A Crash Course on Linear Programs[1]

- In these notes, we do a quick revision of basic concepts from the theory of linear programs which are often used in approximation algorithms.

- A linear program on n variables looks as follows.

$$\begin{aligned} \text{minimize} \quad & \underbrace{\mathbf{c}^\top \mathbf{x}}_{:=\mathsf{lp}(\mathbf{x})} = \sum_{i=1}^{n} c_i x_i & & \text{(Linear Program)} \\ & A\mathbf{x} \geq \mathbf{b}, \qquad A \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

  Note that the m constraints above may also include constraints of the form $x_i \geq 0$ and $x_i \leq 1$ (represented as $-x_i \geq -1$).

  Geometrically, one should think of each constraint $\mathbf{a}_i^\top \mathbf{x} \geq \mathbf{b}_i$, where $\mathbf{a}_i$ is the ith row of A, as a *half space* which partitions $\mathbb{R}^n$ into two parts, one of which satisfies $\mathbf{a}_i^\top \mathbf{x} \geq \mathbf{b}_i$, and the other with $< \mathbf{b}_i$. Any vector $\mathbf{x}$ which satisfies all the constraints are called *feasible* solutions. The set of feasible solutions, $\mathcal{F} := \{\mathbf{x} : \mathbf{x} \text{ is feasible}\}$, therefore is an *intersection of half-spaces*.

  Here is a simple fact which is easily checked.

  Fact 1. If $\mathbf{x}$ and $\mathbf{x}'$ are feasible, then so is $\theta\mathbf{x} + (1-\theta)\mathbf{x}'$ for any $0 \leq \theta \leq 1$.

  Geometrically, the above fact states that the intersection of half-spaces forms a *convex set* : given any two feasible points, the line-segment connecting them fully lies in $\mathcal{F}$.

- ***Basic Feasible Solutions.*** By definition, any feasible solution $\mathbf{x} \in \mathcal{F}$ satisfies the m-inequalities $A\mathbf{x} \geq \mathbf{b}$. Some of these inequalities are *tight*, that is, we have equality. The rest of the inequalities have slack. Let $T := \{i \in [m] \ : \ \mathbf{a}_i^\top \mathbf{x} = \mathbf{b}_i\}$ be the tight rows induced by $\mathbf{x}$. Let $B \subseteq T$ be a collection of *linearly independent* rows. Let $\mathbf{b}_B$ be the corresponding entries of $\mathbf{b}$. Note that $B\mathbf{x} = \mathbf{b}_B$.

  A feasible solution $\mathbf{x}$ is a *basic feasible solution* (bfs) if and only if $|B| = n$, that is, it satisfies exactly n linearly independent rows with equality. Geometrical, a bfs is the point obtained by the intersection of n linearly independent *hyperplanes*. Note that once the linearly independent collection of rows is fixed, then there is a unique solution satisfying them with equality; it is $\mathbf{x} = B^{-1}\mathbf{b}_B$.

  The following is another simple, but important, fact which shows why bfs are also called *extreme point* solutions or *vertex* solutions.

[1] *Lecture notes by Deeparnab Chakrabarty. Last modified : 20th Jan, 2022*
These have not gone through scrutiny and may contain errors. If you find any, or have any other comments, please email me at deeparnab@dartmouth.edu. Highly appreciated!

Fact 2. If $\mathbf{x}$ is a basic feasible solution, then there cannot exist feasible solutions $\mathbf{x}_1$ and $\mathbf{x}_2$, and parameter $0 < \theta < 1$ such that $\mathbf{x} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$.

Geometrically, a bfs $\mathbf{x}$ can't lie in any non-trivial line segment between two other feasible solutions. Thus, a bfs is a "corner" of the feasible set $\mathcal{F}$.

Proof. Let B be the set of linearly independent rows which are satisfied with equality. Thus, $B\mathbf{x} = \mathbf{b}_B$, and $\mathbf{x}$ is the unique such solution. Suppose that $\mathbf{x}_1$ and $\mathbf{x}_2$ exist such that $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 = \mathbf{x}$ for some $1 > \theta > 0$. Now, since $\mathbf{x}_1 \neq \mathbf{x}$ and since $\mathbf{x}_1$ is feasible, there must exist some $i \in B$ such that $\mathbf{a}_i^\top \mathbf{x}_1 > \mathbf{b}_i$. That is, some inequality in B which is slack at $\mathbf{x}_1$. We also know that $\mathbf{a}_i^\top \mathbf{x}_2 \geq \mathbf{b}_i$ and thus $(1-\theta)\mathbf{a}_i^\top \mathbf{x}_2 \geq (1-\theta)\mathbf{b}_i$. Now we get a contradiction:

$$\mathbf{b}_i = \mathbf{a}_i^\top \mathbf{x} = \mathbf{a}_i^\top \left(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2\right) > \theta\mathbf{b}_i + (1-\theta)\mathbf{b}_i > \mathbf{b}_i \qquad \square$$

Indeed, a converse is also true.

