Part II

Bipartite Matching and Covering

Part II: Bipartite Matching and Covering

A second classical area of combinatorial optimization is formed by bipartite matching. The area gives rise to a number of basic problems and techniques, and has an abundance of applications in various forms of assignment and transportation.

Work of Frobenius in the 1910s on the decomposition of matrices formed the incentive to Kőnig to study matchings in bipartite graphs. An extension by Egerváry in the 1930s to weighted matchings inspired Kuhn in the 1950s to design the 'Hungarian method' for the assignment problem (which is equivalent to finding a minimum-weight perfect matching in a complete bipartite graph).

Parallel to this, Tolstoĭ, Kantorovich, Hitchcock, and Koopmans had investigated the transportation problem. It motivated Kantorovich and Dantzig to consider more general problems, culminating in the development of linear programming. It led in turn to solving the assignment problem by linear programming, and thus to a polyhedral approach.

Several variations and extensions of bipartite matching, like edge covers, factors, and transversals, can be handled similarly. Major explanation is the total unimodularity of the underlying matrices.

Bipartite matching and transportation can be considered as special cases of disjoint paths and of transshipment, studied in the previous part — just consider a bipartite graph as a directed graph, by orienting all edges from one colour class to the other. It was however observed by Hoffman and Orden that this can be turned around, and that disjoint paths and transshipment problems can be reduced to bipartite matching and transportation problems. So several results in this part on bipartite matching are matched by results in the previous part on paths and flows. Viewed this way, the present part forms a link between the previous part and the next part on *nonbipartite* matching, where the underlying matrices generally are not totally unimodular.

Chapters:

Chapter 16

Cardinality bipartite matching and vertex cover

'Cardinality matching' deals with maximum-size matchings. In this chapter we give the theorems of Frobenius on the existence of a perfect matching in a bipartite graph, and the extension by Kőnig on the maximum size of a matching in a bipartite graph. We also discuss finding a maximum-size matching in a bipartite graph algorithmically.
We start with an easy but fundamental theorem relating maximum-size matchings and M-alternating paths, that applies to any graph and that will also be important for nonbipartite matching.
In this chapter, graphs can be assumed to be simple.

16.1. M-augmenting paths

Let $G = (V, E)$ be an undirected graph. A *matching* in G is a set of disjoint edges. An important concept in finding a maximum-size matching, both in bipartite and in nonbipartite graphs, is that of an 'augmenting path' (introduced by Petersen [1891]).

Let M be a matching in a graph $G = (V, E)$. A path P in G is called M*-augmenting* if P has odd length, its ends are not covered by M, and its edges are alternatingly out of and in M.

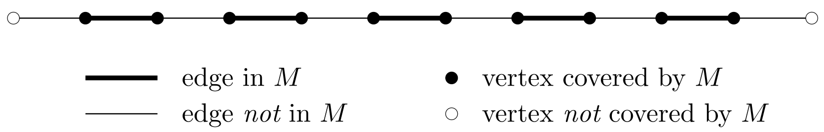


Figure 16.1
An M-augmenting path

Clearly, if P is an M-augmenting path, then

$$M' := M \triangle EP \tag{16.1}$$

is again a matching and satisfies $|M'| = |M| + 1$.[1] In fact, it is not difficult to show (Petersen [1891]):

Theorem 16.1. *Let $G = (V, E)$ be a graph and let M be a matching in G. Then either M is a matching of maximum size or there exists an M-augmenting path.*

Proof. If M is a maximum-size matching, there cannot exist an M-augmenting path P, since otherwise $M \triangle EP$ would be a larger matching.

Conversely, if M' is a matching larger than M, consider the components of the graph $G' := (V, M \cup M')$. Then G' has maximum degree two. Hence each component of G' is either a path (possibly of length 0) or a circuit. Since $|M'| > |M|$, at least one of these components should contain more edges in M' than in M. Such a component forms an M-augmenting path. ∎

So in any graph, if we have an algorithm finding an M-augmenting path for any matching M, then we can find a maximum-size matching: we iteratively find matchings $M_0, M_1, \ldots$, with $|M_i| = i$, until we have a matching M_k such that there exists no M_k-augmenting path. (Also this was observed by Petersen [1891].)

16.2. Frobenius' and Kőnig's theorems

A classical min-max relation due to Kőnig [1931] characterizes the maximum size of a matching in a bipartite graph. To this end, call a set C of vertices of a graph G a *vertex cover* if each edge of G intersects C. Define

(16.2) $\nu(G) :=$ the maximum size of a matching in G,
$\tau(G) :=$ the minimum size of a vertex cover in G.

These numbers are called the *matching number* and the *vertex cover number* of G, respectively. It is easy to see that, for any graph G,

(16.3) $\nu(G) \leq \tau(G)$,

since any two edges in any matching contain different vertices in any vertex cover. The graph K_3 has strict inequality in (16.3). However, if G is bipartite, equality holds, which is the content of Kőnig's matching theorem (Kőnig [1931]). It can be seen to be equivalent to a theorem of Frobenius [1917] (Corollary 16.2a below).

Theorem 16.2 (Kőnig's matching theorem)**.** *For any bipartite graph $G = (V, E)$ one has*

(16.4) $\nu(G) = \tau(G)$.

[1] EP denotes the set of edges in P. $\triangle$ denotes symmetric difference.

That is, the maximum size of a matching in a bipartite graph is equal to the minimum size of a vertex cover.

Proof. By (16.3) it suffices to show that $\nu(G) \geq \tau(G)$. We may assume that G has at least one edge. Then:

(16.5) G has a vertex u covered by each maximum-size matching.

To see this, let $e = uv$ be any edge of G, and suppose that there are maximum-size matchings M and N missing u and v respectively[2]. Let P be the component of $M \cup N$ containing u. So P is a path with end vertex u. Since P is not M-augmenting (as M has maximum size), P has even length, and hence does not traverse v (otherwise, P ends at v, contradicting the bipartiteness of G). So $P \cup e$ would form an N-augmenting path, a contradiction (as N has maximum size). This proves (16.5).

Now (16.5) implies that for the graph $G' := G - u$ one has $\nu(G') = \nu(G) - 1$. Moreover, by induction, G' has a vertex cover C of size $\nu(G')$. Then $C \cup \{u\}$ is a vertex cover of G of size $\nu(G') + 1 = \nu(G)$. ∎

(This proof is due to De Caen [1988]. For Kőnig's original, algorithmic proof, see the proof of Theorem 16.6. Note that also Menger's theorem implies Kőnig's matching theorem (using the construction given in the proof of Theorem 16.4 below). For a proof based on showing that any minimum bipartite graph with a given vertex cover number is a matching, see Lovász [1975d]. For another proof (of Rizzi [2000a]), see Section 16.2c. As we will see in Chapter 18, Kőnig's matching theorem also follows from the total unimodularity of the incidence matrix of a bipartite graph. (Flood [1960] and Entringer and Jackson [1969] gave proofs similar to Kőnig's proof.))

A consequence of Theorem 16.2 is a theorem of Frobenius [1917] that characterizes the existence of a perfect matching in a bipartite graph. (A matching is *perfect* if it covers all vertices.) Actually, this theorem motivated Kőnig to study matchings in graphs, and in turn it can be seen to imply Kőnig's matching theorem.

Corollary 16.2a (Frobenius' theorem)**.** *A bipartite graph $G = (V, E)$ has a perfect matching if and only if each vertex cover has size at least $\frac{1}{2}|V|$.*

Proof. Directly from Kőnig's matching theorem, since G has a perfect matching if and only if $\nu(G) \geq \frac{1}{2}|V|$. ∎

This implies an earlier theorem of Kőnig [1916] on *regular* bipartite graphs:

---

[2] M *misses* a vertex u if $u \notin \bigcup M$. Here $\bigcup M$ denotes the union of the edges in M; that is, the set of vertices covered by the edges in M.

Corollary 16.2b. *Each regular bipartite graph (of positive degree) has a perfect matching.*

Proof. Let $G = (V, E)$ be a k-regular bipartite graph. So each vertex is incident with k edges. Since $|E| = \frac{1}{2}k|V|$, we need at least $\frac{1}{2}|V|$ vertices to cover all edges. Hence Corollary 16.2a implies the existence of a perfect matching. ∎

Let A be the $V \times E$ incidence matrix of the bipartite graph $G = (V, E)$. Kőnig's matching theorem (Theorem 16.2) states that the optima in the linear programming duality equation

(16.6) $$\max\{\mathbf{1}^{\mathsf{T}}x \mid x \geq \mathbf{0}, Ax \leq \mathbf{1}\} = \min\{y^{\mathsf{T}}\mathbf{1} \mid y \geq \mathbf{0}, y^{\mathsf{T}}A \geq \mathbf{1}^{\mathsf{T}}\}$$

are attained by integer vectors x and y. This can also be derived from the total unimodularity of A — see Section 18.3.

16.2a. Frobenius' proof of his theorem

The proof method given by Frobenius [1917] of Corollary 16.2a is in terms of matrices, but can be formulated in terms of graphs as follows. Necessity of the condition being easy, we prove sufficiency. Let U and W be the colour classes of G. As both U and W are vertex covers, and hence have size at least $\frac{1}{2}|V|$, we have $|U| = |W| = \frac{1}{2}|V|$.

Choose an edge $e = \{u, w\}$ with $u \in U$ and $w \in W$. We may assume that $G - u - w$ has no perfect matching. So, inductively, $G - u - w$ has a vertex cover C' with $|C'| < |U| - 1$. Then $C := C' \cup \{u, w\}$ is a vertex cover of G, with $|C| \leq |U|$, and hence $|C| = |U|$.

Now $U \triangle C$ and $W \triangle C$ partition V (where $\triangle$ denotes symmetric difference). If both $U \triangle C$ and $W \triangle C$ are matchable[3], then G has a perfect matching. So, by symmetry, we may assume that $U \triangle C$ is not matchable. Now $U \triangle C \neq V$ as $u \notin U \triangle C$. Hence we can apply induction, giving that $G[U \triangle C]$ has a vertex cover D with $|D| < \frac{1}{2}|U \triangle C|$. Then the set $D \cup (U \cap C)$ is a vertex cover of G (since each edge of G intersects both U and C, and hence it either intersects $U \cap C$, or is contained in $U \triangle C$ and hence intersects D). However, $|D| + |U \cap C| < \frac{1}{2}|U \triangle C| + |U \cap C| = \frac{1}{2}(|U| + |C|) = \frac{1}{2}|V|$, a contradiction.

(This is essentially also the proof method of Rado [1933] and Dulmage and Halperin [1955].)

16.2b. Linear-algebraic proof of Frobenius' theorem

Frobenius [1917] was motivated by a determinant problem, namely by the following direct consequence of his theorem. Let $A = (a_{i,j})$ be an $n \times n$ matrix in which each entry $a_{i,j}$ is either 0 or a variable $x_{i,j}$ (where the variables $x_{i,j}$ are independent). Then Frobenius' theorem is equivalent to: $\det A = 0$ if and only if A has a $k \times l$ all-zero submatrix with $k + l > n$. (Earlier, Frobenius [1912] showed that for such

[3] A set T of vertices is called *matchable* if there exists a matching M with $T = \bigcup M$.

a matrix A, $\det A$ is reducible (that is, there exist nonconstant polynomials p and q with $\det A = p \cdot q$) if and only if A has a $k \times l$ all-zero submatrix with $k + l = n$ and $k, l \geq 1$.)

Edmonds [1967b] showed that the argumentation can be applied also the other way around. This gives the following linear-algebraic proof of Frobenius' theorem (implying linear-algebraic proofs also of other bipartite matching theorems).

Let $G = (V, E)$ be a bipartite graph not having a perfect matching. Let U and W be the colour classes of G. We may assume that $|U| = |W|$ (otherwise the smaller colour class is a vertex cover of size less than $\frac{1}{2}|V|$).

Make a $U \times W$ matrix $A = (a_{u,w})$, where $a_{u,w} = 0$ if u and w are not adjacent, and $a_{u,w} = x_{u,w}$ otherwise, where the $x_{u,w}$ are independent variables.

As G has no perfect matching, we know that $\det A = 0$, and hence the columns of A are linearly dependent. Let $W' \subseteq W$ be the index set of a minimal set of linearly dependent columns of A. Then there is a subset U' of U with $|U'| = |W'| - 1$ such that the $U' \times W'$ submatrix A' of A has rank $|U'|$. Hence there is a vector y such that $A'y = \mathbf{0}$ and such that each entry in y is a nonzero polynomial in those variables $x_{u,w}$ that occur in A'. Let A'' be the $U \times W'$ submatrix of A. Then $A''y = \mathbf{0}$, and hence all entries in the $(U \setminus U') \times W'$ submatrix of A are 0. Hence the rows in U' and columns in $W \setminus W'$ cover all nonzeros. As $|U'| + |W \setminus W'| < |W|$, we have Frobenius' theorem.

16.2c. Rizzi's proof of Kőnig's matching theorem

Rizzi [2000a] gave the following short proof of Kőnig's matching theorem. Let $G = (V, E)$ be a counterexample with $|V| + |E|$ minimal. Then G has a vertex u of degree at least 3. Let v be a neighbour of u. By the minimality of G, $G - v$ has a vertex cover U of size $\nu(G - v)$. Then $U \cup \{v\}$ is a vertex cover of G. As G is a counterexample, we have $|U \cup \{v\}| \geq \nu(G) + 1$, and so $\nu(G - v) = |U| \geq \nu(G)$. Therefore, G has a maximum-size matching M not covering v. Let $f \in E \setminus M$ be incident with u and not with v. Then $\nu(G - f) \geq |M| = \nu(G)$. Let W be a vertex cover of $G - f$ of size $\nu(G - f) = \nu(G)$. Then $v \notin W$, since v is not covered by M. Hence $u \in W$, as W covers edge uv of $G - f$. Therefore, W also covers f, and hence it is a vertex cover of G of size $\nu(G)$.

16.3. Maximum-size bipartite matching algorithm

We now focus on the problem of finding a maximum-size matching in a bipartite graph algorithmically. In view of Theorem 16.1, this amounts to finding an augmenting path. In the bipartite case, this can be done by finding a directed path in an auxiliary directed graph. This method is essentially due to van der Waerden [1927] and Kőnig [1931].

Matching augmenting algorithm for bipartite graphs

input: a bipartite graph $G = (V, E)$ and a matching M,
output: a matching M' satisfying $|M'| > |M|$ (if there is one).

description of the algorithm: Let G have colour classes U and W. Make a directed graph D_M by orienting each edge $e = \{u, w\}$ of G (with $u \in U, w \in W$) as follows:

(16.7) if $e \in M$, then orient e from w to u,
if $e \notin M$, then orient e from u to w.

Let U_M and W_M be the sets of vertices in U and W (respectively) missed by M.

Now an M-augmenting path (if any) can be found by finding a directed path in D_M from U_M to W_M. This gives a matching larger than M. ∎

The correctness of this algorithm is immediate. Since a directed path can be found in time $O(m)$, we can find an augmenting path in time $O(m)$. Hence we have the following result (implicit in Kuhn [1955b]):

Theorem 16.3. *A maximum-size matching in a bipartite graph can be found in time $O(nm)$.*

Proof. Note that we do at most n iterations, each of which can be done in time $O(m)$ by breadth-first search (Theorem 6.3). ∎

16.4. An $O(n^{1/2}m)$ algorithm

Hopcroft and Karp [1971,1973] and Karzanov [1973b] proved the following sharpening of Theorem 16.3, which we derive from the (equivalent) result of Karzanov [1973a], Tarjan [1974e], and Even and Tarjan [1975] on the complexity of finding a maximum number of vertex-disjoint paths (Corollary 9.7a).

Theorem 16.4. *A maximum-size matching in a bipartite graph can be found in $O(n^{1/2}m)$ time.*

Proof. Let $G = (V, E)$ be a bipartite graph, with colour classes U and W. Make a directed graph $D = (V, A)$ as follows. Orient all edges from U to W. Moreover, add a new vertex s, with arcs (s, u) for all $u \in U$, and a new vertex t, with arcs (w, t) for all $w \in W$. Then the maximum number of internally vertex-disjoint $s - t$ paths in D is equal to the maximum size of a matching in G. The result now follows from Corollary 9.7a. ∎

In fact, the factor $n^{1/2}$ can be reduced to $\nu(G)^{1/2}$ (as before, $\nu(G)$ and $\tau(G)$ denote the maximum size of a matching and the minimum size of a vertex cover, respectively):

Theorem 16.5. *A maximum-size matching in a bipartite graph G can be found in $O(\nu(G)^{1/2}m)$ time.*

Proof. Similar to the proof of Theorem 16.4, using Theorem 9.8 and the fact that $\nu(G) = \tau(G)$. ∎

Gabow and Tarjan [1988a] observed that the method of Corollary 9.7a applied to the bipartite matching problem implies that for each k one can find in time $O(km)$ a matching of size at least $\nu(G) - \frac{n}{k}$.

16.5. Finding a minimum-size vertex cover

From a maximum-size matching in a bipartite graph, one can derive a minimum-size vertex cover. The method gives an alternative proof of Kőnig's matching theorem (in fact, this is the original proof of Kőnig [1931]):

Theorem 16.6. *Given a bipartite graph G and a maximum-size matching M in G, we can find a minimum-size vertex cover in G in time $O(m)$.*

Proof. Make D_M, U_M, and W_M as in the matching-augmenting algorithm, and let R_M be the set of vertices reachable in D_M from U_M. So $R_M \cap W_M = \emptyset$. Then each edge uw in M is either contained in R_M or disjoint from R_M (that is, $u \in R_M \iff w \in R_M$). Moreover, no edge of G connects $U \cap R_M$ and $W \setminus R_M$, as no arc of D_M leaves R_M. So $C := (U \setminus R_M) \cup (W \cap R_M)$ is a vertex cover of G. Since C is disjoint from $U_M \cup W_M$ and since no edge in M is contained in C, we have $|C| \leq |M|$. Therefore, C is a minimum-size vertex cover. ∎

Hence:

Corollary 16.6a. *A minimum-size vertex cover in a bipartite graph can be found in $O(n^{1/2}m)$ time.*

Proof. Directly from Theorems 16.4 and 16.6. ∎

16.6. Matchings covering given vertices

The following theorem characterizes when one of the colour classes of a bipartite graph can be covered by a matching, and is a direct consequence of Kőnig's matching theorem (where $N(S)$ denotes the set of vertices not in S that have a neighbour in S):

Theorem 16.7. *Let $G = (V, E)$ be a bipartite graph with colour classes U and W. Then G has a matching covering U if and only if $|N(S)| \geq |S|$ for each $S \subseteq U$.*

Proof. Necessity being trivial, we show sufficiency. By Kőnig's matching theorem (Theorem 16.2) it suffices to show that each vertex cover C has $|C| \geq |U|$. This indeed is the case, since $N(U \setminus C) \subseteq C \cap W$, and hence

$$|C| = |C \cap U| + |C \cap W| \geq |C \cap U| + |N(U \setminus C)| \geq |C \cap U| + |U \setminus C| = |U|. \tag{16.8}$$

▮

This can be extended to general subsets of V. First, Hoffman and Kuhn [1956b] and Mendelsohn and Dulmage [1958a] showed:

Theorem 16.8. *Let $G = (V, E)$ be a bipartite graph with colour classes U and W and let $R \subseteq V$. Then there exists a matching covering R if and only if there exist a matching M covering $R \cap U$ and a matching N covering $R \cap W$.*

Proof. Necessity being trivial, we show sufficiency. We may assume that G is connected, that $E = M \cup N$, and that neither M nor N covers R. This implies that there is a $u \in R \cap U$ missed by N and a $w \in R \cap W$ missed by M. So G is an even-length $u - w$ path, a contradiction, since $u \in U$ and $w \in W$. ▮

(This theorem goes back to theorems of F. Bernstein (cf. Borel [1898] p. 103), Banach [1924], and Knaster [1927] on injective mappings between two sets.)

Theorem 16.8 implies a characterization of sets that are covered by some matching:

Corollary 16.8a. *Let $G = (V, E)$ be a bipartite graph with colour classes U and W and let $R \subseteq V$. Then there is a matching covering R if and only if $|N(S)| \geq |S|$ for each $S \subseteq R \cap U$ and for each $S \subseteq R \cap W$.*

Proof. Directly from Theorems 16.7 and 16.8. ▮

It also gives the following exchange property:

Corollary 16.8b. *Let $G = (V, E)$ be a bipartite graph, with colour classes U and W, let M and N be maximum-size matchings, let U' be the set of vertices in U covered by M, and let W' be the set of vertices in W covered by N. Then there exists a maximum-size matching covering $U' \cup W'$.*

Proof. Directly from Theorem 16.8: the matching found is maximum-size since $|U'| = |W'| = \nu(G)$. ▮

Notes. These results also are special cases of the exchange results on paths discussed in Section 9.6c. Perfect [1966] gave the following linear-algebraic argument for Corollary 16.8b. Make a $U \times W$ matrix A with $a_{u,w} = x_{u,w}$ if $uw \in E$ and $a_{u,w} := 0$ otherwise, where the $x_{u,w}$ are independent variables. Let U' be any maximum-size subset of U covered by some matching and let W' be any maximum-size subset of W covered by some matching. Then U' gives a maximum-size set of

linearly independent rows of A and W' gives a maximum-size set of linearly independent columns of A. Then the $U' \times W'$ submatrix of A is nonsingular, hence of nonzero determinant. It implies (by the definition of determinant) that G has a matching covering $U' \cup W'$.
(Related work includes Perfect and Pym [1966], Pym [1967], Brualdi [1969b,1971b], and Mirsky [1969].)

16.7. Further results and notes

16.7a. Complexity survey for cardinality bipartite matching

Complexity survey for cardinality bipartite matching ($*$ indicates an asymptotically best bound in the table):

	$O(nm)$	Kőnig [1931], Kuhn [1955b]
	$O(\sqrt{n}\, m)$	Hopcroft and Karp [1971,1973], Karzanov [1973a]
$*$	$\widetilde{O}(n^{\omega})$	Ibarra and Moran [1981]
	$O(n^{3/2}\sqrt{\frac{m}{\log n}})$	Alt, Blum, Mehlhorn, and Paul [1991]
$*$	$O(\sqrt{n}\, m \log_n(n^2/m))$	Feder and Motwani [1991,1995]

Here ω is any real such that any two $n \times n$ matrices can be multiplied by $O(n^{\omega})$ arithmetic operations (e.g. $\omega = 2.376$).

Goldberg and Kennedy [1997] described a bipartite matching algorithm based on the push-relabel method, of complexity $O(\sqrt{n}\, m \log_n(n^2/m))$. Balinski and Gonzalez [1991] gave an alternative $O(nm)$ bipartite matching algorithm (not using augmenting paths).

16.7b. Finding perfect matchings in regular bipartite graphs

By Kőnig's matching theorem, each k-regular bipartite graph has a perfect matching (if $k \geq 1$). One can use the regularity also to find quickly a perfect matching. This will be used in Chapter 20 on bipartite edge-colouring.

First we show the following result of Cole and Hopcroft [1982] (which will not be used any further in this book):

Theorem 16.9. *A perfect matching in a regular bipartite graph can be found in $O(m \log n)$ time.*

Proof. We first describe an $O(m \log n)$-time algorithm for the following problem:

(16.9) given: a k-regular bipartite graph $G = (V, E)$ with $k \geq 2$,
find: a nonempty proper subset F of E with (V, F) regular.

Let G have colour classes U and W. First let k be even. Then find an Eulerian orientation of the edges of G (this can be done in $O(m)$ time (Theorem 6.7)). Let F be the set of edges oriented from U to W.

Next let k be odd. Call a subset F of E *almost regular* if $|\deg_F(v)-\deg_F(u)| \leq 1$ for all $u, v \in V$. (Here $\deg_F(v)$ is the degree of v in the graph (V, F).)

Moreover, let $\text{odd}(F)$ and $\text{even}(F)$ denote the sets of vertices v with $\deg_F(v)$ odd and even, respectively, and let $\Delta(F)$ denote the maximum degree of the graph (V, F). We give an $O(m)$ algorithm for the following problem:

(16.10) given: an almost regular subset F of E with $\Delta(F) \geq 2$,
find: an almost regular subset F' of E with $\Delta(F') \geq 2$ and $|\text{odd}(F')| \leq \frac{1}{2}|\text{odd}(F)|$.

In time $O(m)$ we can find a subset F'' of F such that

(16.11) $$\lfloor \tfrac{1}{2} \deg_F(v) \rfloor \leq \deg_{F''}(v) \leq \lceil \tfrac{1}{2} \deg_F(v) \rceil$$

for each vertex v: make an Eulerian orientation in the graph obtained from (V, F) by adding edges so as to make all degrees even, and choose for F'' the subset of all edges oriented from U to W. So F'' and $F \setminus F''$ are almost regular.

We choose F'' such that

(16.12) $$|\text{odd}(F'') \cap \text{odd}(F)| \leq \tfrac{1}{2}|\text{odd}(F)|$$

(otherwise replace F'' by $F \setminus F''$). Let $2l$ be the degree of the even-degree vertices of (V, F). We consider two cases.

Case 1: *l* is even. Define $F' := F''$. By (16.11), F' is almost regular. Moreover, as l is even, $\text{odd}(F') \subseteq \text{odd}(F)$, implying (with (16.12)) that $|\text{odd}(F')| \leq \frac{1}{2}|\text{odd}(F)|$. Finally, $\Delta(F') \geq 2$, since otherwise $\Delta(F) \leq 3$ and hence $l = 0$, implying $\Delta(F) \leq 1$, a contradiction.

Case 2: *l* is odd. Define $F' := F'' \cup (E \setminus F)$. Then F' is almost regular, since each $\deg_{F'}(v)$ is either $\lfloor \frac{1}{2} \deg_F(v) \rfloor + k - \deg_F(v) = k - \lceil \frac{1}{2} \deg_F(v) \rceil$ or $\lceil \frac{1}{2} \deg_F(v) \rceil + k - \deg_F(v) = k - \lfloor \frac{1}{2} \deg_F(v) \rfloor$.

Since k is odd, one also has (by definition of F'): $\deg_{F'}(v)$ is odd $\iff$ $\deg_{F''}(v) + k - \deg_F(v)$ is odd $\iff$ $\deg_{F''}(v) \equiv \deg_F(v) \pmod 2$ $\iff$ $v \in \text{odd}(F'') \cap \text{odd}(F)$ (since $\text{even}(F) \subseteq \text{odd}(F'')$, as l is odd). So $|\text{odd}(F')| = |\text{odd}(F'') \cap \text{odd}(F)| \leq \frac{1}{2}|\text{odd}(F)|$, by (16.12).

Finally, suppose that $\Delta(F') \leq 1$. Choose $v \in \text{odd}(F) \setminus \text{odd}(F')$. So $v \in \text{even}(F')$, hence $\deg_{F'}(v) = 0$, implying $\deg_{F''}(v) = 0$ and $\deg_F(v) = k$. But then $0 = \lfloor \frac{1}{2} k \rfloor$, and so $k \leq 1$, a contradiction.

This describes the $O(m)$-time algorithm for problem (16.10). It implies that one can find an almost regular subset F of E with $\Delta(F) \geq 2$ and $\text{odd}(F) = \emptyset$ in $O(m \log n)$ time. So (V, F) is a regular subgraph of G, and we have solved (16.9).

This implies an $O(m \log n)$ algorithm for finding a perfect matching: First find a subset F of E as in (16.9). Without loss of generality, $|F| \leq \frac{1}{2}|E|$. Recursively, find a perfect matching in (V, F). The time is bounded by $O((m + \frac{1}{2}m + \frac{1}{4}m + \cdots) \log n) = O(m \log n)$. ∎

In fact, as was shown by Cole, Ost, and Schirra [2001], one can find a perfect matching in a regular bipartite graph in $O(m)$ time. To explain this algorithm, we first describe an algorithm that finds a perfect matching in a k-regular bipartite graph in $O(km)$ time (Schrijver [1999]). So for each fixed degree k one can find a perfect matching in a k-regular graph in linear time, which is also a consequence of an $O(n2^{2^{O(k)}})$-time algorithm of Cole [1982].

Theorem 16.10. *A perfect matching in a k-regular bipartite graph can be found in time $O(km)$.*

Proof. Let $G = (V, E)$ be a k-regular bipartite graph. For any function $w : E \to \mathbb{Z}_+$, define $E_w := \{e \in E \mid w_e > 0\}$.

Initially, set $w_e := 1$ for each $e \in E$. Next apply the following iteratively:

(16.13) Find a circuit C in E_w. Let $C = M \cup N$ for matchings M and N with $w(M) \geq w(N)$. Reset $w := w + \chi^M - \chi^N$.

Note that at any iteration, the equation $w(\delta(v)) = k$ is maintained for all v.

To see that the process terminates, note that at any iteration the sum

$$(16.14) \qquad \sum_{e \in E} w_e^2$$

increases by

$$(16.15) \qquad \sum_{e \in M}((w_e+1)^2 - w_e^2) + \sum_{e \in N}((w_e-1)^2 - w_e^2) = 2w(M) + |M| - 2w(N) + |N|,$$

which is at least $|M| + |N| = |C|$. Since $w_e \leq k$ for each $e \in E$, (16.14) is bounded, and hence the process terminates. We now estimate the running time.

