ORF 522

Linear Programming and Convex Analysis

The geometry of the optimal solutions of an LP.

Marco Cuturi
Reminder: Convexity

• basic notions of convexity
  o **Convex** set $C$: $\forall x_1, x_2 \in C, [x_1, x_2] = \{\lambda x_1 + (1 - \lambda)x_2, 0 < \lambda < 1\} \subset C$.
  o **Boundary** point: $\forall r > 0, B_r(x) \cap C \neq \emptyset, B_r(x) \cap \mathbb{R}^n \setminus C \neq \emptyset$.
  o **x extreme point** of a convex set: $x = \frac{a+b}{2} \Rightarrow a = b = x$

• hyperplanes and carathéodory
  o **isolation**; $C$ convex: $\forall y \notin C, \exists c \in \mathbb{R}^n, z \in \mathbb{R} | C \subset H_{c,z}$
  o **supporting hyperplace**: $y \in \partial C$: $\exists c \in \mathbb{R}^n, z \in \mathbb{R} | C \subset \overline{H_{c,z}}$, $y \in \overline{H_{c,z}}$
  o **$C$ convex & bounded from below** : every supporting hyperplane of $C$ contains an extreme point of $C$.
  o **convex hull** $\langle A \rangle$ of a set $A$ is the minimal convex set that contains $A$.
  o See also convex hull = **all convex combinations**.
  o **Carathéodory**: $S \subset \mathbb{R}^n$, $\langle S \rangle = \bigcup_{C \subset S, \text{card}(C) = n+1} \langle C \rangle$. 

ORF-522
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 $C = \{x_1, x_2, x_3, x_4, x_5\} \subset \mathbb{R}^2$ and its hull $\langle C \rangle$.

\[ x_1, \quad x_5 \]
\[ x_2, \quad C \]
\[ x_3, \quad x_4 \]

\[ \implies \]

\[ y_1 \] can be written as a convex combination of $x_1, x_2, x_3$ (or $x_1, x_2, x_5$);
\[ y_2 \] can be written as a convex combination of $x_1, x_3, x_4$;
\[ y_3 \] can be written as a convex combination of $x_1, x_4, x_5$;

• For a set $C$ of 5 points in $\mathbb{R}^2$ there seems to be always a way to write a point $y \in \langle C \rangle$ 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.
  - \((\subset)\): by induction on \(k\). if \(k = 1\) then \(B_1 = \langle\{x_1\}\rangle \subseteq A_1 = \{x_1\}\).
  - Suppose we have

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

  - Let now \(x \in A_k\) such that \(x = \sum_{i=1}^{k} \alpha_i x_i\).
  - If \(x = x_k\) then trivially \(x \in B_k\). If \(x \neq x_k\) then

\[
x = (1 - \alpha_k) \sum_{i=1}^{k-1} \frac{\alpha_i}{1 - \alpha_k} x_i + \frac{\alpha_k}{1 - \alpha_k} x_k = (1 - \alpha_k)y + \alpha_k x_k.
\]

  - \(y \in B_{k-1} \subset B_k\) and \(x_k \in B_k\). \(B_k\) convex, hence \(x \in B_k\).
Theorem 2. The convex hull $\langle S \rangle$ of a set of points $S$ is the union of all convex combinations of $k$ points $\{x_1, \ldots, x_k\} \in S$, $k \in \mathbb{N}$.

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

- **Proof** $\supset$ is obvious.

- **C**
  - 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, $\langle S \rangle$. 

...generalized for convex hull of (infinite) sets
Theorem 3. Let $S \subset \mathbb{R}^n$. Then every point $x$ of the convex hull $\langle S \rangle$ can be represented as a convex combination of $n + 1$ points from $S$,

$$x = \alpha_1 x_1 + \cdots + \alpha_{n+1} x_{n+1}, \quad \sum_{i=1}^{n+1} \alpha_i = 1, \alpha_i \geq 0.$$ 

Alternative formulation:

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

- **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 $\mathbb{R}^m$. 

Proof.

- $(\supset)$ is direct.

