

1 Bipartite matching

A bipartite graph is a graph $G = (V = V_1 \cup V_2, E)$ with disjoint V_1 and V_2 and $E \subseteq V_1 \times V_2$. The graph may optionally have weights given by $w : E \rightarrow \mathbb{Q}_+$. The bipartite matching problem is one where, given a bipartite graph, we seek a matching $M \subseteq E$ (a set of edges such that no two share an endpoint) of maximum cardinality or weight. We call a matching M a perfect matching if $\deg_M(v) = 1$ for all $v \in V$.

Theorem 1 *Bipartite matching is in P.*

There are a variety of polynomial-time algorithms for finding an optimal bipartite matching. A common algorithm is one that reduces the matching problem into an equivalent max-flow instance.

One explanation of what makes the bipartite matching problem “easy” is that if we take all feasible solutions M and represent each by a vector $\chi_M \in \{0, 1\}^E$ (where $\chi_M(e) = 1$ iff $e \in M$), then the convex hull of these vectors is a “nice polytope”.

We denote by $P_{\text{match}}(G)$ the polytope $\text{conv}\{\chi_M : M \text{ is a matching in } G\}$ and by $P_{\text{perfect-match}}(G)$ the polytope $\text{conv}\{\chi_M : M \text{ is a perfect matching in } G\}$. In the following, we use the notation $\delta(v)$ to denote all edges incident to v , and $x(\delta(v)) = \sum_{e \in \delta(v)} x_e$ to denote the sum of the respective variables. To denote the entire vector $(x_e)_{e \in E}$, we use the boldface notation \mathbf{x} .

Theorem 2 *Suppose $G = (V, E)$ is a bipartite graph, and let*

$$Q(G) = \left\{ \mathbf{x} \in \mathbb{R}_+^E : \forall v \in V, \sum_{e \in \delta(v)} x_e \leq 1 \right\}.$$

Then $P_{\text{match}}(G) = Q(G)$.

Proof: The fact that $P_{\text{match}}(G) \subseteq Q(G)$ is rather straightforward. If $\mathbf{x} \in P_{\text{match}}(G)$ then it is a convex linear combination of matchings χ_M , i.e. $\mathbf{x} = \sum_M \alpha_M \chi_M$. In any such matching, $\chi_M(e)$ is 1 for at most one $e \in \delta(v)$ and 0 for all others, so $\sum_{e \in \delta(v)} \chi_M(e) \leq 1$. Hence

$$\begin{aligned} \sum_{e \in \delta(v)} x_e &= \sum_{e \in \delta(v)} \left(\sum_M \alpha_M \chi_M(e) \right) \\ &= \sum_M \alpha_M \sum_{e \in \delta(v)} \chi_M(e) \\ &\leq \sum_M \alpha_M \\ &= 1. \end{aligned}$$

Now we prove $Q(G) \subseteq P_{\text{match}}(G)$. Take any $\mathbf{x} \in Q(G)$. Define

$$F = \text{supp}(\mathbf{x}) = \{e \in E : x_e > 0\}$$

The proof proceeds by induction on $|F|$. For the base case, if $|F| \leq 1$ then $\mathbf{x} \in P_{\text{match}}(G)$ since both the empty matching and the single-edge matching for any edge are valid matchings in G .

We split the rest of the proof into cases.

Case 1: F contains a cycle.

Since G is bipartite, the cycle is even. Let $\mathbf{d} = (\cdots + 1 \cdots - 1 \cdots + 1 \cdots - 1 \cdots)$, where the ± 1 appear in the positions of the edges around the cycle, the sign for such an entry is chosen by the parity of that edge's appearance along the cycle, and the remaining entries are 0.

We know that for some $\epsilon > 0$, $\mathbf{x} \pm \epsilon \mathbf{d} \in Q(G)$, since \mathbf{d} has nonzero entries only where \mathbf{x} has nonzero entries, and for any such ϵ , the incident-edge constraint of Q remains unchanged. That is, for all $v \in V$,

$$\sum_{e \in \delta(v)} x_e = \sum_{e \in \delta(v)} (\mathbf{x} + \epsilon \mathbf{d})_e = \sum_{e \in \delta(v)} (\mathbf{x} - \epsilon \mathbf{d})_e.$$