At termination, we have that the set E_w contains no circuit, and hence is a perfect matching (since $w(\delta(v)) = k$ for each vertex v). So at termination, the sum (16.14) is equal to $\frac{1}{2}nk^2 = km$.

Now we can find a circuit C in E_w in $O(|C|)$ time on average. Indeed, keep a path P in E_w such that $w_e < k$ for each e in P. Let v be the last vertex of P. Then there is an edge $e = vu$ not occurring in P, with $0 < w_e < k$. Reset $P := P \cup \{e\}$. If P is not a path, it contains a circuit C, and we can apply (16.13) to C, after which we reset $P := P \setminus C$. We continue with P.

Concluding, as each step increases the sum (16.14) by at least $|C|$, and takes $O(|C|)$ time on average, the algorithm terminates in $O(km)$ time. ∎

The bound given in this theorem was improved to linear time *independent of* the degree, by Cole, Ost, and Schirra [2001]. Their method forms a sharpening of the method described in the proof of Theorem 16.10, utilizing the fact that when breaking a circuit, the path segments left ('chains') can be used in the further path search to extend the path by chains, rather than just edge by edge. To this end, these chains need to be supplied with some extra data structure, the 'self-adjusting binary trees', in order to avoid that we have to run through the chain to find an end of the chain where it can be attached to the path. The basic operation is the 'splay'.

The main technique of Cole, Ost, and Schirra's theorem is contained in the proof of the following theorem. For any graph $G = (V, E)$ call a ('weight') function $w : E \to \mathbb{R}$ *k-regular* if $w(\delta(v)) = k$ for all $v \in V$.

Theorem 16.11. *Given a bipartite graph $G = (V, E)$ and a k-regular $w : E \to \mathbb{Z}_+$, for some $k \geq 2$, a perfect matching in G can be found in time $O(m \log^2 k)$.*

Proof. I. *Conceptual outline.* We first give a conceptual description, as extension of the algorithm described in the previous proof. First delete all edges e with $w_e = 0$.

We keep a set F of edges such that each component of (V, F) is a path (possibly a singleton) with at most k^2 vertices, and we keep a path

(16.16) $$Q = (P_0, e_1, P_1, \ldots, e_t, P_t),$$

where each P_j is a (path) component of (V, F). Let v be the last vertex of Q and let $e = vu$ be an edge in $E \setminus F$ incident with v with $w_e < k$. Let P be the component of (V, F) containing u.

If u is not on Q, let R be a longest segment of P starting from u. Delete the first edge of the other segment of P (if any) from F. If $|P_t| + |R| \leq k^2$, add e to F, and reset P_t to P_t, e, R. (Here and below, $|X|$ denotes the number of vertices of a path X.) Otherwise, extend Q by e, R.

If u is on Q, then:

(16.17) split Q into a part Q_1 from the beginning to u, and a part Q_2 from u to the end;
split the circuit Q_2, e into two matchings M and N, such that $w(M) \geq w(N)$;
let α be the minimum of the weights in N;
reset $w := w + \alpha(\chi^M - \chi^N)$;
delete the edges g with $w(g) = 0$ or $w(g) = k$ (in the latter case, also delete the two ends of g);
delete the first edge of Q_2 from F if it was in F;
reset $Q := Q_1$;
iterate.

If v is incident with no edge $e \in E \setminus F$ satisfying $w_e < k$, start Q in a new vertex that is incident with an edge e with $w_e < k$. If no such vertex exists, we are done: the edges left form a perfect matching.

II. *Data structure.* In order to make profit of storing paths, we need additional data structure (based on 'self-adjusting binary trees', analyzed by Sleator and Tarjan [1983b,1985], cf. Tarjan [1983]).

We keep a collection $\mathcal{P}$ of paths (possibly singletons), each being a subpath of a component of F, such that

(16.18)
(i) each component of F itself is a path in $\mathcal{P}$;
(ii) $\mathcal{P}$ is *laminar*, that is, any two paths in $\mathcal{P}$ are vertex disjoint, or one is a subpath of the other;
(iii) any nonsingleton path $P \in \mathcal{P}$ has an edge e_P such that the two components of $P - e_P$ again belong to $\mathcal{P}$.

With any path $P \in \mathcal{P}$ we keep the following information:

(16.19)
(i) the number $|P|$ of vertices in P;
(ii) a list ends(P) of the ends of P (so ends(P) contains one or two vertices);

(iii) if P is not a singleton, the edge e_P, and a list subpaths(P) of the two components of $P - e_P$;
(iv) the smallest path parent(P) in $\mathcal{P}$ that properly contains P (null if there is no such path).

Then for each edge $e \in F$ there is a unique path $P_e \in \mathcal{P}$ traversing e such that both components of $P_e - e$ again belong to $\mathcal{P}$ (that is, $e_{P_e} = e$). We keep with any $e \in F$ the path P_e.

Call a path $P \in \mathcal{P}$ a *root* if parent(P) = null. So the roots correspond to the components of the graph (V, F). Along a path $P \in \mathcal{P}$ we call edges alternatingly *odd* and *even in* P in such a way that e_P is odd.

We also store information on the current values of the w_e. Algorithmically, we only reset explicitly those w_e for which e is not in F. For $e \in F$, these values are stored implicitly, such that it takes only $O(1)$ time to update w_e for all e in a root when adding α to the odd edges and $-\alpha$ to the even edges in it. This can be done as follows.

If P is a root, we store $w(e_P)$ at P. If P has a parent Q, we store

(16.20) $\quad w(e_P) \pm w(e_Q)$

at P, where $\pm$ is $-$ if e_P is odd in Q, and $+$ otherwise.

We also need the following values for any $P \in \mathcal{P}$ with $EP \neq \emptyset$:

(16.21) $\quad$ minodd$(P) := \min\{w_e \mid e$ odd in $P\}$, mineven$(P) := \min\{w_e \mid e$ even in $P\}$,
diffsum$(P) := \sum(w_e \mid e$ odd in $P) - \sum(w_e \mid e$ even in $P)$

(taking a minimum ∞ if the range is empty). When storing these data, we relate them to $w(e_P)$, again so as to make them invariant under updates. Thus we store

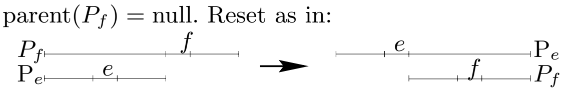
(16.22) $\quad \text{diffsum}(P) - |EP|w(e_P), \text{minodd}(P) - w(e_P), \text{mineven}(P) + w(e_P)$

at P. So for any root P we have diffsum(P), minodd(P), and mineven(P) ready at hand, as we know $w(e_P)$.

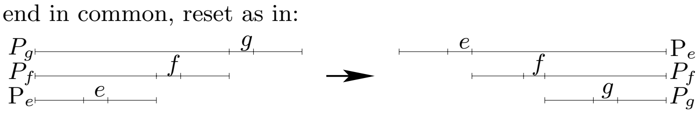
III. *The splay.* We now describe *splaying* an edge $e \in F$. It changes the data structure so that P_e becomes a root, keeping F invariant. It modifies the tree associated with the laminar family through three generations at a time, so as to attain efficiency on average. (The adjustments make future searches more efficient.)

The splay is as follows. While parent$(P_e) \neq$ null, do the following:

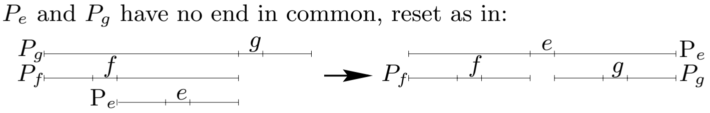
(16.23) $\quad$ Let $P_f :=$ parent(P_e).
Case 1: parent(P_f) = null. Reset as in:



Case 2: parent$(P_f) \neq$ null. Let $P_g :=$ parent(P_f). If P_e and P_g have an end in common, reset as in:



If P_e and P_g have no end in common, reset as in:



Note that Case 1 applies only in the last iteration of the while loop. It is straightforward to check that the data associated with the paths can be restored in $O(1)$ time at any iteration.

IV. *Running time of one splay.* To estimate the running time of a splay, define:

(16.24) $$\gamma := \sum_{P \in \mathcal{P}} \log |P|,$$

taking logarithms with base 2 (again, $|P|$ denotes the number of vertices of P).

For any splay of e one has (adding $'$ to parameters after the splay):

(16.25) the number of iterations of (16.23) is at most $\gamma - \gamma' + 3(\log |P'_e| - \log |P_e|) + 1$.

To show this, consider any iteration (16.23) (adding $'$ to parameters after the iteration).

If Case 1 applies, then

(16.26) $$\begin{aligned}&\gamma - \gamma' + 3(\log |P'_e| - \log |P_e|) + 1\\ &= \log |P_e| + \log |P_f| - \log |P'_e| - \log |P'_f| + 3\log |P'_e| - 3\log |P_e| + 1\\ &= 3\log |P_f| - \log |P'_f| - 2\log |P_e| + 1 \geq 1,\end{aligned}$$

since $P'_e = P_f$ and since P'_f and P_e are subpaths of P_f. If Case 2 applies, then

(16.27) $$\begin{aligned}&\gamma - \gamma' + 3(\log |P'_e| - \log |P_e|) = \log |P_e| + \log |P_f| + \log |P_g|\\ &- \log |P'_e| - \log |P'_f| - \log |P'_g| + 3\log |P'_e| - 3\log |P_e|\\ &= 3\log |P_g| + \log |P_f| - \log |P'_f| - \log |P'_g| - 2\log |P_e| \geq 1.\end{aligned}$$

The last equality follows from $P'_e = P_g$. The last inequality holds since P_e is a subpath of P_f, and P'_f, P'_g, and P_e are subpaths of P_g, and since, if the first alternative in Case 2 holds, then P_e and P'_g are vertex-disjoint (implying $2\log |P_g| \geq \log |P_e| + \log |P'_g| + 1$), and, if the second alternative in Case 2 holds, then P'_f and P'_g are vertex-disjoint (implying $2\log |P_g| \geq \log |P'_f| + \log |P'_g| + 1$).

(16.26) and (16.27) imply (16.25).

V. *The algorithm.* Now we use the splay to perform the conceptual operations described in the conceptual outline (proof section I above). Thus, let v be the last vertex of the current path Q (cf. (16.16)) and let $e = vu$ be an edge in $E \setminus F$ incident with u. Determine the root $P \in \mathcal{P}$ containing u (possibly by splaying an edge in F incident with u).

Case A: P is not on Q. (We keep a pointer to indicate if a root belongs to Q.) Find a root R as follows. If u is incident with no edge in F, then $R := \{u\}$. If u is incident with exactly one edge $f \in F$, splay f and let $R := P_f$. If u is incident with two edges in F, by splaying find $f \in F$ incident with u such that (after splaying f) subpaths$(P_f) = \{R, R'\}$ where $u \in$ ends(R) and $|R| > |R'|$; then delete P_f from $\mathcal{P}$, and f from F.

This determines R. If $|P_t| + |R| \leq k^2$, add e to F, let P_e be the join of P_t, e, and R, and reset P_t in Q to P_e. If $|P_t| + |R| > k^2$, extend Q by $e, P_{t+1} := R$.

Case B: P is on Q, say $P = P_j$. By (possibly) splaying, we can decide if u is at the end of P_j or not. In the former case, reset $Q := P_0, e_1, P_1, \ldots, e_j, P_j$ and let $C := e_{j+1}, P_{j+1}, \ldots, P_t, e$. In the latter case, split P_j to P'_j, f, P''_j in such a way that

$Q := P_0, e_1, P_1, \ldots, e_j, P'_j$ is the initial segment of the original Q ending at u, and let $C := f, P''_j, e_{j+1}, P_{j+1}, \ldots, P_t, e$.

Determine the difference of the sum of the w_e over the odd edges in C and that over the even edges in C. As we know diffsum(S) for any root S, this can be done in time $O(t - j + 1)$. Depending on whether this difference is positive or not, we know (implicitly) which splitting of the edges on C into matchings M and N gives $w(M) \geq w(N)$. From the values of minodd and mineven for the paths $P \in \mathcal{P}$ on C and from the values of w_e for the edges $e_{j+1}, \ldots, e_t$, e on C (and possibly f), we can find the maximum decrease α on the edges in N, and reset the parameters.

Next, for any $P \in \mathcal{P}$ on C with minodd$(P) = 0$ or mineven$(P) = 0$, determine the edges on P of weight 0, delete them after splaying, and decompose P accordingly. Delete any edge e_i on C with $w(e_i) = 0$ (similarly f).

This describes the iteration.

VI. *Running time of the algorithm.* We finally estimate the running time. In any iteration, let γ be the number of roots of $\mathcal{P}$ that are not on Q. Initially, $\gamma \leq n$. During the algorithm, γ only increases when we are in Case B and break a circuit C, in which case γ increases by at most

$$(16.28) \qquad 2\frac{L_C}{k^2} + m_C + 2,$$

where L_C is the length of C in G (that is, the number of edges e_i plus the sum of the lengths of the paths P_i in C), and where m_C is the number of edges of weight 0 deleted at the end of the iteration. Bound (16.28) uses the fact that the sizes of any two consecutive paths along C sum up to more than k^2, except possibly at the beginning and the end of the circuit, and that any edge of weight 0 can split a root into two new roots.

Now if we sum bound (16.28) over all circuits C throughout the iterations, we have

$$(16.29) \qquad \sum_C (2\frac{L_C}{k^2} + m_C + 2) = O(m),$$

since $\sum_C L_C \leq nk^2$, like in the proof of the previous theorem (note that $m_C \geq 1$ for each C, so the term 2 is absorbed by m_C). So the number of roots created throughout the Case B iterations is $O(m)$. Now at each Case A iteration, we split off a part of a root of size less than half the size of the root; the split off part can be used again by Q some time in later iterations. Hence any root can be split at most $\log k^2$ times, and therefore, the number of Case A iterations is $O(m \log k)$. In particular, the number of times we join two paths in $\mathcal{P}$ and make a new path is $O(m \log k)$.

Next consider γ as defined in (16.24). Note that at any iteration except for joins and splays, γ does not increase. At any join, γ increases by at most $\log k^2$, and hence the total increase of γ during joins is $O(m \log^2 k)$.

Now the number of splays during any Case A iteration is $O(1)$, and during any Case B iteration $O(L_C/k^2 + m_C + 1)$. Hence by (16.29), the total number of splays is $O(m \log k)$. By (16.25), each splay takes time $O(\delta + \log k)$, where δ is the decrease of γ (possibly $\delta < 0$). The sum of δ over all splays is $O(m \log^2 k)$, as this is the total increase of γ during joins. So all splays take time $O(m \log^2 k)$. As the number of splits is proportional to the number of splays, and each takes $O(1)$ time, we have the overall time bound of $O(m \log^2 k)$. ∎

This implies a linear-time perfect matching algorithm for regular bipartite graphs:

Corollary 16.11a. *A perfect matching in a regular bipartite graph can be found in linear time.*

Proof. Let $G = (V, E)$ be a k-regular bipartite graph. We keep a weight function $w : E \to \mathbb{Z}_+$, with the property that $w(\delta(v)) = k$ for each $v \in V$. Throughout the algorithm, let G_i be the subgraph of G consisting of those edges e of G with $w_e = 2^i$ (for $i = 1, \ldots$).

Initially, define a weight $w_e := 1$ for each edge e. For $i = 0, 1, \ldots, \lfloor \log_2 k \rfloor$ do the following. Perform a depth-first search in G_i. If we meet a circuit C in G_i, then split C arbitrarily into matchings M and N, reset $w := w + 2^i(\chi^M - \chi^N)$, delete the edges in N, and update G_i (that is, delete the edges of C from G_i).

As G_i has at most $m/2^i$ edges (since $w(E) = \frac{1}{2}kn = m$), and as depth-first search can be done in time linear in the number of edges, this can be done in $O(m + \frac{1}{2}m + \frac{1}{4}m + \cdots) = O(m)$ time.

For the final G and w, all weights are a power of 2 and each graph G_i has no circuits, and hence has at most $|V| - 1$ edges. So G has at most $|V| \log_2 k$ edges. As w is k-regular, by Theorem 16.11 we can find a perfect matching in G in time $O(|V| \log^3 k)$, which is linear in the number of edges of the original graph G. ∎

This result will be used in obtaining a fast edge-colouring algorithm for bipartite graphs (Section 20.9a).

Notes. Alon [2000] gave the following easy $O(m \log m)$-time method for finding a perfect matching in a regular bipartite graph $G = (V, E)$. Let k be the degree, and choose t with $2^t \geq kn$. Let $\alpha := \lfloor 2^t/k \rfloor$ and $\beta := 2^t - k\alpha$. So $\beta < k$. Let H be the graph obtained from G by replacing each edge by α parallel edges, and by adding a β-regular set F of (new) edges, consisting of $\frac{1}{2}n$ disjoint classes, each consisting of β parallel edges. So H is 2^t-regular.

Iteratively, split H into two regular graphs of equal degree (by determining an Eulerian orientation), and reset H to the graph that has a least number of edges in F.

As $|F| = \frac{1}{2}\beta n < 2^t$, after $\log_2 |F| < t$ iterations, H contains no edge in F. Hence after t iterations we have a perfect matching in H not intersecting F; that is, we have a perfect matching in G.

This gives an $O(m \log m)$-time method, provided that we do not display the graph H fully, but handle the parallel edges implicitly (by the sizes as a function of the underlying edges).

Note that $O(m \log m) = O(nk(\log k + \log n))$. An $O(nk + n \log n \log k)$-time algorithm finding a perfect matching in a k-regular bipartite graph was given by Rizzi [2002].

(Csima and Lovász [1992] described a space-efficient $O(n^2 k \log k)$-time algorithm for finding a perfect matching in a k-regular bipartite graph.)

16.7c. The equivalence of Menger's theorem and Kőnig's theorem

We have seen that Kőnig's matching theorem can be derived from Menger's theorem (by the construction given in the proof of Theorem 16.4) — in fact it forms the induction basis in Menger's proof. The interrelation however is even stronger, as was noticed by Hoffman [1960] (cf. Orden [1955], Ford and Fulkerson [1958c], Hoffman and Markowitz [1963], Ingleton and Piff [1973]): in turn Menger's theorem (in the form of Theorem 9.1) can be derived from Kőnig's matching theorem by a direct (noninductive) construction.

Let $D = (V, A)$ be a directed graph and let $S, T \subseteq V$. We may assume that $S \cap T = \emptyset$. For each $v \in V \setminus S$ introduce a vertex v' and for each $v \in V \setminus T$ introduce a vertex v''. Let E be the set of pairs $\{u', v''\}$ with $u \in V \setminus S$ and $v \in V \setminus T$ with the property that $(u, v) \in A$ or $u = v$. This makes the bipartite graph G, containing the matching

(16.30) $$M := \{\{v', v''\} \mid v \in V \setminus (S \cup T)\}.$$

For any $X \subseteq V$, let $X' := \{v' \mid v \in X\}$ and $X'' := \{v'' \mid v \in X\}$.

Now let M' be a matching in G of size $\nu(G)$. For each component of $M \triangle M'$ having more than one vertex, we may assume that it is an M-augmenting path (since any other component K has an equal number of edges in M and in M', and hence we can replace M' by $M' \triangle K$). Each M-augmenting path is an $S'' - T'$ path. Hence there exist $|M'| - |M| = \nu(G) - |V \setminus (S \cup T)|$ vertex-disjoint $S - T$ paths.

Let $U \subseteq V \setminus T$ and $W \subseteq V \setminus S$ be such that $D := U'' \cup W'$ is a vertex cover of G, with $|U| + |W| = \tau(G)$. Then

(16.31) $$C := (U \cap S) \cup (U \cap W) \cup (W \cap T)$$

intersects each $S - T$ path in D. Indeed, suppose $P = (v_0, v_1, \ldots, v_k)$ is an $S - T$ path not intersecting C. We may assume that P intersects S and T only at v_0 and v_k, respectively. Now

(16.32) $$Q := (v_0'', v_1', v_1'', \ldots, v_{k-1}', v_{k-1}'', v_k')$$

is a path in G of odd length $2k - 1$. Hence D intersects Q in at least k vertices. Therefore, $v_0'' \in D$ (hence $v_0 \in U \cap S \subseteq C$), or $v_k' \in D$ (hence $v_k \in W \cap T \subseteq C$), or $v_i', v_i'' \in D$ for some $i \in \{1, \ldots, k-1\}$ (hence $v_i \in U \cap W \subseteq C$). So C intersects each $S - T$ path in D.

As

(16.33) $$\begin{aligned}|C| &= |U \cap S| + |U \cap W| + |W \cap T| = |U \cap S| + |U| + |W| - |U \cup W| + |W \cap T| \\ &= |U| + |W| - |V \setminus (S \cup T)|\end{aligned}$$

(since $(U \cup W) \setminus (S \cup T) = V \setminus (S \cup T)$), and as $|U| + |W| = \tau(G) = \nu(G)$, we have that the size of C is at most the number of disjoint $S - T$ paths found above.

The converse construction (described by Kuhn [1956]) also applies. Let be given a bipartite graph $G = (V, E)$, with colour classes U and W, and a matching M in G. Orient each edge from U to W, and next contract all edges in M. This gives a directed graph $D = (V', A)$. Let S and T be the sets of vertices in U and W missed by M. Then the maximum number of vertex-disjoint $S - T$ paths in D is equal to $\nu(G) - |M|$.

These constructions also imply:

Theorem 16.12. *For any function $\phi(n,m)$ one has: the bipartite matching problem with n vertices and m edges is solvable in time $O(\phi(n,m)) \iff$ the disjoint $s-t$ paths problem with n vertices and m arcs is solvable in time $O(\phi(n,m))$.*

Proof. See above. ∎

16.7d. Equivalent formulations in terms of matrices

Frobenius [1917] proved his theorem (Corollary 16.2a) in terms of matrices, in the following form:

(16.34) Each diagonal of an $n \times n$ matrix has product 0 if and only if M has a $k \times l$ all-zero submatrix with $k + l > n$.

Similarly, Kőnig's matching can be formulated in matrix terms as follows:

(16.35) In a matrix, the maximum number of nonzero entries with no two in the same line (=row or column) is equal to the minimum number of lines that include all nonzero entries.

An equivalent form of Kőnig's theorem on the existence of a perfect matching in a regular bipartite graph (Corollary 16.2b) is:

(16.36) If in a nonnegative matrix each row and each column has the same positive sum, then it has a diagonal with positive entries.

16.7e. Equivalent formulations in terms of partitions

Bipartite graphs can be studied also as unions of two partitions of a given set. Indeed, let $G = (V, E)$ be a bipartite graph. Then the family $(\delta(v) \mid v \in V)$ is a union of two partitions of E. Since each union of two partitions arises in this way, we can formulate theorems on bipartite graphs equivalently as theorems on unions of two partitions of a set.

The following equivalent form of Frobenius' theorem (Corollary 16.2a) was given by Maak [1936]:

(16.37) Let $\mathcal{A}$ and $\mathcal{B}$ be two partitions of the finite set X. Then there is a subset Y of X intersecting each set in $\mathcal{A} \cup \mathcal{B}$ in exactly one element if and only if for each natural number k, the union of any collection of k classes of $\mathcal{A}$ intersects at least k classes of $\mathcal{B}$.

This implies the following equivalent form of Corollary 16.2b, given by van der Waerden [1927] (with short proof by Sperner [1927] — see Section 22.7d):

(16.38) Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be two partitions of a finite set X with $|A_1| = \cdots = |A_n| = |B_1| = \cdots = |B_n|$. Then there is a subset Y of X intersecting each A_i and each B_i in exactly one element.

Some of the matching results can be formulated in terms of (common) transversals. We will discuss this more extensively in Chapters 22 and 23.

16.7f. On the complexity of bipartite matching and vertex cover

In a *bipartite* graph we can derive a minimum-size vertex cover from a maximum-size matching in linear time (for general graphs this would imply NP=P) — see Theorems 16.6.

So knowing a maximum-size matching in a bipartite graph gives us a minimum-size vertex cover in linear time. The reverse, however, is unlikely, unless there would exist an algorithm to find a perfect matching in a bipartite graph in linear time. To see this, suppose that there is an algorithm $\mathcal{A}$ to derive from a minimum-size vertex cover a maximum-size matching in linear time. Now let $G = (V, E)$ be a bipartite graph in which we want to find a perfect matching. Then we may *assume* that G has a perfect matching. So we may assume by Frobenius' theorem that the colour classes U and W are minimum-size vertex covers. Then apply $\mathcal{A}$ to G and U. Then either we obtain a perfect matching if U indeed is a minimum-size vertex cover, or else (if our assumption is wrong) the algorithm gets stuck, in which case we may conclude that G has no perfect matching.

16.7g. Further notes

Extensions of Frobenius' and Kőnig's theorems to the infinite case were considered by Kőnig and Valkó [1925], Shmushkovich [1939], de Bruijn [1943], Rado [1949b], Brualdi [1971f], Aharoni [1983b,1984b], and Aharoni, Magidor, and Shore [1992].

Itai, Rodeh, and Tanimoto [1978] showed that, given a bipartite graph $G = (V, E)$, $F \subseteq E$, and $k \in \mathbb{Z}_+$, one can find a perfect matching M with $|M \cap F| \leq k$ (or decide that no such perfect matching exists) in time $O(nm)$. (This amounts to a minimum-cost flow problem.)

Karp, Vazirani, and Vazirani [1990] gave an optimal on-line bipartite matching algorithm. Motwani [1989,1994] investigated the expected running time of matching algorithms.

The following question was posed by A. Frank: Given a bipartite graph $G = (V, E)$ whose edges are coloured red and blue, and given k and l; when does there exist a matching containing k red edges and l blue edges? This problem is NP-complete, but for complete bipartite graphs it was characterized by Karzanov [1987c].

An extension of Frobenius' theorem to more general matrices than described in Section 16.2b was given by Hartfiel and Loewy [1984].

Dulmage and Mendelsohn [1958] study minimum-size vertex covers in a bipartite graph as a lattice. For maintaining perfect matchings 'in the presence of failure', see Sha and Steiglitz [1993]. Lovász [1970a] gave a generalization of Kőnig's matching theorem — see Section 60.1a. Uniqueness of a maximum-size matching in a bipartite graph was investigated by Cechlárová [1991], and related work was reported by Costa [1994]. A variant of Kőnig's matching theorem was given by de Werra [1984].

For surveys on matching algorithms, see Galil [1983,1986a,1986b]. For surveys on bipartite matching, see Woodall [1978a,1978b]. Books discussing bipartite matching include Ford and Fulkerson [1962], Ore [1962], Dantzig [1963], Christofides [1975], Lawler [1976b], Even [1979], Papadimitriou and Steiglitz [1982], Tarjan [1983], Tutte [1984], Halin [1989], Cook, Cunningham, Pulleyblank, and Schrijver

[1998], Jungnickel [1999], Mehlhorn and Näher [1999], and Korte and Vygen [2000].

16.7h. Historical notes on bipartite matching

The fundaments of matching theory in bipartite graphs were laid by Frobenius (in terms of matrices and determinants) and Kőnig. In his article *Über Matrizen aus nicht negativen Elementen* (On matrices with nonnegative elements), Frobenius [1912] investigated the decomposition of matrices:

> In §11 dehne ich die Untersuchung auf zerlegbare Matrizen aus, und in §12 zeige ich, daß eine solche nur auf eine Art in unzerlegbare Teile zerfällt werden kann. Dabei ergibt sich der merkwürdige Determinantensatz:
> I. *Die Elemente einer Determinante nten Grades seien n^2 unabhängige Veränderliche. Man setze einige derselben Null, doch so, daß die Determinante nicht identisch verschwindet. Dann bleibt sie eine irreduzible Funktion, außer wenn für einen Wert $m < n$ alle Elemente verschwinden, die m Zeilen mit $n - m$ Spalten gemeinsam haben.*[4]

Frobenius gave a combinatorial and an algebraic proof.

In a reaction to Frobenius' paper, Kőnig [1915] ('presented to Class III of the Hungarian Academy of Sciences on 16 November 1914') next gave a proof of Frobenius' result with the help of graph theory:

> A graphok alkalmazásával e tételnek egyszerű és szemléletes új bizonyitását adjuk a következőkben.[5]

He introduced a now quite standard construction of making a bipartite graph from a matrix $(a_{i,j})$: for each row index i there is a vertex A_i and for each column index j there is a vertex B_j; then vertices A_i and B_j are connected by an edge if and only if $a_{i,j} \neq 0$.

Kőnig was interested in graphs because of his interest in set theory, especially cardinal numbers (cf. footnotes in Kőnig [1916]). In proving Schröder-Bernstein type results on the equivalence of sets, graph-theoretic arguments (in particular: matchings) can be illustrative. This led Kőnig to studying graphs (in particular bipartite graphs) and its applications in other areas of mathematics.