- $(\subset)$ any $x \in \langle S \rangle$ can be written as a convex combination of $p$ points, $x = \alpha_1 x_1 + \cdots + \alpha_p x_p$. We can assume $\alpha_i > 0$ for $i = 1, \cdots, p$.
  - If $p < n + 1$ then we add terms $0x_{p+1} + 0x_{p+2} + \cdots$ to get $n + 1$ terms.
  - If $p > n + 1$, we build a new combination with one term less:
    - Let $A = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ 1 & 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n+1 \times p}$.
    - The key here is that since $p > n + 1$ there exists a solution $\eta \in \mathbb{R}^m \neq 0$ to $A\eta = 0$.
    - By the last row of $A$, $\eta_1 + \eta_2 + \cdots + \eta_m = 0$, thus $\eta$ has both $+$ and $-$ coordinates.
    - Let $\tau = \min\{\frac{\alpha_i}{\eta_i}, \eta_i > 0\} = \frac{\tilde{\alpha}_{i_0}}{\eta_{i_0}}$.
    - Let $\tilde{\alpha}_i = \alpha_i - \tau \eta_i$. Hence $\tilde{\alpha}_i \geq 0$, $\tilde{\alpha}_{i_0} = 0$ and
      $\tilde{\alpha}_1 + \cdots + \tilde{\alpha}_p = (\alpha_1 + \cdots + \alpha_p) - \tau(\eta_1 + \cdots + \eta_p) = 1$.
    - $\tilde{\alpha}_1 x_1 + \cdots + \tilde{\alpha}_p x_p = \alpha_1 x_1 + \cdots + \alpha_p x_p - \tau(\eta_1 x_1 + \cdots + \eta_p x_p) = x$.
    - Thus $x = \sum_{i \neq i_0} \alpha_i x_i$ of $p - 1$ points $\{x_i, i \neq i_0\}$.
    - Iterate this procedure until $x$ is a convex combin. of $n + 1$ points of $S$. 

ORF-522 9
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

  \[
  \begin{align*}
  \text{maximize} & \quad c^T x \\
  \text{subject to} & \quad Ax \leq b, \\
  & \quad x \geq 0.
  \end{align*}
  \]

  \(b \geq 0 \Rightarrow \text{feasible} \) canonical form. Initial feasible point: \(x = 0\).

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

\[
\begin{align*}
\text{maximize} & \quad c^T x \\
\text{subject to} & \quad Ax = b, \\
& \quad x \geq 0.
\end{align*}
\]

- Easy to go from one to the other but dimensions of \(x, c, A, b\) 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**.
∃ feasible solutions ⇒ ∃ 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 \( x \) is such that \( \sum_{i \in I} x_i a_i = p \) and where \( \text{card}(I) > m \) then we show we can have an expansion of \( x \) 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.
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 $
abla \mathbf{x} = (x_1, \ldots, x_p, 0, \ldots, 0)$ and $\sum_{i=1}^{p} x_i a_i = b$.

- If $\{a_i\}_{i=1}^{p}$ 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 $\{a_i\}_{i=1}^{p}$ is linearly dependent.
  - Assume all $a_i$, $i \leq p$ are non-zero. If there is a zero vector we can remove it from the start. Hence we have $\sum_{i=1}^{p} \alpha_i a_i = 0$ with $\alpha \neq 0$.
  - Let $\alpha_r \neq 0$, hence $a_r = \sum_{j=1, j\neq r}^{p} \left( -\frac{\alpha_j}{\alpha_r} \right) a_j$, which, when substituted in $\mathbf{x}$'s expansion,
    $$\sum_{j=1, j\neq r}^{p} \left( x_j - x_r \frac{\alpha_j}{\alpha_r} \right) a_j = b,$$
    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
    $$x_j - x_r \frac{\alpha_j}{\alpha_r} \geq 0, j = 1, 2, \ldots, p.$$ (4)
Proof

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

\[
\frac{x_j}{\alpha_j} - \frac{x_r}{\alpha_r} \geq 0 \quad \text{for } \alpha_j > 0, \tag{5}
\]

\[
\frac{x_j}{\alpha_j} - \frac{x_r}{\alpha_r} \leq 0 \quad \text{for } \alpha_j < 0, \tag{6}
\]

- if $\alpha_r > 0$, (6) is Ok, we set $r = \arg \min_j \left\{ \frac{x_j}{\alpha_j} \mid \alpha_j > 0 \right\}$ for (5)

- if $\alpha_r < 0$, (5) is Ok, we set $r = \arg \min_j \left\{ \frac{x_j}{\alpha_j} \mid \alpha_j < 0 \right\}$ 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 $a_i \Rightarrow$ only need to reiterate.
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 $x_I$ is uniquely defined by $B_I$’s invertibility, that is $x_I$ is uniquely defined as $B^{-1}_I b$. This helps to prove that $x$ is extreme.
• Suppose $x$ is a basic feasible solution, that is with proper reordering $x$ has the form $x = \begin{bmatrix} x_B \\ 0 \end{bmatrix}$ with $x_B = B^{-1}b$ and $B \in \mathbb{R}^{m \times m}$ an invertible matrix made of l.i. columns of $A$.