Fact 3. If $\mathbf{x}$ is a feasible solution which is *not* basic, then there exists two feasible solutions $\mathbf{x}_1$ and $\mathbf{x}_2$, and a parameter $0 < \theta < 1$ such that $\mathbf{x} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$.

Proof. Let T be the set of tight linear inequalities at $\mathbf{x}$. So, for all $j \notin T$, we have $\mathbf{a}_j^\top \mathbf{x} > \mathbf{b}_j$. Let $\delta := \min_{j \notin T} \left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right) > 0$.

Since $\mathbf{x}$ is not basic feasible, the rank of T is *not* n-dimensional. That is, $\mathcal{V} := \{T\mathbf{v} : \mathbf{v} \in \mathbb{R}^n\}$ is not an n-dimensional vector space. Which implies, the *null-space* is non-trivial. In particular, there exists a non-zero vector $\mathbf{v} \in \mathbb{R}^n$ such that $T\mathbf{v} = 0$. Define $M := \max_{j \notin T} \left|\mathbf{a}_j^\top \mathbf{v}\right|$ and let $\varepsilon := \frac{1}{2M}$ if $M > 0$ and $\varepsilon := 0$ otherwise. In any case, $-\varepsilon\delta M > -\delta/2$. Now consider the vectors

$$\mathbf{x}_1 := \mathbf{x} + \varepsilon\delta \cdot \mathbf{v} \;\text{ and }\; \mathbf{x}_2 := \mathbf{x} - \varepsilon\delta \cdot \mathbf{v}$$

Note that $\mathbf{x} = (\mathbf{x}_1 + \mathbf{x}_2)/2$. We also claim that $\mathbf{x}_1, \mathbf{x}_2$ are feasible. Indeed, $T\mathbf{x}_1 = T\mathbf{x}_2 = T\mathbf{x} = \mathbf{b}_T$. For any $j \notin T$, we see that

$$\mathbf{a}_j^\top \mathbf{x}_1 = \mathbf{a}_j^\top \mathbf{x} + \varepsilon\delta \cdot \mathbf{a}_j^\top \mathbf{v} \geq \mathbf{b}_j + \delta - \varepsilon\delta M \geq \mathbf{b}_j + \frac{\delta}{2}$$

So, $\mathbf{x}_1$ is feasible. Similarly, one argues that $\mathbf{x}_2$ is feasible, completing the proof. $\square$

Remark: *In fact, if one observes the proof carefully, one can in fact see that one can find feasible solutions $\mathbf{x}_1$ and $\mathbf{x}_2$ such that if T_1 and T_2 are the tight inequalities for them respectively, then (a) $T \subseteq T_1 \cap T_2$, and (b) at least one of them another row $j \notin T$ which is* not *in the row-space of T. This one gets by defining ε correctly. Continuing thus, one can then write any feasible solution as a* convex combination *of basic feasible solutions.*

Exercise: *Given any feasible solution* $\mathbf{x}$*, describe a procedure which finds basic feasible solutions* $\mathbf{x}_1, \ldots, \mathbf{x}_t$ *and positive coefficients* $\theta_1, \ldots, \theta_t$ *such that* $\sum_i \theta_i = 1$ *and* $\mathbf{x} = \sum_i \theta_i \mathbf{x}_i$*.*

- ***Optimal Basic Feasible Solutions.*** The following fact is quite useful. No matter the objective function, there always exists an optimum solution which is a basic feasible solution. Geometrically, if we look at the "parallel hyperplanes" $H_\gamma := \{\mathbf{c}^\top \mathbf{x} = \gamma\}$ for some real γ, we want to figure out the smallest γ for which $H_\gamma \cap \mathcal{F} \neq \emptyset$. For such a smallest γ, the intersection must contain a corner. If one pauses and thinks a bit, at least in three dimensions this should make sense : a body resting on any plane must have a corner touching the plane.

  Fact 4. For any linear program of the form (Linear Program), there exists an optimal solution $\mathbf{x}^*$ which is basic feasible.