Kőnig's work on matchings in regular bipartite graphs

Earlier, on 7 April 1914, Kőnig had presented the following theorem at the *Congrès de Philosophie mathématique* in Paris (cf. Kőnig [1923]):

> *A. Chaque graphe régulier à circuits pairs possède un facteur du premier degré.*[6]

[4] In §11, I extend the investigation to decomposable matrices, and in §12, I show that such a matrix can be decomposed in only one way into indecomposable parts. With that, the [following] curious determinant theorem comes up:
I. *Let the elements of a determinant of degree n be n^2 independent variables. One sets some of them equal to zero, but such that the determinant does not vanish identically. Then it remains an irreducible function, except when for some value $m < n$ all elements vanish that have m rows in common with $n - m$ columns.*

[5] In the following we will give a simple and clear new proof by applying graphs to this theorem.

[6] *A. Each regular graph with even circuits has a factor of the first degree.*

That is, every regular bipartite graph has a perfect matching (= factor of degree 1). As a corollary, Kőnig derived:

> *B. Chaque graphe régulier à circuits pairs est le produit de facteurs du premier degré; le nombre de ces facteurs est égal au degré du graphe.*[7]

That is, each k-regular bipartite graph is k-edge-colourable (cf. Chapter 20). Kőnig did not give a proof of the theorem in the Paris paper, but expressed the hope to give a complete proof 'at another occasion'.

This occasion came in Kőnig [1916] ('presented to Class III of the Hungarian Academy of Sciences on 15 November 1915') where next to the above mentioned Theorems A and B, Kőnig gave the following result:

> C) *Ha egy páros körüljárású graph bármelyik csúcsába legfeljebb k-számú él fut, akkor minden éléhez oly módon lehet k-számú index valamelyikét hozzárendelni, hogy ugyanabba a csúcsba futó két élhez mindenkor két különböző index legyen rendelve.*[8]

In other words, the edge-colouring number of a bipartite graph is equal to its maximum degree. Kőnig gave a proof of result C), and derived A and B. (See the proof of Theorem 20.1 below of Kőnig's proof.)

In §2 of Kőnig [1916], applications of his results to matrices and determinants are studied. First:

> D) *Ha egy nem negativ [egész számú] elemekből álló determináns minden sora és minden oszlopa ugyanazt a positiv összeget adja, akkor van a determinánsnak legalább egy el nem tünő tagja.*[9]

Next:

> E) *Ha egy determináns minden sorában és oszlopában pontosan k-számú el nem tünő elem van, akkor legalább k-számú determinánstag nem tünik el.*[10]

Third:

> F) *Ha egy* n^2 *mezejű quadratikus táblán kn-számú figura úgy van elhelyezve (ugyanazon a mezőn több figura is lehet), hogy minden sorban és oszlopban pontosan k-számú figura fordul elő, akkor e konfiguráczió mindig mint k-számú ugyancsak* n^2 *mezejű oly konfiguráczió egyesítése keletkeztethető, melyek mindegyikében egy-egy figura van minden sorban és minden oszlopban.*[11]

---

[7] *B. Each regular graph with even circuits is the product of factors of the first degree; the number of these factors is equal to the degree of the graph.*

[8] C) *If in each vertex of an even circuit graph at most k edges meet, then one can assign to each of the edges of the graph one from k indices in such a way that two edges that meet in a point always obtain different indices.*

[9] D) *If in a determinant of nonnegative [integer] numbers each row and each column yield the same positive sum, then at least one member of the determinant is different from zero.*

[10] E) *If the number of nonvanishing elements in each row and column of a determinant is exactly equal to k, then there are at least k nonvanishing determinant members.*

[11] F) *If kn pieces are placed on a quadratic board with* n^2 *fields (where several pieces may stand in the same field), such that each row and each column contains exactly k pieces, then this configuration always arises by joining k such configurations with also* n^2 *fields, in which each row and each column contains exactly one piece.*

Frobenius' theorem

Chronologically next is a paper of Frobenius [1917]. In order to give an elementary proof of his result in Frobenius [1912] quoted above, he proved the following 'Hilfssatz':

> II. *Wenn in einer Determinante nten Grades alle Elemente verschwinden, welche p $(\leq n)$ Zeilen mit $n-p+1$ Spalten gemeinsam haben, so verschwinden alle Glieder der entwickelten Determinante.*
> *Wenn alle Glieder einer Determinante nten Grades verschwinden, so verschwinden alle Elemente, welche p Zeilen mit $n-p+1$ Spalten gemeinsam haben für $p=1$ oder $2, \cdots$ oder n.*[12]

That is, if $A = (a_{i,j})$ is an $n \times n$ matrix, and if $\prod_{i=1}^{n} a_{i,j} = 0$ for each permutation π of $\{1, \ldots, n\}$, then for some p there exist p rows and $n-p+1$ columns of A such that each element that is both in one of these rows and in one of these columns, is equal to 0.

In other words, a bipartite graph $G = (V, E)$ with colour classes V_1 and V_2 satisfying $|V_1| = |V_2| = n$ has a perfect matching if and only if one cannot select p vertices in V_1 and $n-p+1$ vertices in V_2 such that no edge is connecting two of these vertices.

Frobenius noticed with respect to Kőnig's work:

> Aus dem Satze II ergibt sich auch leicht ein Ergebnis der Hrn. DÉNIS KŐNIG, *Uber Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre, Math. Ann. Bd. 77.*
> *Wenn in einer Determinante aus nicht negativen Elementen die Größen jeder Zeile und jeder Spalte dieselbe, von Null verschiedene Summe haben, so können ihre Glieder nicht sämtlich verschwinden.*[13]

Frobenius gave a short combinatorial proof of his theorem — see Section 16.2a. His proof is in terms of determinants, and he offered his opinion on graph-theoretic methods:

> Die Theorie der Graphen, mittels deren Hr. KŐNIG den obigen Satz abgeleitet hat, ist nach meiner Ansicht ein wenig geeignetes Hilfsmittel für die Entwicklung der Determinantentheorie. In diesem Falle führt sie zu einem ganz speziellen Satze von geringem Werte. Was von seinem Inhalt Wert hat, ist in dem Satze II ausgesprochen.[14]

(See Schneider [1977] for some comments.)

---

[12] II. *If in a determinant of the nth degree all elements vanish that p $(\leq n)$ rows have in common with $n-p+1$ columns, then all members of the expanded determinant vanish.*
If all members of a determinant of degree n vanish, then all elements vanish that p rows have in common with $n-p+1$ columns for $p = 1$ or $2, \cdots$ or n.

[13] From Theorem II, a result of Mr DÉNIS KŐNIG, *Uber Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre, Math. Ann. Vol. 77* follows also easily.
If in a determinant of nonnegative elements the quantities of each row and of each column have the same nonzero sum, then its members cannot vanish altogether.

[14] The theory of graphs, by which Mr KŐNIG has derived the theorem above, is to my opinion of little appropriate help for the development of determinant theory. In this case it leads to a very special theorem of little value. What from its contents has value, is enunciated in Theorem II.

Equivalent formulations in terms of partitions

In October 1926, van der Waerden [1927] presented the following theorem at the *Mathematisches Seminar* in Hamburg:

> *Es seien zwei Klasseneinteilungen einer endlichen Menge $\mathcal{M}$ gegeben. Die eine soll die Menge in μ zueinander fremde Klassen $\mathcal{A}_1, \ldots, \mathcal{A}_\mu$ zu je n Elementen zerlegen, die andere ebenfalls in μ fremde Klassen $\mathcal{B}_1, \ldots, \mathcal{B}_\mu$ zu je n Elementen. Dann gibt es ein System von Elementen $x_1, \ldots, x_\mu$, derart, daß jede A-Klasse und ebenso jede B-Klasse under den x_i durch ein Element vertreten wird.*[15]

The proof of van der Waerden is based on an augmenting path argument. Moreover, van der Waerden remarked that E. Artin has communicated orally to him that the result can be sharpened to the existence of n disjoint such common transversals.

In the article, the following note is added in proof:

> **Zusatz bei der Korrektur.** Ich bemerke jetzt, daß der hier bewiesene Satz mit einem Satz von Dénes Kőnig über reguläre Graphen äquivalent ist.[16]

The article of van der Waerden is followed by an article of Sperner [1927] (presented at the *Mathematisches Seminar* in Januari 1927), which gives a 'simple proof' of van der Waerden's result — we quote the full paper in Section 22.7d.

Kőnig's matching theorem

At the meeting of 26 March 1931 of the Eötvös Loránd Matematikai és Fizikai Társulat (Loránd Eötvös Mathematical and Physical Society) in Budapest, Kőnig [1931] presented a new result that formed the basis for Menger's theorem:

> Páros körüljárású graphban az éleket kimerítő szögpontok minimális száma megegyezik a páronként közös végpontot nem tartalmazó élek maximális számával.[17]

In other words, the maximum size of a matching in a bipartite graph is equal to the minimum number of vertices needed to cover all edges. As we discussed in Section 9.6e, Kőnig's proof formed the missing basis for Menger's theorem. Kőnig also referred to the work of Frobenius (but did not notice that his theorem can be derived from Frobenius' theorem).

The proof of Kőnig [1931] is based on an augmenting path argument. A German version of it was published in Kőnig [1932] (stating that another proof was given by L. Kalmár), in which paper he described several other results as consequences of the theorem. First he derived his theorem on the existence of a perfect matching in a regular bipartite graph:

[15] Let be given two partitions of a finite set $\mathcal{M}$. One of them should decompose the set into μ mutually disjoint classes $\mathcal{A}_1, \ldots, \mathcal{A}_\mu$ each of n elements, the other likewise in μ disjoint classes $\mathcal{B}_1, \ldots, \mathcal{B}_\mu$ each of n elements. Then there exists a system of elements $x_1, \ldots, x_\mu$ such that each $\mathcal{A}$-class and likewise each $\mathcal{B}$-class is represented by one element among the x_i.

[16] **Note added in proof.** I now notice that the theorem proved here is equivalent to a theorem of Dénes Kőnig on regular graphs.

[17] In an even circuit graph, the minimal number of vertices that exhaust the edges agrees with the maximal number of edges that pairwise do not contain any common end point.

> Um die Tragweite dieses Satzes zu beleuchten, wollen wir noch zeigen, daß ein von mir schon vor längerer Zeit bewiesener Satz über die Faktorenzerlegung von regulären endlichen paaren Graphen aus Satz 13 unmittelbar abgeleitet werden kann.
> Der betreffende Satz lautet:
> 14. *Jeder endliche paare reguläre Graph besitzt einen Faktor ersten Grades.*[18]

In a footnote, Kőnig mentioned:

> Später wurden für diesen Satz, bzw. für seine Interpretation in der Determinantentheorie und in der Kombinatorik verschiedene Beweise gegeben, so durch Frobenius, Sainte-Laguë, van der Waerden, Sperner, Skolem, Egerváry.[19]

Another consequence is a graph-theoretic variant of the result of Frobenius [1912] on reducible determinants:

> 16. *Im (paaren) Graphen G soll jede Kante einen der Punkte von $\Pi_1 = (P_1, P_2, \ldots, P_n)$ mit einem der Punkte von $\Pi_2 = (Q_1, Q_2, \ldots, Q_n)$ verbinden $(P_i \neq Q_j)$ und diejenigen Kanten von G, die in einem Faktor ersten Grades von G enthalten sind, sollen einen nichtzusammenhängenden Graphen G^* bilden. Dann kann man $r(> 0)$ Punkte aus Π_1 und $n - r(> 0)$ Punkte aus Π_2 so auswählen, daß keine Kante von G zwei ausgewählte Punkte verbinde.*[20]

As consequences in matrix theory, Kőnig [1932] gave:

> 17. *Verschwinden sämtliche Entwicklungsglieder aller Unterdeterminanten n-ter Ordnung einer Matrix von p Zeilen und q Spalten (wo $n \leq p$, $n \leq q$ ist), so verschwinden alle Elemente, welche r Zeilen mit $(p + q - n + 1) - r$ Spalten gemeinsam haben für $r = 1$, oder $2, \ldots,$ oder p.*[21]

and

> 18. *Die Minimalzahl der Reihen (Zeilen und Spalten), welche in ihr Gesamtheit jedes nicht-verschwindende Element einer Matrix enthalten, ist gleich der Maximalzahl von nicht-verschwindenden Elementen, welche paarweise verschiedenen Zeilen und verschiedenen Spalten angehören.*[22]

Again, a footnote is added:

---

[18] To illustrate the bearing of this theorem, we want to show that a theorem, proved by me already long ago, on the factorization of regular finite bipartite graphs, can be derived immediately from Theorem 13.
The theorem referred to reads:
14. *Every finite bipartite regular graph possesses a factor of first degree.*

[19] Later, several proofs were given for this theorem, respectively for its interpretation in determinant theory and in combinatorics, so by Frobenius, Sainte-Laguë, van der Waerden, Sperner, Skolem, Egerváry.

[20] 16. *Let every edge in the (bipartite) graph G connect a vertex of $\Pi_1 = (P_1, \ldots, P_n)$ with a vertex of $\Pi_2 = (Q_1, \ldots, Q_n)$ $(P_i \neq Q_j)$, and let those edges of G that are contained in a factor of first degree form a disconnected graph G^*. Then one can choose $r(> 0)$ vertices in Π_1 and $n - r(> 0)$ vertices in Π_2 such that no edge of G connects two of the chosen vertices.*

[21] 17. *If all expansion terms of all underdeterminants of the order n of a matrix with p rows and q columns vanish (where $n \leq p$, $n \leq q$), then all entries vanish that r rows have in common with $(p + q - n + 1) - r$ columns, for $r = 1$, or $2, \ldots,$ or p.*

[22] 18. *The minimum number of lines (rows and columns) that together contain each nonvanishing entry of a matrix, is equal to the maximum number of nonvanishing entries that pairwise belong to different rows and different columns.*

> Die Sätze 17 und 18 hat der Verfasser, mit den hier gegebenen Beweisen, am 26. März 1931 in der Budapester Mathematischen und Physikalischen Gesellschaft vorgetragen, s. [6]. Hieran anschließend hat dann E. Egerváry [1] für den Satz 18 einen anderen Beweis und eine interessante Verallgemeinerung gegeben.[23]

(We note that references [6] and [1] in Kőnig's article correspond to our references Kőnig [1931] and Egerváry [1931].)

Kőnig also derived the theorems of Frobenius [1912,1917] mentioned above:

> 19. *Wenn alle Glieder einer Determinante n-ter Ordnung verschwinden, so verschwinden alle Elemente, welche r Zeilen mit $n-r+1$ Spalten gemeinsam haben, für $r = 1$ oder $2, \ldots,$ oder n.*[24]

> 20. *In einer Determinante n-ter Ordnung D seien die nichtverschwindenden Elemente unabhängige Veränderliche. Ist D eine reduzible Funktion ihrer (nichtverschwindenden) Elemente, so verschwinden alle Elemente von D, welche r Zeilen mit $n - r$ Spalten gemeinsam haben für $r = 1$ oder $2, \ldots,$ oder $n - 1$.*[25]

With respect to Frobenius [1912], Kőnig noticed in a footnote:

> Dort wird dieser Satz "aus verborgenen Eigenschaften der Determinanten mit nichtnegativen Elementen" durch komplizierte Betrachtungen bewiesen. Ich gab dann in 1915 in meiner Arbeit [4] einen elementaren graphentheoretischen Beweis (welcher hier durch einen noch einfacheren ersetzt wird). In 1917 hat dann auch Frobenius [3] einen elementaren Beweis publiziert, und zwar nach dem ich ihm meinen Beweis (in deutscher Übersetzung) zugeschickt hatte. Frobenius hat es dort unterlassen, diese Tatsache, sowie überhaupt meine Arbeit [4] zu erwähnen. Jedoch zitiert er meine Arbeit [5] und zwar mit folgender Bemerkung: "Die Theorie der Graphen, mittels deren Hr. Kőnig den obigen Satz [dies ist die determinantentheoretische Interpretation von Satz 14] abgeleitet hat, ist nach meiner Ansicht ein wenig geeignetes Hilfsmittel für die Entwicklung der Determinantentheorie. In diesem Falle führt sie zu einem ganz speziellen Satz vom geringem Werte. Was von seinem Inhalt Wert hat, ist in dem Satze II [dies ist der Frobeniussche Satz 19] ausgesprochen."
> Es ist wohl natürlich, daß der Verfasser vorliegender Abhandlung diese Meinung nicht unterschreiben wird. Die Gründe, die man für oder gegen den Wert oder Unwert eines Satzes oder eine Methode anführen könnte, haben stets, mehr oder weniger, einen subjektiven Character, so daß es vom geringen wissenschaftlichen Wert wäre, wenn wir hier den Standpunkt von Frobenius zu bekämpfen versuchten. Wollte aber Frobenius seine verwerfende Kritik über die Anwendbarkeit der Graphen auf Determinantentheorie damit begründen, daß sein tatsächlich "wertvoller" Satz 19 nicht graphentheoretisch bewiesen werden kann, so ist seine Begründung—wie wir gesehen haben—sicherlich nicht stichhaltig. Der graphentheoretische Beweis, den wir für Satz 19 gegeben haben, scheint uns ein einfacher und anschaulicher Beweis zu sein, der dem *kombinatorischen* Character der Satzes in natürlicher Weise entspricht und auch zu einer bemerkenswerten Verallgemeinerung (Satz 17) führt.

[23] The author has presented Theorems 17 and 18, with the proofs given here, on 26 March 1931 to the Budapest Mathematical and Physical Society, see [6]. Following this, E. Egerváry [1] has next given another proof for Theorem 18 and an interesting generalization.

[24] 19. *When all members of a determinant of the order n vanish, then all elements vanish that have r rows in common with $n - r + 1$ columns, for $r = 1$ or $2, \ldots,$ or n.*

[25] 20. *Let, in a determinant D of order n, the nonvanishing entries be independent variables. If D is a reducible function of its (nonvanishing) entries, then all entries of D vanish that have r rows in common with $n - r$ columns for $r = 1$ or $2, \ldots,$ or $n - 1$.*

> Es sei noch erwähnt, daß wir oben, im §2, beim Beweis des Satzes 16 einen Gedanken von FROBENIUS benützt haben, den er bei seiner Zurückführung des Satzes 20 auf Satz 19 angewendet hat.[26]

(We note that Kőnig's quotation 'aus verborgenen Eigenschaften der Determinanten mit nichtnegativen Elementen' is from Frobenius [1917]. The references [3], [4], and [5] in Kőnig's article correspond to our references Frobenius [1917], Kőnig [1915], and Kőnig [1916], respectively.)

In terms of transversals, the theorems of Frobenius and Kőnig have been rediscovered by Hall [1935] — see the historical notes on transversals in Section 22.7d. Other developments are mentioned in Section 19.5a.

[26] This theorem was proved there 'from hidden properties of determinants with nonnegative elements' by complicated arguments. Next, I gave in 1915, in my work [4], an elementary, graph-theoretic proof (which was replaced here by an even simpler one). Next, in 1917, also FROBENIUS [3] has published an elementary proof, and that after I had sent him my proof (in German translation). FROBENIUS has refrained from mentioning this fact there, as well as my work [4] at all. Yet, he quotes my work [5], and that with the following remark: 'The theory of graphs, by which Mr. KŐNIG has derived the theorem above [this is the determinant-theoretic interpretation of Theorem 14], is, to my opinion, of little appropriate help for the development of determinant theory. In this case it leads to a very special theorem of little value. What from its contents has value, is expressed in Theorem II [this is Theorem 19 of Frobenius]'.

Obviously, the author of the present treatise will not subscribe to this opinion. The arguments that one can produce for or against the value or valuelessness of a theorem or a method, have always, more or less, a subjective character, so that it would be of little scientific value when we here tried to fight the point of view of FROBENIUS. But if FROBENIUS wants to base his rejecting criticism about the applicability of graphs to determinant theory on the fact that his actually 'more valuable' Theorem 19 cannot be proved graph-theoretically, then his ground is—as we have seen—certainly not solid. The graph-theoretic proof that we have given for Theorem 19 seems to us to be a simple and illustrative proof, that corresponds naturally to the *combinatorial* character of the theorem and also leads to a remarkable generalization (Theorem 17).

Let it finally be mentioned that above, in §2, in the proof of Theorem 16, we have used an idea of FROBENIUS, which he has applied at his reduction of Theorem 20 to Theorem 19.

Chapter 17

Weighted bipartite matching and the assignment problem

The methods and results of the previous chapter can be extended to handle maximum-*weight* matchings. Egerváry's theorem is the weighted version of Kőnig's matching theorem. It led Kuhn to develop the 'Hungarian method' for the assignment problem. This problem is equivalent to finding a minimum-weight perfect matching in a complete bipartite graph.

17.1. Weighted bipartite matching

For bipartite graphs, Egerváry [1931] characterized the maximum weight of a matching by the following duality relation:

Theorem 17.1 (Egerváry's theorem)**.** *Let $G = (V, E)$ be a bipartite graph and let $w : E \to \mathbb{R}_+$ be a weight function. Then the maximum weight of a matching in G is equal to the minimum value of $y(V)$, where $y : V \to \mathbb{R}_+$ is such that*

$$y_u + y_v \geq w_e \tag{17.1}$$

for each edge $e = uv$. If w is integer, we can take y integer.

Proof. The maximum is not more than the minimum, since for any matching M and any $y \in \mathbb{R}_+^V$ satisfying (17.1) for each edge $e = uv$, one has

$$w(M) \leq \sum_{e=uv \in M} (y_u + y_v) \leq \sum_{v \in V} y_v. \tag{17.2}$$

To see equality, choose a $y \in \mathbb{R}_+^V$ attaining the minimum value. Let F be the set of edges e having equality in (17.1) and let R be the set of vertices v with $y_v > 0$.

If F contains a matching M covering R, we have equality throughout in (17.2), showing that the maximum is equal to the minimum value.

So we may assume that no such matching exists. Then by Corollary 16.8a there exists a stable set $S \subseteq R$ containing no edge and such that $|N(S)| < |S|$.

Then there is an $\alpha > 0$ such that decreasing y_v by α for $v \in S$ and increasing y_v by α for $v \in N(S)$ gives a better y — a contradiction.

If w is integer we can keep y integer, by taking $\varepsilon = 1$ throughout. ∎

(This is essentially the proof method of Egerváry [1931].)

We can formulate Egerváry's theorem in combinatorial terms. Let $G = (V, E)$ be a graph and let $w \in \mathbb{Z}_+^E$. A *w-vertex cover* is a vector $y \in \mathbb{Z}_+^V$ such that

$$y_u + y_v \geq w_e \tag{17.3}$$

for each edge $e = uv$ of G. The *size* of any vector $y \in \mathbb{R}^V$ is the sum of its components.

Corollary 17.1a. *Let $G = (V, E)$ be a bipartite graph and let $w : E \to \mathbb{Z}_+$ be a weight function. Then the maximum weight of a matching in G is equal to the minimum size of a w-vertex cover.*

Proof. The corollary is a reformulation of the integer part of Egerváry's theorem (Theorem 17.1). ∎

Let A be the $V \times E$ incidence matrix of G. Egerváry's theorem states that for $w \in \mathbb{Z}_+^V$, the optima in the linear programming duality equation

$$\max\{w^{\mathsf{T}} x \mid x \geq \mathbf{0}, Ax \leq \mathbf{1}\} = \min\{y^{\mathsf{T}} \mathbf{1} \mid y \geq \mathbf{0}, y^{\mathsf{T}} A \geq w^{\mathsf{T}}\} \tag{17.4}$$

are attained by integer vectors x and y. This also follows from the total unimodularity of A — see Section 18.3.

17.2. The Hungarian method

We describe the *Hungarian method* for the maximum-weight matching problem. In its basic form it is due to Kuhn [1955b], based on Egerváry's proof above. Sharpenings were given by Munkres [1957] (yielding a polynomial-time method), Iri [1960], Edmonds and Karp [1970], and Tomizawa [1971].

Let $G = (V, E)$ be a bipartite graph, with colour classes U and W, and let $w : E \to \mathbb{Q}$ be a weight function.

We start with matching $M = \emptyset$. If we have found a matching M, let D_M be the directed graph obtained from G by orienting each edge e in M from W to U, with length $l_e := w_e$, and orienting each edge e not in M from U to W, with length $l_e := -w_e$. Let U_M and W_M be the set of vertices in U and W, respectively, missed by M. If there is a $U_M - W_M$ path, find a shortest such path, P say, and reset $M' := M \triangle EP$.

We iterate until no $U_M - W_M$ path exists in D_M (whence M is a maximum-*size* matching). The maximum-weight matching among the matchings found, has maximum weight among all matchings.

To see this, call a matching M *extreme* if it has maximum weight among all matchings of size $|M|$. Then, inductively:

Theorem 17.2. *Each matching M found is extreme.*

Proof. This is clearly true if $M = \emptyset$. Suppose next that M is extreme, and let P and M' be the path and matching found in the iteration. Consider any extreme matching N of size $|M|+1$. As $|N| > |M|$, $M \cup N$ has a component Q that is an M-augmenting path. As P is a shortest M-augmenting path, we know $l(Q) \geq l(P)$. As $N \triangle Q$ is a matching of size $|M|$, and as M is extreme, we have $w(N \triangle Q) \leq w(M)$. Hence $w(N) = w(N \triangle Q) - l(Q) \leq w(M) - l(P) = w(M')$. ▮

If M is extreme, then D_M has no negative-length circuit C (otherwise $M \triangle C$ is a matching of size $|M|$ and larger weight than M). So by the theorem, we can find with the Bellman-Ford method a shortest $U_M - W_M$ path in time $O(nm)$, yielding an $O(n^2 m)$ method overall (Iri [1960]).

But in fact one may apply Dijkstra's method (Edmonds and Karp [1970], Tomizawa [1971]) and obtain a better time bound:

Theorem 17.3. *The method can be performed in time $O(n(m + n \log n))$.*

Proof. Let R_M denote the set of vertices reachable in D_M from U_M. We show that along with M we can keep a potential p for the subgraph $D_M[R_M]$ of D_M induced by R_M (with respect to the length function l defined above).[27]

When $M = \emptyset$ we take $p(v) := \max\{w_e \mid e \in E\}$ if $v \in U$ and $p(v) := 0$ if $v \in W$.

Suppose next that for given extreme M we have a potential p for $D_M[R_M]$. Then define $p'(v) := \mathrm{dist}_l(U_M, v)$ for each $v \in R_M$. Note that having p, one can determine p' in $O(m + n \log n)$ time (cf. Section 8.2).

Then p' is a potential for $D_{M'}[R_{M'}]$. To see this, let P be the path in D_M with $M' = M \triangle EP$. Trivially, $U_{M'} \subseteq U_M$. Moreover, $R_{M'} \subseteq R_M$. Indeed, otherwise some arc of $D_{M'}$ leaves R_M. As no arc of D_M leaves R_M, this implies that P has an arc entering R_M. So P has an arc leaving R_M, contradicting the definition of R_M. Concluding, $R_{M'} \subseteq R_M$.

Finally consider an arc (u, v) of $D_{M'}[R_{M'}]$. If (u, v) is also an arc of D_M, then $p'(v) \leq p'(u) + l(u, v)$. If (u, v) is not an arc of D_M, then (v, u) belongs to P, and hence (as P is shortest) $p'(u) = p'(v) + l(v, u)$. So $p'(v) - p'(u) = -l(v, u) = l(u, v)$. ▮

Observe that in the Hungarian method one can stop as soon as matching M' has no larger weight than M; that is, D_M has no $U_M - W_M$ path of negative length. For let N be a matching with $w(N) > w(M)$. So $|N| > |M|$

[27] A *potential* for a digraph $D = (V, A)$ with respect to a length function $l : A \to \mathbb{R}$ is a function $p : V \to \mathbb{R}$ satisfying $p(v) - p(u) \leq l(a)$ for each arc $a = (u, v)$.

(since all matchings of size $\leq |M|$ have weight $\leq w(M)$). Choose N with $|N \triangle M|$ minimal. By similar arguments as used in the proof of Theorem 17.2, we may assume that $N \triangle M$ has $|N| - |M|$ nontrivial components, each having one more edge in N than in M. So each component gives a $U_M - W_M$ path in D_M. As none of them have negative length, we have $w(N) \leq w(M)$, a contradiction.

Hence we can reduce the factor n in the time bound:

Theorem 17.4. *In a weighted bipartite graph, a maximum-weight matching can be found in time $O(n'(m + n \log n))$, where n' is the minimum size of a maximum-weight matching.*

Proof. See above. ∎

17.3. Perfect matching and assignment problems

The methods described above also find a maximum-weight perfect matching in a bipartite graph. This follows from the fact that a maximum-weight perfect matching is an extreme matching of size $\frac{1}{2}|V|$.

By multiplying all weights by -1, this problem can be seen to be equivalent to finding a *minimum*-weight perfect matching. Hence:

Corollary 17.4a. *A minimum-weight perfect matching can be found in time $O(n(m + n \log n))$.*

Proof. Directly from the above. ∎

This in turn gives an algorithm for the *assignment problem*:

(17.5) given: a rational $n \times n$ matrix $A = (a_{i,j})$;

find: a permutation π of $\{1, \ldots, n\}$ minimizing $\displaystyle\sum_{i=1}^{n} a_{i,\pi(i)}$.