• Suppose $\exists x_1, x_2$ s.t. $x = \frac{x_1 + x_2}{2}$.

• Write $x_1 = \begin{bmatrix} u_1 \\ v_1 \end{bmatrix}$, $x_2 = \begin{bmatrix} u_2 \\ v_2 \end{bmatrix}$

• since $v_1, v_2 \geq 0$ and $\frac{v_1 + v_2}{2} = 0$ necessarily $v_1 = v_2 = 0$.

• Since $x_1$ and $x_2$ are feasible, $Bu_1 = b$ and $Bu_2 = b$ hence $u_1 = u_2 = B^{-1}b = x_B$ which proves that $x_1 = x_2 = x$. 
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 $x$ 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 $\sum_{i=1}^{r} x_i a_i = b$.
- Let us prove by contradiction that $\{a_i\}_{i=1}^{r}$ are linearly independent.
- if not, $\exists (\alpha_1, \cdots, \alpha_r) \neq 0$ such that $\sum_{i=1}^{r} \alpha_i a_i = 0$. We show how to use the family $\alpha$ to create two distinct feasible points $x_1$ and $x_2$ such that $x$ 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 $x_1 = x + \varepsilon \alpha$ and $x_2 = x - \varepsilon \alpha$ with $\alpha = (\alpha_1, \cdots, \alpha_r, 0, \cdots, 0) \in \mathbb{R}^n$.
- $x_1, x_2$ are feasible: by definition of $\varepsilon$, $x_1, x_2 \geq 0$. Furthermore, $Ax_1 = Ax_2 = Ax \pm \varepsilon A\alpha = b$ since $A\alpha = 0$
- We have $\frac{x_1 + x_2}{2} = x$ 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 $c^T x$.

- Any optimal solution $x$ is necessarily in the boundary of the feasible region. If not, $\exists \varepsilon > 0$ such that $x + \varepsilon c$ is still feasible, and $c^T (x + \varepsilon c) = z^* + \varepsilon |c|^2 > z^*$.

- The set of solutions is the intersection of $H_{c,z^*}$ and the feasible region $C$ which is convex & bounded from below. $H_{c,z^*}$ is a supporting plane of $C$ on the boundary point $x$, thus $H_{c,z^*}$ contains an extreme point (Thm. 3, lecture 3).

\[\square\]

... 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 $P_1 = \{Ax \leq b, x \geq 0\} \subset \mathbb{R}^d$ the feasible set of a canonical form. We can draw it when $d = 3$, whatever $m$.

- We augment it to $P_2 = \{[A, I]x' = b, x' \geq 0\} \subset \mathbb{R}^n$ 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.
Great, we can start drawing in 2D and 3D.

Consider the set in $\mathbb{R}^2$ defined by

\[
\begin{align*}
    x_1 + \frac{8}{3}x_2 &\leq 4 \\
    x_1 + x_2 &\leq 2 \\
    2x_1 &\leq 3 \\
    x_1, x_2 &\geq 0.
\end{align*}
\]

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

\[
\begin{align*}
    x_1 + \frac{8}{3}x_2 + x_3 &\leq 4 \\
    x_1 + x_2 + x_4 &\leq 2 \\
    2x_1 + x_5 &\leq 3 \\
    x_1, x_2, x_3, x_4, x_5 &\geq 0.
\end{align*}
\]

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 $x_2, x_4, x_5$

\[
\begin{align*}
\frac{8}{3} x_2 &= 4 \\
2x_2 + x_4 &= 2 \\
x_5 &= 3
\end{align*}
\]

- 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 $\left(\frac{5}{3}\right) = 10$ 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!)
<table>
<thead>
<tr>
<th>extreme points</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-basics</td>
<td>$x_1, x_3$</td>
<td>$x_4, x_3$</td>
<td>$x_4, x_5$</td>
<td>$x_2, x_5$</td>
<td>$x_1, x_2$</td>
<td>$x_1, x_5$</td>
</tr>
</tbody>
</table>
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 $x_I = B_I^{-1}b$ 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.