Advanced Heuristic and Approximation Algorithms Grenoble-INP (ORCO)

Lecture 1

Lecturer: Alantha Newman *October 5, 2018*

1 Introduction to Convex Relaxations

Given a bipartite graph $G = (V, E)$ with edge weights w_e, let us consider the problem of finding a maximum weight matching. A matching is set of edges $M \subseteq E$ such that each vertex V is adjacent to at most one edge in M. If each vertex in V is adjacent to exactly one edge in M, then M is a *perfect matching*.

Finding maximum cardinality and maximum weight matchings in bipartite graphs are fundamental discrete optimization problems. One method for finding a maximum *cardinality* matching (i.e., a maximum weight matching when all edges have unit weight, $w_e = 1$) is the (combinatorial) *augmenting path* algorithm. A method for finding a maximum weight matching is via a reduction to the maximum flow problem. Here we consider the approach of solving the problem via a convex relaxation.

1.1 Polytopes

A polytope can provide a (sometimes compact) description of the solutions to a discrete optimization problem. A matching M in a graph $G = (V, E)$ can be represented by a vector in $\{0, 1\}^E$, where there is an entry in the vector for each edge $e \in E$, and the value for that entry is a 1 if edge $e \in M$ and 0 otherwise. If we take the convex hull of all such points corresponding to the set of matchings, then we have the *matching polytope*. Some basic definitions:

Definition 1. *A set $S \in \mathbb{R}^n$ is* convex *if for all $\mathbf{x}, \mathbf{y} \in S$, and any $\lambda \in [0, 1]$, $\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} \in S$.*

Definition 2. *A* polytope *is the convex hull of a finite number of points in $\mathbb{R}^n$.*

A polytope is a convex set.

1.2 Polyhedra

A polytope is a polyhedron, and this connection yields an alternate definition and description of a polytope. The following geometric concepts are necessary to define and understand polyhedra.

Let n be a positive integer, let $\mathbf{a} \in \mathbb{R}^n$ be a vector and let $b \in \mathbb{R}$ be a scalar.

Definition 3. *The set $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^\intercal\mathbf{x} = b\}$ is called a* hyperplane*.*

Definition 4. *The set $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^\intercal\mathbf{x} \geq b\}$ is called a* halfspace*.*

Note that a halfspace consists of all points that are one side of a hyperplane.

Definition 5. *A* polyhedron *is the intersection of a finite number of halfspaces.*

Like a polytope, a polyhedron is also a convex set. We say a polyhedron is *bounded* if it does not contain a line or a half-line. The set $P \in \mathbb{R}^n$ is a polytope if and only if it is a bounded polyhedron.[1]

[1] This can be proven using the Weyl-Minkowski Theorem.

1.3 Vertices of a Polyhedron

Consider the polyhedron shown in Figure 1. Notice that the optimal solution occurs at a "corner" of the polyhedron. There are several different ways of formally defining these special points.