  Proof. If $\mathbf{x}^*$ is not basic feasible, then it can be written as a convex combination of basic feasible solutions. That is, one can find bfs $\mathbf{x}_1, \ldots, \mathbf{x}_t$ and non-negative coefficients $\theta_1, \ldots, \theta_t$ with $\sum_i \theta_i = 1$ such that $\mathbf{x}^* = \sum_i \theta_i \mathbf{x}_i$. Since

  $$\mathbf{c}^\top \mathbf{x}^* = \sum_{i=1}^{t} \theta_i \left(\mathbf{c}^\top \mathbf{x}_i\right) \geq \sum_{i=1}^{t} \theta_i \left(\mathbf{c}^\top \mathbf{x}^*\right) = \mathbf{c}^\top \mathbf{x}^*$$

  we must have equality every where. That is, *each* $\mathbf{x}_i$ is an optimal solution and they are, by definition, basic feasible. $\square$

  The above fact is often used in the analysis of approximation algorithms in the following way. Suppose an LP relaxation of some problem looks like

  $$\text{minimize} \quad \underbrace{\mathbf{c}^\top \mathbf{x}}_{:=\mathsf{lp}(\mathbf{x})} \quad : \; \widetilde{A}\mathbf{x} \geq b, \quad 0 \leq \mathbf{x}_i \leq 1$$

  where in $\widetilde{A}$ we have pulled out the so-called non-trivial constraints. Now suppose the number of linearly independent rows in $\widetilde{A}$ is r. Or say, we can somehow prove that in any subset of rows of A that can ever hold simultaneously with equality, the maximum number of linearly independent rows is r. Then, we can assert that at least $(n-r)$ of the $\mathbf{x}_i$'s in a basic feasible solution must be set to $\{0,1\}$. Or, put differently, the number of *fractional* variables in any basic feasible solution is at most r. From a rounding perspective, only these r variables need to be rounded. This has been used in the design and analysis of many approximation algorithms.

- ***Local Optimum = Global Optimum.*** In this bullet point we give a very high-level feel of why linear programs are tractable, and how one may solve them. One key observation is that local optima, in a sense we define below, are indeed global optima.

  Recall, that any basic feasible solution $\mathbf{x}$ is associated with a subset B of n linearly independent rows of A such that $B\mathbf{x} = \mathbf{b}_B$ and $A\mathbf{x} \geq b$. For the purpose of this discussion, which is high-level as it is, we assume that the problem given is *non-degenerate* in the following sense : for every row $\mathbf{a}_j \notin B$, we in fact have $\mathbf{a}_j^\top \mathbf{x} > \mathbf{b}_j$. That is, the inequality holds *strictly*. We also assume that the feasible region is bounded, that is, there is no infinite *ray* which lies in the feasible region.

  We now establish the following fact about basic feasible solutions which allow us to move from one to the other.

Fact 5. Let $\mathbf{x}$ be a basic feasible solution to $A\mathbf{x} \geq b$ and let B be the corresponding basis. Then, for every row $i \in B$ there exists a row $j \in A \setminus B$ such that $B' := B = i + j$ is a basis and $\mathbf{x}^{(i)} := (B')^{-1}\mathbf{b}_{B'}$ is a *feasible* solution (and thus also a basic feasible soluton). Furthermore, $B \cdot (\mathbf{x}^{(i)} - \mathbf{x}) = \gamma_i \mathbf{e}_i$ for some $\gamma_i > 0$, and in particular, the vectors $(\mathbf{x}^{(i)} - \mathbf{x})$ span $\mathbb{R}^n$.

The basic feasible solutions $\mathbf{x}$ and $\mathbf{x}'$ in the above fact statement are called *neighboring* bfs's.

Proof. The ith unit vector $\mathbf{e}_i \in \mathbb{R}^n$ is the n-dimensional vector with 1 in the first coordinate and 0 everywhere else. Consider the vector $\mathbf{d} := B^{-1}\mathbf{e}_i$. Notice that $B \cdot (\mathbf{x} + \theta\mathbf{d}) = \mathbf{b}_B + \theta\mathbf{e}_i$. In particular, if $\theta > 0$, we have $B \cdot (\mathbf{x} + \theta\mathbf{d}) \geq \mathbf{b}_B$, with equalities everywhere but the ith row. Furthermore, for any $j \notin B$, we can choose θ small enough such that $\mathbf{a}_j^\top(\mathbf{x} + \theta\mathbf{d}) > \mathbf{b}_j$. This is where we are using the non-degeneracy assumption that $\mathbf{a}_j^\top \mathbf{x} > \mathbf{b}_j$ to begin with. In particular, $\mathbf{x} + \theta\mathbf{d}$ is feasible.