Corollary 17.4b. *The assignment problem can be solved in time $O(n^3)$.*

Proof. Take $G = K_{n,n}$ in Corollary 17.4a. ∎

The following characterization of the minimum weight of a perfect matching can be derived from Egerváry's theorem — we however give a direct proof that might be illuminating:

Theorem 17.5. *Let $G = (V, E)$ be a bipartite graph having a perfect matching and let $w : E \to \mathbb{Q}$ be a weight function. The minimum weight of a perfect matching is equal to the maximum value of $y(V)$ taken over $y : V \to \mathbb{Q}$ with*

(17.6) $y_u + y_v \leq w_e$ *for each edge* $e = uv$.

If w *is integer, we can take* y *integer.*

Proof. Clearly, the minimum is not less than the maximum, since for any perfect matching M and any $y \in \mathbb{Q}^V$ satisfying (17.6) one has

$$w(M) = \sum_{e \in M} w_e \geq \sum_{v \in V} y_v = y(V). \tag{17.7}$$

To see the reverse inequality, let M be a minimum-weight perfect matching. Make a digraph $D = (V, A)$, with length function, as follows. Orient any edge e of G from one colour class, U say, to the other, W say, with length w_e. Moreover, add for each edge e in M an arc parallel to e oriented from W to U, with length $-w_e$. As M is minimum-weight, the digraph has no negative-weight directed circuits (otherwise we could make a perfect matching of smaller weight). Hence, by Theorem 8.2, there exists a function $p : V \to \mathbb{Q}$ such that $w(a) \geq p(v) - p(u)$ for each arc $a = (u, v)$ of D. Defining $y_v := -p(v)$ for $v \in U$ and $y_v := p(v)$ for $v \in W$, we obtain a function y satisfying (17.6). For each edge $e = uv$ in M, the arcs (u, v) and (v, u) form a zero-length directed circuit in D, and therefore $w_e = y_u + y_v$. This gives equality in (17.7).

If w is integer, we can take p and hence y integer. ∎

17.4. Finding a minimum-size w-vertex cover

Given a maximum-weight matching M in a bipartite graph $G = (V, E)$ with weight $w : E \to \mathbb{Z}_+$, we can find a minimum-size w-vertex cover as follows. Let U and W be the colour classes of G. As before, define $U_M := U \setminus \bigcup M$ and $W_M := W \setminus \bigcup M$.

For any edge $e = uv$, with $u \in U$, $v \in W$, make an arc $a = (u, v)$, of length $l(a) := -w_e$. If $e \in M$, make also an arc $a' = (v, u)$, of length $l(a') := w_e$. We obtain a directed graph $D = (V, A)$ without negative-length directed circuits and no negative-length directed path from $U_M \cup (W \setminus W_M)$ to $W_M \cup (U \setminus U_M)$ (otherwise we can improve M). Then we can find a potential $p : V \to \mathbb{Z}$ such that $l(a) \geq p(v) - p(u)$ for each arc $a = (u, v)$ of D and such that $p(v) = 0$ for each $v \in U_M \cup W_M$, $p(v) \geq 0$ for each $v \in U$, and $p(v) \leq 0$ for each $v \in W$. To see this, add an extra vertex r, and arcs (r, v) for each $v \in U_M \cup (W \setminus W_M)$ and (v, r) for each $v \in W_M \cup (U \setminus U_M)$. Let the new arcs have length 0. Then the extended digraph D' has no negative-length circuits. Let p be a potential for D'. By translating, we can assume $p(r) = 0$. Resetting $p(v)$ to 0 if $v \in U_M \cup W_M$ maintains that p is a potential. This gives a potential for D as described.

Now set $y_v := -p(v)$ if $v \in U$ and $y_v := p(v)$ if $v \in W$. Then y is a w-vertex cover of size $w(M)$, and hence it is a minimum-size w-vertex cover. Therefore (Iri [1960]):

Theorem 17.6. *A minimum-size w-vertex cover in a bipartite graph can be found in $O(n(m + n \log n))$ time.*

Proof. See above. ∎

17.5. Further results and notes

17.5a. Complexity survey for maximum-weight bipartite matching

Complexity survey for the maximum-weight bipartite matching ($*$ indicates an asymptotically best bound in the table):

	$O(nW \cdot \mathrm{VC}(n,m))$	Egerváry [1931] (implicitly)
	$O(2^n n^2)$	Easterfield [1946]
	$O(nW \cdot \mathrm{DC}(n,m,W))$	Robinson [1949]
	$O(n^4)$	Kuhn [1955b], Munkres [1957][28] Hungarian method
	$O(n^2 m)$	Iri [1960]
	$O(n^3)$	Dinits and Kronrod [1969]
$*$	$O(n \cdot \mathrm{SP}_+(n,m,W))$	Edmonds and Karp [1970], Tomizawa [1971]
	$O(n^{3/4} m \log W)$	Gabow [1983b,1985a,1985b]
$*$	$O(\sqrt{n}\, m \log(nW))$	Gabow and Tarjan [1988b,1989] (cf. Orlin and Ahuja [1992])
	$O(\sqrt{n}\, mW)$	Kao, Lam, Sung, and Ting [1999]
$*$	$O(\sqrt{n}\, mW \log_n(n^2/m))$	Kao, Lam, Sung, and Ting [2001]

Here $W := \|w\|_\infty$ (assuming w to be integer-valued). Moreover, $\mathrm{SP}_+(n,m,W)$ is the time needed to find a shortest path in a directed graph with n vertices and m arcs, with nonnegative integer lengths on the arcs, each at most W. Similarly, $\mathrm{DC}(n,m,W)$ is the time required to find a negative-length directed circuit in a directed graph with n vertices and m arcs, with integer lengths on the arcs, each at most W in absolute value. Moreover, $\mathrm{VC}(n,m)$ is the time required to find a minimum-size vertex cover in a bipartite graph with n vertices and m edges.

Dinits [1976] gave an algorithm for finding a minimum-weight matching in $K_{p,q}$ of size p, with time bound $O(|p|^3 + pq)$ (taking $p \le q$).

17.5b. Further notes

Simplex method. Finding a maximum-weight matching in a bipartite graph is a special case of a linear programming problem (see Chapter 18), and hence linear programming methods like the simplex method apply.

[28] Munkres showed that Kuhn's 'Hungarian method' takes $O(n^4)$ time.

Gassner [1964] studied cycling of the simplex method when applied to the assignment problem. Using the 'strongly feasible' trees of Cunningham [1976], Roohy-Laleh [1980] showed that a version of the simplex method solves the assignment problem in less than n^3 pivots (cf. Hung [1983], Orlin [1985], and Akgül [1993]; the last paper gives a method with $O(n^2)$ pivots, yielding an $O(n(m + n \log n))$ algorithm).

Balinski [1985] (cf. Goldfarb [1985]) showed that a version of the dual simplex method (the *signature method*) solves the assignment problem in strongly polynomial time ($O(n^2)$ pivots, yielding an $O(n^3)$ algorithm). More can be found in Dantzig [1963], Barr, Glover, and Klingman [1977], Balinski [1986], Ahuja and Orlin [1988,1992], Akgül [1988], Paparrizos [1988], and Akgül and Ekin [1991].

For further algorithmic studies of the assignment problem, consult Flood [1960], Kurtzberg [1962], Hoffman and Markowitz [1963], Balinski and Gomory [1964], Tabourier [1972], Carpaneto and Toth [1980a,1983,1987], Hung and Rom [1980], Karp [1980], Bertsekas [1981,1987,1992] ('auction method'), Engquist [1982], Avis [1983], Avis and Devroye [1985], Derigs [1985b,1988a], Carraresi and Sodini [1986], Derigs and Metz [1986a], Glover, Glover, and Klingman [1986], Jonker and Volgenant [1986], Kleinschmidt, Lee, and Schannath [1987], Avis and Lai [1988], Bertsekas and Eckstein [1988], Motwani [1989,1994], Kalyanasundaram and Pruhs [1991, 1993], Khuller, Mitchell, and Vazirani [1991,1994], Goldberg and Kennedy [1997] (push-relabel), and Arora, Frieze, and Kaplan [1996,2002].

For computational studies, see Silver [1960], Florian and Klein [1970], Barr, Glover, and Klingman [1977] (simplex method), Gavish, Schweitzer, and Shlifer [1977] (simplex method), Bertsekas [1981], Engquist [1982], McGinnis [1983], Lindberg and Ólafsson [1984] (simplex method), Glover, Glover, and Klingman [1986], Jonker and Volgenant [1987], Bertsekas and Eckstein [1988], and Goldberg and Kennedy [1995] (push-relabel). Consult also Johnson and McGeoch [1993].

Linear-time algorithm for weighted bipartite matching problems satisfying a quadrangle or other inequality were given by Karp and Li [1975], Buss and Yianilos [1994,1998], and Aggarwal, Bar-Noy, Khuller, Kravets, and Schieber [1995].

For generating *all* minimum-weight perfect matchings, see Fukuda and Matsui [1992]. For studies of the 'most vital' edges in a weighted bipartite graph, see Hung, Hsu, and Sung [1993].

Aráoz and Edmonds [1985] gave an example showing that iterative dual improvements in the linear programming problem dual to the assignment problem, need not converge for irrational data.

For the 'bottleneck' assignment problem, see Gross [1959] and Garfinkel [1971]. An algebraic approach to assignment problems was described by Burkard, Hahn, and Zimmermann [1977].

For surveys on matching algorithms, see Galil [1983,1986a,1986b]. Books covering the weighted bipartite matching and assignment problems include Ford and Fulkerson [1962], Dantzig [1963], Christofides [1975], Lawler [1976b], Bazaraa and Jarvis [1977], Burkard and Derigs [1980], Papadimitriou and Steiglitz [1982], Gondran and Minoux [1984], Rockafellar [1984], Derigs [1988a], Bazaraa, Jarvis, and Sherali [1990], Cook, Cunningham, Pulleyblank, and Schrijver [1998], Jungnickel [1999], Mehlhorn and Näher [1999], and Korte and Vygen [2000].

17.5c. Historical notes on weighted bipartite matching and optimum assignment

Monge: optimum assignment

The assignment problem is one of the first studied combinatorial optimization problems. It was investigated by Monge [1784], albeit camouflaged as a continuous problem, and often called a transportation problem.

Monge was motivated by transporting earth, which he considered as the discontinuous, combinatorial problem of transporting molecules:

> Lorsqu'on doit transporter des terres d'un lieu dans un autre, on a coutime de donner le nom de *Déblai* au volume des terres que l'on doit transporter, & le nom de *Remblai* à l'espace qu'elles doivent occuper après le transport.
> Le prix du transport d'une molécule étant, toutes choses d'ailleurs égales, proportionnel à son poids & à l'espace qu'on lui fait parcourir, & par conséquent le prix du transport total devant être proportionnel à la somme des produits des molécules multipliées chacune par l'espace parcouru, il s'ensuit que le déblai & le remblai étant donnés de figure & de position, il n'est pas indifférent que telle molécule du déblai soit transportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine distribution à faire des molécules du premier dans le second, d'après laquelle la somme de ces produits sera la moindre possible, & le prix du transport total sera un *minimum.*[29]

Monge described an interesting geometric method to solve this problem. Consider a line that is tangent to both areas, and move the molecule m touched in the first area to the position x touched in the second area, and repeat, until all earth has been transported. Monge's argument that this would be optimum is simple: if molecule m would be moved to another position, then another molecule should be moved to position x, implying that the two routes traversed by these molecules cross, and that therefore a shorter assignment exists:

> Étant données sur un même plan deux aires égales $ABCD$, & $abcd$, terminées par des contours quelconques, continus ou discontinus, trouver la route que doit suivre chaque molécule M de la premiere, & le point m où elle doit arriver dans la seconde, pour que tous les points étant semblablement transportés, ils replissent exactement la seconde aire, & que la somme des produits de chaque molécule multipliée par l'espace parcouru soit un *minimum.*
> Si par un point M quelconque de la première aire, on mène une droite Bd, telle que le segment BAD soit égal au segment bad, je dis que pour satisfaire à la question, il faut que toutes les molécules du segment BAD, soient portées sur le segment bad, & que par conséquent les molécules du segment BCD soient portées

[29] When one must transport earth from one place to another, one usually gives the name of *Déblai* to the volume of earth that one must transport, & the name of *Remblai* to the space that they should occupy after the transport.

The price of the transport of one molecule being, if all the rest is equal, proportional to its weight & to the distance that one makes it covering, & hence the price of the total transport having to be proportional to the sum of the products of the molecules each multiplied by the distance covered, it follows that, the déblai & the remblai being given by figure and position, it makes difference if a certain molecule of the déblai is transported to one or to another place of the remblai, but that there is a certain distribution to make of the molecules from the first to the second, after which the sum of these products will be as little as possible, & the price of the total transport will be a *minimum.*

sur le segment égal *bcd*; car si un point K quelconque du segment BAD, étoit porté sur un point k de *bcd*, il faudroit nécessairement qu'un point égal L, pris quelque part dans BCD, fût transporté dans un certain point l de *bad*, ce qui ne pourroit pas se faire sans que les routes Kk, Ll, ne se coupassent entre leurs extrémités, & la somme des produits des molécules par les espaces parcourus ne seroit pas un *minimum*. Pareillement, si par un point M' infiniment proche du point M, on mène la droite $B'd'$, telle qu'on ait encore le segment $B'A'D'$, égal au segment $b'a'd'$, il faut pour que la question soit satisfaite, que les molécules du segment $B'A'D'$ soient transportées sur $b'a'd'$. Donc toutes les molécules de l'élément $BB'D'D$ doivent être transportées sur l'élément égal $bb'd'd$. Ainsi en divisant le déblai & le remblai en une infinité d'élémens par des droites qui coupent dans l'un & dans l'autre des segmens égaux entr'eux, chaque élément du déblai doit être porté sur l'élément correspondant du remblai.
Les droites Bd & $B'd'$ étant infiniment proches, il est indifférent dans quel ordre les molécules de l'élément $BB'D'D$ se distribuent sur l'élément $bb'd'd$; de quelque manière en effet que se fasse cette distribution, la somme des produits des molécules par les espaces parcourus, est toujours la même, mais si l'on remarque que dans la pratique il convient de débleyer premièrement les parties qui se trouvent sur le passage des autres, & de n'occuper que les dernières les parties du remblai qui sont dans le même cas; la molécule MM' ne devra se transporter que lorsque toute la partie $MM'D'D$ qui la précêde, aura été transportée en $mm'd'd$; donc dans cette hypothèse, si l'on fait $mm'd'd = MM'D'D$, le point m sera celui sur lequel le point M sera transporté.[30]

Although geometrically intuitive, the method is however not fully correct, as was noted by Appell [1928]:

[30] Being given, in the same plane, two equal areas $ABCD$ & *abcd*, bounded by arbitrary contours, continuous or discontinuous, find the route that every molecule M of the first should follow & the point m where it should arrive in the second, so that, all points being transported likewise, they fill precisely the second area & so that the sum of the products of each molecule multiplied by the distance covered, is *minimum*.

If one draws a straight line Bd through an arbitrary point M of the first area, such that the segment BAD is equal to the segment *bad*, I assert that, in order to satisfy the question, all molecules of the segment BAD should be carried on the segment *bad*, & hence the molecules of the segment BCD should be carried on the equal segment *bcd*; for, if an arbitrary point K of segment BAD, is carried to a point k of *bcd*, then necessarily some point L somewhere in BCD is transported to a certain point l in *bad*, which cannot be done without that the routes Kk, Ll cross each other between their end points, & the sum of the products of the molecules by the distances covered would not be a *minimum*. Likewise, if one draws a straight line $B'd'$ through a point M' infinitely close to point M, in such a way that one still has that segment $B'A'D'$ is equal to segment $b'a'd'$, then in order to satisfy the question, the molecules of segment $B'A'D'$ should be transported to $b'a'd'$. So all molecules of the element $BB'D'D$ must be transported to the equal element $bb'd'd$. Dividing the déblai & the remblai in this way into an infinity of elements by straight lines that cut in the one & in the other segments that are equal to each other, every element of the déblai must be carried to the corresponding element of the remblai.

The straight lines Bd & $B'd'$ being infinitely close, it does not matter in which order the molecules of element $BB'D'D$ are distributed on the element $bb'd'd$; indeed, in whatever manner this distribution is being made, the sum of the products of the molecules by the distances covered is always the same; but if one observes that in practice it is convenient first to dig off the parts that are in the way of others, & only at last to cover similar parts of the remblai; the molecule MM' must be transported only when the whole part $MM'D'D$ that precedes it will have been transported to $mm'd'd$; hence with this hypothesis, if one has $mm'd'd = MM'D'D$, point m will be the one to which point M will be transported.

Il est bien facile de faire la figure de manière que les chemins suivis par les deux parcelles dont parle Monge ne se croisent pas.[31]

(cf. Taton [1951]).

Egerváry

Egerváry [1931] published a weighted version of Kőnig's theorem:

Ha az $\|a_{ij}\|$ n-edrendű matrix elemei adott nem negatív egész számok, úgy a

$$\lambda_i + \mu_j \geq a_{ij}, \qquad (i, j = 1, 2, ...n),$$

(λ_i, μ_j nem negatív egész számok)

feltételek mellett

$$\min . \sum_{k=1}^{n} (\lambda_k + \mu_k) = \max . (a_{1\nu_1} + a_{2\nu_2} + \cdots + a_{n\nu_n}).$$

hol $\nu_1, \nu_2, ...\nu_n$ az $1, 2, ...n$ számok összes permutációit befutják.[32]

The proof method of Egerváry is essentially algorithmic. Assume that the $a_{i,j}$ are integer. Let λ_i^*, μ_j^* attain the minimum. If there is a permutation ν of $\{1, \ldots, n\}$ with $\lambda_i^* + \mu_{\nu_i}^* = a_{i,\nu_i}$ for all i, then this permutation attains the maximum, and we have the required equality. If no such permutation exists, by Frobenius' theorem there are subsets I, J of $\{1, \ldots, n\}$ such that

$$\lambda_i^* + \mu_j^* > a_{i,j} \text{ for all } i \in I, j \in J \tag{17.8}$$

and such that $|I| + |J| = n + 1$. Resetting $\lambda_i^* := \lambda_i^* - 1$ if $i \in I$ and $\mu_j^* := \mu_j^* + 1$ if $j \notin J$, would give feasible values for the λ_i and μ_j, however with their total sum being decreased. This is a contradiction.

Translated into an algorithm, it consists of applying $O(nW)$ times a cardinality bipartite matching algorithm, where W is the maximum weight. So its running time is $O(nW \cdot B(n))$, where $B(n)$ is a bound on the running time of any algorithm finding a maximum-size matching and a minimum-size vertex cover in a bipartite graph with n vertices.

This method forms the basis for the *Hungarian method* of Kuhn [1955b,1956] — see below.

---

[31] It is very easy to make the figure in such a way that the routes followed by the two particles of which Monge speaks, do not cross each other.

[32] *If the elements of the matrix $\|a_{ij}\|$ of order n are given nonnegative integers, then under the assumption*

$$\lambda_i + \mu_j \geq a_{ij}, \qquad (i, j = 1, 2, ...n),$$

(λ_i, μ_j nonnegative integers)

we have

$$\min . \sum_{k=1}^{n} (\lambda_k + \mu_k) = \max . (a_{1\nu_1} + a_{2\nu_2} + \cdots + a_{n\nu_n}).$$

where $\nu_1, \nu_2, ...\nu_n$ run over all possible permutations of the numbers $1, 2, ...n$.

The 1940s

The first algorithm for the assignment problem might have been published by Easterfield [1946], who described his motivation as follows:

> In the course of a piece of organisational research into the problems of demobilisation in the R.A.F., it seemed that it might be possible to arrange the posting of men from disbanded units into other units in such a way that they would not need to be posted again before they were demobilised; and that a study of the numbers of men in the various release groups in each unit might enable this process to be carried out with a minimum number of postings. Unfortunately the unexpected ending of the Japanese war prevented the implications of this approach from being worked out in time for effective use. The algorithm of this paper arose directly in the course of the investigation.

Easterfield seems to have worked without knowledge of the existing literature. He formulated and proved a theorem equivalent to Hall's marriage theorem (see Section 22.1a) and he described a primal-dual type method for the assignment problem from which Egerváry's result given above follows. The idea of the method can be described as follows.

Let $A = (a_{i,j})$ be an $n \times n$ matrix and let for each column index j, I_j be the set of row indices i for which $a_{i,j}$ is minimum among all entries in row i. If the collection $(I_1, \ldots, I_n)$ has a transversal, say $i_1, \ldots, i_n$ (with $i_j \in I_j$), then $i_j \to j$ is an optimum assignment.

If $(I_1, \ldots, I_n)$ has no transversal, let $\mathcal{J}$ be the collection of subsets J of $\{1, \ldots, n\}$ for which $(I_j \mid j \in J)$ has a transversal. Select an inclusionwise minimal set J that is not in $\mathcal{J}$. Then there exists an $\varepsilon > 0$ such that subtracting ε from each entry in each of the columns in J extends $\mathcal{J}$ by (at least) J. (This can be seen using Hall's condition.)

Easterfield described an implementation (including scanning all subsets in lexicographic order), that has running time $O(2^n n^2)$. (This is better than scanning all permutations, which takes time $\Omega(n!)$.) The algorithm was explained again by Easterfield [1960].

Birkhoff [1946] derived from Hall's marriage theorem that each doubly stochastic matrix is a convex combination of permutation matrices. Birkhoff's motivation was:

> Estas matrices son interesantes para la probabilidad, y los cuadrados mágicos son múltiplos escalares de estas matrices.[33]

A breakthrough in solving the assignment problem came when Dantzig [1951a] showed that the assignment problem can be formulated as a linear programming problem that automatically has an integer optimum solution. Indeed, by Birkhoff's theorem, minimizing a linear functional over the set of doubly stochastic matrices (which is a linear programming problem) gives a permutation matrix, being the optimum assignment. So the assignment problem can be solved with the simplex method.

In an address delivered on 9 September 1949 at a meeting of the American Psychological Association at Denver, Colorado, Thorndike [1950] studied the problem of the 'classification' of personnel:

---

[33] These matrices are interesting because of the probability, and the magic squares are scalar multiples of these matrices.

> The past decade, and particularly the war years, have witnessed a great concern about the classification of personnel and a vast expenditure of effort presumably directed towards this end.

He exhibited little trust in mathematicians:

> There are, as has been indicated, a finite number of permutations in the assignment of men to jobs. When the classification problem as formulated above was presented to a mathematician, he pointed to this fact and said that from the point of view of the mathematician there was no problem. Since the number of permutations was finite, one had only to try them all and choose the best. He dismissed the problem at that point. This is rather cold comfort to the psychologist, however, when one considers that only ten men and ten jobs mean over three and a half million permutations. Trying out all the permutations may be a mathematical solution to the problem, it is not a practical solution.

Thorndike next presented three heuristics for the assignment problem, the *Method of Divine Intuition*, the *Method of Daily Quotas*, and the *Method of Predicted Yield*.

In a RAND Report dated 5 December 1949, Robinson [1949] reported that an 'unsuccessful attempt' to solve the traveling salesman problem, led her to the following 'cycle-cancelling' method for the optimum assignment problem.

Let matrix $(a_{i,j})$ be given, and consider any permutation π. Define for all i, j a 'length' $l_{i,j}$ by: $l_{i,j} := a_{j,\pi(i)} - a_{i,\pi(i)}$ if $j \neq \pi(i)$ and $l_{i,\pi(i)} = \infty$. If there exists a negative-length directed circuit, there is a straightforward way to improve π. If there is no such circuit, then π is an optimal permutation.

This clearly is a finite method. Robinson remarked:

> I believe it would be feasible to apply it to as many as 50 points provided suitable calculating equipment is available.

The early 1950s

Von Neumann considered the complexity of the assignment problem. In a talk in the Princeton University Game Seminar on 26 October 1951, he showed that the assignment problem can be reduced to finding an optimum column strategy in a certain zero-sum two-person game, and that it can be found by a method given by Brown and von Neumann [1950]. We give first the mathematical background.

A zero-sum two-person game is given by a matrix A, the 'pay-off matrix'. The interpretation as a game is that a 'row player' chooses a row index i and a 'column player' chooses simultaneously a column index j. After that, the column player pays the row player $A_{i,j}$. The game is played repeatedly, and the question is what is the best strategy.

Let A have order $m \times n$. A *row strategy* is a vector $x \in \mathbb{R}^m_+$ satisfying $\mathbf{1}^\mathsf{T} x = 1$. Similarly, a *column strategy* is a vector $y \in \mathbb{R}^n_+$ satisfying $\mathbf{1}^\mathsf{T} y = 1$. Then

$$\max_x \min_j (x^\mathsf{T} A)_j = \min_y \max_i (Ay)_i, \tag{17.9}$$

where x ranges over row strategies, y over column strategies, i over row indices, and j over column indices. Equality (17.9) follows from LP-duality.

It implies that the best strategy for the row player is to choose rows with distribution an optimum x in (17.9). Similarly, the best strategy for the column player is to choose columns with distribution an optimum y in (17.9). The average pay-off then is the value of (17.9).

The method of Brown [1951] to determine the optimum strategies is that each player chooses in turn the line that is best with respect to the distribution of the lines chosen by the opponent so far. It was proved by Robinson [1951] that this converges to optimum strategies. The method of Brown and von Neumann [1950] is a continuous version of this, and amounts to solving a system of linear differential equations.

Now von Neumann noted that the following reduces the assignment problem to the problem of finding an optimum column strategy. Let $C = (c_{i,j})$ be an $n \times n$ cost matrix, as input for the assignment problem. We may assume that C is positive. Consider the following pay-off matrix A, of order $2n \times n^2$, with columns indexed by ordered pairs (i, j) with $i, j = 1, \ldots, n$. The entries of A are given by: $A_{i,(i,j)} := 1/c_{i,j}$ and $A_{n+j,(i,j)} := 1/c_{i,j}$ for $i, j = 1, \ldots, n$, and $A_{k,(i,j)} := 0$ for all i, j, k with $k \neq i$ and $k \neq n + j$. Then any minimum-cost assignment, of cost γ say, yields an optimum column strategy y by: $y_{(i,j)} := c_{i,j}/\gamma$ if i is assigned to j, and $y_{(i,j)} := 0$ otherwise. Any optimum column strategy is a convex combination of strategies obtained this way from optimum assignments. So an optimum assignment can in principle be found by finding an optimum column strategy.

According to a transcript of the talk (cf. von Neumann [1951,1953]), von Neumann noted the following on the number of steps:

> It turns out that this number is a moderate power of n, i.e., considerably smaller than the "obvious" estimate n! mentioned earlier.

However, no further argumentation is given. (Related observations were given by Dulmage and Halperin [1955] and Koopmans and Beckmann [1955,1957].)

Beckmann and Koopmans [1952] studied the quadratic assignment problem, and they noted that the traveling salesman problem is a special case. In a Cowles Commission Discussion Paper of 2 April 1953, Beckmann and Koopmans [1953] mentioned applying polyhedral methods to solve the assignment problem, and noted:

> It should be added that in all the assignment problems discussed, there is, of course, the obvious brute force method of enumerating all assignments, evaluating the maximand at each of these, and selecting the assignment giving the highest value. This is too costly in most cases of practical importance, and by a method of solution we have meant a procedure that reduces the computational work to manageable proportions in a wider class of cases.

Geometric methods were proposed by Lord [1952] and Dwyer [1954] (the 'method of optimal regions') and other heuristics by Votaw and Orden [1952] and Törnqvist [1953]. A survey of developments on the assignment problem until 1955 was given by Motzkin [1956].

Computational results of the early 1950s

In a paper presented at the Symposium on Linear Inequalities and Linear Programming (14–16 June 1951 in Washington, D.C.), Votaw and Orden [1952] mentioned that solving a 10×10 transportation problem took 3 minutes on the SEAC (National Bureau of Standards Eastern Automatic Computer). However, in a later paper (submitted 1 November 1951), Votaw [1952] said that solving a 10×10 assignment problem with the simplex method on the SEAC took 20 minutes.

Moreover, in his reminiscences, Kuhn [1991] mentioned:

> The story begins in the summer of 1953 when the National Bureau of Standards and other US government agencies had gathered an outstanding group of combinatorialists and algebraists at the Institute for Numerical Analysis (INA) located on the campus of the University of California at Los Angeles. Since space was tight, I shared an office with Ted Motzkin, whose pioneering work on linear inequalities and related systems predates linear programming by more than ten years. A rather unique feature of the INA was the presence of the Standards Western Automatic Computer (SWAC), the entire memory of which consisted of 256 Williamson cathode ray tubes. The SWAC was faster but smaller than its sibling machine, the Standards Eastern Automatic Computer (SEAC), which boasted a liquid mercury memory and which had been coded to solve linear programs.
>
> During the summer, C.B. Tompkins was attempting to solve 10 by 10 assignment problems by programming the SWAC to enumerate the $10! = 3,628,800$ permutations of 10 objects. He never succeeded in this project.
>
> Thus, the 10 by 10 assignment problem is a linear program with 100 nonnegative variables and 20 equation constraints (of which only 19 are needed). In 1953, there was no machine in the world that had been programmed to solve a linear program this large!

If 'the world' includes the Eastern Coast of the U.S.A., there seems to be some discrepancy with the remarks of Votaw [1952] mentioned above.

On 23 April 1954, Gleyzal [1955] wrote that a code of his algorithm for the transportation problem, for the special case of the assignment problem with an 8×8 matrix, had just been composed for the SWAC.

