

ORF 522

Linear Programming and Convex Analysis

The geometry of the optimal solutions of an LP.

Marco Cuturi

Reminder: Convexity

- basic notions of convexity
 - **Convex** set C : $\forall \mathbf{x}_1, \mathbf{x}_2 \in C, [\mathbf{x}_1, \mathbf{x}_2] = \{\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2, 0 < \lambda < 1\} \subset C$.
 - **Boundary** point: $\forall r > 0, B_r(\mathbf{x}) \cap C \neq \emptyset, B_r(\mathbf{x}) \cap \mathbf{R}^n \setminus C \neq \emptyset$.
 - \mathbf{x} **extreme point** of a convex set: $\mathbf{x} = \frac{\mathbf{a} + \mathbf{b}}{2} \Rightarrow \mathbf{a} = \mathbf{b} = \mathbf{x}$
- hyperplanes and carathéodory
 - **isolation**; C convex: $\forall \mathbf{y} \notin C, \exists \mathbf{c} \in \mathbf{R}^n, z \in \mathbf{R} \mid C \subset H_{\mathbf{c},z}^+$
 - **supporting hyperplane**: $\mathbf{y} \in \partial C: \exists \mathbf{c} \in \mathbf{R}^n, z \in \mathbf{R} \mid C \subset \overline{H_{\mathbf{c},z}^+}, \mathbf{y} \in \overline{H_{\mathbf{c},z}^+}$
 - C **convex & bounded from below** : every supporting hyperplane of C contains an extreme point of C .
 - **convex hull** $\langle A \rangle$ of a set A is the minimal convex set that contains A .
 - See also convex hull = **all convex combinations**.
 - **Carathéodory**: $S \subset \mathbf{R}^n, \langle S \rangle = \bigcup_{C \subset S, \text{card}(C)=n+1} \langle C \rangle$.

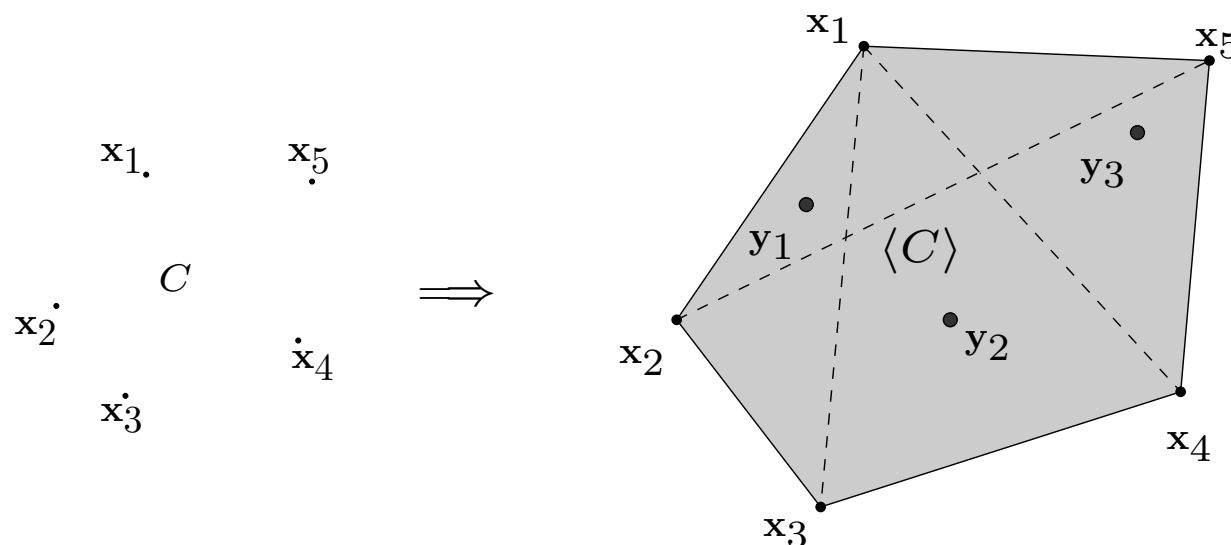
Today

- Carathéodory theorem proof
- Some important theorems
 - (i) Existence of one feasible solution \Rightarrow Existence of a **basic feasible** solution;
 - (ii) **basic feasible** solutions \Leftrightarrow **extreme** points of the feasible region;
 - (iii) **Optimum** of an LP occurs at an **extreme** point of the feasible region;
- A comment on polyhedra in canonical form and polyhedra in standard form.