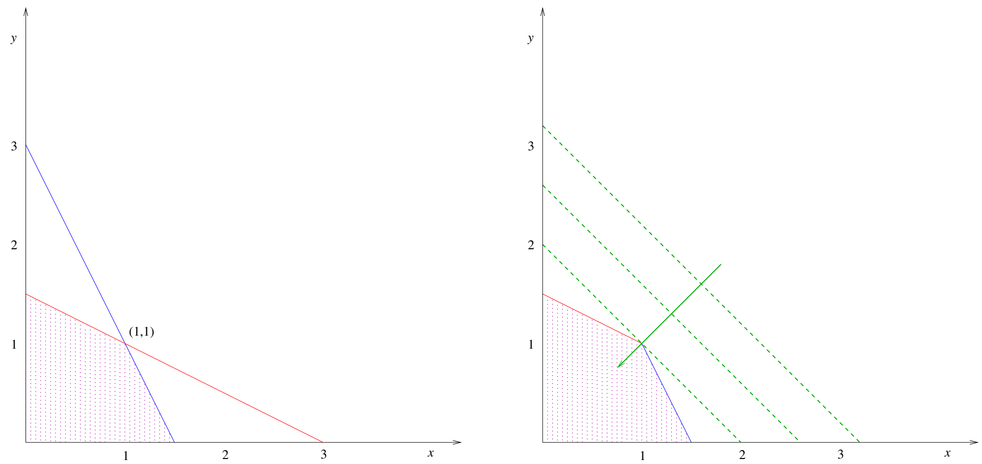

Figure 1: The polyhedron shown in these figures is the intersection of the halfspaces: $\{-x - 2y \geq -3, -2x - y \geq -3, x \geq 0, y \geq 0\}$. The objective function is $\min -x - y$. Consider the hyperplane $-x - y = z$ and slide it (i.e. increase the value of z) towards the polyhedron. The first point where it touches the polyhedron is $(1, 1)$, where $-x - y = -2$.

We denote a polyhedron $P \subset \mathbb{R}^n$ as follows: $P = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$.

Definition 6. *Let P be a polyhedron. A vector $\mathbf{x} \in P$ is a* vertex *of P if there exists some $\mathbf{c}$ such that $\mathbf{c}'\mathbf{x} < \mathbf{c}'\mathbf{y}$ for all $\mathbf{y} \in P$, $\mathbf{y} \neq \mathbf{x}$.*

Another way to define a vertex: $\mathbf{x}$ is a vertex of P iff P is on one side of the hyperplane $\{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{c}'\mathbf{y} = \mathbf{c}'\mathbf{x}\}$, which meets P only at the point $\mathbf{x}$.

Definition 7. *Let P be a polyhedron. A vector $\mathbf{x} \in P$ is an* extreme point *of P if we cannot find two vectors $\mathbf{y}, \mathbf{z} \in P$ ($\mathbf{y} \neq \mathbf{x}$, $\mathbf{z} \neq \mathbf{x}$) and a scalar $\lambda \in [0, 1]$ such that $\mathbf{x} = \lambda\mathbf{y} + (1 - \lambda)\mathbf{z}$.*

A point $\mathbf{x} \in \mathbb{R}^n$ is a vertex of P if and only if it is an extreme point of P.

1.4 Formal Description of a Linear Program

A *linear program* is the problem of optimizing a linear objective function over the points in a polyhedron. For instance, let $\mathbf{A}$ be an $m \times n$ matrix (i.e., m rows and n columns), and let $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{c} \in \mathbb{R}^n$ be vectors. The vector $\mathbf{x} \in \mathbb{R}^n$ represents the vector of the n variables $x_1, \ldots, x_n$. Then the following is a (generic) linear program.

$$\begin{aligned} \min \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{subject to:} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}. \end{aligned} \qquad (P)$$

We often require that the variables are nonnegative and write a linear program as follows.

$$\begin{aligned} \min \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{subject to:} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b} \\ & \mathbf{x} \geq 0. \end{aligned} \qquad (P)$$

An optimal solution to (P) is a point $\mathbf{y} \in P$ such that the objective value, $\mathbf{c}^\intercal \mathbf{y}$, is minimum.

Note that writing an integer program (and consequently a linear programming relaxation) for a discrete optimization problem (e.g., maximum weight matching, minimum weight spanning tree) is not always obvious or easy.

1.5 Integrality Gap

One way of measuring the quality of a linear programming relaxation is known as the *integrality gap*. Given a set $\mathcal{I}$ of instances for a minimization problem, let $OPT(I)$ denote the optimal (integral) solution of a fixed integer program on instance I, and let $OPT_f(I)$ denote the optimal fractional solution for the linear programming relaxation of this integer program. Then the integrality gap of the linear programming relaxation is defined as:

$$\sup_{I \in \mathcal{I}} \frac{OPT(I)}{OPT_f(I)}.$$

For a maximization problem, it is defined as:

$$\sup_{I \in \mathcal{I}} \frac{OPT_f(I)}{OPT(I)}.$$

The integrality gap establishes a limit on the usefulness of a linear programming relaxation. Specifically, if an approximation algorithm is solely based on a linear program, then it cannot be used to obtain an approximation ratio better than the integrality gap of this linear program.