Tompkins [1956] mentioned the following 'branch-and-bound' approach to the assignment problem:

> Benjamin Handy, on the suggestion of D.H. Lehmer and with advice from T.S. Motzkin [1], coded this problem for SWAC; he used exhaustive search including rejection of blocks of permutations when the first few elements of the trace led to a hopelessly low contribution. The problem worked for a problem whose matrix had 12 rows and 12 columns and was composed of random three-digit numbers. The solution in this case took three hours. Some restrictions which had been imposed concerning the types of problems to which the code should be applicable led to some inefficiencies; however, the simplex method of G.B. Dantzig [7] and various other methods of solution of this problem seem greatly superior to this method of exhaustive search;

(References [1] and [7] in this quotation are Motzkin [1956] and Dantzig [1951b].)

Kuhn, Munkres: the Hungarian method

Kuhn [1955b,1956] developed a new combinatorial procedure for solving the assignment problem. The method is based on the work of Egerváry [1931], and therefore Kuhn introduced the name *Hungarian method* for it. (According to Kuhn [1955b], the algorithm is 'latent in work of D. Kőnig and J. Egerváry'.) The method was sharpened by Munkres [1957].

In an article *On the origin of the Hungarian method*, Kuhn [1991] presented the following reminiscences on the Hungarian method, from the time starting Summer 1953:

> During this period, I was reading Kőnig's classical book on the theory of graphs and realized that the matching problem for a bipartite graph on two sets of n vertices was exactly the same as an n by n assignment problem with all $a_{ij} = 0$ or 1. More significantly, Kőnig had given a combinatorial algorithm (based on augmenting paths) that produces optimal solutions to the matching problem and its combinatorial (or linear programming) dual. In one of the several formulations given by Kőnig (p. 240, Theorem D), given an n by n matrix $A = (a_{ij})$ with all $a_{ij} = 0$ or 1, the maximum number of 1's that can be chosen with no two in the same line (horizontal row or vertical column) is equal to the minimum number of lines that contain all of the 1's. Moreover, the algorithm seemed to be 'good' in a sense that will be made precise later. The problem then was: how could the general assignment problem be reduced to the 0-1 special case?
> Reading Kőnig's book more carefully, I was struck by the following footnote (p. 238, footnote 2): "... Eine Verallgemeinerung dieser Sätze gab Egerváry, Matrixok kombinatorius tulajdonságairól (Über kombinatorische Eigenschaften von Matrizen), Matematikai és Fizikai Lapok, *38*, 1931, S. 16-28 (ungarisch mit einem deutschen Auszug) ..." This indicated that the key to the problem might be in Egerváry's paper. When I returned to Bryn Mawr College in the fall, I obtained a copy of the paper together with a large Hungarian dictionary and grammar from the Haverford College library. I then spent two weeks learning Hungarian and translated the paper [1]. As I had suspected, the paper contained a method by which a general assignment problem could be reduced to a finite number of 0-1 assignment problems.
> Using Egerváry's reduction and Kőnig's maximum matching algorithm, in the fall of 1953 I solved several 12 by 12 assignment problems (with 3-digit integers as data) by hand. Each of these examples took under two hours to solve and I was convinced that the combined algorithm was 'good'. This must have been one of the last times when pencil and paper could beat the largest and fastest electronic computer in the world.

(Reference [1] is the English translation of the paper of Egerváry [1931].)

The method described by Kuhn is a sharpening of the method of Egerváry sketched above, in two respects: (i) it gives an (augmenting path) method to find either a perfect matching or sets I and J as required, and (ii) it improves the λ_i and μ_j not by 1, but by the largest value possible.

Kuhn [1955b] described the method in terms of matrices — in terms of graphs it amounts to the following algorithm for the maximum weighted perfect matching problem in a complete bipartite graph $G = (V, E)$, with weight function $w : E \to \mathbb{Z}_+$. Let U and W be the colour classes of G. Throughout there is a function $p : V \to \mathbb{Z}$ satisfying

(17.10) $\qquad p(u) + p(v) \geq w(uv)$ for each edge uv

and a matching M in the subgraph $G' = (V, E')$ of G consisting of those edges having equality in (17.10).

If M is not a perfect matching, orient each edge in M from W to U, and every other edge of G' from U to W, giving graph D'_M. Let U_M and W_M be the sets of vertices in U and W missed by M.

Kuhn [1955b] described a depth-first search to find the set R_M of vertices that are reachable by a directed path in D'_M from U_M. (In a subsequent paper, Kuhn [1956] described a breadth-first search, starting at only one vertex in U_M.)

Case 1: $R_M \cap W_M \neq \emptyset$. We have an M-augmenting path in G', by which we increase M.

Case 2: $R_M \cap U_M = \emptyset$. Determine

(17.11) $\mu := \min\{p(u) + p(v) - w(u,v) \mid u \in U \cap R_M, v \in W \setminus R_M\}.$

This number is positive, since no edge of G' connects $U \cap R_M$ and $W \setminus R_M$. Decrease $p(u)$ by μ if $u \in U \cap R_M$ and increase $p(v)$ by μ if $v \in W \cap R_M$. Then (17.10) is maintained, while the sum $\sum_{v \in V} p(v)$ decreases (as $|U \cap R_M| > |W \cap R_M|$).

After this we iterate, until we have a perfect matching M in G', which is a maximum-weight perfect matching.

Kuhn [1955b] contented himself with stating that the number of iterations is finite (since the number of iterations where Case 2 applies is finite (as $\sum_v p(v)$ is nonnegative)).

It was observed by Munkres [1957] that the method runs in strongly polynomial time, since, between any two occurrences of Case 1, the number of iterations where Case 2 applies is at most n, as at each such iteration $R_M \cap W$ increases (namely by all vertices v that attain the minimum (17.11)).

So the number of iterations is at most n^2 (since M can increase at most n times). As the (depth- or breadth-first) search takes $O(n^2)$ this gives an $O(n^4)$ algorithm.

Munkres [1957] observed also that after an occurrence of Case 2 one can continue the search of the previous iteration, since edges of G' traversed in the search from U_M, remain edges of the new graph G'. Hence between any two occurrences of Case 1, the depth-first search takes time $O(n^2)$. This still gives an $O(n^4)$ algorithm, since calculating the minimum (17.11) takes $O(n^2)$ time. (Munkres claimed that his algorithm takes $O(n^3)$ operations, but he takes 'scanning a line' (that is, considering all edges incident with a given vertex) as one operation.)

(However, all Case 2-iterations can be combined to one iteration, by finding distances from U_M, with respect to the length function w in the oriented G'. It amounts to including a Dijkstra-like labeling, yielding an $O(n^3)$ time bound. This is the method we described in Section 17.2. This principle was noticed by Edmonds and Karp [1970] and Tomizawa [1971].)

Ford and Fulkerson [1955,1957b] (cf. Ford and Fulkerson [1956c,1956d]) extended the Hungarian method to general transportation problems. They state in Ford and Fulkerson [1956c,1956d]:

> Large systems involving hundreds of equations in thousands of unknowns have been successfully solved by hand using the simplex computation. The procedure of this paper has been compared with the simplex method on a number of randomly chosen problems and has been found to take roughly half the effort for small problems. We believe that as the size of the problem increases, the advantages of the present method become even more marked.

In a footnote, the authors add as to the assignment problem:

> The largest example tried was a 20×20 optimal assignment problem. For this example, the simplex method required well over an hour, the present method about thirty minutes of hand computation.

Chapter 18

Linear programming methods and the bipartite matching polytope

The weighted matching problem for bipartite graphs discussed in the previous chapter is related to the 'matching polytope' and the 'perfect matching polytope', and can be handled with linear programming methods by the total unimodularity of the incidence matrix of a bipartite graph.
In this chapter, graphs can be assumed to be simple.

18.1. The matching and the perfect matching polytope

Let $G = (V, E)$ be a graph. The *perfect matching polytope* $P_{\text{perfect matching}}(G)$ of G is defined as the convex hull of the incidence vectors of perfect matchings in G. So $P_{\text{perfect matching}}(G)$ is a polytope in $\mathbb{R}^E$.

The perfect matching polytope is a polyhedron, and hence can be described by linear inequalities. The following are clearly valid inequalities:

$$\text{(18.1)} \qquad \begin{array}{rll} \text{(i)} & x_e \geq 0 & \text{for each edge } e, \\ \text{(ii)} & x(\delta(v)) = 1 & \text{for each vertex } v. \end{array}$$

These inequalities are generally not enough (for instance, not for K_3). However, as Birkhoff [1946] showed, for bipartite graphs they are enough:

Theorem 18.1. *If G is bipartite, the perfect matching polytope of G is determined by* (18.1).

Proof. Let x be a vertex of the polytope determined by (18.1). Let F be the set of edges e with $x_e > 0$. Suppose that F contains a circuit C. As C has even length, $EC = M \cup N$ for two disjoint matchings M and N. Then for ε close enough to 0, both $x + \varepsilon(\chi^M - \chi^N)$ and $x - \varepsilon(\chi^M - \chi^N)$ satisfy (18.1), contradicting the fact that x is a vertex of the polytope. So (V, F) is a forest, and hence by (18.1), F is a perfect matching. ∎

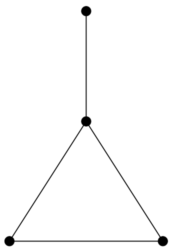


Figure 18.1

The implication cannot be reversed, as is shown by the graph in Figure 18.1.

Theorem 18.1 was shown by Birkhoff in the terminology of doubly stochastic matrices. A matrix A is called *doubly stochastic* if A is nonnegative and each row sum and each column sum equals 1. A *permutation matrix* is an integer doubly stochastic matrix (so it is $\{0,1\}$-valued, and has precisely one 1 in each row and in each column). Then:

Corollary 18.1a (Birkhoff's theorem)**.** *Each doubly stochastic matrix is a convex combination of permutation matrices.*

Proof. Directly from Theorem 18.1, by taking $G = K_{n,n}$. ∎

Theorem 18.1 also implies a characterization of the matching polytope for bipartite graphs. For any graph $G = (V, E)$, the *matching polytope* $P_{\text{matching}}(G)$ of G is the convex hull of the incidence vectors of matchings in G. So again it is a polytope in $\mathbb{R}^E$. The following are valid inequalities for the matching polytope:

$$(18.2) \qquad \begin{array}{rll} \text{(i)} & x_e \geq 0 & \text{for each edge } e, \\ \text{(ii)} & x(\delta(v)) \leq 1 & \text{for each vertex } v. \end{array}$$

Then:

Corollary 18.1b. *The matching polytope of G is determined by* (18.2) *if and only if G is bipartite.*

Proof. To see necessity, suppose that G is not bipartite, and let C be an odd circuit in G. Define $x_e := \frac{1}{2}$ if $e \in C$ and $x_e := 0$ otherwise. Then x satisfies (18.2) but does not belong to the matching polytope of G.

To see sufficiency, let G be bipartite and let x satisfy (18.2). Let G' and x' be a copy of G and x, and add edges vv', where v' is the copy of $v \in V$. Define $y(vv') := 1 - x(\delta(v))$. Then x, x', y satisfy (18.1) with respect to the new graph, and hence by Theorem 18.1, it is a convex combination of incidence vectors of perfect matchings in the new graph. Hence x is a convex combination of incidence vectors of matchings in G. ∎

Notes. Birkhoff derived Corollary 18.1a from Hall's marriage theorem (Theorem 22.1), which is equivalent to Kőnig's matching theorem. (Also Dulmage and Halperin [1955] derived Birkhoff's theorem from Kőnig's matching theorem.) Other proofs were given by von Neumann [1951,1953], Dantzig [1952], Hoffman and Wielandt [1953], Koopmans and Beckmann [1955,1957], Hammersley and Mauldon [1956] (a polyhedral proof based on total unimodularity), Tompkins [1956], Mirsky [1958], and Vogel [1961]. A survey was given by Mirsky [1962]. More can be found in Johnson, Dulmage, and Mendelsohn [1960], Nishi [1979], and Brualdi [1982].

18.2. Totally unimodular matrices from bipartite graphs

In this section we show that the results on matchings discussed above can also be derived from linear programming duality with total unimodularity (Hoffman [1956b]).

Let A be the $V \times E$ incidence matrix of a graph $G = (V, E)$. The matrix A generally is not totally unimodular. E.g., if G is the complete graph K_3 on three vertices, then the determinant of A is equal to $+2$ or -2.

However, the following can be proved (necessity can also be derived directly from the total unimodularity of the incidence matrix of a directed graph (Theorem 13.9) — we give a direct proof):

Theorem 18.2. *A graph $G = (V, E)$ is bipartite if and only if its incidence matrix A is totally unimodular.*

Proof. *Sufficiency.* Assume that A is totally unimodular and G is not bipartite. Then G has a circuit of odd length, t say. The submatrix of A induced by the vertices and edges in C is a $t \times t$ matrix with exactly two ones in each row and each column. As t is odd, the determinant of this matrix is ± 2, contradicting the total unimodularity of A.

Necessity. Let G be bipartite. We show that A is totally unimodular. Let B be a square submatrix of A, of order $t \times t$ say. We show that $\det B$ equals 0 or ± 1 by induction on t. If $t = 1$, the statement is trivial. So let $t > 1$. We distinguish three cases.

Case 1: B has a column with only 0*'s.* Then $\det B$=0.

Case 2: B has a column with exactly one 1. In that case we can write (possibly after permuting rows or columns):

$$(18.3) \qquad B = \begin{pmatrix} 1 & b^{\mathsf{T}} \\ \mathbf{0} & B' \end{pmatrix},$$

for some matrix B' and vector b, where $\mathbf{0}$ denotes the all-zero vector in $\mathbb{R}^{t-1}$. By the induction hypothesis, $\det B' \in \{0, \pm 1\}$. Hence, by (18.3), $\det B \in \{0, \pm 1\}$.

Case 3. Each column of B contains exactly two 1's. Then, since G is bipartite, we can write (possibly after permuting rows):

(18.4) $$B = \begin{pmatrix} B' \\ B'' \end{pmatrix},$$

in such a way that each column of B' contains exactly one 1 and each column of B'' contains exactly one 1. So adding up all rows in B' gives the all-one vector, and also adding up all rows in B'' gives the all-one vector. The rows of B therefore are linearly dependent, and hence $\det B{=}0$. ∎

18.3. Consequences of total unimodularity

Let $G = (V, E)$ be a bipartite graph and let A be its $V \times E$ incidence matrix. Consider Kőnig's matching theorem (Theorem 16.2): the maximum size of a matching in G is equal to the minimum size of a vertex cover in G. This can be derived from the total unimodularity of A as follows. By Corollary 5.20a, both optima in the LP-duality equation

(18.5) $$\max\{\mathbf{1}^{\mathsf{T}}x \mid x \geq \mathbf{0}, Ax \leq \mathbf{1}\} = \min\{y^{\mathsf{T}}\mathbf{1} \mid y \geq \mathbf{0}, y^{\mathsf{T}}A \geq \mathbf{1}^{\mathsf{T}}\}$$

have integer optimum solutions x^* and y^*. Now x^* necessarily is the incidence vector of a matching and y^* is the incidence vector of a vertex cover. So we have Kőnig's matching theorem.

One can also derive the weighted version of Kőnig's matching theorem, Egerváry's theorem (Theorem 17.1): for any weight function $w : E \to \mathbb{Z}_+$, the maximum weight of a matching in G is equal to the minimum value of $\sum_{v \in V} y_v$, where y ranges over all $y : V \to \mathbb{Z}_+$ with $y_u + y_v \geq w_e$ for each edge $e = uv$ of G. To derive this, consider the LP-duality equation

(18.6) $$\max\{w^{\mathsf{T}}x \mid x \geq \mathbf{0}, Ax \leq \mathbf{1}\} = \min\{y^{\mathsf{T}}\mathbf{1} \mid y \geq \mathbf{0}, y^{\mathsf{T}}A \geq w^{\mathsf{T}}\}.$$

By the total unimodularity of A, these optima are attained by integer x^* and y^*, and we have the theorem.

The min-max relation for minimum-weight *perfect* matching (Theorem 17.5) follows similarly.

One can also derive the characterizations of the matching polytope and perfect matching polytope of a bipartite graph (Theorem 18.1 and Corollary 18.1b) from the total unimodularity of the incidence matrix of a bipartite graph. This amounts to the fact that the polyhedra

(18.7) $$\{x \mid x \geq \mathbf{0}, Ax \leq \mathbf{1}\}$$

and

(18.8) $$\{x \mid x \geq \mathbf{0}, Ax = \mathbf{1}\}$$

are integer polyhedra, by the total unimodularity of A.

18.4. The vertex cover polytope

One can similarly derive, from the total unimodularity, a description of the vertex cover polytope of a bipartite graph. The *vertex cover polytope* of a graph G is the convex hull of the incidence vectors of vertex covers. It is a polytope in $\mathbb{R}^V$.

For bipartite graphs, it is determined by:

$$\text{(18.9)} \qquad \begin{array}{rll} \text{(i)} & 0 \leq y_v \leq 1 & \text{for each } v \in V, \\ \text{(ii)} & y_u + y_v \geq 1 & \text{for each } e = uv \in E. \end{array}$$

In fact, this characterizes bipartiteness:

Theorem 18.3. *A graph G is bipartite if and only if the vertex cover polytope of G is determined by* (18.9).

Proof. Necessity follows from the total unimodularity of the incidence matrix of A (Theorem 18.2). Sufficiency can be seen as follows. Suppose that G contains an odd circuit C. Define $y_v := \frac{1}{2}$ for each $v \in V$. Then y satisfies (18.9) but does not belong to the vertex cover polytope, as each vertex cover contains more than $\frac{1}{2}|VC|$ vertices in C. ∎

The total unimodularity of A also yields descriptions of the edge cover and stable set polytopes of a bipartite graph — see Section 19.5.

18.5. Further results and notes

18.5a. Derivation of Kőnig's matching theorem from the matching polytope

We note here that Kőnig's matching theorem quite easily follows from description (18.2) of the matching polytope of a bipartite graph.

Since the matching polytope of a bipartite graph $G = (V, E)$ is determined by (18.2), the maximum size of a matching in G is equal to the minimum value of $\sum_{v \in V} y_v$ where $y_v \geq 0$ ($v \in V$) such that $y_u + y_v \leq 1$ for each edge $e = uv$.

Now consider any vertex u with $y_u > 0$. Then by complementary slackness, each maximum-size matching covers u. That is, we have (16.5), which (as we saw) directly implies Kőnig's matching theorem, by applying induction to $G - u$.

18.5b. Dual, primal-dual, primal?

The Hungarian method is considered as the first so-called 'primal-dual' method. It maintains a feasible dual solution, and tries to build up a feasible primal solution fulfilling the complementary slackness conditions. We will show that in a certain sense the method can also be considered as just dual or just primal.

We consider the problem of finding a minimum-weight perfect matching in a bipartite graph $G = (V, E)$, with weight function $w : E \to \mathbb{Q}_+$. Let U and W be the colour classes of G, with $|U| = |W|$. The corresponding LP-duality equation is

(18.10) $$\min\{w^{\mathsf{T}}x \mid x \geq \mathbf{0}, Ax = \mathbf{1}\} = \max\{y^{\mathsf{T}}\mathbf{1} \mid y^{\mathsf{T}}A \leq w^{\mathsf{T}}\},$$

where A is the $V \times E$ incidence matrix of G.

To describe the Hungarian method as a purely dual method one can start with $y = \mathbf{0}$. So y satisfies

(18.11) $$y_u + y_v \leq w_e$$

for each edge $e = uv$ of G. Consider the subset

(18.12) $$F := \{e = uv \in E \mid y_u + y_v = w_e\}$$

of E. If F contains a perfect matching M, then M is a minimum-weight perfect matching, by complementary slackness applied to (18.10). If F contains no perfect matching, by Frobenius' theorem (Corollary 16.2a) there exist $U' \subseteq U$ and $W' \subseteq W$ such that each edge in F intersecting U' also intersects W' and such that $|W'| < |U'|$. Now we can reset

(18.13) $$y_v := \begin{cases} y_v + \alpha & \text{if } v \in U', \\ y_v - \alpha & \text{if } v \in W', \end{cases}$$

choosing α as large as possible while maintaining (18.11). That is, α is equal to the minimum of $w_e - y_u - y_v$ over all edges $e = uv \in E$ with $u \in U'$ and $v \notin W'$. So $\alpha > 0$, and hence $y^{\mathsf{T}}\mathbf{1}$ increases. After that we iterate.

Described in this way it is a purely *dual method*, since only in the last iteration we see a primal solution. In each iteration we test the existence of a perfect matching from scratch. We could, however, remember our work of the previous iteration in our search for a perfect matching in F.

To this end, we keep at any iteration a maximum-size matching M in F. Let D_M be the directed graph obtained from (V, F) by orienting each edge in M from W to U and each edge in $F \setminus M$ from U to W. Let U_M and W_M be the set of vertices in U and W, respectively, missed by M. We also keep, throughout the iterations, the set R_M of vertices reachable in D_M from U_M.

Then we can take $U' := U \cap R_M$ and $W' := W \cap R_M$. Resetting (18.13) of y increases R_M, since at least one edge connecting U' and $W \setminus W'$ is added to F, while all edges in F that were contained in $U' \cup W'$ remain in F. So after at most n iterations, R_M contains a vertex in W_M, in which case we can augment M.

Described in this way it is a *primal-dual method*. Throughout the iterations we keep a feasible dual solution y and a partially feasible primal solution M.

We could however combine all updates of y, between any two augmentations of M, by taking $l_e := w_e - y_u - y_v$ as a length function, and by determining, for each vertex v, the distance $d(v)$ from v to W_M in D_M with respect to length function l. Resetting

(18.14) $$y_v := \begin{cases} y_v + d(v) & \text{if } v \in U, \\ y_v - d(v) & \text{if } v \in W, \end{cases}$$

maintains (18.10), while the new F contains an M-augmenting path (namely, any shortest $U_M - W_M$ path in D_M). Note that this updating of y is the same as the aggregated updating of y (in (18.13)) between any two matching augmentations.

This still is a primal-dual method, since we keep sequences of vectors y and matchings M. It enables us to apply Dijkstra's method to find the distances and the shortest path, since the length function l is nonnegative. We can however do without y, at the cost of an increase in the complexity, since we then must use the Bellman-Ford method (like in our description in Section 17.2). We can use this method since D_M has no negative-length directed circuit, because M is an extreme matching (that is, a matching of minimum weight among all matchings M' with $|M'| = |M|$).

Indeed, we can define the length function l by $l_e := w_e$ if $e \in E \setminus M$ and $l_e := -w_e$ if $e \in M$. Then D_M has no negative-length directed circuits. Any shortest $U_M - W_M$ path is an M-augmenting path yielding an extreme matching M' with $|M'| = |M| + 1$.

Described in this way we have a purely *primal method*, since we keep no vector $y \in \mathbb{Q}^V$ anymore.

18.5c. Adjacency and diameter of the matching polytope

Clearly, for each perfect matching M, the incidence vector χ^M is a vertex of the perfect matching polytope. Adjacency is also easily characterized (Balinski and Russakoff [1974]):

Theorem 18.4. *Let M and N be perfect matchings in a graph $G = (V, E)$. Then χ^M and χ^N are adjacent vertices of the perfect matching polytope if and only if $M \triangle N$ is a circuit.*

Proof. To see necessity, let χ^M and χ^N be adjacent. Then $M \triangle N$ is the vertex-disjoint union of circuits $C_1, \ldots, C_k$. If $k = 1$ we are done so assume $k \geq 2$. Let $M' := M \triangle C_1$ and $N' := N \triangle C_1$. Then $\frac{1}{2}(\chi^M + \chi^N) = \frac{1}{2}(\chi^{M'} + \chi^{N'})$. This contradicts the adjacency of χ^M and χ^N.

To see sufficiency, define a weight function $w : E \to \mathbb{R}$ by $w_e := 0$ if $e \in M \cup N$ and $w_e := 1$ otherwise. Then M and N are the only two perfect matchings in G of minimum weight. Hence χ^M and χ^N are adjacent. ∎

This gives for the diameter:

Corollary 18.4a. *The perfect matching polytope of a graph $G = (V, E)$ has diameter at most $\frac{1}{2}|V|$. If G is simple, the diameter is at most $\frac{1}{4}|V|$.*

Proof. Let M and N be perfect matchings of G. Let $M \triangle N$ be the vertex-disjoint union of circuits $C_1, \ldots, C_k$. Define $M_i := M \triangle (C_1 \cup \cdots \cup C_i)$, for $i = 0, \ldots, k$. Then $M = M_0$, $N = M_k$, and M_i and M_{i+1} give adjacent vertices of the perfect matching polytope of G (by Theorem 18.4). As each C_i has at least two vertices, we have $k \leq \frac{1}{2}|V|$. If G is simple, each C_i has at least four vertices, and hence $k \leq \frac{1}{4}|V|$. ∎

For complete bipartite graphs, this bound can be strengthened. The *assignment polytope* is the perfect matching polytope of a complete bipartite graph $K_{n,n}$. So in matrix terms, it is the polytope of the $n \times n$ doubly stochastic matrices. Balinski and Russakoff [1974] showed:

Theorem 18.5. *The diameter of the assignment polytope is* 2 *(if $n \geq 4$).*

Proof. Let U and W be the two colour classes of $K_{n,n}$. Let M and N be two distinct perfect matchings in $K_{n,n}$. Assume that $M \neq N$ and that M and N are not adjacent. Let $M \triangle N$ be the vertex-disjoint union of the circuits $C_1, \ldots, C_k$. As M and N are not adjacent, $k \geq 2$. For each $i = 1, \ldots, k$, choose an edge $u_i w_i \in C_i \cap M$, with $u_i \in U$ and $w_i \in W$. Let C be the circuit

(18.15) $$C := \{u_1w_1, u_2w_1, u_2w_2, u_3w_2, \ldots, u_nw_n, u_1w_n\}$$

and let $L := M \triangle C$. As $M \triangle L = C$, L is a perfect matching adjacent to M. Now L is adjacent also to N as well, since $N \triangle L = (C_1 \cup \cdots \cup C_k) \triangle C$, which is a circuit. ∎

Naddef [1982] characterized the dimension of the perfect matching polytope of a bipartite graph (cf. Lovász and Plummer [1986]):

Theorem 18.6. *Let $G = (V, E)$ be a bipartite graph with at least one perfect matching. Then the dimension of the perfect matching polytope of G is equal to $|E_0| - |V| + k$, where E_0 is the set of edges contained in at least one perfect matching and where k is the number of components of the graph (V, E_0).*

Proof. It is easy to see that we may assume that $E_0 = E$ and that G is connected and has at least four vertices. Let T be the edge set of a spanning tree in G. So $|E \setminus T| = |E| - |V| + 1$. Now for any $x \in P_{\text{perfect matching}}(G)$, the values x_e with $e \in T$ are determined by the values x_e with $e \in E \setminus T$. Hence $\dim(P_{\text{perfect matching}}(G)) \leq |E \setminus T| = |E| - |V| + 1$.

To see the reverse inequality, choose a vector x in the relative interior of $P_{\text{perfect matching}}(G)$. So $0 < x_e < 1$ for each $e \in E$ (as each edge is contained in some perfect matching and is missed by some perfect matching). Then any small enough change of x_e for any $e \in E \setminus T$ can be corrected by changing values of $x(e')$ with $e' \in T$. Therefore $\dim(P_{\text{perfect matching}}(G)) \geq |E \setminus T|$. ∎

Rispoli [1992] showed that the 'monotonic diameter' (that is, the maximum length of a shortest path on the polytope where a given objective function is monotonically increasing) of the assignment polytope is equal to $\lfloor \frac{n}{2} \rfloor$. More can be found in Balinski and Russakoff [1974], Padberg and Rao [1974], Brualdi and Gibson [1976,1977a,1977b,1977c], Roohy-Laleh [1980], Hung [1983], Balinski [1985], and Goldfarb [1985].

18.5d. The perfect matching space of a bipartite graph

The *perfect matching space* of a graph $G = (V, E)$ is the linear hull of the incidence vectors of perfect matchings:

(18.16) $$S_{\text{perfect matching}}(G) := \text{lin.hull}\{\chi^M \mid M \text{ perfect matching in } G\}.$$

(Here lin.hull denotes linear hull.)

Note that Theorem 18.6 directly implies the dimension of the perfect matching space of a bipartite graph:

Corollary 18.6a. *Let $G = (V, E)$ be a bipartite graph with at least one perfect matching. Then the dimension of the perfect matching space of G is equal to $|E_0| -$*

$|V|+k+1$, where E_0 is the set of edges contained in at least one of perfect matching, and where k is the number of components of the graph (V, E_0).

Proof. The dimension of the perfect matching space is 1 more than the dimension of the perfect matching polytope (as $\mathbf{0}$ does not belong to the affine hull of the incidence vectors of perfect matchings). So the Corollary follows from Theorem 18.6. ∎

With the help of the description of the perfect matching polytope we can similarly describe the perfect matching space in terms of equations:

Theorem 18.7. *The perfect matching space of a bipartite graph $G = (V, E)$ is equal to the set of vectors $x \in \mathbb{R}^E$ such that*

(18.17) (i) $x_e = 0$ if e is contained in no perfect matching,
(ii) $x(\delta(u)) = x(\delta(v))$ for all $u, v \in V$.