Carathéodory Theorem

The intuition

- Start with the example of $C = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\} \subset \mathbf{R}^2$ and its hull $\langle C \rangle$.



- \mathbf{y}_1 can be written as a convex combination of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ (or $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5$);
 - \mathbf{y}_2 can be written as a convex combination of $\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4$;
 - \mathbf{y}_3 can be written as a convex combination of $\mathbf{x}_1, \mathbf{x}_4, \mathbf{x}_5$;
- For a set C of 5 points in \mathbf{R}^2 there seems to be always a way to write a point $\mathbf{y} \in \langle C \rangle$ as the convex combination of $2 + 1 = 3$ of such points.
- Is this result still valid for **general hulls** $\langle S \rangle$ (not necessarily polytopes but also balls etc..) and **higher dimensions**?

Remember for convex hull of finite sets...

- We had proved this for finite sets,

Theorem 1. *The smallest convex set that contains a finite set of points is the set of all their convex combinations.*

- **Proof reminder**, (\supseteq) is obvious.

- (\subseteq) : by induction on k . if $k = 1$ then $B_1 = \langle \{\mathbf{x}_1\} \rangle \subseteq A_1 = \{\mathbf{x}_1\}$.
- Suppose we have

$$\{\text{convex combinations of } x_1, \dots, x_{k-1}\} = A_{k-1} \subseteq B_{k-1} = \langle \{x_1, \dots, x_{k-1}\} \rangle.$$

- Let now $\mathbf{x} \in A_k$ such that $\mathbf{x} = \sum_{i=1}^k \alpha_i \mathbf{x}_i$.
- If $\mathbf{x} = \mathbf{x}_k$ then trivially $\mathbf{x} \in B_k$. If $\mathbf{x} \neq \mathbf{x}_k$ then

$$\mathbf{x} = (1 - \alpha_k) \sum_{i=1}^{k-1} \frac{\alpha_i}{1 - \alpha_k} \mathbf{x}_i + \frac{\alpha_k}{1 - \alpha_k} \mathbf{x}_k = (1 - \alpha_k) \mathbf{y} + \alpha_k \mathbf{x}_k.$$

- $\mathbf{y} \in B_{k-1} \subset B_k$ and $\mathbf{x}_k \in B_k$. B_k convex, hence $\mathbf{x} \in B_k$.

...generalized for convex hull of (infinite) sets

Theorem 2. *The convex hull $\langle S \rangle$ of a set of points S is the union of all convex combinations of k points $\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \in S, k \in \mathbf{N}$.*

$$\langle S \rangle = \bigcup_{k \geq 1} \bigcup_{C \subset S, \text{card}(C)=k} \langle C \rangle.$$

- **Proof** \supset is obvious.
- \subset
 - easy to show that the union on the right hand side is a **convex** set. **Not because it is a union** but because taking two points in this union we can show that their segment is included in the union.
 - hence the **RHS is convex** and **contains** S .
 - Hence it also contains the minimal convex set, $\langle S \rangle$.

Carathéodory's Theorem

Theorem 3. *Let $S \subset \mathbf{R}^n$. Then every point \mathbf{x} of the convex hull $\langle S \rangle$ can be represented as a convex combination of $n + 1$ points from S ,*

$$\mathbf{x} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_{n+1} \mathbf{x}_{n+1}, \sum_{i=1}^{n+1} \alpha_i = 1, \alpha_i \geq 0.$$

alternative formulation:

$$\langle S \rangle = \bigcup_{C \subset S, \text{card}(C)=n+1} \langle C \rangle.$$

- **Proof strategy:** show that when a point is written as a combination of m points and $m > n + 1$, it is possible to write it as a combination of $m - 1$ points by solving a homogeneous linear equation of $n + 1$ equations in \mathbf{R}^m .

Proof.

