \lambda_k x_k + \left(1 - \sum_{i=1}^{k} \lambda_i\right) x_{k+1}.$$

Define $z_{k-1} = \frac{\sum_{i=1}^{k-1} \lambda_i}{\sum_{i=1}^{k} \lambda_i} z_{k-2} + \frac{\lambda_k}{\sum_{i=1}^{k} \lambda_i} x_k$, then we can rewrite y as $y = \sum_{i=1}^{k} \lambda_i z_{k-1} + \left(1 - \sum_{i=1}^{k} \lambda_i\right) x_{k+1}$. Since $z_{k-2}, x_k \in \tilde{C}$, the sum of weights is 1, and all weights are greater than 0, we have $z_{k-1} \in \tilde{C}$, which implies that $y \in \tilde{C}$, concluding the proof. □

1.2 Half-spaces, Polyhedra, and Polytopes

Let us now consider several canonical examples of convex sets.

Definition 7 (Half-space)**.** A *half-space* is a set $\{x \in \mathbb{R}^n \mid a^\top x \geq b\}$, with $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

Geometrically, a hyperplane $\left\{x \in \mathbb{R}^n \mid a^\top x = b\right\}$ divides the space $\mathbb{R}^n$ into two half-spaces. Algebraically, a hyperplane corresponds to an equality constraint, and a half-space corresponds to an inequality constraint.

Definition 8 (Polyhedron)**.** A *polyhedron* is the intersection of *finite* numbers of half-spaces.

Definition 9 (Polytope)**.** A *polytope* is a bounded polyhedron. In other words, if a polyhedron can be fitted into a finite-radius ball, it is also a polytope.

Note that the above three geometries (half-space, polyhedron, polytope) are all convex sets, but not necessarily vice versa, that is, we can have a convex set which is neither of these definitions. An example is a ball (prove this formally, but intuitively note that a ball is the intersection of infinitely many half-spaces, since a ball is very smooth).

Next, we show some examples of polyhedra.

Example 10. *A box $[-1,1]^n$ is a polyhedron. See Figure 4.*

This is a hyper-box of width two in each dimension. We can easily check that it is a polyhedron with two inequality constraints in each dimension: $\forall i \in [n]$, $x_i \leq 1$, $x_i \geq -1$. In other words, it is the intersection of $2n$ half-spaces.

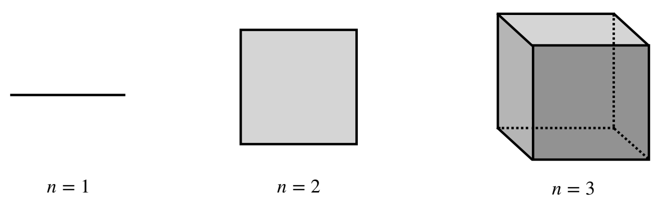

Figure 4: The n-dimensional box $[-1,1]^n$

Example 11. *A crosspolytope $\{x \in \mathbb{R}^n \mid \sum_{i=1}^n |x_i| \leq 1\}$ is a polyhedron. See Figure 5.*

This example is less obvious. We can describe it as the intersection of 2^n half-spaces:

$$\forall S \subseteq [n], \sum_{i \in S} x_i - \sum_{i \notin S} x_i \leq 1.$$

Here a question arises: can we reduce the number of constraints to describe an n-dimensional crosspolytope at a higher-dimensional space?

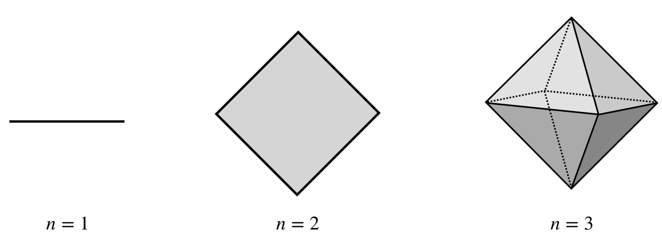

Figure 5: The n-dimensional crosspolytope $\{x \in \mathbb{R}^n \mid \sum_{i=1}^n |x_i| \leq 1\}$

The above examples are simple geometric polytopes. There is an entire field of combinatorial optimization, called *polyhedron tiling*, where we seek polytopes associated with particular combinatorial objects. An example is *Birkhoff polytope*, which corresponds to the feasible region of the bipartite perfect matching LP.