Proof. (18.17) clearly is a necessary condition for each vector x in the perfect matching space. To see sufficiency, let $x \in \mathbb{R}^E$ satisfy (18.17). We can assume that G has at least one perfect matching.

By adding sufficiently many incidence vectors of perfect matchings to x, we can achieve that $x_e \geq 0$ for all $e \in E$. By scaling we can achieve that $x(\delta(v)) = 1$ for each $v \in V$. Then x belongs to the perfect matching polytope of G, and hence to the perfect matching space. ∎

This theorem has as direct consequence a characterization of the linear space orthogonal to the perfect matching space:

Corollary 18.7a. *Let $G = (V, E)$ be a bipartite graph and let $w \in \mathbb{R}^E$. Then $w(M) = 0$ for each perfect matching M if and only if there exists a vector $b \in \mathbb{R}^V$ with $b(V) = 0$ such that $w_e = b_u + b_v$ for each edge $e = uv$ contained in at least one perfect matching.*

Proof. Directly by orthogonality from Theorem 18.7. ∎

18.5e. Up and down hull of the perfect matching polytope

Fulkerson [1970b] studied the up hull of the perfect matching polytope of a graph $G = (V, E)$, that is,

(18.18) $P^{\uparrow}_{\text{perfect matching}}(G) = P_{\text{perfect matching}}(G) + \mathbb{R}^E_+$.

Any x in this polyhedron satisfies:

(18.19) (i) $x_e \geq 0$ for each $e \in E$,
(ii) $x(E[S]) \geq |S| - \frac{1}{2}|V|$ for each $S \subseteq V$.

Here $E[S]$ denotes the set of edges spanned by S. Inequality (18.19)(ii) follows from the fact that any perfect matching M has at most $|V \setminus S|$ edges not contained in S, and hence at least $\frac{1}{2}|V| - |V \setminus S| = |S| - \frac{1}{2}|V|$ edges contained in S.

Fulkerson [1970b] showed that for bipartite graphs these inequalities are enough to characterize polyhedron (18.18):

Theorem 18.8. *If G is bipartite, then $P^{\uparrow}_{\text{perfect matching}}(G)$ is determined by* (18.19).

Proof. Let U and W be the colour classes of G. Let $x \in \mathbb{R}^E$ satisfy (18.19). Note that this implies that $|U| = |W| = \frac{1}{2}|V|$, for if (say) $|U| > \frac{1}{2}|V|$, then (18.19) implies that $0 = x(E[U]) \geq |U| - \frac{1}{2}|V| > 0$, a contradiction.

We must show that there exists a vector y such that $\mathbf{0} \leq y \leq x$ and such that $y(\delta(v)) = 1$ for each $v \in V$. This can be shown quite directly with flow theory, for instance with Gale's theorem (Corollary 11.2g): Make a directed graph by orienting each edge from U to W. Then by Gale's theorem (taking $b(v) := -1$ if $v \in U$ and $b(v) := 1$ if $v \in W$), it suffices to show that $|W'| - |U'| \leq x(\delta^{\text{in}}(U' \cup W'))$ for each $U' \subseteq U$ and $W' \subseteq W$. Let $S := (U \setminus U') \cup W'$. Then $\delta^{\text{in}}(U' \cup W') = E[S]$ and $|W'| - |U'| = |S| - \frac{1}{2}|V|$, giving the required inequality. ∎

(Fulkerson [1970b] derived Theorem 18.8 from an earlier result in Fulkerson [1964b], which is Corollary 20.9a below. Related results were given by O'Neil [1971,1975], Cruse [1975], and Houck and Pittenger [1979].)

Note that the theorem gives also a characterization of the convex hull of the incidence vectors of edge sets containing a perfect matching in a bipartite graph:

Corollary 18.8a. *Let $G = (V, E)$ be a bipartite graph. Then the convex hull of the incidence vectors of edge sets containing a perfect matching is determined by* (18.19) *together with $x_e \leq 1$ for each $e \in E$.*

Proof. Directly from Theorem 18.8. ∎

One can similarly characterize the convex hull of the incidence vectors of subsets of perfect matchings in a bipartite graph. Consider:

(18.20) (i) $x_e \geq 0$ for each $e \in E$,
(ii) $x(E[S]) \leq |S| - \frac{1}{2}|V|$ for each vertex cover S.

Theorem 18.9. *The convex hull of the incidence vectors of subsets of perfect matchings in a bipartite graph is determined by* (18.20).

Proof. Similar to the proof of Theorem 18.8. ∎

(Alternative proofs of Theorems 18.8 and 18.9 were given by Cunningham and Green-Krótki [1986].)

See Section 20.6a for more results on $P^{\uparrow}_{\text{perfect matching}}(G)$.

18.5f. Matchings of given size

Let $G = (V, E)$ be a graph and let $k, l \in \mathbb{Z}_+$ with $k \leq l$. It is easy to derive from the description of the matching polytope, a description of the convex hull of incidence vectors of matchings M satisfying $k \leq |M| \leq l$. To this end we show:

Theorem 18.10. *Let $G = (V, E)$ be an undirected graph and let $x \in P_{\text{matching}}(G)$. Then x is a convex combination of incidence vectors of matchings M satisfying*

(18.21) $$\lfloor \mathbf{1}^{\mathsf{T}} x \rfloor \leq |M| \leq \lceil \mathbf{1}^{\mathsf{T}} x \rceil.$$

Proof. Write $x = \sum_M \lambda_M \chi^M$, where M ranges over all matchings in G and where $\lambda_M \geq 0$ with $\sum_M \lambda_M = 1$. Assume that we have chosen the λ_M such that

(18.22) $$\sum_M \lambda_M |M|^2$$

is as small as possible. We show that if M and N are matchings with $\lambda_M > 0$ and $\lambda_N > 0$, then $\big||M| - |N|\big| \leq 1$. This implies the theorem.

Suppose that $|M| \geq |N| + 2$. Let P be a component of $M \cup N$ having more elements in M than in N. Let $M' := M \triangle EP$ and $N' := N \triangle EP$. Then $\chi^{M'} + \chi^{N'} = \chi^M + \chi^N$ and $|M'|^2 + |N'|^2 < |M|^2 + |N|^2$. So decreasing λ_M and λ_N by ε, and increasing $\lambda_{M'}$ and $\lambda_{N'}$ by ε, where $\varepsilon := \min\{\lambda_M, \lambda_N\}$, would decrease sum (18.22), contradicting our assumption. ∎

This implies that certain slices of the matching polytope are again integer polytopes:

Corollary 18.10a. *Let $G = (V, E)$ be an undirected graph and let $k, l \in \mathbb{Z}_+$ with $k \leq l$. Then the convex hull of the incidence vectors of matchings M satisfying $k \leq |M| \leq l$ is equal to the set of those vectors x in the matching polytope of G satisfying $k \leq \mathbf{1}^{\mathsf{T}} x \leq l$.*

Proof. Directly from Theorem 18.10. ∎

A special case is the following result of Mendelsohn and Dulmage [1958b]. Call a matrix a *subpermutation matrix* if it is a $\{0, 1\}$-valued matrix with at most one 1 in each row and in each column. Then:

Corollary 18.10b. *A matrix M belongs to the convex hull of the subpermutation matrices of rank r if and only if M is nonnegative, each row and column sum is at most 1, and the sum of the entries in M is equal to r.*

Proof. Directly from Theorem 18.10. ∎

18.5g. Stable matchings

Let $G = (V, E)$ be a graph and let for each $v \in V$, $\leq_v$ be a total order on $\delta(v)$. Put $e \preceq f$ if e and f have a vertex v in common with $e \leq_v f$. Call a set M of edges *stable* if for each $e \in E$ there exists an $f \in M$ with $e \preceq f$.

In general, stable matchings need not exist (e.g., generally not for K_3). However, Gale and Shapley [1962] showed that if G is bipartite, they do exist:

Theorem 18.11 (Gale-Shapley theorem). *If G is bipartite, then there exists a stable matching.*

Proof. Let U and W be the colour classes of G. For each edge $e = uw$ with $u \in U$ and $w \in W$, let $\phi(e)$ be the height of e in $(\delta(w), \leq_w)$. (The *height* of e is the maximum size of a chain with maximum e.) Choose a matching M in G such that for each edge $e = uw$ of G, with $u \in U$ and $w \in W$,

(18.23) if $f \leq_u e$ for some $f \in M$, then $e \leq_w g$ for some $g \in M$,

and such that $\sum_{e \in M} \phi(e)$ is as large as possible. (Such a matching exists, since $M = \emptyset$ satisfies (18.23).) We show that M is stable.

Choose $e = uw \in E$ with $u \in U$ and $w \in W$ and suppose that there is no $e' \in M$ with $e \preceq e'$. Choose e largest in $\leq_u$ with this property. Then by (18.23) there is no $f \in M$ with $f \leq_u e$; and moreover, there is no $f \in M$ with $e \leq_u f$. Hence u is missed by M.

Since also there is no $g \in M$ with $e \leq_w g$, we can remove any edge in M incident with w and add e to M, so as to obtain a matching satisfying (18.23) with larger $\sum_{e \in M} \phi(e)$, a contradiction. ∎

This proof also gives a polynomial-time algorithm to find a stable matching[34].

The following fact was shown by McVitie and Wilson [1970]:

Theorem 18.12. *Each two stable matchings cover the same set of vertices.*

Proof. Let M and N be two stable matchings, and suppose that there exists a vertex v covered by M but not by N. Let P be the path component of $M \cup N$ starting at v. Denote $P = (v_0, e_1, v_1, e_2, \ldots, e_k, v_k)$ with $v = v_0$. As v_0 is missed by N, $e_1 <_{v_1} e_2$. As M and N are stable, if $e_{i-1} <_{v_{i-1}} e_i$, then $e_i <_{v_i} e_{i+1}$ for each $i < k$. So $e_{k-1} <_{v_{k-1}} e_k$. However, as v_k is missed by M or N, $e_k <_{v_{k-1}} e_{k-1}$. So we have a contradiction. ∎

In particular:

Corollary 18.12a. *All stable matchings have the same size.*

Proof. Directly from Theorem 18.12. ∎

In order to find a maximum-*weight* stable matching, we consider the *stable matching polytope* $P_{\text{stable matching}}(G)$ of G, which is defined as the convex hull of the incidence vectors of the stable matchings. Vande Vate [1989] (also Rothblum [1992]) characterized the inequalities determining the stable matching polytope if G is bipartite. In that case it suffices to add the following inequalities to the system defining the matching polytope:

(18.24) $$\sum_{f \succeq e} x(f) \geq 1 \text{ for each } e \in E.$$

Theorem 18.13. *If G is bipartite, then $x \in P_{\text{stable matching}}(G)$ if and only if $x \in P_{\text{matching}}(G)$ and x satisfies* (18.24).

[34] It was noted by Roth [1984] that this algorithm is in fact in use in practice since 1951 in the U.S., to match hospitals and medical students (cf. Roth and Sotomayor [1990]).

Proof. Necessity is easy, since the incidence vector of any stable matching satisfies (18.24). To see sufficiency, let x be a vertex of the polytope of all vectors in $P_{\text{matching}}(G)$ satisfying (18.24). Define E^+ to be the set of edges e with $x_e > 0$, and V^+ the set of vertices covered by E^+. For each $v \in V^+$, let e_v be the maximum element of $(\delta(v) \cap E^+, \leq_v)$.

We first show that for each $v \in V^+$, with say $e_v = vv'$,

(18.25) e_v is the minimum element in $(\delta(v') \cap E^+, \leq_{v'})$ and that $x(\delta(v')) = 1$.

Indeed, (18.24) implies (writing $e := e_v$):

(18.26) $$1 \leq \sum_{f \succeq e} x(f) = \sum_{f \geq_{v'} e} x(f) = x(\delta(v')) - \sum_{f <_{v'} e} x(f) \leq 1 - \sum_{f <_{v'} e} x(f).$$

Hence we have equality throughout in (18.26). This implies that $x(f) = 0$ for each $f <_{v'} e$ and that $x(\delta(v')) = 1$. This proves (18.25).

It follows that for each $v' \in V^+$ there is exactly one $v \in V^+$ with $e_v = vv'$. Now let U and W be the colour classes of G. The sets $M := \{e_v \mid v \in U \cap V^+\}$ and $N := \{e_v \mid v \in W \cap V^+\}$ are matchings covering V^+. Consider the vector $x' = x + \varepsilon\chi^M - \varepsilon\chi^N$, with ε close enough to 0 (positive or negative). It is easy to see that x' again belongs to the matching polytope. To see that x' satisfies (18.24) for ε close enough to 0, let e be an edge of G attaining equality in (18.24). We show that $e \preceq f$ for exactly one $f \in M$. If $e \in M$, this is trivial, so assume that $e \notin M$. Let $e = uw$ with $u \in U$ and $w \in W$. Then

(18.27) there is an $f \in M$ with $e <_u f \iff \sum_{f >_u e} x(f) > 0 \iff \sum_{g \geq_w e} x(g) < 1$
$\iff$ there is no $g \in M$ with $e <_w g$.

Similarly, $e \preceq f$ for exactly one $f \in N$. Concluding,

(18.28) $$\sum_{f \succeq e} x'(f) = \sum_{f \succeq e} x(f) = 1$$

if ε is close enough to 0. So x' again satisfies (18.24). Since x is a vertex, we have $\chi^M = \chi^N$, that is, $M = N$. So $E^+ = M$, and hence $x = \chi^M$, and therefore x is $\{0,1\}$-valued. ∎

As for algorithms, this theorem directly implies:

Corollary 18.13a. *A maximum-weight stable matching can be found in polynomial time.*

Proof. This follows from the fact that Theorem 18.13 transforms the problem to a linear programming problem. ∎

For surveys and further results, see Wilson [1972a], Knuth [1976], Itoga [1978, 1981], Roth [1982], Gale and Sotomayor [1985], Irving [1985], Gusfield [1987b,1988], Irving, Leather, and Gusfield [1987], Blair [1988], Gusfield and Irving [1989], Ng [1989], Knuth, Motwani, and Pittel [1990a,1990b], Ng and Hirschberg [1990], Ronn [1990], Roth and Sotomayor [1990], Khuller, Mitchell, and Vazirani [1991,1994], Tan [1991], Feder [1992], Roth, Rothblum, and Vande Vate [1993], Abeledo and Rothblum [1994], Feder, Megiddo, and Plotkin [1994,2000], Subramanian [1994], Abeledo and Blum [1996], Balinski and Ratier [1997], Teo and Sethuraman [1997, 1998], Teo, Sethuraman, and Tan [1999], Fleiner [2001a], and Aharoni and Fleiner [2002].

18.5h. Further notes

Perfect and Mirsky [1965] characterized which patterns can occur as the support of a doubly stochastic matrix. It is equivalent to characterizing matching-covered bipartite graphs (that is, bipartite graphs in which each edge belongs to at least one perfect matching).

Frank and Karzanov [1992] gave a polynomial-time combinatorial algorithm to determine the Euclidean distance of the perfect matching polytope of a bipartite graph to the origin.

Chapter 19

Bipartite edge cover and stable set

While matchings cover each vertex *at most* once, edge covers are required to cover each vertex *at least* once. Most edge cover results can be proved similarly to matching results, but in fact, they often can be reduced to matching results, by a method of Gallai.
In this chapter, graphs can be assumed to be simple.

19.1. Matchings, edge covers, and Gallai's theorem

Let $G = (V, E)$ be a graph. An *edge cover* is a subset F of E such that for each vertex v there exists an edge $e \in F$ satisfying $v \in e$. Note that an edge cover can exist only if G has no isolated vertices.

A *stable set* is a subset S of V such that no two vertices in S are adjacent. So for any $U \subseteq V$:

(19.1) $\quad S$ is a stable set $\iff$ $V \setminus S$ is a vertex cover.

Define:

(19.2) $\quad \alpha(G) :=$ the maximum size of a stable set in G,
$\rho(G) :=$ the minimum size of an edge cover in G.

These numbers are called the *stable set number* and the *edge cover number*, respectively.

It is not difficult to show that:

(19.3) $\quad \alpha(G) \leq \rho(G)$.

The triangle K_3 shows that strict inequality is possible. Recall that for the matching number $\nu(G)$ and the vertex cover number $\tau(G)$ we have

(19.4) $\quad \nu(G) \leq \tau(G)$.

In fact, equality in one of the relations (19.3) and (19.4) implies equality in the other, as Gallai [1959a] proved the following[35]:

[35] Gallai mentioned that he had formulated and proved this theorem in 1932 (cf. also Erdős [1982]), and that to his knowledge also D. Kőnig had known this theorem.

Theorem 19.1 (Gallai's theorem)**.** *For any graph $G = (V, E)$ without isolated vertices one has*

(19.5) $$\alpha(G) + \tau(G) = |V| = \nu(G) + \rho(G).$$

Proof. The first equality follows directly from (19.1).

To see the second equality, let M be a maximum-size matching and let U be the set of vertices missed by M. For each vertex $v \in U$, choose an edge e_v containing v. Then $F = M \cup \{e_v \mid v \in U\}$ is an edge cover of size

(19.6) $$|F| = |M| + |U| = |M| + (|V| - 2|M|) = |V| - |M| = |V| - \nu(G).$$

So $\rho(G) \leq |V| - \nu(G)$.

To see the reverse inequality, let F be a minimum-size edge cover. Let M be an inclusionwise maximal matching contained in F. Let U be the set of vertices missed by M. Since U spans no edge in F, we have $|U| \leq |F \setminus M|$. Hence $|V| - 2|M| = |U| \leq |F \setminus M| = |F| - |M|$. This implies $\nu(G) \geq |M| \geq |V| - |F| = |V| - \rho(G)$. ∎

This proof method implies the following theorem (observed by Gallai [1959a] and Norman and Rabin [1959]):

Theorem 19.2. *Let $G = (V, E)$ be a graph without isolated vertices. Then every maximum-size matching is contained in a minimum-size edge cover, and every minimum-size edge cover contains a maximum-size matching.*

Proof. See above. ∎

Moreover, there is the following complexity result, observed by Norman and Rabin [1959]:

Theorem 19.3. *Let $G = (V, E)$ be an undirected graph with n vertices and m edges. If we have a maximum-size matching in G, we can find a minimum-size edge cover in time $O(m)$, and vice versa.*

Proof. See the proof of Gallai's theorem (Theorem 19.1). ∎

This gives:

Corollary 19.3a. *A minimum-size edge cover and a maximum-size stable set in a bipartite graph can be found in time $O(n^{1/2}m)$.*

Proof. By Theorems 16.4 and 19.3 and Corollary 16.6a. ∎

Short proof of Gallai's theorem. For any partition Π of V into edges and singletons, let $f(\Pi)$ be the number of edges in Π. So $f(\Pi) + |\Pi| = |V|$. Then $\nu(G)$ is equal to the maximum of $f(\Pi)$ over all such partitions, and $\rho(G)$ is equal to the minimum of $|\Pi|$ over all such partitions. Hence $\nu(G) + \rho(G) = |V|$.

19.2. The Kőnig-Rado edge cover theorem

Combination of Theorems 19.1 and 16.2 yields the following theorem, which Gallai [1958a,1958b] attributes to oral communication from D. Kőnig in 1932. In a different but equivalent form it was stated by Rado [1933] — see Section 19.5a. (Hoffman [1956b] called it a 'well-known theorem'.)

Theorem 19.4 (Kőnig-Rado edge cover theorem)**.** *For any bipartite graph $G = (V, E)$ without isolated vertices one has*

(19.7) $\alpha(G) = \rho(G)$.

That is, the maximum size of a stable set in a bipartite graph is equal to the minimum size of an edge cover.

Proof. Directly from Theorems 19.1 and 16.2, as $\alpha(G) = |V| - \tau(G) = |V| - \nu(G) = \rho(G)$. ∎

By representing a bipartite graph as a partially ordered set, the Kőnig-Rado edge cover theorem can be derived also from Dilworth's decomposition theorem (Theorem 14.2).

19.3. Finding a minimum-weight edge cover

There is a straightforward reduction of the minimum-weight edge cover problem to the minimum-weight perfect matching problem. Indeed, let $G = (V, E)$ be a graph without isolated vertices, and let $w : E \to \mathbb{Q}_+$. Let $G' = (V', E')$ be the graph obtained from G by adding a disjoint copy $\widetilde{G} = (\widetilde{V}, \widetilde{E})$ of G, and adding for each vertex v of G an edge $v\tilde{v}$ connecting v with its copy $\tilde{v}$. Let w' be the weight function on E' defined by:

(19.8) $w'(e) := w'(\tilde{e}) := w(e)$ for each $e \in E$ (where $\tilde{e}$ is the copy of e);
$w'(v\tilde{v}) := 2\mu(v)$ for each $v \in V$, where $\mu(v)$ is the minimum weight of the edges of G incident with v.

Then a minimum-weight perfect matching M in G' yields a minimum-weight edge cover F in G: replace any edge $v\tilde{v}$ in M by an edge e_v of minimum weight of G incident with v, and delete all edges in $M \cap \widetilde{E}$. Then $w(F) = \frac{1}{2}w'(M)$. Conversely, any edge cover F' of G gives by a reverse construction a perfect matching M' in G' with $w'(M') \leq 2w(F')$. Hence $w(F) = \frac{1}{2}w'(M) \leq \frac{1}{2}w'(M') \leq w(F')$. So F is a minimum-weight edge cover in G.

Note that if G is bipartite, then also G' is bipartite. Hence:

Corollary 19.4a. *A minimum-weight edge cover in a bipartite graph can be found in time $O(n(m + n \log n))$.*

Proof. From the above, using Theorem 17.3. ∎

19.4. Bipartite edge covers and total unimodularity

Similarly to Kőnig's matching theorem, also the Kőnig-Rado edge cover theorem (Theorem 19.4) can be derived from the total unimodularity of the $V \times E$ incidence matrix of a bipartite graph $G = (V, E)$. This follows by considering the LP-duality equation

(19.9) $$\min\{\mathbf{1}^\mathsf{T}x \mid x \geq \mathbf{0}, Ax \geq \mathbf{1}\} = \max\{y^\mathsf{T}\mathbf{1} \mid y \geq \mathbf{0}, y^\mathsf{T}A \leq \mathbf{1}^\mathsf{T}\}.$$

More generally, we can derive the analogue of Egerváry's theorem:

Theorem 19.5. *Let $G = (V, E)$ be a bipartite graph and let $w : E \to \mathbb{R}_+$ be a weight function on E. Then the minimum weight of an edge cover in G is equal to the maximum value of $y(V)$, where y ranges over all functions $y : V \to \mathbb{R}_+$ with $y_u + y_v \leq w_e$ for each edge $e = uv$ of G. If w is integer, we can restrict y to be integer.*

Proof. Again, let A be the $V \times E$ incidence matrix of G. Then the statement is equivalent to the statement that the minimum in

(19.10) $$\min\{w^\mathsf{T}x \mid x \geq \mathbf{0}, Ax \geq \mathbf{1}\} = \max\{y^\mathsf{T}\mathbf{1} \mid y \geq \mathbf{0}, y^\mathsf{T}A \leq w^\mathsf{T}\}$$

has an integer optimum solution x. This fact follows from the total unimodularity of A. If w is integer, we can take also y integer. ∎

The integer part of this theorem can be formulated as follows. For any graph $G = (V, E)$ and $w \in \mathbb{Z}_+^E$, a *w-stable set* is a function $y \in \mathbb{Z}_+^V$ with $y_u + y_v \leq w_e$ for each edge $e = uv$. So if $w = \mathbf{1}$ and G has no isolated vertices, w-stable sets coincide with the incidence vectors of stable sets.

The *size* of a vector $y \in \mathbb{R}^V$ is equal to $y(V)$. Then:

Corollary 19.5a. *Let $G = (V, E)$ be a bipartite graph and let $w : E \to \mathbb{Z}_+$ be a weight function on E. Then the minimum weight of an edge cover in G is equal to the maximum size of a w-stable set.*

Proof. Directly from Theorem 19.5. ∎

19.5. The edge cover and stable set polytope

Like in Sections 18.3 and 18.4, the total unimodularity of the incidence matrix of a bipartite graph yields descriptions of the edge cover and the stable set polytope for bipartite graphs.

The *edge cover polytope* $P_{\text{edge cover}}(G)$ of a graph is the convex hull of the incidence vectors of the edge covers in G. For any graph, each vector x in $P_{\text{edge cover}}(G)$ satisfies:

(19.11)
(i) $0 \leq x_e \leq 1$ for each $e \in E$,
(ii) $x(\delta(v)) \geq 1$ for each $v \in V$.

Theorem 19.6. *If G is bipartite, the edge cover polytope is determined by* (19.11).

Proof. Directly from the total unimodularity of the constraint matrix in (19.11). ∎

This implication cannot be turned around, as is shown by the graph in Figure 18.1.

The *stable set polytope* $P_{\text{stable set}}(G)$ of a graph $G = (V, E)$ is the convex hull of the incidence vectors of the stable sets in G. For any graph G, each vector x in $P_{\text{stable set}}(G)$ satisfies:

(19.12) (i) $0 \le x_v \le 1$ for each $v \in V$,
(ii) $x_u + x_v \le 1$ for each edge $e = uv \in E$.

Theorem 19.7. *The stable set polytope is determined by* (19.12) *if and only if G is bipartite.*

Proof. Sufficiency follows from the total unimodularity of the incidence matrix of a bipartite graph. Necessity follows from the fact that if C is an odd circuit in G, then defining $x_v := \frac{1}{2}$ for each $v \in V$, we obtain a vector x satisfying (19.12) but not belonging to the stable set polytope of G, since any stable set intersects C in at most $\frac{1}{2}|VC| - \frac{1}{2}$ vertices. ∎

In fact, there is an easy direct proof of sufficiency in Theorem 19.7. Let x satisfy (19.12) and let U and W be the colour classes of G. For any $\lambda \in [0, 1]$, define

(19.13) $$S_\lambda := \{u \in U \mid x_u > \lambda\} \cup \{w \in W \mid x_w > 1 - \lambda\}.$$

Then S_λ is a stable set, and

(19.14) $$x = \int_0^1 \chi^{S_\lambda} d\lambda.$$

This describes x as a convex combination of incidence vectors of stable sets.

19.5a. Some historical notes on bipartite edge covers

Gallai [1958a,1958b,1959a] wrote that the edge cover theorem (Theorem 19.4) was orally communicated to him by Kőnig in 1932. In the latter paper, Gallai also mentioned that he found Theorem 19.1 in 1932, and that, to his knowledge, also D. Kőnig knew this theorem. Together with Theorem 16.2 of Kőnig [1931] it implies Theorem 19.4.

The oldest written version of Theorem 19.4 seems to be the paper of Rado [1933] entitled *Bemerkungen zur Kombinatorik im Anschluß an Untersuchungen von Herrn D. Kőnig*[36]. The investigations referred to in the title are those of Kőnig [1916] on matchings in *regular* bipartite graphs.

[36] Remarks on combinatorics in connection to investigations of Mr D. Kőnig.

Rado formulated the edge cover theorem in terms of partitions:

> *Es seien* $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$ *endlich viele nicht leere, paarweise elementenfremde Mengen. Ebenso seien* $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ *endlich viele nicht leere, paarweise elementenfremde Mengen. Alle Mengen* $\mathcal{A}_\mu$ und $\mathcal{B}_\nu$ *seien Teilmengen einer Menge* $\mathcal{M}$*. Unter dieser Annahme gilt: Dann und nur dann sind die Mengen*
> (26) $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m, \mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$
> *durch* k *Elemente von* $\mathcal{M}$ *zu repräsentieren, wenn es unter den Mengen* (26) *keine* $k+1$ *zu einander fremde Mengen gibt.*[37]

The proof of Rado is based on a decomposition similar to that used by Frobenius (see Section 16.2a). The equivalence with Theorem 19.4 follows with the construction described in Section 16.7e. (A theorem similar to Rado's was published by Kreweras [1946].)

[37] *Let* $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$ *be finitely many nonempty, pairwise disjoint sets. Similarly, let* $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$ *be finitely many nonempty, pairwise disjoint sets. All sets* $\mathcal{A}_\mu$ and $\mathcal{B}_\nu$ *are subsets of a set* $\mathcal{M}$*. Under this condition the following holds: The sets*
(26) $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m, \mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$
can be represented by k *elements of* $\mathcal{M}$*, if and only if there are no* $k+1$ *disjoint sets among the sets* (26).

Chapter 20

Bipartite edge-colouring

Edge-colouring means partitioning the edge set into matchings. While for general graphs, finding a minimum edge-colouring is NP-complete, another fundamental theorem of Kőnig gives a min-max relation for bipartite edge-colouring, and his proof method yields a polynomial-time algorithm. Also the capacitated case and the 'dual' problem of partitioning the edge set into edge covers are tractable for bipartite graphs.

20.1. Edge-colourings of bipartite graphs

For any graph $G = (V, E)$, an *edge-colouring* or *k-edge-colouring* is a partition $\Pi = (M_1, \ldots, M_k)$ of the edge set E into matchings. Each of the M_i is called a *colour*. If $e \in M_i$ we say that e *has* colour i.