- (\supset) is direct.
- (\subset) any $\mathbf{x} \in \langle S \rangle$ can be written as a convex combination of p points, $\mathbf{x} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_p \mathbf{x}_p$. We can assume $\alpha_i > 0$ for $i = 1, \dots, p$.
 - If $p < n + 1$ then we add terms $0\mathbf{x}_{p+1} + 0\mathbf{x}_{p+2} + \cdots$ to get $n + 1$ terms.
 - If $p > n + 1$, we build a new combination with one term less:
 - ▷ let $A = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ 1 & 1 & \cdots & 1 \end{bmatrix} \in \mathbf{R}^{n+1 \times p}$.
 - ▷ **The key here is that since $p > n + 1$ there exists a solution $\eta \in \mathbf{R}^m \neq \mathbf{0}$ to $A\eta = \mathbf{0}$.**
 - ▷ By the last row of A , $\eta_1 + \eta_2 + \cdots + \eta_m = 0$, thus η has both $+$ and $-$ coordinates.
 - ▷ Let $\tau = \min\{\frac{\alpha_i}{\eta_i}, \eta_i > 0\} = \frac{\alpha_{i_0}}{\eta_{i_0}}$.
 - ▷ **Let $\tilde{\alpha}_i = \alpha_i - \tau\eta_i$.** Hence $\tilde{\alpha}_i \geq 0$, $\tilde{\alpha}_{i_0} = 0$ and $\tilde{\alpha}_1 + \cdots + \tilde{\alpha}_p = (\alpha_1 + \cdots + \alpha_p) - \tau(\eta_1 + \cdots + \eta_p) = 1$.
 - ▷ $\tilde{\alpha}_1 \mathbf{x}_1 + \cdots + \tilde{\alpha}_p \mathbf{x}_p = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_p \mathbf{x}_p - \tau(\eta_1 \mathbf{x}_1 + \cdots + \eta_p \mathbf{x}_p) = \mathbf{x}$.
 - ▷ Thus $\mathbf{x} = \sum_{i \neq i_0} \tilde{\alpha}_i \mathbf{x}_i$ of $p - 1$ points $\{\mathbf{x}_i, i \neq i_0\}$.
 - ▷ Iterate this procedure until \mathbf{x} is a convex combin. of $n + 1$ points of S .

Basic Solutions, Extreme Points and Optima of Linear Programs

Terminology

- A linear program is a **mathematical program** with **linear objectives** and **linear constraints**.
- A linear program in **canonical** form is the program

$$\begin{array}{ll} \text{maximize} & \mathbf{c}^T \mathbf{x} \\ \text{subject to} & \mathbf{Ax} \leq \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

$\mathbf{b} \geq \mathbf{0} \Rightarrow$ **feasible** canonical form. Initial feasible point: $\mathbf{x} = \mathbf{0}$.

- In broad terms:
 - In resource allocation problems canonical is more adapted,
 - in flow problems standard is usually more natural.
- However our algorithms work in **standard** form.

Terminology

- A linear program in **standard** form is the program

$$\text{maximize } \mathbf{c}^T \mathbf{x} \quad (1)$$

$$\text{subject to } \mathbf{Ax} = \mathbf{b}, \quad (2)$$

$$\mathbf{x} \geq \mathbf{0}. \quad (3)$$

- Easy to go from one to the other but dimensions of \mathbf{x} , \mathbf{c} , A , \mathbf{b} may change.
- Ultimately, **all** LP can be written in **standard form**.

Terminology

- Definition 1.** (i) A **feasible solution** to an LP in standard form is a vector \mathbf{x} that satisfies constraints (2)(3).
- (ii) The set of all feasible solutions is called the **feasible set** or **feasible region**.
- (iii) A feasible solution to an LP is an **optimal solution** if it maximizes the objective function of the LP.
- (iv) A feasible solution to an LP in standard form is said to be a **basic feasible solution (BFS)** if it is a basic solution with respect to Equation (2).
- (v) If a basic solution is non-degenerate, we call it a **non-degenerate basic feasible solution**.
- note that an optimal solution may not be unique, but the optimal value of the problem is.
 - Anytime “**basic**” is quoted, we are implicitly using the **standard form**.