1.3 Relationship to Linear Programming

Based on the above definitions, we now revisit linear programming from a geometric point of view. Let P be a *general* LP:

$$\max_{x \in \mathbb{R}^n} c^\top x \quad \text{subject to} \quad Ax \leq b,$$

with $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^n$. Any inequality constraint $a_i^\top x \le b_i$ is a half-space, which is of course a convex set. The set of all feasible solutions of P is the intersection of m half-spaces, which by Claim 2 is also a convex set, and particularly, a polyhedron. Thus, in LP we want to optimize a linear objective function over the convex polyhedron.

Below we have another different interpretation of linear programming. One can show that any polytope is a convex hull of a certain number of points. Thus, we can see linear programming as given a set of points, optimize a linear function over their convex hull. In the more general case, *convex optimization* is to optimize a convex function over a convex set. Linear programming is a special case of it. We will revisit convex optimization later in this course as well.

2 Vertices, Extremal Points, and Basic Feasible Solutions

For any polyhedron S, we give the following three definitions of "corners" of the polyhedron S (note that we expect a polyhedron to be "pointy").

Definition 12 (Vertices, Extremal Points, and Basic Feasible Solutions)**.**

(1) A **vertex** of a polyhedron S is a unique maximizer of some linear function over S.
Precisely, $x \in S$ is a vertex if and only if there exists $c \in \mathbb{R}^n$, such that $\forall y \ne x \in S, c^\top x > c^\top y$.

(2) $x \in S$ is an **extremal point** if there does not exist $y \ne z \in S$ and $t \in (0,1)$, s.t. $x = ty + (1-t)z$.

(3) Let polyhedron $S = \{x \in \mathbb{R}^n \mid Ax \le b\}$ with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Its **basic feasible solutions (BFSs)** are $x \in S$ that make n linearly independent constraints tight (i.e., hold with equality).[a]

[a] We have defined BFSs previously in this course as well.

The above three definitions somehow capture the notion of "corner", but it is not clear at all that they should all refer to the same thing. The following proposition proves this important result.

Proposition 13. *The three definitions in Definition 12 are equal.*

Proof. We prove it by showing that (1) $\implies$ (2), (2) $\implies$ (3), and (3) $\implies$ (1).

(1) $\implies$ (2): We first show that if x is a vertex of S, then x is also an extremal point.

We prove by contradiction that there are two different points $y \ne z \in S$ and $t \in (0,1)$ such that $x = ty + (1-t)z$. We know from (1) that there is $c \in \mathbb{R}^n$ such that x is a *unique* maximizer over S. Thus,

$$c^\top x = t \cdot c^\top y + (1-t) \cdot c^\top z < t \cdot c^\top x + (1-t) \cdot c^\top x = c^\top x,$$

a contradiction.

(2) $\implies$ (3): Assuming x is an extremal point, we show that x is also a basic feasible solution.

Let $T \subseteq [m]$ be the set of tight constraints of x, i.e., the set of rows of A such that $A_T x = b_T$. If the rank of A_T is n, then we are done by the definition of BFS. Let us assume that the rank of A_T is less than n.

Since A_T is not full rank, it has a non-empty kernel and thus there exists a vector $w \ne 0^n \in \mathbb{R}^n$ such that $A_T w = 0$. Let $\epsilon > 0$ be sufficiently small (the choice will become clear) and take $y = x + \varepsilon w$ and $z = x - \varepsilon w$. Notice that $x = (y+z)/2$, namely, a convex combination of y and z. Thus, we only need to show that y, z belong to the polyhedron S to say that x is not ax extremal point, a contradiction.

We have $A_T y = A_T(x + \varepsilon w) = b_T + 0 = b_T$, and also $A_T z = b_T$. For the remaining part of A we have $A_{-T} y = A_{-T} x + \varepsilon A_{-T} y \le b_{-T}$, since ε is small enough. Similarly, we get $A_{-T} z \le b_{-T}$. Combining together we have $Ay \le b$ and $Az \le b$, which implies that $y, z \in S$, which concludes the proof.