We can increase ϵ until $(\mathbf{x} \pm \epsilon \mathbf{d})_e$ becomes 0 for some e where originally $x_e > 0$. If we increase it further the resulting vector will no longer be in $Q(G)$. Let's take $\epsilon_1 > 0$ and $\epsilon_2 < 0$ of absolute value as large as possible such that $\mathbf{x}_1 = \mathbf{x} + \epsilon_1 \mathbf{d} \in Q(G)$ and $\mathbf{x}_2 = \mathbf{x} + \epsilon_2 \mathbf{d} \in Q(G)$. We can now write

$$\mathbf{x} = \frac{\epsilon_2}{\epsilon_2 - \epsilon_1} \mathbf{x}_1 + \frac{\epsilon_1}{\epsilon_1 - \epsilon_2} \mathbf{x}_2$$

We know $|\text{supp}(\mathbf{x}_1)|$ and $|\text{supp}(\mathbf{x}_2)|$ are both smaller than $|\text{supp}(\mathbf{x})|$, so by the inductive hypothesis, $\mathbf{x}_1, \mathbf{x}_2 \in \text{conv}\{\chi_M : M \text{ is a matching in } G\}$. Since \mathbf{x} is a convex combination of \mathbf{x}_1 and \mathbf{x}_2 , \mathbf{x} is in $\text{conv}\{\chi_M : M \text{ is a matching in } G\} = P_{\text{match}}(G)$.

Case 2: F is a forest, but not a matching.

Let \tilde{P} be a maximum-length (leaf-to-leaf) path in F . Since F is not a matching, \tilde{P} has at least two edges. Let $\mathbf{d} = (\cdots + 1 \cdots - 1 \cdots + 1 \cdots - 1 \cdots)$ be as in Case 1, with ± 1 entries for edges along the path and sign determined by parity of the edge's placement along the path.

We know again that for some $\epsilon > 0$, $\mathbf{x} \pm \epsilon \mathbf{d} \in Q(G)$. This is because \mathbf{d} has nonzero entries only where \mathbf{x} has nonzero entries and also every edge in \tilde{P} has $x_e < 1$ since the neighboring edge in \tilde{P} must have $x_e > 0$ (being in F). The vertices of degree 2 along \tilde{P} remain unchanged in terms of the incident-edge constraints (as in Case 1). With this established, the argument proceeds identically as in Case 1.

Case 3: F is a matching.

It follows rather directly that \mathbf{x} is a convex combination of matching vectors χ_M . The construction for such a convex combination is as follows. Assume that $\text{supp}(\mathbf{x}) = \{1, 2, \dots, k\}$ and the elements are ordered so that $x_1 \leq x_2 \leq \dots \leq x_k$. Now write

$$\mathbf{x} = x_1 \chi_{F_1} + (x_2 - x_1) \chi_{F_2} + \cdots + (x_k - x_{k-1}) \chi_{F_k} + (1 - x_k) \chi_\emptyset$$

where $F_i = \{i, i + 1, \dots, k\}$ for each $1 \leq i \leq k$. □

Exercise 3 Suppose we solve the LP variant of maximum matching $\max\{w^T x : x \in P_{\text{match}}(G)\}$ and the resulting $x \in P_{\text{match}}(G)$ is a fractional solution. How do we find a (non-fractional) matching M such that $w(M) \geq w^T x$? (Hint: Follow the proof of Theorem 2.)

Theorem 4 Suppose $G = (V, E)$ is a bipartite graph, and let

$$Q(G) = \left\{ \mathbf{x} \in \mathbb{R}_+^E : \forall v \in V, \sum_{e \in \delta(v)} x_e = 1 \right\}$$

Then $P_{\text{perfect-match}}(G) = Q(G)$.

Proof: The fact that $P_{\text{perfect-match}}(G) \subseteq Q(G)$ is straightforward as in the proof of Theorem 2, wherein this time, for a fixed $v \in V$, $\chi_M(e)$ is 1 for exactly one $e \in \delta(v)$ and 0 for all others.

Now we prove $Q(G) \subseteq P_{\text{perfect-match}}(G)$. Take any $\mathbf{x} \in Q(G)$. We know $\mathbf{x} \in P_{\text{match}}(G)$ by Theorem 2, hence \mathbf{x} is a convex combination of matchings. That is, denoting by \mathcal{M} the set of all matchings in G ,

$$\mathbf{x} = \sum_{M \in \mathcal{M}} \alpha_M \chi_M$$

where $\sum_{M \in \mathcal{M}} \alpha_M = 1$ and $\alpha_M \geq 0$ for every matching M . Suppose by way of contradiction that $\alpha_{M'} > 0$ for some matching M' that is not perfect, and let v be a node it fails to cover. Then

$$\sum_{e \in \delta(v)} x_e = \sum_{M \text{ covering } v} \alpha_M < 1$$

which contradicts the definition of $Q(G)$. Therefore \mathbf{x} is a convex combination of perfect matchings only, so $\mathbf{x} \in P_{\text{perfect-match}}(G)$. □

Theorem 5 The vertices of $P_{\text{perfect-match}}(G)$ are exactly $\{\chi_M : M \text{ is a perfect matching in } G\}$.