\exists feasible solutions $\Rightarrow \exists$ basic feasible solutions

Theorem 4. *The feasible region to an LP is **convex, closed, bounded from below**.*

Theorem 5. *If there is a feasible solution to a LP in standard form, then there is a **basic feasible** solution.*

- **Proof idea:**

- if \mathbf{x} is such that $\sum_{i \in I} x_i \mathbf{a}_i = \mathbf{p}$ and where $\text{card}(I) > m$ then we show we can have an expansion of \mathbf{x} with a smaller family I' .
- Eventually by making I smaller we turn it into a basis \mathbf{I} .
- Some of the simplex's algorithm ideas are contained in the proof.

- **Remarks:**

- Finding an **initial** feasible solution might be a **problem to solve by itself**.
- We **assume** in the next slides **we have one**. More on this later.

Proof

Assume \mathbf{x} is a solution with $p \leq n$ positive variables. Up to a reordering and for convenience, assume that such variables are the p first variables, hence $\mathbf{x} = (x_1, \dots, x_p, 0, \dots, 0)$ and $\sum_{i=1}^p x_i \mathbf{a}_i = \mathbf{b}$.

- if $\{\mathbf{a}_i\}_{i=1}^p$ is linearly independent, then necessarily $p \leq m$. If $p = m$ then the solution is *basic*. If $p < m$ it is *basic and degenerate*.
- Suppose $\{\mathbf{a}_i\}_{i=1}^p$ is *linearly dependent*.
 - Assume all $\mathbf{a}_i, i \leq p$ are non-zero. If there is a zero vector we can remove it from the start. Hence we have $\sum_{i=1}^p \alpha_i \mathbf{a}_i = \mathbf{0}$ with $\alpha \neq \mathbf{0}$.
 - Let $\alpha_r \neq 0$, hence $\mathbf{a}_r = \sum_{j=1, j \neq r}^p \left(-\frac{\alpha_j}{\alpha_r}\right) \mathbf{a}_j$, which, when substituted in \mathbf{x} 's expansion,

$$\sum_{j=1, j \neq r}^p \left(x_j - x_r \frac{\alpha_j}{\alpha_r}\right) \mathbf{a}_j = \mathbf{b},$$

with has now no more than $p - 1$ **non-zero** variables.

- **non-zero** is not enough, since we need **feasibility**. We show how to choose r such that

$$x_j - x_r \frac{\alpha_j}{\alpha_r} \geq 0, j = 1, 2, \dots, p. \quad (4)$$

Proof

- For indexes j such that $\alpha_j = 0$ the condition (4) is ok. For those $\alpha_j \neq 0$, (4) becomes

$$\frac{x_j}{\alpha_j} - \frac{x_r}{\alpha_r} \geq 0 \quad \text{for } \alpha_j > 0, \quad (5)$$

$$\frac{x_j}{\alpha_j} - \frac{x_r}{\alpha_r} \leq 0 \quad \text{for } \alpha_j < 0, \quad (6)$$

- ▷ if $\alpha_r > 0$, (6) is Ok, we set $r = \operatorname{argmin}_j \left\{ \frac{x_j}{\alpha_j} \mid \alpha_j > 0 \right\}$ for (5)
- ▷ if $\alpha_r < 0$, (5) is Ok, we set $r = \operatorname{argmin}_j \left\{ \frac{x_j}{\alpha_j} \mid \alpha_j < 0 \right\}$ for (6)

In both cases r has been chosen suitably, such that (4) is always satisfied

- **Finally:** when $p > m$, we can show that there exists a **feasible** solution which can be written as a combination of $p - 1$ vectors $\mathbf{a}_i \Rightarrow$ only need to reiterate.

Basic feasible solutions of an LP \subset Extreme points of the feasible region

Theorem 6. *The **basic feasible** solutions of an LP in standard form are **extreme points** of the corresponding feasible region.*

- **Proof idea:** basic solutions means that \mathbf{x}_I is uniquely defined by B_I 's invertibility, that is \mathbf{x}_I is uniquely defined as $B_I^{-1}\mathbf{b}$. This helps to prove that \mathbf{x} is extreme.