The *edge-colouring number* $\chi(G)$ of G is the minimum number of colours in an edge-colouring of G.

Let $\Delta(G)$ denote the maximum degree of (the vertices of) G. Clearly,

(20.1) $\qquad \chi(G) \geq \Delta(G),$

since at each vertex v, the edges incident with v should have different colours. The triangle K_3 has strict inequality in (20.1). Kőnig [1916] showed that for bipartite graphs the two numbers are equal:

Theorem 20.1 (Kőnig's edge-colouring theorem)**.** *For any bipartite graph $G = (V, E)$,*

(20.2) $\qquad \chi(G) = \Delta(G).$

That is, the edge-colouring number of a bipartite graph is equal to its maximum degree.

Proof. Let $M_1, \ldots, M_{\Delta(G)}$ be a collection of disjoint matchings covering a maximum number of edges. If all edges are covered, we are done. So suppose that edge $e = uv$, say, is not covered. Then (since $\deg(u) \leq \Delta(G)$) some M_i misses u and (similarly) some M_j misses v. If $i = j$ we can extend M_i to $M_i \cup \{e\}$. If $i \neq j$, $M_i \cup M_j \cup \{e\}$ makes a bipartite graph of maximum degree at most two. Hence there exist matchings M and N with $M_i \cup M_j \cup \{e\} = M \cup N$.

So replacing M_i and M_j by M and N, increases the number of edges covered, contradicting our assumption. ∎

This proof, due to Kőnig [1916] (using a simplification of Skolem [1927]), also gives a polynomial-time algorithm to find a $\Delta(G)$-edge-colouring with $\Delta(G)$ colours. In fact, if G is simple, it gives an $O(nm)$ algorithm for edge-colouring. This bound can be achieved also for bipartite multigraphs using an appropriate data-structure — see Section 20.9a.

20.1a. Edge-colouring regular bipartite graphs

Kőnig's edge-colouring theorem is directly equivalent to the special case of regular bipartite graphs (since any bipartite graph of maximum degree Δ is a subgraph of a Δ-regular bipartite graph (Kőnig [1932])). Rizzi [1997,1998] gave the following very elegant short argument for the k-edge-colourability of k-regular bipartite graphs. (A similar proof in terms of common transversals of two partitions of a set into equally sized classes was given by Sperner [1927] — see Section 22.7d.)

Let G be a counterexample with fewest edges. So G has no perfect matching. Choose an edge $e = uv$. Then we can extend the graph $G - u - v$ to a k-regular bipartite graph H by adding at most $k - 1$ new edges. As H has fewer edges than G, H has a k-edge-colouring. Since less than k new edges have been added, there is a colour M that uses none of the new edges. Then $M \cup \{e\}$ is a perfect matching in G, a contradiction.

20.2. The capacitated case

Egerváry [1931] observed that the following capacitated version directly follows from Kőnig's edge-colouring theorem:

Corollary 20.1a. *Let $G = (V, E)$ be a bipartite graph and let $c : E \to \mathbb{Z}_+$ be a capacity function. Then the minimum size of a family of matchings such that each edge e is in at least c_e of them is equal to the maximum of $c(\delta(v))$ over all $v \in V$.*

Proof. Directly from Kőnig's edge-colouring theorem, by replacing each edge e by c_e parallel edges. ∎

This reduction being easy, it might not be satisfactory algorithmically. It would not yield a polynomial-time reduction for the following problem:

(20.3) given: a bipartite graph $G = (V, E)$ and a capacity function $c : E \to \mathbb{Z}_+$;
find: matchings $M_1, \ldots, M_k$ and nonnegative integers $\lambda_1, \ldots, \lambda_k$ such that $\sum_{i=1}^{k} \lambda_i \chi^{M_i} = c$ and such that $\sum_{i=1}^{k} \lambda_i$ is minimized.

However, there is an easy strongly polynomial-time algorithm for this problem: Let F be the subset of edges e of G with $c_e > 0$. Find a matching M in F covering all vertices v of G that maximize $c(\delta(v))$. Let $\lambda := \min\{c_e \mid e \in M\}$, and replace c by $c - \lambda\chi^M$. Next iterate this.

Since in each iteration the number of edges e with $c_e > 0$ decreases, there are at most $|E|$ iterations. Since a matching covering a given set R of vertices can be found in time $O(|R||E|)$, this gives an $O(nm^2)$ algorithm. However, by starting in each iteration with the matching left from the previous iteration, one can do better (Gonzalez and Sahni [1976]):

Theorem 20.2. *Problem* (20.3) *can be solved in time* $O(m^2)$.

Proof. We may assume that $c(\delta(v))$ is equal for all v, by duplicating G and connecting each vertex with its copy, giving the new edges appropriate capacities. We can also assume that $c_e > 0$ for each edge e.

First we find a perfect matching in G, which can be done in time $O(nm)$, since we can apply $O(n)$ matching-augmenting iterations to find a perfect matching.

In any further iteration, let M be the matching obtained in the previous iteration. Suppose that after resetting c, there exist α edges e in M with $c_e = 0$. Delete these edges. Then in α matching-augmenting steps we can obtain a perfect matching M' in the new graph. So the iteration takes $O(\alpha m)$ time. Since over all iterations the α add up to $|E|$, we have the time bound $O(m^2)$. ∎

20.3. Edge-colouring polyhedrally

Polyhedrally, edge-colouring can be studied with the help of the 'substar polytope' of an undirected graph $G = (V, E)$. Call a set F of edges of G a *substar* if $F \subseteq \delta(v)$ for some $v \in V$. The *substar polytope* $P_{\text{substar}}(G)$ of G is the convex hull of the incidence vectors of substars. So it is a polytope in $\mathbb{R}^E$.

Each vector x in the substar polytope trivially satisfies

$$(20.4) \qquad \begin{array}{lll} \text{(i)} & x_e \geq 0 & \text{for each } e \in E, \\ \text{(ii)} & x(M) \leq 1 & \text{for each matching } M. \end{array}$$

The following is direct from the description of the bipartite matching polytope (Corollary 18.1b) with the theory of antiblocking polyhedra:

Theorem 20.3. *The substar polytope of a bipartite graph is determined by* (20.4).

Proof. By Corollary 18.1b, the matching polytope is the antiblocking polyhedron of the substar polytope. Hence the substar polytope is the antiblocking polyhedron of the matching polytope (cf. Section 5.9), which is the content of the theorem. ∎

What Kőnig's edge-colouring theorem adds to it is:

Theorem 20.4. *System* (20.4) *is TDI.*

Proof. This is equivalent to Corollary 20.1a. ∎

Note that Kőnig's edge-colouring theorem also can be derived easily from the characterization of the matching polytope. For any bipartite graph $G = (V, E)$, the vector $\Delta(G)^{-1} \cdot \mathbf{1}$ belongs to the matching polytope (where $\mathbf{1}$ is the all-one vector in $\mathbb{R}^E$), and hence it is a convex combination of matchings. Each of these matchings should cover each maximum-degree vertex. So there exists a matching M covering all maximum-degree vertices. Hence $\Delta(G - M) = \Delta(G) - 1$, and we can apply induction.

Also, the integer decomposition property of the matching polytope is equivalent to Kőnig's edge-colouring theorem. (The integer decomposition property follows from the total unimodularity of the incidence matrix of G.)

20.4. Packing edge covers

A theorem 'dual' to Kőnig's edge-colouring theorem was shown by Gupta [1967,1978]. The edge-colouring number $\chi(G)$ of a graph G is the minimum number of matchings needed to cover the edges of a G. Dually, one can define the *edge cover packing number* $\xi(G)$ of a graph by:

(20.5) $\quad \xi(G) :=$ the maximum number of disjoint edge covers in G.

So, in terms of colours, $\xi(G)$ is the maximum number of colours that can be used in colouring the edges of G in such a way that at each vertex all colours occur. Hence, if $\delta(G)$ denotes the minimum degree of G, then

(20.6) $\quad \xi(G) \leq \delta(G)$.

The triangle K_3 again is an example having strict inequality. For bipartite graphs however Gupta [1967,1978] showed:

Theorem 20.5. *For any bipartite graph* $G = (V, E)$*:*

(20.7) $\quad \xi(G) = \delta(G)$.

That is, the maximum number of disjoint edge covers is equal to the minimum degree.

Proof. We give a reduction to Kőnig's edge-colouring theorem (Theorem 20.1).

One may derive from G a bipartite graph H, each vertex of which has degree $\delta(G)$ or 1, by repeated application of the following procedure:

(20.8) for any vertex v of degree larger than $\delta(G)$, add a new vertex u, and replace one of the edges incident with v, $\{v, w\}$ say, by $\{u, w\}$.

So there is a one-to-one correspondence between the edges of the final graph H and the edges of G. Since H has maximum degree $\delta(G)$, by Theorem 20.1 the edges of H can be coloured with $\delta(G)$ colours such that no two edges of the same colour intersect. So at any vertex of H of degree $\delta(G)$, all colours occur. This gives a colouring of the edges of G with $\delta(G)$ colours such that at any vertex of G all colours occur. ∎

Gupta [1974,1978] gave the following common generalization of Theorems 20.1 on edge-colouring and 20.5 on disjoint edge covers:

Theorem 20.6. *Let $G = (V, E)$ be a bipartite graph and let $k \in \mathbb{Z}_+$. Then E can be partitioned into classes $E_1, \ldots, E_k$ such that each vertex v is covered by at least* $\min\{k, \deg_G(v)\}$ *of the E_i.*

Proof. Like in the proof of Theorem 20.5, split off edges from vertices of degree larger than k, until each vertex has degree at most k. Applying Kőnig's edge-colouring theorem to the final graph yields a partitioning of the original edge set as required. ∎

Call a set F of edges of a graph $G = (V, E)$ a *superstar* if $F \supseteq \delta(v)$ for some $v \in V$. The *superstar polytope* $P_{\text{superstar}}(G)$ of G is the convex hull of the incidence vectors of superstars in G. Consider

(20.9) (i) $0 \leq x_e \leq 1$ for each $e \in E$,
(ii) $x(F) \geq 1$ for each edge cover F.

Theorem 20.7. *If G is bipartite, system* (20.9) *determines the superstar polytope and is TDI.*

Proof. With the theory of blocking polyhedra, Theorem 19.6 implies that the superstar polytope is determined by (20.9). Total dual integrality of (20.9) is equivalent to the capacitated version of Theorem 20.5. ∎

20.5. Balanced colours

McDiarmid [1972] and de Werra [1970,1972] showed the following generalization of Kőnig's edge-colouring theorem (in fact, it is a special case of a theorem of Folkman and Fulkerson [1969] (see Theorem 20.10 below), and also it is a consequence of the result in Dulmage and Mendelsohn [1969]):

Theorem 20.8. *Let $G = (V, E)$ be a bipartite graph and let $k \geq \Delta(G)$. Then E can be partitioned into matchings $M_1, \ldots, M_k$ such that*

(20.10) $$\lfloor |E|/k \rfloor \leq |M_i| \leq \lceil |E|/k \rceil$$

for each $i = 1, \ldots, k$.

Proof. As $k \geq \Delta(G)$, by Kőnig's edge-colouring theorem, E can be partitioned into matchings $M_1, \ldots, M_k$ (possibly empty). Choose $M_1, \ldots, M_k$ such that

(20.11) $$\sum_{i=1}^{k} |M_i|^2$$

is minimized.

Suppose that (20.10) is violated. Then there exist M_i and M_j with $|M_i| \geq |M_j|+2$. Then $M_i \cup M_j$ has at least one component K containing more edges in M_i than in M_j. Let $M_i' := M_i \triangle K$ and $M_j' := M_j \triangle K$. Then $|M_i'|^2 + |M_j'|^2 = (|M_i|-1)^2 + (|M_j|+1)^2 = |M_i|^2 + |M_j|^2 - 2|M_i| + 2|M_j| + 2 < |M_i|^2 + |M_j|^2$. So replacing M_i and M_j by M_i' and M_j' decreases the sum (20.11), contradicting our minimality assumption. ∎

Related results can be found in Dulmage and Mendelsohn [1969], Folkman and Fulkerson [1969], Brualdi [1971b], and de Werra [1971,1976].

20.6. Packing perfect matchings

Packing perfect matchings seems less directly reducible to partitioning into matchings or edge covers. It can be handled with the following more general result of Folkman and Fulkerson [1969] on packing matchings of a fixed size p, which is proved by reduction to Menger's theorem:

Theorem 20.9. *Let $G = (V, E)$ be a bipartite graph and let $k, p \in \mathbb{Z}_+$. Then there exist k disjoint matchings of size p if and only if each subset X of V spans at least $k(p + |X| - |V|)$ edges.*

Proof. To see necessity, let $X \subseteq V$ and consider a matching M in G of size p. Since at most $|V| - |X|$ edges in M intersect $V \setminus X$, at least $|M| - (|V| - |X|) = p + |X| - |V|$ edges of M are spanned by X. So k disjoint matchings of size p have at least $k(p + |X| - |V|)$ edges spanned by X.

To see sufficiency, let U and W be the colour classes of G. Orient all edges from U to W. Moreover, add vertices s and t, and, for each $u \in U$, add k parallel arcs from s to u, and, for each $w \in W$, add k parallel arcs from w to t. Let D be the directed graph arising.

We show with Menger's theorem that D contains kp arc-disjoint $s-t$ paths. Consider any $s-t$ cut $\delta^{\text{out}}(Y)$, with $s \in Y$, $t \notin Y$. Let $X := (U \cap Y) \cup (W \setminus Y)$. Then

(20.12) $\quad |\delta^{\text{out}}(Y)| = k|U \setminus Y| + k|W \cap Y| + |E[X]| = k(|V| - |X|) + |E[X]|,$

where $E[X]$ is the set of edges spanned by X. As $|E[X]| \geq k(p + |X| - |V|)$, it follows that $|\delta^{\text{out}}(Y)| \geq kp$.

So D contains kp arc-disjoint $s-t$ paths. The edges of G that belong to these paths form a subgraph of G with kp edges, of maximum degree at most k. So by Theorem 20.8, G has k disjoint matchings of size p. ∎

This implies the following theorem of Fulkerson [1964b] on the maximum number of disjoint perfect matchings (in fact equivalent to a result of Ore [1956], see Corollary 20.9b below):

Corollary 20.9a. *Let $G = (V, E)$ be a bipartite graph and let $k \in \mathbb{Z}_+$. Then G has k disjoint perfect matchings if and only if each subset X of V spans at least $k(|X| - \frac{1}{2}|V|)$ edges.*

Proof. Directly by taking $p := \frac{1}{2}|V|$ in Theorem 20.9. ∎

(Lebensold [1977] and Murty [1978] gave other proofs of this corollary.)

Note that, by Kőnig's edge-colouring theorem, a bipartite graph $G = (V, E)$ has k disjoint perfect matchings if and only if G has a k-factor. (A *k-factor* is a subset F of E with the graph (V, F) k-regular.)

So Corollary 20.9a is equivalent to the following result of Ore [1956]:

Corollary 20.9b. *A bipartite graph $G = (V, E)$ has a k-factor if and only if each subset X of V spans at least $k(|X| - \frac{1}{2}|V|)$ edges.*

Proof. Directly from Corollary 20.9a. ∎

20.6a. Polyhedral interpretation

We can interpret these results polyhedrally. In Theorem 18.8 we saw that for any bipartite graph $G = (V, E)$, the up hull of the perfect matching polytope of G,

(20.13) $\quad P^{\uparrow}_{\text{perfect matching}}(G) = P_{\text{perfect matching}}(G) + \mathbb{R}^E_+$

is determined by the inequalities

(20.14)
(i) $x_e \geq 0$ for each $e \in E$,
(ii) $x(E[S]) \geq |S| - \frac{1}{2}|V|$ for each $S \subseteq V$.

Then Corollary 20.9a implies that for each $k \in \mathbb{Z}_+$, each integer vector $w \in k \cdot P^{\uparrow}_{\text{perfect matching}}(G)$ is the sum of k vectors in $P^{\uparrow}_{\text{perfect matching}}(G)$. In other words:

Corollary 20.9c. *$P^{\uparrow}_{\text{perfect matching}}(G)$ has the integer decomposition property.*

Proof. From Corollary 20.9a, by replacing each edge by $w(e)$ parallel edges. ∎

We can view this also in terms of the blocking polyhedron of $P^{\uparrow}_{\text{perfect matching}}(G)$, which is the polyhedron Q determined by

$$(20.15) \qquad \begin{array}{rll} \text{(i)} & x_e \geq 0 & \text{for each } e \in E, \\ \text{(ii)} & x(M) \geq 1 & \text{for each perfect matching } M. \end{array}$$

Since $P^{\uparrow}_{\text{perfect matching}}(G)$ is determined by (20.14), the theory of blocking polyhedra gives that Q is equal to the up hull of the convex hull of the vectors

$$(20.16) \qquad \frac{1}{|S|-\frac{1}{2}|V|}\chi^{E[S]}$$

where $S \subseteq V$ with $|S| > \frac{1}{2}|V|$.

So the minimum value of $\mathbf{1}^{\mathsf{T}}x$ over Q is equal to

$$(20.17) \qquad \min\{\frac{|E[S]|}{|S|-\frac{1}{2}|V|} \mid S \subseteq V, |S| > \tfrac{1}{2}|V|\}.$$

By LP-duality, this is equal to the maximum value of $\sum_M \lambda_M$, where M ranges over perfect matchings and where $\lambda_M \geq 0$ such that $\sum_M \lambda_M \chi^M \leq \mathbf{1}$. So Corollary 20.9a states: the maximum number of disjoint perfect matchings in a bipartite graph is equal to

$$(20.18) \qquad \lfloor \max\{\sum_M \lambda_M \mid \lambda_M \geq 0, \sum_M \lambda_M \chi^M \leq \mathbf{1}\}\rfloor.$$

As we can directly extend this to a weighted version, one has:

Corollary 20.9d. *System* (20.15) *has the integer rounding property.*

Proof. See above. ∎

20.6b. Extensions

The results of Sections 20.5 and 20.6 can be extended as follows, as was shown by Folkman and Fulkerson [1969]. It is based on the following theorem:

Theorem 20.10. *Let $G = (V, E)$ be a bipartite graph, let $k \geq \Delta(G)$, and let $p \geq |E|/k$. Then G has a k-edge-colouring in which l colours have size p if and only if G has l disjoint matchings of size p.*

Proof. Necessity being trivial, we show sufficiency. Let G have l disjoint matchings of size p. We must show that there exist l disjoint matchings of size p such that at each vertex v at most $k - l$ edges incident with v are in none of these matchings (since then the edges not contained in the matchings can be properly coloured by $k - l$ colours).

That is, by Theorem 20.8 it suffices to show that there exists a subset F of E such that

$$(20.19) \qquad \begin{array}{rl} \text{(i)} & \deg_F(v) \leq l \text{ and } \deg_{E\setminus F}(v) \leq k - l \text{ for each vertex } v; \\ \text{(ii)} & |F| = lp. \end{array}$$

Let F be any subset of E satisfying (20.19)(i), with $|F| \leq lp$, and with $|F|$ as large as possible. Such an F exists, since by Theorem 20.8 we can k-edge-colour G such that each colour has size at most $\lceil |E|/k \rceil \leq p$. Any l of the colours gives F as required.

If $|F| = lp$ we are done, so assume that $|F| < lp$. Since G has l disjoint matchings of size p, E has a subset F' of size lp with $\deg_{F'}(v) \leq l$ for each vertex v. Choose F' with $F' \setminus F$ as small as possible.

Consider an orientation D of the graph $(V, F \triangle F')$, where each edge in $F \setminus F'$ is oriented from colour class U (say) to colour class W (say), and where each edge in $F' \setminus F$ is oriented from W to U. If D contains a directed circuit C, we can reduce $F' \setminus F$, by replacing F' by $F' \triangle C$. So D is acyclic, and hence we can partition $F \triangle F'$ into directed paths, where each path starts at a vertex v with $\deg^{\text{out}}_D(v) > \deg^{\text{in}}_D(v)$ and ends at a vertex v with $\deg^{\text{in}}_D(v) > \deg^{\text{out}}_D(v)$. As $|F'| > |F|$, at least one of these paths, P say, has more edges in F' than in F. Now replacing F by $F \triangle EP$ does not violate (20.19)(i), since $\deg_{F \triangle EP}(v) = \deg_F(v) + 1 \leq \deg_{F'}(v) \leq l$ if v is an end of P and $\deg_{F \triangle EP}(v) = \deg_F(v)$ for any other vertex v. As this increases $|F|$, it contradicts our maximality assumption. ∎

This implies the following result of Folkman and Fulkerson [1969], generalizing Theorems 20.8 and 20.9 (by taking $p_2 = 1$):

Corollary 20.10a. *Let $G = (V, E)$ be a bipartite graph and let $k_1, k_2, p_1, p_2 \in \mathbb{Z}_+$ be such that $k_1 + k_2 \geq \Delta(G)$, $k_1p_1 + k_2p_2 = |E|$, and $p_1 \geq p_2$. Then E can be partitioned into k_1 matchings of size p_1 and k_2 matchings of size p_2 if and only if each subset X of V spans at least $k_1(p_1 + |X| - |V|)$ edges.*

Proof. Necessity being easy, we prove sufficiency. By Theorem 20.9, G has k_1 disjoint matchings of size p_1. Let $k := k_1 + k_2$. Since $p_1 \geq p_2$, we have $p_1 \geq (p_1k_1 + p_2k_2)/k = |E|/k$. Hence, by Theorem 20.10, G has k_1 disjoint matchings of size p_1, such that the uncovered edges form a subgraph of maximum degree at most k_2. As this subgraph has $|E| - p_1k_1 = p_2k_2$ edges, by Theorem 20.8 we can split its edge set into k_2 matchings of size p_2. ∎

These results relate to simple b-matchings — see Corollary 21.29a.

20.7. Covering by perfect matchings

A series of results similar to those in Section 20.6 can be derived for covering by perfect matchings and for the down hull of the perfect matching polytope. Brualdi [1979] showed the covering analogue of Corollary 20.9a:

Theorem 20.11. *Let $G = (V, E)$ be a bipartite graph and let $k \in \mathbb{Z}_+$. Then E can be covered by k perfect matchings if and only if any vertex cover X spans at most $k(|X| - \frac{1}{2}|V|)$ edges.*

Proof. *Necessity.* Let G be covered by k perfect matchings and let X be a vertex cover. Each perfect matching contains $|V \setminus X|$ edges not spanned by X, and hence $\frac{1}{2}|V| - |V \setminus X| = |X| - \frac{1}{2}|V|$ edges spanned by X. This proves necessity.

Sufficiency. Assume that the condition holds. This implies that both colour classes of G have size $\frac{1}{2}|V|$, since each of them is a vertex cover X

spanning no edge, implying $|X| \geq \frac{1}{2}|V|$. It also implies that the maximum degree of G is at most k, since for each vertex v the set $U \cup \{v\}$ (where U is the colour class of G not containing v) spans at most k edges.

For each vertex v, let $b_v := k - \deg(v)$. Split each vertex v into b_v vertices, and replace any edge uv by $b_u b_v$ edges connecting the b_u copies of u with the b_v copies of v. This yields the bipartite graph H, with $k|V| - 2|E|$ vertices.

Now H has a perfect matching, as follows from Frobenius' theorem: if Y is a vertex cover in H, then the set X of vertices v of G for which all copies in H belong to Y, is a vertex cover in G. Now by the condition, X spans at most $k(|X| - \frac{1}{2}|V|)$ edges of G. Hence

(20.20) $$|Y| \geq \sum_{v \in X}(k - \deg(v)) = k|X| - |E| - |E[X]| \geq \tfrac{1}{2}k|V| - |E|.$$

So Y is not smaller than half the number of vertices of H. Therefore, by Frobenius' theorem, H has a perfect matching M.

For each edge e of G, add parallel edges to e as often as a copy of e occurs in M. We obtain a k-regular bipartite graph G'. By Kőnig's edge-colouring theorem, the edges of G' can be partitioned into k perfect matchings. This gives k perfect matchings in G covering E. ∎

(This proof method in fact consists of showing that G has a perfect b-matching — see Chapter 21.)

The result is equivalent to characterizing bipartite graphs that are k-regularizable. A graph $G = (V, E)$ is *k-regularizable* if we can replace each edge by a positive number of parallel edges so as to obtain a k-regular graph. Then:

Corollary 20.11a. *Let $G = (V, E)$ be a bipartite graph and let $k \in \mathbb{Z}_+$. Then G is k-regularizable if and only if any vertex cover X spans at most $k(|X| - \frac{1}{2}|V|)$ edges.*

Proof. Directly from Theorem 20.11. ∎

20.7a. Polyhedral interpretation

Again we can interpret Theorem 20.11 polyhedrally. In Theorem 18.9 we saw that for a bipartite graph $G = (V, E)$, the down hull of the perfect matching polytope of G,

(20.21) $$P^{\downarrow}_{\text{perfect matching}}(G) = (P_{\text{perfect matching}}(G) - \mathbb{R}^E_+) \cap \mathbb{R}^E_+$$

is determined by the inequalities

(20.22)
(i) $x_e \geq 0$ for each $e \in E$,
(ii) $x(E[S]) \leq |S| - \frac{1}{2}|V|$ for each vertex cover S.

Then Theorem 20.11 implies that for each $k \in \mathbb{Z}_+$, each integer vector $w \in k \cdot P^{\downarrow}_{\text{perfect matching}}(G)$ is a sum of k integer vectors in $P^{\downarrow}_{\text{perfect matching}}(G)$. That is:

Corollary 20.11b. *$P^{\downarrow}_{\text{perfect matching}}(G)$ has the integer decomposition property.*

Proof. See above. ∎

We can view this result also in terms of the antiblocking polyhedron of $P^{\downarrow}_{\text{perfect matching}}(G)$, which is the polyhedron Q determined by

$$\text{(20.23)} \qquad \begin{array}{lll} \text{(i)} & x_e \geq 0 & \text{for each } e \in E, \\ \text{(ii)} & x(M) \leq 1 & \text{for each perfect matching } M. \end{array}$$

By the theory of antiblocking polyhedra, Q is equal to the down hull of the convex hull of the vectors

$$\text{(20.24)} \qquad \frac{1}{|S|-\frac{1}{2}|V|}\chi^{E[S]}$$

where S is a vertex cover with $|S| > \frac{1}{2}|V|$.

So the maximum value of $\mathbf{1}^{\mathsf{T}}x$ over Q is equal to

$$\text{(20.25)} \qquad \max\{\frac{|E[S]|}{|S|-\frac{1}{2}|V|} \mid S \text{ vertex cover, } |S| > \tfrac{1}{2}|V|\}.$$

By LP-duality, this is equal to the minimum value of $\sum_M \lambda_M$, where M ranges over perfect matchings and where $\lambda_M \geq 0$ with $\sum_M \lambda_M \chi^M \geq \mathbf{1}$. So Theorem 20.11 states: the minimum number of perfect matchings needed to cover all edges in a bipartite graph is equal to

$$\text{(20.26)} \qquad \lceil \min\{\sum_M \lambda_M | \lambda_M \geq 0, \sum_M \lambda_M \chi^M \geq \mathbf{1}\}\rceil.$$

As we can directly extend this to a weighted version, one has:

Corollary 20.11c. *The polyhedron determined by* (20.23) *has the integer rounding property.*

Proof. See above. ∎

20.8. The perfect matching lattice of a bipartite graph

The *perfect matching lattice* (often briefly the *matching lattice*) of a graph $G = (V, E)$ is the lattice generated by the incidence vectors of perfect matchings in G; that is,

$$\text{(20.27)} \qquad L_{\text{perfect matching}}(G) := \text{lattice}\{\chi^M \mid M \text{ perfect matching in } G\}.$$

With the help of Kőnig's edge-colouring theorem, it is not difficult to characterize the perfect matching lattice of a bipartite graph (cf. Lovász [1985]). Recall that the *perfect matching space* of a graph G is the linear hull of the incidence vectors of the perfect matchings in G (cf. Section 18.5d).

Theorem 20.12. *The perfect matching lattice of a bipartite graph $G = (V, E)$ is equal to the set of integer vectors in the perfect matching space of G.*

Proof. Obviously, each vector in the perfect matching lattice is integer and belongs to the perfect matching space. To see the reverse inclusion, let x be an integer vector in the perfect matching space. So $x_e = 0$ for each edge covered by no perfect matching, and $x(\delta(u)) = x(\delta(v))$ for all $u, v \in V$. By adding to x incidence vectors of perfect matchings, we can assume that $x_e \geq 0$ for all $e \in E$.

Replace any edge e by x_e parallel copies. We obtain a k-regular bipartite graph H, with $k := x(\delta(v))$ for any $v \in V$. Hence, by Kőnig's edge-colouring theorem, H is k-edge-colourable. As each colour is a perfect matching in H, we can decompose x as a sum of k incidence vectors of perfect matchings in G. So x belongs to the perfect matching lattice of G. ▮

This gives a characterization of the perfect matching lattice for matching-covered bipartite graphs (which will be used in the characterization of the perfect matching lattice of an arbitrary graph in Chapter 38). A graph is called *matching-covered* if each edge belongs to a perfect matching.