2 Matching Polytope in Bipartite Graphs

For each matching M in G, let $\chi^M \in \{0,1\}^{|E|}$ be the indicator vector of the matching. Let $\mathcal{M}(G)$ denote the convex hull of all such indicator vectors of matchings in G.

$$\mathcal{M}(G) \quad = \quad \text{convex-hull}\{\chi^M \mid M \text{ is a matching}\}.$$

Consider the following polyhedron consisting of valid matching constraints:

$$Q^{\mathrm{M}}_{\mathrm{BIP}}(G) \quad = \quad \{\mathbf{x} \mid \sum_{e \in \delta(v)} x_e \leq 1, \forall v \in V;\ x_e \geq 0, \forall e \in E\}.$$

We observe that for any graph G containing at least one edge, $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$ is a *bounded* polyhedron (since it does not contain a line) and is therefore a polytope.

Lemma 8. $\mathcal{M}(G) \subseteq Q^{M}_{\mathrm{BIP}}(G)$.

Proof. Each indicator vector of a matching obeys the constraints in $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$, since the constraints in $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$ are valid for any matching in G. □

Lemma 9. *If G is bipartite, then $Q^{M}_{\mathrm{BIP}}(G) \subseteq \mathcal{M}(G)$.*

Proof. Consider the smallest graph G (in terms of $|V| + |E|$) such that $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$ has a vertex that is not a matching, i.e. is not integral. Consider such a vertex $\mathbf{x}^\star \in Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$. Note that $\mathbf{x}^\star$ has no integer entries, as that would allow us to delete the corresponding edges from G and consider an even smaller graph. Now let us consider two cases:

(i) Suppose that G contains a cycle C. Since G is bipartite, C is an even cycle. Define:

1. $\alpha = \min_{e \in C} x^\star_e$,
2. $\beta = \max_{e \in C} x^\star_e$.

Let $\epsilon = \min\{\alpha, 1-\beta\}$ and let $\mathbf{z} = \{1, -1, 1, \ldots, -1\}$, where $\mathbf{z}$ has length $|C|$. Then we obtain the following two vectors:

$$\mathbf{x}' = \mathbf{x}^\star + \epsilon\mathbf{z}, \qquad \mathbf{x}'' = \mathbf{x}^\star - \epsilon\mathbf{z}.$$

Both $\mathbf{x}'$ and $\mathbf{x}''$ satisfy the constraints in $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$, so $\mathbf{x}', \mathbf{x}'' \in Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$. Since $\mathbf{x}^\star = (\mathbf{x}' + \mathbf{x}'')/2$, it follows that $\mathbf{x}^\star$ is not an extreme point of $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$, which is a contradiction.

(ii) Suppose G does not contain a cycle. Then it contains a path P. Define:

1. $\alpha = \min_{e \in S} x^\star_e$,
2. $\beta = \max_{e \in S} x^\star_e$.

Let $\epsilon = \min\{\alpha, 1-\beta\}$ and let $\mathbf{z} = \{1, -1, 1, \ldots, -1\}$, where $\mathbf{z}$ has length $|P|$. As in case (i), we obtain the following two vectors:

$$\mathbf{x}' = \mathbf{x}^\star + \epsilon\mathbf{z}, \qquad \mathbf{x}'' = \mathbf{x}^\star - \epsilon\mathbf{z}.$$

Both $\mathbf{x}'$ and $\mathbf{x}''$ satisfy the constraints in $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$, so $\mathbf{x}', \mathbf{x}'' \in Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$. Since $\mathbf{x}^\star = (\mathbf{x}' + \mathbf{x}'')/2$, it follows that $\mathbf{x}^\star$ is not an extreme point of $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$, which is a contradiction. □

Theorem 10 follows from Lemmas 8 and 9. It implies that if we can find an extreme point in the polyhedron $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$ that maximizes a specified objective function (i.e., using the simplex algorithm), then this solution will be integral.