Proof

- Suppose \mathbf{x} is a basic feasible solution, that is with proper reordering \mathbf{x} has the form $\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{0} \end{bmatrix}$ with $\mathbf{x}_B = B^{-1}\mathbf{b}$ and $B \in \mathbf{R}^{m \times m}$ an invertible matrix made of l.i. columns of A .
- Suppose $\exists \mathbf{x}_1, \mathbf{x}_2$ s.t. $\mathbf{x} = \frac{\mathbf{x}_1 + \mathbf{x}_2}{2}$.
- Write $\mathbf{x}_1 = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{v}_1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} \mathbf{u}_2 \\ \mathbf{v}_2 \end{bmatrix}$
- since $\mathbf{v}_1, \mathbf{v}_2 \geq 0$ and $\frac{\mathbf{v}_1 + \mathbf{v}_2}{2} = \mathbf{0}$ necessarily $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{0}$.
- Since \mathbf{x}_1 and \mathbf{x}_2 are feasible, $B\mathbf{u}_1 = \mathbf{b}$ and $B\mathbf{u}_2 = \mathbf{b}$ hence $\mathbf{u}_1 = \mathbf{u}_2 = B^{-1}\mathbf{b} = \mathbf{x}_B$ which proves that $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}$.

Basic feasible solutions of an LP \supset Extreme points of the feasible region

Theorem 7. *The **extreme points** of the feasible region of an LP in standard form are **basic feasible** solutions of the LP.*

- **Proof idea:** Similar to the previous proof, the fact that a point is extreme helps show that it only has m or less non-zero components.

Proof

Let \mathbf{x} be an extreme point of the feasible region of an LP, with $r \leq n$ zero variables. We reorder variables such that $x_i, i \leq r$ are positive and $x_i = 0$ for $r + 1 \leq i \leq n$.

- As usual $\sum_{i=1}^r x_i \mathbf{a}_i = \mathbf{b}$.
- Let us prove by contradiction that $\{\mathbf{a}_i\}_{i=1}^r$ are linearly independent.
- if not, $\exists(\alpha_1, \dots, \alpha_r) \neq \mathbf{0}$ such that $\sum_{i=1}^r \alpha_i \mathbf{a}_i = \mathbf{0}$. We show how to use the family α to create two distinct feasible points \mathbf{x}_1 and \mathbf{x}_2 such that \mathbf{x} is their center.
- Let $0 < \varepsilon < \min_{\alpha_i \neq 0} \frac{x_i}{|\alpha_i|}$. Then $x_i \pm \varepsilon \alpha_i > 0$ for $i \leq r$ and set $\mathbf{x}_1 = \mathbf{x} + \varepsilon \alpha$ and $\mathbf{x}_2 = \mathbf{x} - \varepsilon \alpha$ with $\alpha = (\alpha_1, \dots, \alpha_r, 0, \dots, 0) \in \mathbf{R}^n$.
- $\mathbf{x}_1, \mathbf{x}_2$ are feasible: by definition of $\varepsilon, \mathbf{x}_1, \mathbf{x}_2 \geq 0$. Furthermore, $A\mathbf{x}_1 = A\mathbf{x}_2 = A\mathbf{x} \pm \varepsilon A\alpha = \mathbf{b}$ since $A\alpha = \mathbf{0}$
- We have $\frac{\mathbf{x}_1 + \mathbf{x}_2}{2} = \mathbf{x}$ which is a contradiction.

∃ **extreme point** in the set of optimal solutions.

Theorem 8. *The **optimal** solution to an LP in standard form occurs at an **extreme point** of the feasible region.*

Proof. Suppose the optimal value of an LP is z^* and suppose the objective is to maximize $\mathbf{c}^T \mathbf{x}$.

- Any optimal solution \mathbf{x} is necessarily in the boundary of the feasible region. If not, $\exists \varepsilon > 0$ such that $\mathbf{x} + \varepsilon \mathbf{c}$ is still feasible, and $\mathbf{c}^T (\mathbf{x} + \varepsilon \mathbf{c}) = z^* + \varepsilon |\mathbf{c}|^2 > z^*$.
- The set of solutions is the intersection of $H_{\mathbf{c}, z^*}$ and the feasible region C which is *convex & bounded from below*. $H_{\mathbf{c}, z^*}$ is a supporting plane of C on the boundary point \mathbf{x} , thus $H_{\mathbf{c}, z^*}$ contains an extreme point (Thm. 3, lecture 3).