In other words, for any perfect matching M , χ_M cannot be expressed as a convex combination of other matchings.

Proof: Consider any vector \mathbf{t} giving us a line through χ_M for a perfect matching M . For some edge e and any $\epsilon > 0$, either $(\chi_M + \epsilon \mathbf{t})_e$ increases relative to $(\chi_M)_e$ and $(\chi_M - \epsilon \mathbf{t})_e$ decreases relative to $(\chi_M)_e$, or vice versa. Since $(\chi_M)_e \in \{0, 1\}$, we cannot move in both directions along \mathbf{t} from χ_M , for any direction \mathbf{t} . This implies that χ_M is a vertex. □

2 The Birkhoff - von Neumann Theorem

Definition 6 (Doubly stochastic matrix) An $n \times n$ matrix \mathbf{X} is doubly stochastic if

- Every entry satisfies $x_{ij} \geq 0$.
- For any row i , $\sum_{j=1}^n x_{ij} = 1$.
- For any column j , $\sum_{i=1}^n x_{ij} = 1$.

A matrix \mathbf{X} can be interpreted as an assignment of coefficients x_{ij} to the edges of the complete bipartite graph $K_{n,n}$. By the definition of a doubly stochastic matrix, we see that $\mathbf{x} \in \left\{ \mathbf{x} \in \mathbb{R}_+^E : \forall v \in V, \sum_{e \in \delta(v)} x_e = 1 \right\}$. Theorem 4 tells us that $\mathbf{x} \in P_{\text{perfect-match}}(K_{n,n})$, being a convex combination of perfect matchings in $K_{n,n}$, which can be viewed as permutation matrices. This result is known as Birkhoff's Theorem.

Theorem 7 (Birkhoff - von Neumann) *Every doubly stochastic matrix is a convex combination of permutation matrices.*

3 Using LP duality with bipartite matching

Suppose we have a bipartite graph $G = (V, E)$ and \mathbf{A} is its incidence matrix (a matrix in $\{0, 1\}^{V \times E}$ where $a_{ij} = 1$ iff edge j is incident to vertex i). Applying the LP duality theorem to the bipartite matching problem on G , we see that

$$\begin{aligned} \max \{ \mathbf{1}^T \mathbf{x} : \mathbf{x} \in P_{\text{match}}(G) \} &= \max \left\{ \mathbf{1}^T \mathbf{x} : \mathbf{x} \geq \mathbf{0} \text{ and } \forall v \in V, \sum_{e \in \delta(v)} x_e \leq 1 \right\} \\ &= \max \{ \mathbf{1}^T \mathbf{x} : \mathbf{x} \geq \mathbf{0}, \mathbf{A}\mathbf{x} \leq \mathbf{1} \} \\ &= \min \{ \mathbf{1}^T \mathbf{y} : \mathbf{A}^T \mathbf{y} \geq \mathbf{1}, \mathbf{y} \geq \mathbf{0} \} \\ &= \min \{ \mathbf{1}^T \mathbf{y} : \mathbf{y} \geq \mathbf{0} \text{ and } \forall \{r, s\} \in E, y_r + y_s \geq 1 \} \end{aligned}$$

which is precisely the LP formulation of the minimum vertex cover problem in G . This leads us to König's Theorem.

Theorem 8 (König) *For any bipartite graph $G = (V, E)$, the cardinality of the maximum matching $\text{max-matching}(G)$ and the cardinality of the minimum vertex cover $\text{min-vertex-cover}(G)$ are the same.*

Proof: The proof proceeds by induction on $|V|$. For $|V| \leq 2$, the claim is trivially verifiable.

Let \mathbf{y}^* be an optimal solution to the dual LP stated above. If $E = \emptyset$, we are done. Otherwise, we know $y_v > 0$ for some $v \in V$. Complementary slackness then dictates that for any optimal primal solution, $x^*(\delta(v)) = 1$. In particular, since maximum matchings are optimal primal solutions, every maximum matching covers v . Hence $\text{max-matching}(G - \{v\}) = \text{max-matching}(G) - 1$. By the inductive hypothesis, $\text{min-vertex-cover}(G - \{v\}) = \text{max-matching}(G - \{v\})$. Adding v to a vertex cover in $G - \{v\}$ gives a vertex cover in G , so

$$\begin{aligned} \text{min-vertex-cover}(G) &\leq \text{min-vertex-cover}(G - \{v\}) + 1 \\ &= \text{max-matching}(G - \{v\}) + 1 \\ &= \text{max-matching}(G) \end{aligned}$$

On the other hand, it is clear that $\text{min-vertex-cover}(G) \geq \text{max-matching}(G)$, since a vertex cover must include at least one vertex from the endpoints of each edge in any matching. \square

4 Equivalent definitions of a vertex

Let's finish with the following statement which will be useful next time.