Now, there must exist a smallest θ at which one $j \notin B$ gets "tight", that is, $\mathbf{a}_j^\top(\mathbf{x} + \theta\mathbf{d}) = \mathbf{b}_j$. Otherwise, the infinite ray will lie in the feasible region, and we are assuming the feasible region is bounded. This is the j we are looking for and γ_i is this θ. Indeed, let $\mathbf{x}^{(i)} := \mathbf{x} + \gamma_i\mathbf{d}$, and let $B' := B - i + j$. By design, $B'\mathbf{x}^{(i)} = \mathbf{b}_{B'}$ since all but the ith inequality in B is satisfied with equality, and so does the jth inequality now. Also, $B \cdot (\mathbf{x}^{(i)} - \mathbf{x}) = B \cdot (\gamma_i d) = \gamma_i \mathbf{e}_i$. It is also easy to see B' is a basis. To show this we only need to show $\mathbf{a}_j$ is linearly independent to the rows of $B - i$. This is simply because $\mathbf{a}_j^\top \mathbf{d} \neq 0$ (since the jth inequality went from being not tight to tight) but $\mathbf{a}_\ell^\top \mathbf{d} = 0$ for all $\mathbf{a}_\ell \in B - i$, by design of $\mathbf{d}$. □

We say a bfs $\mathbf{x}$ is a *local optimum* solution if for any neighboring bfs $\mathbf{x}'$, we have $\mathsf{lp}(\mathbf{x}) \leq \mathsf{lp}(\mathbf{x}')$.

Fact 6. A locally optimum basic feasible solution is a global optimum.

Proof. Again, we prove it under the assumption of non-degeneracy and boundedness; both of these can be removed with some extra work. Under these assumptions, Fact 5 implies that for any bfs $\mathbf{x}$ with basis B, there are n different neighboring bfs's $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ such that the vectors $(\mathbf{x}^{(i)} - \mathbf{x})$ span $\mathbb{R}^n$ and $B(\mathbf{x}^{(i)} - \mathbf{x}) = \gamma_i \mathbf{e}_i$ for some $\gamma_i > 0$.

Let $\mathbf{x}^*$ be a global optimum bfs. We can write $\mathbf{x}^* - \mathbf{x} = \sum_{i=1}^n \alpha_i \cdot (\mathbf{x}^{(i)} - \mathbf{x})$ for some $\alpha_i \in \mathbb{R}$ since the $(\mathbf{x}^{(i)} - \mathbf{x})$'s span $\mathbb{R}^n$. Since $B\mathbf{x} = \mathbf{b}_B$ and since $B\mathbf{x}^* \geq \mathbf{b}_B$, we get that $B(\mathbf{x}^* - \mathbf{x}) \geq 0$. Substituting, we get that $\sum_{i=1}^n \alpha_i \gamma_i \mathbf{e}_i \geq 0$. Therefore, since $\gamma_i > 0$, we get $\alpha_i \geq 0$ for all $1 \leq i \leq n$. Therefore,

$$\mathsf{lp}(\mathbf{x}^*) - \mathsf{lp}(\mathbf{x}) = \mathbf{c}^\top(\mathbf{x}^* - \mathbf{x}) = \sum_{i=1}^n \alpha_i \cdot \left(\mathbf{c}^\top \mathbf{x}^{(i)} - \mathbf{c}^\top \mathbf{x}\right) = \sum_{i=1}^n \alpha_i \cdot \left(\mathsf{lp}(\mathbf{x}^{(i)}) - \mathsf{lp}(\mathbf{x})\right)$$

Since $\mathbf{x}$ is a local optimum, each parenthesized item in the RHS is ≥ 0 and since $\alpha_i \geq 0$, we get the RHS is ≥ 0. Which implies $\mathsf{lp}(\mathbf{x}) \leq \mathsf{lp}(\mathbf{x}^*)$, completing the proof. □

Remark: *The above discussion forms the seeds of the* simplex *method which was the first impactful algorithm designed for solving LPs. At a very high level, instead of doing a local search, the algorithm chooses which row to swap in the basis using a so-called "pivot rule". The Simplex Algorithm is still used extensively everywhere LPs are solved, which is in almost all sectors of engineering. On the other hand, there is no "pivot rule" which is known to terminate in polynomial time! For, I think, almost every pivot rule designed, researchers have come up with pathological polytopes where that rule takes super-polynomial time.*

- ***Duality.***
- ***Solving LPs via Cutting Planes.***

Notes

Since this is not a course on linear programming, my notes will be short because the alternative is to be extremely long. All I will say is that everyone who studies linear programming has a favorite source which enlightened them. For me it was this beautiful text [1] by Bertsimas and Tsitsiklis.

References

[1] D. Bertsimas and J. Tsitsiklis. *Introduction to Linear Programming*. Athena-Scientific, 1997.