■

... *but some solutions that are not extreme points might be optimal.*

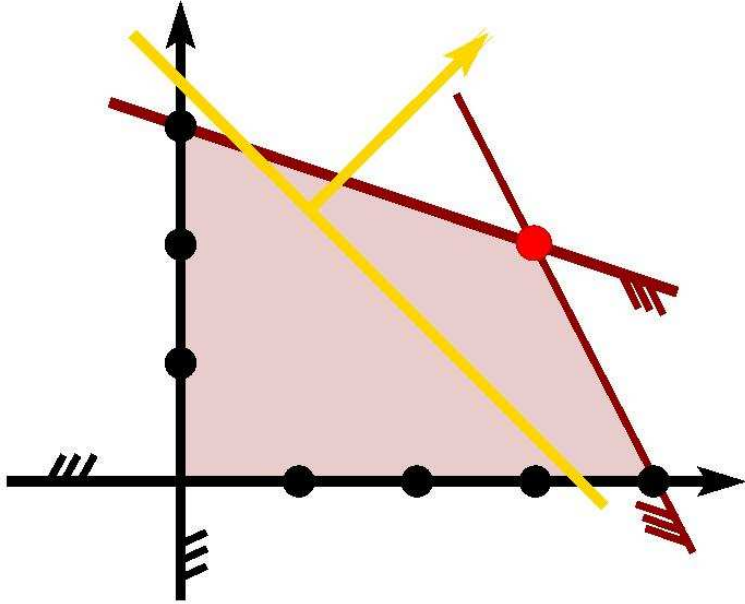
Wrap-up

- (i) a feasible solution exists \Rightarrow we know how to turn it into a **basic feasible** solution;
- (ii) **basic feasible** solutions \Leftrightarrow **extreme** points of the feasible region;
- (iii) **Optimum** of an LP occurs at an **extreme** point of the feasible region;

A Comment on Polyhedra in Canonical and Standard Form

Some extra bits of rigor

- Often you are shown this kind of image for LP's:



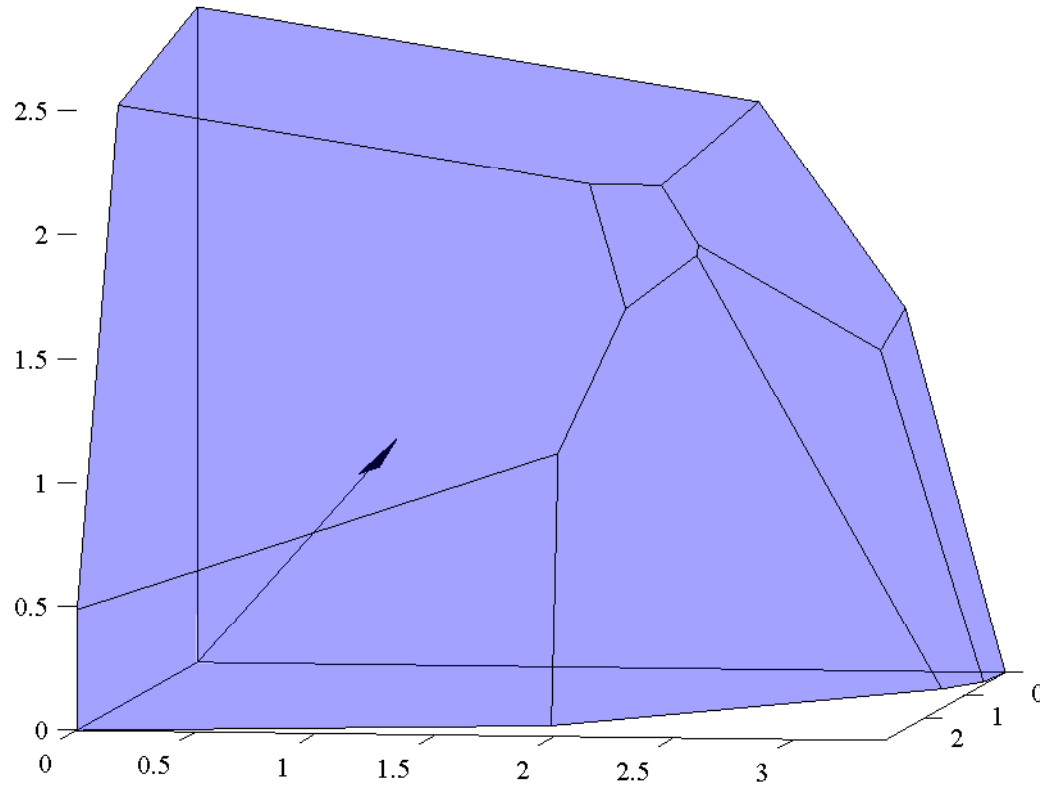
- And we think
 1. we have an LP,
 2. we can plot the feasible polyhedron (grey zone),
 3. the solutions need to be extreme points,
 4. and actually we can see it that they indeed are.
- Yet something's wrong in our logic. What?