Lemma 9 *Let $P \subset \mathbb{R}^n$ be a polyhedron and let $\mathbf{v} \in P$. The following properties are equivalent:*

- (1) *There is a hyperplane $H = \{\mathbf{x} : \mathbf{w}^T \mathbf{x} = \lambda\}$ such that $\mathbf{w}^T \mathbf{x} \leq \lambda$ for all $\mathbf{x} \in P$ and $P \cap H = \{\mathbf{v}\}$.*
- (2) *\mathbf{v} cannot be expressed as a convex combination of points in $P \setminus \{\mathbf{v}\}$.*
- (3) *There is no nonzero vector \mathbf{d} such that $\mathbf{v} \pm \mathbf{d} \in P$.*
- (4) *There are n constraints $\mathbf{a}_j^T \mathbf{x} \leq b_j$ valid for P , which are tight at \mathbf{v} , i.e. $\mathbf{a}_j^T \mathbf{v} = b_j$ for $1 \leq j \leq n$, and $\mathbf{a}_1, \dots, \mathbf{a}_n$ are linearly independent.*

Proof:

- (1) \Rightarrow (2) Suppose $H = \{\mathbf{x} : \mathbf{w}^T \mathbf{x} = \lambda\}$ is a hyperplane such that $P \cap H = \{\mathbf{v}\}$. If \mathbf{v} is a convex combination of other points in P , $\mathbf{v} = \sum \alpha_i \mathbf{v}_i$ where $\sum \alpha_i = 1$ and $\alpha_i \geq 0$, then we have

$$\mathbf{w}^T \mathbf{v} = \sum \alpha_i (\mathbf{w}^T \mathbf{v}_i) = \lambda.$$

However, $\mathbf{w}^T \mathbf{v}_i \leq \lambda$ for each i . The only way the equality above can be achieved is that $\mathbf{w}^T \mathbf{v}_i = \lambda$ for each i . This means that \mathbf{v} is not the only point in $H \cap P$.

- (2) \Rightarrow (3) If $\mathbf{v} \pm \mathbf{d} \in P$, then $\mathbf{v} = \frac{1}{2}(\mathbf{v} + \mathbf{d}) + \frac{1}{2}(\mathbf{v} - \mathbf{d})$ is a convex combination of other points in P , a contradiction.
- (3) \Rightarrow (4) We have $\mathbf{v} \in P = \{\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$. Consider $T = \{\mathbf{a}_j : \mathbf{a}_j^T \mathbf{v} = b_j\}$, the normal vectors to all constraints which are tight at \mathbf{v} . If $\dim(\text{span}(T)) < n$, then there exists $\mathbf{d} \neq \mathbf{0}$ such that \mathbf{d} is orthogonal to $\text{span}(T)$. I.e. for all $\mathbf{a}_j \in T$, $\mathbf{a}_j^T \mathbf{d} = 0$ and $\mathbf{a}_j^T (\mathbf{v} \pm \epsilon \mathbf{d}) = \mathbf{a}_j^T \mathbf{v} = b_j$. For all other constraints, we have $\mathbf{a}_i^T \mathbf{v} < b_i$, so there is some small $\epsilon > 0$ such that $\mathbf{a}_i^T (\mathbf{v} \pm \epsilon \mathbf{d}) \leq b_i$ for all i . This means that $\mathbf{v} \pm \epsilon \mathbf{d} \in P$.
- (4) \Rightarrow (1) Assume that $\mathbf{a}_j^T \mathbf{x} \leq b_j, 1 \leq j \leq n$ are valid constraints for P , the vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ are linearly independent, and $\mathbf{a}_j^T \mathbf{v} = b_j$ for each j . Consider the hyperplane

$$H = \{\mathbf{x} : \sum_{j=1}^n \mathbf{a}_j^T \mathbf{x} = \sum_{j=1}^n b_j\}.$$

For any $\mathbf{x} \in P$, since $\mathbf{a}_j^T \mathbf{x} \leq b_j$ for each j , we have $\sum_{j=1}^n \mathbf{a}_j^T \mathbf{x} \leq \sum_{j=1}^n b_j$. Furthermore, the only way we can get an equality is that $\mathbf{a}_j^T \mathbf{x} = b_j$ for each j . However, this is a rank n system of linear equations in dimension n , which has a unique solution - the point \mathbf{v} .

□