(3) $\implies$ (1): We now prove the last part. Let x be a BFS of S, then we show that x is also a vertex of S.

Let B be the set of tight linearly independent constraints in x (its basis). Set $c = (\sum_{i \in B} a_i)$, where a_i is the i-th row of A. Since x is tight for the constraints B, we have $c^\top x = \sum_{i \in B} b_i$. And in the set $S = \{x \mid Ax \leqslant b\}$ all the points y have $c^\top y \leqslant \sum_{i \in B} b_i$. And as A_B is full-ranked, x is the only solution makes $c^\top x \leqslant \sum_{i \in B} b_i$ tight. To make the statement formal, in order to show that $c^\top y < c^\top x$ for all $y \neq x \in S$, prove by contradiction that there exists $c^\top y = c^\top x$ for some $y \neq x \in S$, which implies that $A_B y = b_B$ and $A_B(y - x) = 0$. Then A_B is not full-ranked and contradicts to the definition of BFSs. $\square$

Remark. It is worth noting that the definition of *vertex* and *extremal point* can be generalized to the case S is any convex set. In this case, the proof of (1) $\Rightarrow$ (2) still holds, i.e, a vertex is an extremal point. But, the converse is not true, namely, we can have extremal points that are not vertices; see the example below.

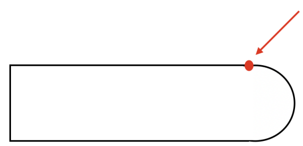

Figure 6: An example of a convex set (a circle attached to a rectangle) and a point (the connecting point of circle and rectangle) which is an extremal point but is not a vertex (the only linear function on the circle side that is uniquely maximized by this point is a tangent to the circle at that point; that line intersects the side of the rectangle fully and thus cannot uniquely maximizes this point on the rectangle).

Having such a equivalence between three different definitions is quite helpful. In the rest of this lecture, we use this equivalence to prove some new results about LPs.

2.1 Optimal Solutions and BFSs in General LPs

Recall that in Lecture 2, we saw that optimal solutions of LPs in the equational form happen on BFSs, and in Lecture 5, we mentioned that this is true for general LPs as well, not only the ones in the equational form (barring some "trivial" exceptions). We now prove this result.

Theorem 14 (Proposition 7 in Lecture 5)**.** *Consider a linear program in the general form P:*

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & c^\top x \\ \text{subject to} \quad & Ax \leq b. \end{aligned}$$

If P has an optimal solution and P has a BFS, then P has an optimal solution which is a BFS.

To prove this theorem we need the following geometric result (which we omit its proof for now).

Proposition 15. *A polyhedron P has an extremal point iff P contains no infinite* line *inside. Here a line is defined as $\{x + ty \mid t \in \mathbb{R}\}$ where $x \in P$ and $y \neq 0^n \in \mathbb{R}^n$.*

Using Proposition 15, we can prove Theorem 14 easily as follows.

Proof of Theorem 14. Let v be the optimal value of P and $Q = \{x \in P \mid c^\top x = v\}$ be the set of all optimal solutions. By the assumption in the theorem, we know that Q is non-empty. Moreover, by Proposition 15 we know that since P has an extremal point (by Proposition 13 since it has a BFS which is an extremal point), it cannot contain an infinite line inside. So, it means that its subset Q also has no infinite line, which, again by proposition 15, implies that Q has an extremal point. By Proposition 13, this means that Q also has a BFS. Let x be any BFS/extremal point in Q.