Theorem 10. *If G is bipartite, then $Q^{M}_{\mathrm{BIP}}(G) = \mathcal{M}(G)$.*

2.1 Perfect Matchings

Analogously, we can define the convex hull of perfect matchings. For each perfect matching M in G, let $\chi^M \in \{0,1\}^{|E|}$ be the indicator vector of the matching. Let $\mathcal{PM}(G)$ denote the convex hull of all such indicator vectors of matchings in G.

$$\mathcal{PM}(G) \quad = \quad \text{convex-hull}\{\chi^M \mid M \text{ is a perfect matching}\}.$$

We can also define a polyhedron consisting of constraints valid for perfect matchings:

$$Q^{\mathrm{PM}}_{\mathrm{BIP}}(G) \quad = \quad \{\mathbf{x} \mid \sum_{e \in \delta(v)} x_e = 1, \forall v \in V;\ x_e \geq 0, \forall e \in E\}.$$

Theorem 11. *If G is bipartite, then $Q^{PM}_{\mathrm{BIP}}(G) = \mathcal{PM}(G)$.*

Proof. $\mathcal{PM}(G) \subseteq Q^{\mathrm{PM}}_{\mathrm{BIP}}(G)$, since each indicator vector of a perfect matching obeys the constraints in $Q^{\mathrm{PM}}_{\mathrm{BIP}}(G)$, since the constraints in $Q^{\mathrm{PM}}_{\mathrm{BIP}}(G)$ are valid for any perfect matching in G. To show that $Q^{\mathrm{PM}}_{\mathrm{BIP}}(G) \subseteq \mathcal{PM}(G)$, it is sufficient to consider case (i) in the proof of Lemma 9. □

Theorems 10 and 11 do not hold when G is not bipartite. Let G be an odd cycle and define $x_e = 1/2$ for each edge in the cycle. Then $\mathbf{x}$ is in $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$, but $\mathbf{x}$ is not a convex combination of matchings. In other words, if G is an odd cycle, then $Q^{\mathrm{M}}_{\mathrm{BIP}}(G)$ is not an integer polytope.

Note that for the weighted matching problem in bipartite graphs, the integrality gap of the linear program Q^{M} is 1. We now look at examples of integer programs for some NP-hard discrete optimization problems and their respective linear programming relaxations for which the integrality gap is not 1 (i.e., the relaxation is not exact).

3 Vertex Cover

Given a graph $G = (V, E)$, the *vertex cover problem* is to find a subset of vertices $S \subset V$ such that for each edge $ij \in E$, at least one endpoint belongs to S. In other words, either i or j belongs to S or, alternatively, both i and j belong to S. The vertex cover problem is NP-hard. We now present a 2-approximation algorithm. Let OPT_{VC} denote that size of a minimum vertex cover for the graph G.

3.1 2-Approximation Algorithm

Since vertex cover is a minimization problem, we first need to find a (preferably large) lower bound. A subset of edges $M \subseteq E$ is a *matching* in G if all edges are vertex disjoint.

Lemma 12. *Let $M \subseteq E$ be a matching in G. Then $|M|$ is a lower bound on the size of a minimum vertex cover in G.*

Proof. Any matching is a set of vertex-disjoint edges. Thus, at least one endpoint of each edge belongs to a vertex cover. Otherwise, there is some edge that is not covered. □

A subset of edge $M \subseteq E$ is a *maximal matching* for G if for every edge ij in E, either i or j (or both) are endpoints of some edge in M. In other words, there are no edges that can be added to M to increase the size (cardinality) of M.