Some extra bits of rigor

- We've only proved that the solutions of an LP in **standard** form are extreme points of the feasible set, which is an **intersection of hyperplanes**. No halfspaces involved.
- We've seen it before, **standard form** is poor when it comes to **visualization**: in 3D, 3 variables, 2 constraints, 1 line for the feasible set... that's the max.
- So to visualize we often use 2D or 3D, but in **canonical form**. That makes a more interesting polyhedron...
- Yet what are the connections between the **BFS** of the corresponding standard form and the **extreme points** in the canonical form?

Some extra bits of rigor

- In other words, does it make sense to think that vertices are relevant here?



- it does.

Some extra bits of rigor

- Suppose we have $P_1 = \{A\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq 0\} \subset \mathbf{R}^d$ the feasible set of a **canonical** form. We can **draw** it when $d = 3$, whatever m .
- We augment it to $P_2 = \{[A, I]\mathbf{x}' = \mathbf{b}, \mathbf{x}' \geq 0\} \subset \mathbf{R}^n$ to **run algorithms** in **standard** form.
- We can prove that an extreme point of P_2 , that is a BFS, corresponds to one and only extreme point of the fancy polyhedron in P_1 .
- ...coming as a simple homework question.

Example

- Great, we can start drawing in 2D and 3D.
- Consider the set in \mathbf{R}^2 defined by

$$\begin{array}{rclcl} x_1 & + & \frac{8}{3}x_2 & \leq & 4 \\ x_1 & + & x_2 & \leq & 2 \\ 2x_1 & & & \leq & 3 \\ & & & & x_1, x_2 \geq 0. \end{array}$$

- Here Let's add slack variables to convert it to standard form:

$$\begin{array}{rclclcl} x_1 & + & \frac{8}{3}x_2 & + & x_3 & \leq & 4 \\ x_1 & + & x_2 & & & + & x_4 & \leq & 2 \\ 2x_1 & & & & & & & + & x_5 & \leq & 3 \\ & & & & & & & & & & x_1, x_2, x_3, x_4, x_5 \geq 0. \end{array}$$