If x is also an extremal point of P we are done. Thus, let us assume toward a contradiction that P is not an extremal point, which means, there exists $y \neq z \in P$ and $t \in (0,1)$ such that $x = ty + (1-t)z$. By a similar argument in the proof of proposition 13, we know that $c^\top x = t \cdot c^\top y + (1-t) \cdot c^\top z$ and since $c^\top x = v$ is maximized, we should also have that $c^\top y = c^\top z = v$. But this implies that both $y, z \in Q$ as well, contradicting the fact that x was an extremal point of Q. $\square$

2.2 Integrality of the Bipartite Matching Polytope

Recall the bipartite matching polytope for a graph $G = (V, E)$:

Bipartite Matching Polytope (P):

$$\sum_{e \ni v} x_e \leqslant 1 \qquad \forall v \in V$$
$$x_e \geqslant 0 \qquad \forall e \in E.$$

In the maximum matching problem, our goal is to find an $x \in P$ that maximizes $\sum_e x_e$. Similarly, in the maximum weight matching problem, the goal is to find an $x \in P$ that maximizes $\sum_e w_e \cdot x_e$ where w_e is the weight of the edge $e \in E$. In Lecture 1, we showed that in the integrality gap of the maximum matching LP is 1, namely, any solution to that LP can be rounded to an integral matching without decreasing the objective value. We now prove a stronger result that shows that BFSs of the polytope (P) are all *integral*; combined with Theorem 14, this implies that optimal solutions of any LP with constraints corresponding to the polytope (P) are always integral[1].

Proposition 16. *Any vertex/extremal point/BFS of the bipartite matching polytope (P) is integral.*

Proof. By Proposition 13 we know the equivalence between vertices, extremal points, and BFSs of (P). Let x be an extremal point of (P) and suppose towards a contradiction that x has some fractional values. Let $F \subseteq E$ denote the set of edges in G with fractional values. We consider two cases.

A cycle $C = e_1, \ldots, e_{2k}$ in F. Since G is bipartite, any cycle in F can only be of even length. Let $\varepsilon > 0$ be sufficiently small and consider the vector $w \in \mathbb{R}^m$ defined such that $w_{e_i} = 1$ for $i \in \{1, 3, \ldots, 2k-1\}$ and $w_{e_j} = -1$ for $j \in \{2, 4, \ldots, 2k\}$, and zero everywhere else. Define $y = x + \varepsilon \cdot w$ and $z = x - \varepsilon \cdot w$ and note that $x = (y+z)/2$. We now show that y, z both belong to (P) which contradicts the fact that x was an extremal point.

To see why y belongs to (P), note that we can take $\varepsilon > 0$ sufficiently small so that $x_e + \varepsilon \leqslant 1$ and $x_e - \varepsilon \geqslant 0$, since all the edges in F had fractional value and thus are away from 0 and 1. Moreover, for any vertex $v \in V$, we have,

$$\sum_{e \ni v} y_e = \sum_{e \ni v} x_e + \varepsilon \cdot \sum_{e \ni v} w_e = (\sum_{e \ni v} x_e) + \varepsilon - \varepsilon \leqslant 1$$

where the last inequality is because x is in (P) and the equality before that is because any vertex is incident on either two consecutive edge of C (with $+\varepsilon$ and $-\varepsilon$ value in w) or none at all. This implies that y in in (P). A similar argument also shows z is in (P) as desired.

[1]In fact, it even shows something stronger. For an algorithm like Simplex for LPs that always return a BFS of the given LP, this result implies that we never need a rounding algorithm on top of Simplex because it anyway returns an integral solution to begin with.

A path $Q = e_1, \ldots, e_k$ in F. Let Q be a *maximal* path in that we cannot extend it in either direction in F. Let s and t be the endpoints of this path. Since the edges in F have fractional value and Q is a maximal path, we have that $0 < \sum_{e \ni s} x_e < 1$ and $0 < \sum_{e \ni t} x_e < 1$. Thus, we can again define y and z as before by alternatively increase or decrease the x_e-value of each edge in Q by ε. This way, for any internal vertex $v \in Q$ we have

$$\sum_{e \ni v} y_e = \sum_{e \ni v} z_e = \sum_{e \ni v} x_e \leqslant 1.$$

For the two endpoints s and t, the summations above may increase or decrease by ε compared to that of x, but since there is a slack for these constraints, by taking $\varepsilon > 0$ small enough, we will still have that y and z belong to (P) also.

To summarize, since F always contains either a cycle or a (maximal) path if it is non-empty, we showed that any fractional vector x can be written as a convex combination of two other points in (P), thus ensuring that x cannot be an extremal point. Thus, all vertices/extremal points/BFSs of the matching polytope (P) has to be integral, concluding the proof. $\square$