VERTEX-COVER-MATCHING(G)

1. $S \leftarrow \emptyset$.
2. Find a maximal matching M in G.
3. For each edge $ij \in M$:
   (a) Add i to S.
   (b) Add j to S.
4. Output the set of vertices S.

Theorem 13. VERTEX-COVER-MATCHING(G) *is a 2-approximation algorithm.*

Proof. First, we show that the set of vertices output by the algorithm, S, is a valid vertex cover. Next, we show that the size of S is at most twice the size of a minimum vertex cover.

Suppose S is not a valid vertex cover. Then there exists an edge $ij \in E$ such that neither i nor j belong to S. This implies that neither i nor j belong to M, which contradicts the assumption that M is a *maximal* matching.

By Lemma 12, we know that $|M|$ is a lower bound on the size of minimum vertex cover. Recall that OPT_{VC} denotes that size of a minimum vertex cover for the graph G. Then,

$$|M| \leq OPT_{VC}.$$

It follows that

$$|S| = 2|M| \leq 2 \cdot OPT_{VC}.$$

Thus, we have proved the theorem. □

3.2 Linear Programming Relaxation

The vertex cover problem can be formulated as the following integer program.

$$\min \sum_{i \in V} x_i$$
$$\text{subject to: } x_i + x_j \geq 1, \quad \text{for all } ij \in E,$$
$$x_i \in \{0,1\}. \tag{IP$_{VC}$}$$

Consider the linear programming relaxation of the integer program (IP$_{VC}$).

$$\min \sum_{i \in V} x_i$$
$$\text{subject to: } x_i + x_j \geq 1, \quad \text{for all } ij \in E,$$
$$0 \leq x_i \leq 1. \tag{LP$_{VC}$}$$

Vertex-Cover-LP(G)

1. $S \leftarrow \emptyset$.
2. Find an optimal solution for (LP$_{VC}$).
3. For each vertex $i \in V$:
   (a) If $x_i \geq \frac{1}{2}$, add i to S.
4. Output the set of vertices S.

Theorem 14. Vertex-Cover-LP(G) *is a 2-approximation algorithm.*

Proof. Let $\mathbf{x}^*$ denote an optimal solution for (LP$_{VC}$) and let $OPT_f(G)$ denote the objective value of $\mathbf{x}^*$. Let $S = \{i \in V \mid x_i^* \geq \frac{1}{2}\}$. Note that S is feasible since for each edge $ij \in E$, the constraints from (LP$_{VC}$) imply that either x_i^* or x_j^* (or both) have value at least $\frac{1}{2}$.

How large can $|S|$ be? Each vertex $i \in V$ for which $x_i^* < \frac{1}{2}$ contributes zero to $|S|$, and each vertex for which $x_i^* \geq \frac{1}{2}$ contributes one vertex to $|S|$. Therefore, $|S| \leq 2 \cdot \sum_{i \in V} x_i^*$. So we have,

$$|S| \ \leq \ 2 \cdot \sum_{i \in V} x_i^* \ = \ 2 \cdot OPT_f(G) \ \leq \ OPT_{VC}.$$

the size of S is no more than twice that of a minimum vertex cover. □

For general graphs, we cannot hope to approximate the vertex cover problem to within a factor better than 2 using just this linear program.

Theorem 15. *The integrality gap of* (LP$_{VC}$) *is at least* $2 - \frac{2}{n}$.

Proof. Consider the complete graph K_n on n vertices. If we set $x_i = \frac{1}{2}$ for all $i \in V$, then the value of $OPT_f(K_n)$ is $\frac{n}{2}$. However, the size of a minimum vertex cover in K_n is $n-1$. Thus, the integrality gap is at least $\frac{n-1}{n/2} = 2 - \frac{2}{n}$, i.e. arbitrarily close to 2. □

References

[BT97] Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to linear optimization*. Athena Scientific Belmont, MA, 1997.