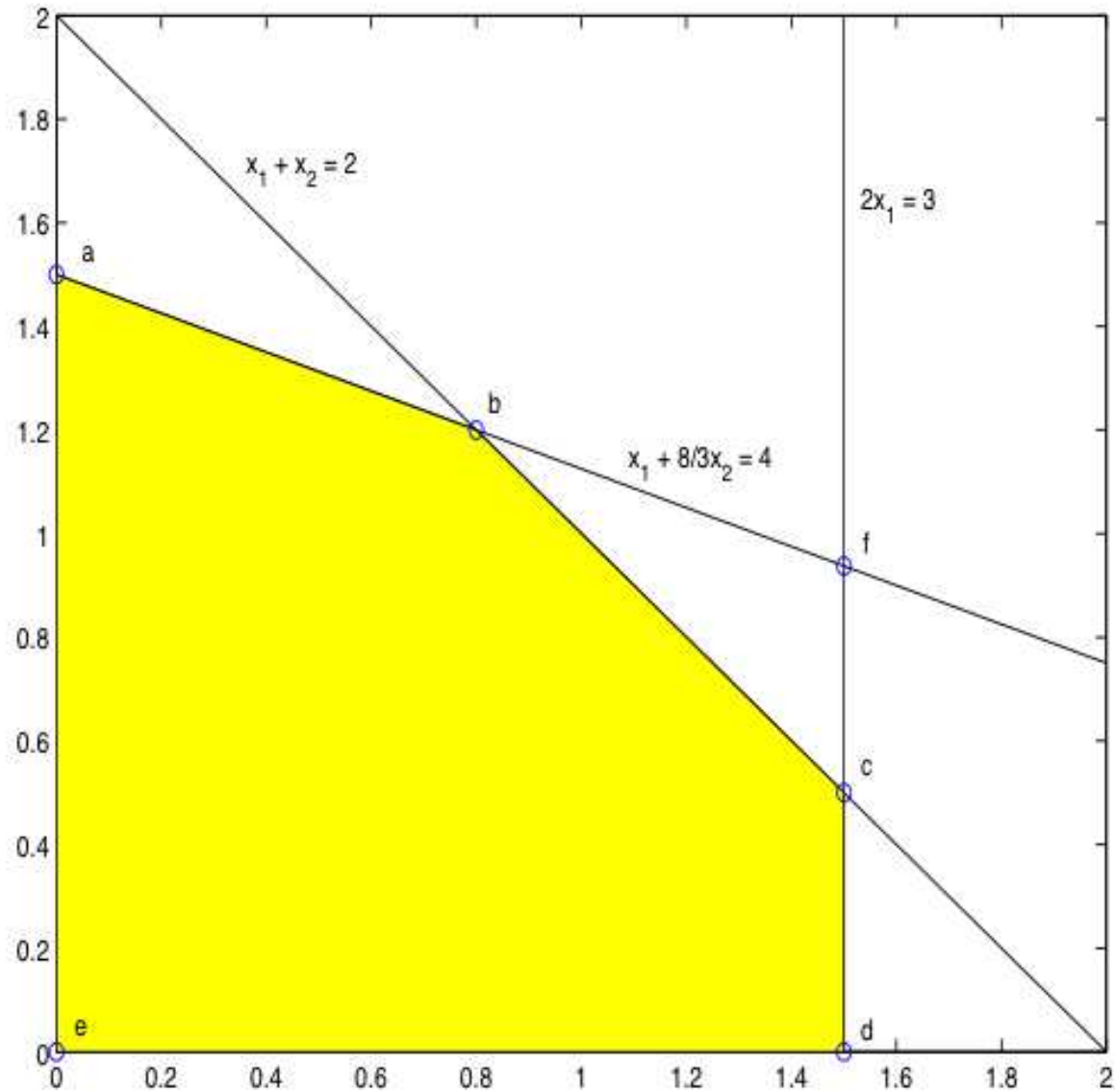
- Here $m = 3$, $n = 5$.

Example

- Check all basic solutions: we set two variables to zero and solve for the remaining three:
- Let $x_1 = x_3 = 0$ and solve for x_2, x_4, x_5

$$\begin{cases} \frac{8}{3}x_2 & & = & 4 \\ x_2 & + & x_4 & = & 2 \\ & & x_5 & = & 3 \end{cases}$$

- This gives the BFS $[0, \frac{3}{2}, 0, \frac{1}{2}, 3]$.
- Look how $[0, \frac{3}{2}]$ is an extreme point in the next figure (a).
- Not all basic solutions are feasible: the maximum is $\binom{5}{3} = 10$ but we typically have far less.
- The exact number is 5 here.
- the number of vertices of the polyhedron (the standard or the canonical, whichever you like!)



extreme points	a	b	c	d	e	f
non-basics	x_1, x_3	x_4, x_3	x_4, x_5	x_2, x_5	x_1, x_2	x_1, x_5

That's it for basic convex analysis and LP's

Major Recap

- A Linear Program is a program with linear constraints and objectives.
- Equivalent formulations for LP's: **canonical** (inequalities) and **standard** (equalities) form.
- Both have feasible **convex** sets that are **bounded from below**.
- **Simplex Algorithm** to solve LP's works in standard form.

- In **standard form**, the optimum occurs on an extreme point of this polyhedron.
- All **extreme points** are **basic feasible solutions**.
- That is, all extreme points are of the type $\mathbf{x}_I = B_I^{-1}\mathbf{b}$ for a subset **I** of coordinates, zero elsewhere.
- Looking for an optimum? **only need to check extreme points/BFS**
- Looking for an optimum? **there exists a basis I which realizes that optimum.**