Corollary 20.12a. *Let $G = (V, E)$ be a matching-covered bipartite graph and let $x \in \mathbb{Z}^E$ be such that $x(\delta(u)) = x(\delta(v))$ for any two vertices u and v. Then x belongs to the perfect matching lattice of G.*

Proof. Directly from Theorems 20.12 and 18.7. ▮

By lattice duality theory, Theorem 20.12 is equivalent to the following.

Corollary 20.12b. *Let $G = (V, E)$ be a bipartite graph and let $w \in \mathbb{R}^E$ be a weight function. Then each perfect matching has integer weight if and only if there exists a vector $b \in \mathbb{R}^V$ with $b(V) = 0$ and with $w_e - b_u - b_v$ integer for each edge $e = uv$ covered by at least one perfect matching.*

Proof. Sufficiency is easy, since if such a b exists, then, for each perfect matching M,

$$w(M) = b(V) + \sum_{e=uv \in M} (w_e - b_u - b_v) = \sum_{e=uv \in M} (w_e - b_u - b_v) \tag{20.28}$$

is an integer.

To see necessity, suppose that $w(M)$ is integer for each perfect matching M. Then (by definition of dual lattice) w belongs to the dual lattice of the perfect matching lattice. Theorem 20.12 implies that the dual lattice is the sum of $\mathbb{Z}^E$ and the linear space orthogonal to the perfect matching space. So $w = w' + w''$, where $w' \in \mathbb{Z}^E$ and w'' is orthogonal to the perfect matching space; that is, $w''(M) = 0$ for each perfect matching M. By Corollary 18.7a, there exists a vector $b \in \mathbb{R}^V$ with $b(V) = 0$ and with $w''_e = b_u + b_v$ for each edge $e = uv$ covered by at least one perfect matching. This is equivalent to the present Corollary. ▮

20.9. Further results and notes

20.9a. Some further edge-colouring algorithms

As mentioned, it is easy to implement an $O(nm)$-time algorithm for finding a $\Delta(G)$-edge-colouring in a simple bipartite graph G. Such an algorithm also exists if G has multiple edges:

Theorem 20.13. *The edges of a bipartite graph G can be coloured with $\Delta(G)$ colours in $O(nm)$ time.*

Proof. Let $\Delta := \Delta(G)$. We update a collection of disjoint matchings $M_1, \ldots, M_\Delta$ (the colours), each stored as a doubly linked list. For each edge e, we keep the i for which $e \in M_i$ ($i = 0$ if e is in no M_i). Initially we set $M_i := \emptyset$ for $i := 1, \ldots, \Delta$. We also store the colour classes U and W as lists.

The algorithm runs along all pairs of vertices $u \in U$ and $w \in W$. Fixing $u \in U$ and $w \in W$, make a list L of edges e connecting u and w (taking $O(\deg(u))$ time, by scanning $\delta(u)$); define $d(u, w) := |L|$; make a list I of $d(u, w)$ indices i for which M_i misses u (taking $O(\deg(u))$ time, by scanning $\delta(u)$); make a list J of $d(u, w)$ indices j for which M_j misses w (taking $O(\deg(w))$ time, by scanning $\delta(w)$); next, while there is an edge e_0 in L:

(20.29) choose $i \in I$ and $j \in J$;
if $i = j$, insert e_0 in M_i, delete e_0 from L, and delete i from I and J;
if $i \neq j$, make for each $v \in V$ a list T_v of edges in $M_i \cup M_j$ incident with v (taking $O(n)$ time, by scanning M_i and M_j);
identify the path component P in $M_i \cup M_j$ starting at u (taking $O(n)$ time, using the T_v);
for each edge e on P, if e is in M_i move e to M_j and if e is in M_j we move e to M_i (taking $O(n)$ time);
insert e_0 in M_j, delete e_0 from L, delete i from I, and delete j from J.

Fixing u and w, the preprocessing takes $O(\deg(u) + \deg(w))$ time, and each of the $d(u, w)$ iterations takes $O(n)$ time. As $\sum_{u \in U} \sum_{w \in W} (\deg(u) + \deg(w) + nd(u, w)) = 2nm$, we obtain an algorithm as required. ∎

From their linear-time perfect matching algorithm for regular bipartite graphs, Cole, Ost, and Schirra [2001] derived (using an idea of Gabow [1976c]):

Theorem 20.14. *A k-regular bipartite graph $G = (V, E)$ can be k-edge-coloured in time $O(m \log k)$.*

Proof. We describe a recursive algorithm, the case $k = 1$ being the basis.

If k is even, find an Eulerian orientation of G, let G' be the $\frac{1}{2}k$-regular graph consisting of all edges oriented from one colour class of G to the other, let G'' be the $\frac{1}{2}k$-regular graph consisting of the remaining edges, and recursively $\frac{1}{2}k$-edge-colour G' and G''. This gives a k-edge-colouring of G.

If k is odd and ≥ 3, find a perfect matching M in G, and recursively $(k-1)$-edge-colour $G - M$. With M, this gives a k-edge-colouring of G.

We show that the running time is $O(m \log k)$. The recursive step takes time $O(m)$, since finding an Eulerian orientation or finding a perfect matching takes $O(m)$ time (Corollary 16.11a). Moreover, in one or two recursive steps, the graph is split into two graphs with half the number of edges. Since $m \log_2 k = m + 2(\frac{1}{2} m \log_2(\frac{1}{2}k))$, the result follows. ∎

Corollary 20.14a. *The edges of a bipartite graph G can be coloured with $\Delta(G)$ colours in $O(m \log \Delta(G))$ time.*

Proof. Let $k := \Delta(G)$. First iteratively merge any two vertices in the same colour class of G if each of them has degree at most $\frac{1}{2}k$. The final graph H will have at most two vertices of degree at most $\frac{1}{2}k$, and moreover, $\Delta(H) = k$ and any k-edge-colouring of H yields a k-edge-colouring of G. Next make a copy H' of H, and join each vertex v of H by $k - \deg_H(v)$ parallel edges with its copy v' in H' (where $\deg_H(v)$ is the degree of v in H). This gives the k-regular bipartite graph G', with $|EG'| = O(|EG|)$.

By Theorem 20.14, we can find a k-edge-colouring of G' in $O(m \log k)$ time. This gives a k-edge-colouring of H and hence a k-edge-colouring of G. ∎

20.9b. Complexity survey for bipartite edge-colouring

	$O(nm)$	Kőnig [1916]
	$O(\sqrt{n}\, m\Delta)$	Hopcroft and Karp [1971,1973] (cf. Gabow and Kariv [1978])
*	$O(\tilde{m}^2)$	Gonzalez and Sahni [1976]
	$O(\sqrt{n}\, m \log \Delta)$	Gabow [1976c]
	$O(m\sqrt{n \log n})$	Gabow and Kariv [1978]
	$O(m\Delta \log n)$	Gabow and Kariv [1978]
	$O((m + n^2) \log \Delta)$	Gabow and Kariv [1978,1982]
	$O(m(\log n)^2 \log \Delta)$	Lev, Pippenger, and Valiant [1981]
	$O(m(\log m)^2)$	Gabow and Kariv [1982]
	$O(m \log m)$	Cole and Hopcroft [1982]
*	$O(n\tilde{m} \log \mu)$	Gabow and Kariv [1982]
	$O((m + n \log n \log^2 \Delta) \log \Delta)$	Cole and Hopcroft [1982]
	$O((m + n \log n \log \Delta) \log \Delta)$	Cole [1982]
	$O(n2^{2^{O(\Delta)}})$	Cole [1982]
	$O((m + n \log n) \log \Delta)$	R. Cole and K. Ost (cf. Ost [1995]), Kapoor and Rizzi [2000]
	$O(m\Delta)$	Schrijver [1999]
	$O(m \log \Delta + n \log n \log \Delta)$	Rizzi [2002]
*	$O(m \log \Delta)$	Cole, Ost, and Schirra [2001]

Here $\tilde{m}$ denotes the number of parallel classes of edges, μ the maximum size of a parallel class, and Δ the maximum degree. As before, $*$ indicates an asymptotically best bound in the table.

Kapoor and Rizzi [2000] showed that a bipartite graph of maximum degree Δ can be Δ-edge-coloured in time $T + O(m \log \Delta)$, where T is the time needed to find a perfect matching in a k-regular bipartite graph with m edges and $k \leq \Delta$. (So this is applied only once!)

20.9c. List-edge-colouring

An interesting extension of Kőnig's edge-colouring theorem was shown by Galvin [1995], which was the 'list-edge-colouring conjecture' for bipartite graphs (cf. Alon [1993], Häggkvist and Chetwynd [1992]). It implies the conjecture of J. Dinitz (1979) that the list-edge-colouring number of the complete bipartite graph $K_{n,n}$ equals n. (This is in fact a special case of the conjecture, formulated by V.G. Vizing in 1975, that the list-edge-colouring number of any graph is equal to its edge-colouring number (see Häggkvist and Chetwynd [1992]).) The proof of Galvin is based on the Gale-Shapley theorem on stable matchings (Theorem 18.11).

Let $G = (V, E)$ be a graph. Then G is *k-list-edge-colourable* if for each choice of finite sets L_e for $e \in E$ with $|L_e| = k$, we can choose $l_e \in L_e$ for $e \in E$ such that $l_e \neq l_f$ if e and f are incident. The smallest k for which G is k-list-edge-colourable is called the *list-edge-colouring number* of G.

Trivially, the list-edge-colouring number of G is at least the edge-colouring number of G, and hence at least the maximum degree $\Delta(G)$ of G. Galvin [1995] showed:

Theorem 20.15. *The list-edge-colouring number of a bipartite graph is equal to its maximum degree.*

Proof. Let $G = (V, E)$ be a bipartite graph, with colour classes U and W, and with maximum degree $k := \Delta(G)$. The theorem follows by applying the following statement to any $\Delta(G)$-edge-colouring $\phi : E \to \{1, \ldots, \Delta(G)\}$ of G.

(20.30) Let $\phi : E \to \mathbb{Z}$ be such that $\phi(e) \neq \phi(f)$ if e and f are incident. For each $e = uw \in E$ with $u \in U$ and $w \in W$, let L_e be a finite set satisfying

$$|L_e| > |\{f \in \delta(u) \mid \phi(f) < \phi(e)\}| + |\{f \in \delta(w) \mid \phi(f) > \phi(e)\}|.$$

Then there exist $l_e \in L_e$ $(e \in E)$ such that $l_e \neq l_f$ if e and f are incident.

So it suffices to prove (20.30), which is done by induction on $|E|$. Choose $p \in \bigcup L_e$ and let $F := \{e \in E \mid p \in L_e\}$. Define for each $v \in V$ a total order $<_v$ on $\delta_F(v)$ by:

(20.31)
$$e \leq_v f \iff \phi(e) \geq \phi(f), \text{ if } v \in U,$$
$$e \leq_v f \iff \phi(e) \leq \phi(f), \text{ if } v \in W,$$

for $e, f \in \delta_F(v)$. By the Gale-Shapley theorem (Theorem 18.11), F contains a stable matching M. So M is a matching such that for each $e \in F$ there is an $f \in M$ with $e \leq_v f$ for some $v \in e$. Hence for each edge $e = uw \in F \setminus M$, with $u \in U$ and

$w \in W$: $\exists f \in M \cap \delta(u) : \phi(f) < \phi(e)$ or $\exists f \in M \cap \delta(w) : \phi(f) > \phi(e)$. So removing M from E and resetting $L_e := L_e \setminus \{p\}$ for each $e \in F \setminus M$, we can apply induction. ∎

(The proof by Slivnik [1996] is similar.) An extension of Galvin's theorem was given by Borodin, Kostochka, and Woodall [1997].

20.9d. Further notes

Edge-colouring relates to timetabling — see Appleby, Blake, and Newman [1960], Gotlieb [1963], Broder [1964], Cole [1964], Csima and Gotlieb [1964], Barraclough [1965], Duncan [1965], Almond [1966], Lions [1966b,1966a,1967], Welsh and Powell [1967], Yule [1967], Dempster [1968,1971], Wood [1968], de Werra [1970,1972], and McDiarmid [1972].

However, most practical timetabling problems require more than just bipartite edge-colouring, and are NP-complete. It is NP-complete to decide if a given partial edge-colouring in a bipartite graph can be extended to a minimum edge-colouring (Even, Itai, and Shamir [1975,1976]). This corresponds to a timetabling problem with 'time windows'. Moreover, the 3-dimensional analogue is NP-complete (Karp [1972b]): given three disjoint sets R, S, and T and a family $\mathcal{F}$ of triples $\{r, s, t\}$ with $r \in R$, $s \in S$, and $t \in T$, colour the sets in $\mathcal{F}$ with a minimum number of colours in such a way that sets of the same colour are disjoint.

Analogues of Kőnig's edge-colouring theorem, in terms of odd paths packing and covering, were given by de Werra [1986,1987]. The edge-colouring number of *almost bipartite graphs* (graphs which have a vertex whose deletion makes the graph bipartite) was characterized by Eggan and Plantholt [1986] and Reed [1999b].

Kőnig [1916] also proved an infinite extension of Theorem 20.1. We refer to Section 16.7h for some historical notes on the fundamental paper Kőnig [1916].

Sainte-Laguë [1923] mentioned (without proof and without reference to Kőnig's work) the result that each k-regular bipartite graph is k-edge-colourable.

Chapter 21

Bipartite b-matchings and transportation

The total unimodularity of the incidence matrix of a bipartite graph leads to general min-max relations, for b-matchings, b-edge covers, w-vertex covers, w-stable sets, and b-factors. The weighted versions of these problems relate to the classical transportation problem.
In this chapter, graphs can be assumed to be simple.

21.1. b-matchings and w-vertex covers

Let $G = (V, E)$ be a graph, with $V \times E$ incidence matrix A. We introduce the concepts of b-matching and w-vertex cover, which will turn out to be dual.

For $b : V \to \mathbb{Z}_+$, a *b-matching* is a function $x : E \to \mathbb{Z}_+$ such that for each vertex v of G:

(21.1) $\qquad x(\delta(v)) \leq b_v,$

where $\delta(v)$ is the set of edges incident with v. In other words, x is a b-matching if and only if x is an integer vector satisfying $x \geq \mathbf{0}$, $Ax \leq b$. So if $b = \mathbf{1}$, then b-matchings are precisely the incidence vectors of matchings.

For $w : E \to \mathbb{Z}_+$, a *w-vertex cover* is a function $y : V \to \mathbb{Z}_+$ such that for each edge $e = uv$ of G:

(21.2) $\qquad y_u + y_v \geq w_e.$

In other words, y is a w-vertex cover if and only if y is an integer vector satisfying $y \geq \mathbf{0}$, $y^{\mathsf{T}} A \geq w^{\mathsf{T}}$. So if $w = \mathbf{1}$, then $\{0, 1\}$-valued w-vertex covers are precisely the incidence vectors of vertex covers.

b-matchings and w-vertex covers are related by the following LP-duality equation:

(21.3) $\qquad \max\{w^{\mathsf{T}} x \mid x \geq \mathbf{0}, Ax \leq b\} = \min\{y^{\mathsf{T}} b \mid y \geq \mathbf{0}, y^{\mathsf{T}} A \geq w^{\mathsf{T}}\}.$

Since A is totally unimodular (Theorem 18.2), both optima are attained by integer vectors. In other words (where the *w-weight* of a vector x equals $w^{\mathsf{T}} x$ and the *b-weight* of a vector y equals $y^{\mathsf{T}} b$):

Theorem 21.1. *Let $G = (V, E)$ be a bipartite graph and let $b : V \to \mathbb{Z}_+$ and $w : E \to \mathbb{Z}_+$. Then the maximum w-weight of a b-matching is equal to the minimum b-weight of a w-vertex cover.*

Proof. See above. ▮

Taking $b = \mathbf{1}$, we obtain Corollary 17.1a. For $w = \mathbf{1}$, we get the following min-max relation for maximum-size b-matching (again, the sum of the entries in a vector is called its *size*):

Corollary 21.1a. *Let $G = (V, E)$ be a bipartite graph and let $b : V \to \mathbb{Z}_+$. Then the maximum size of a b-matching is equal to the minimum b-weight of a vertex cover.*

Proof. This is the special case $w = \mathbf{1}$ of Theorem 21.1. ▮

An alternative way of proving this is by derivation from Kőnig's matching theorem: Split each vertex v into b_v copies, and replace each edge uv by $b_u b_v$ edges connecting the b_u copies of u with the b_v copies of v. (This construction is due to Tutte [1954b].)

Corollary 21.1a implies a characterization of the existence of a perfect b-matching. A b-matching is called *perfect* if equality holds in (21.1) for each vertex v. So a b-matching is perfect if and only if it has size $\frac{1}{2}b(V)$. Hence:

Corollary 21.1b. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{Z}_+^V$. Then there exists a perfect b-matching if and only if $b(C) \geq \frac{1}{2}b(V)$ for each vertex cover C.*

Proof. Directly from Corollary 21.1a. ▮

21.2. The *b*-matching polytope and the *w*-vertex cover polyhedron

The total unimodularity of the incidence matrix also implies characterizations of the corresponding polyhedra.

The *b-matching polytope* is the convex hull of the b-matchings. For bipartite graphs it is determined by:

(21.4)
- (i) $x_e \geq 0$ for each $e \in E$,
- (ii) $x(\delta(v)) \leq b_v$ for each $v \in V$.

Theorem 21.2. *The b-matching polytope of a bipartite graph $G = (V, E)$ is determined by* (21.4).

Proof. Directly from the facts that system (21.4) amounts to $x \geq \mathbf{0}, Ax \leq b$ and that A is totally unimodular, where A is the $V \times E$ incidence matrix of G. By Theorem 5.20, the vertices of the polytope $\{x \geq \mathbf{0} \mid Ax \leq b\}$ are integer, hence they are b-matchings. ∎

This generalizes the sufficiency part of Corollary 18.1b.

Similarly, the *w-vertex cover polyhedron*, being the convex hull of the w-vertex covers, is, for bipartite graphs, determined by:

$$\text{(21.5)} \qquad \begin{array}{rll} \text{(i)} & y_v \geq 0 & \text{for each } v \in V, \\ \text{(ii)} & y_u + y_v \geq w_e & \text{for each } e = uv \in E. \end{array}$$

Theorem 21.3. *The w-vertex cover polyhedron of a bipartite graph is determined by* (21.5).

Proof. Directly from the total unimodularity of the incidence matrix of a bipartite graph. ∎

This generalizes the necessity part in Theorem 18.3.

21.3. Simple b-matchings and b-factors

In the context of b-matchings, call a vector x *simple* if it is $\{0,1\}$-valued. So a simple b-matching is the incidence vector of a set F of edges with $\deg_F(v) \leq b_v$ for each vertex v. We will identify the vector and the subset.

To characterize the maximum size of a simple b-matching, let, for any $X \subseteq V$, $E[X]$ denote the set of edges spanned by X.

Theorem 21.4. *The maximum size of a simple b-matching in a bipartite graph $G = (V, E)$ is equal to the minimum value of $b(V \setminus X) + |E[X]|$ taken over $X \subseteq V$.*

Proof. This can be reduced to the nonsimple case by replacing each edge uv by a path of length 3 connecting u and v (thus introducing two new vertices for each edge), and extending b by defining $b(s) := 1$ for each new vertex s. Then the maximum size of a simple b-matching in the original graph is equal to the maximum size of a b-matching in the new graph minus $|E|$, and we can apply Corollary 21.1a. ∎

(This construction is due to Tutte [1954b].)

The theorem can also be derived from the fact that both optima in the LP-duality equation:

$$\text{(21.6)} \qquad \begin{aligned} &\max\{\mathbf{1}^{\mathsf{T}} x \mid \mathbf{0} \leq x \leq \mathbf{1}, Ax \leq b\} \\ &= \min\{y^{\mathsf{T}} b + z^{\mathsf{T}} \mathbf{1} \mid y \geq \mathbf{0}, z \geq \mathbf{0}, y^{\mathsf{T}} A + z^{\mathsf{T}} \geq \mathbf{1}^{\mathsf{T}}\} \end{aligned}$$

have integer optimum solutions, since A (the incidence matrix of G) is totally unimodular.

Theorem 21.4 implies the following result of Ore [1956] (who formulated it in terms of directed graphs). A *b-factor* is a simple perfect b-matching. So it is a subset F of E with $\deg_F(v) = b_v$ for each $v \in V$ (again identifying a subset of E with its incidence vector in $\mathbb{R}^E$).

Corollary 21.4a. *Let $G = (V, E)$ be a bipartite graph and let $b : V \to \mathbb{Z}_+$. Then G has a b-factor if and only if each subset X of V spans at least $b(X) - \frac{1}{2}b(V)$ edges.*

Proof. Directly from Theorem 21.4. ∎

If b is equal to a constant k, Theorem 21.4 amounts to (with the help of Kőnig's edge-colouring theorem):

Corollary 21.4b. *Let $G = (V, E)$ be a bipartite graph and let $k \in \mathbb{Z}_+$. Then the maximum size of the union of k matchings is equal to the minimum value of $k|V \setminus X| + |E[X]|$ taken over $X \subseteq V$.*

Proof. Apply Theorem 21.4 to $b_v := k$ for all $v \in V$. We obtain a formula for the maximum size of a subset F of E with $\deg_F(v) \leq k$ for all $v \in V$. By Theorem 20.1, this is the union of k matchings. ∎

A *k-factor* in a graph $G = (V, E)$ is a subset F of E with $\deg_F(v) = k$ for each $v \in V$. Then:

Corollary 21.4c. *Let $G = (V, E)$ be a bipartite graph and let $k \in \mathbb{Z}_+$. Then G has a k-factor if and only if each subset X of V spans at least $k(|X| - \frac{1}{2}|V|)$ edges.*

Proof. Directly from Corollary 21.4a. ∎

From this one can derive the result of Fulkerson [1964b] (Corollary 20.9a) that a bipartite graph has k disjoint perfect matchings if and only if each subset X of V spans at least $k(|X| - \frac{1}{2}|V|)$ edges.

By the total unimodularity of the incidence matrix of bipartite graphs, the *simple b-matching polytope* (the convex hull of the simple b-matchings) of a bipartite graph $G = (V, E)$ is determined by:

(21.7)
$$\begin{array}{ll} 0 \leq x_e \leq 1 & \text{for each } e \in E, \\ x(\delta(v)) \leq b_v & \text{for each } v \in V. \end{array}$$

Similarly, the following min-max relation for maximum-weight simple b-matching follows (Vogel [1963]):

Theorem 21.5. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{Z}_+^V$ and $w \in \mathbb{Z}_+^E$. Then the maximum weight $w^\mathsf{T}x$ of a simple b-matching x is equal to the minimum value of*

(21.8) $$\sum_{v \in V} y_v b_v + \sum_{e \in E} z_e$$

where $y \in \mathbb{Z}_+^V$ and $z \in \mathbb{Z}_+^E$ with $y_u + y_v + z_e \geq w_e$ for each edge $e = uv$.

Proof. Directly from the LP-duality equation

(21.9) $$\max\{w^{\mathsf{T}} x \mid \mathbf{0} \leq x \leq \mathbf{1}, Ax \leq b\} \\ = \min\{y^{\mathsf{T}} b + z^{\mathsf{T}} \mathbf{1} \mid y \geq \mathbf{0}, z \geq \mathbf{0}, y^{\mathsf{T}} A + z^{\mathsf{T}} \geq w^{\mathsf{T}}\}$$

(where A is the $V \times E$ incidence matrix of G), using the total unimodularity of A. ∎

Moreover:

Theorem 21.6. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{Z}_+^V$ and $w \in \mathbb{Z}_+^E$. Then the minimum weight $w^{\mathsf{T}} x$ of a b-factor x is equal to the maximum value of*

(21.10) $$\sum_{v \in V} y_v b_v + \sum_{e \in E} z_e$$

where $y \in \mathbb{Z}^V$ and $z \in \mathbb{Z}_+^E$ with $y_u + y_v - z_e \leq w_e$ for each edge $e = uv$.

Proof. Directly from the LP-duality equation

(21.11) $$\min\{w^{\mathsf{T}} x \mid \mathbf{0} \leq x \leq \mathbf{1}, Ax = b\} \\ = \max\{y^{\mathsf{T}} b - z^{\mathsf{T}} \mathbf{1} \mid z \geq \mathbf{0}, y^{\mathsf{T}} A - z^{\mathsf{T}} \leq w^{\mathsf{T}}\}$$

(where A is the $V \times E$ incidence matrix of G), using the total unimodularity of A. ∎

Notes. Hartvigsen [1999] gave a characterization of the convex hull of square-free simple 2-matching in a bipartite graph. (A 2-*matching* is a b-matching with $b = \mathbf{2}$. A simple 2-matching is *square-free* if it contains no circuit of length 4.) It implies that a maximum-weight square-free 2-matching in a bipartite graph can be found in strongly polynomial time.

21.4. Capacitated b-matchings

If we require that a b-matching x satisfies $x \leq c$ for some 'capacity' function $c : E \to \mathbb{Z}_+$, we speak of a *capacitated* b-matching. So simple b-matchings correspond to capacitated b-matchings for $c = \mathbf{1}$.

Theorem 21.7. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{Z}_+^V$ and $c \in \mathbb{Z}_+^E$. Then the maximum size of a b-matching $x \leq c$ is equal to*

(21.12) $$\min_{X \subseteq V} b(V \setminus X) + c(E[X]).$$

Proof. The proof is similar to that of Theorem 21.4. Now we define $b(s) := c_e$ if s is a new vertex on the path connecting the end vertices of e. ∎

Alternatively, we can reduce this theorem to Theorem 21.4, by replacing each edge e by c_e parallel edges, or we can use total unimodularity similarly to (21.6).

Again we have the perfect case as direct consequence:

Corollary 21.7a. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{Z}_+^V$ and $c \in \mathbb{Z}_+^E$. Then there exists a perfect b-matching $x \leq c$ if and only if*

(21.13) $$c(E[X]) \geq b(X) - \tfrac{1}{2}b(V)$$

for each $X \subseteq V$.

Proof. Directly from Theorem 21.7. ∎

Again, by the total unimodularity of the incidence matrix of bipartite graphs, the *c-capacitated b-matching polytope* (the convex hull of the b-matchings $x \leq c$) of a bipartite graph $G = (V, E)$ is determined by:

(21.14) $$\begin{array}{ll} 0 \leq x_e \leq c_e & \text{for each } e \in E, \\ x(\delta(v)) \leq b_v & \text{for each } v \in V. \end{array}$$

Similarly, the following min-max relation for maximum-weight capacitated b-matching follows:

Theorem 21.8. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{Z}_+^V$ and $w, c \in \mathbb{Z}_+^E$. Then the maximum weight $w^\mathsf{T}x$ of a b-matching $x \leq c$ is equal to the minimum value of*

(21.15) $$\sum_{v \in V} y_v b_v + \sum_{e \in E} z_e c_e$$

where $y \in \mathbb{Z}_+^V$ and $z \in \mathbb{Z}_+^E$ satisfy $y_u + y_v + z_e \geq w_e$ for each edge $e = uv$.

Proof. Directly from the LP-duality equation

(21.16) $$\begin{aligned} &\max\{w^\mathsf{T}x \mid \mathbf{0} \leq x \leq c, Ax \leq b\} \\ &= \min\{y^\mathsf{T}b + z^\mathsf{T}c \mid y \geq \mathbf{0}, z \geq \mathbf{0}, y^\mathsf{T}A + z^\mathsf{T} \geq w^\mathsf{T}\} \end{aligned}$$

(where A is the $V \times E$ incidence matrix of G), using the total unimodularity of A. ∎

21.5. Bipartite b-matching and w-vertex cover algorithmically

Algorithmically, optimization problems on b-matchings and w-vertex covers in bipartite graphs can be reduced to minimum-cost flow problems, and hence can be solved in strongly polynomial time.

Theorem 21.9. *Given a bipartite graph $G = (V, E)$, $b : V \to \mathbb{Z}_+$, $c : E \to \mathbb{Z}_+$, and $w : E \to \mathbb{Q}$, a b-matching $x \leq c$ maximizing $w^\mathsf{T} x$ can be found in strongly polynomial time. Similarly, a perfect b-matching $x \leq c$ minimizing $w^\mathsf{T} x$ can be found in strongly polynomial time.*

Proof. Let S and T be the colour classes of G, and orient the edges of G from S to T, giving the digraph D. Then b-matchings in G correspond to integer z-transshipments in D with $0 \leq z(v) \leq b(v)$ if $v \in T$ and $-b(v) \leq z(v) \leq 0$ if $v \in S$. Perfect b-matchings correspond to integer b'-transshipments, where $b'(v) := -b(v)$ if $v \in S$ and $b'(v) := b(v)$ if $v \in T$. Hence this theorem follows from Corollary 12.2d. ∎

Wagner [1958] (cf. Dantzig [1955]) observed that the capacitated version of the minimum-weight perfect b-matching problem can be reduced to the uncapacitated version, by a construction similar to that used in proving Theorem 21.4.

One similarly has for w-vertex covers:

Theorem 21.10. *Given a bipartite graph $G = (V, E)$, $b : V \to \mathbb{Q}_+$, $c : V \to \mathbb{Z}_+$, and $w : E \to \mathbb{Z}_+$, a w-vertex cover $y \leq c$ minimizing $y^\mathsf{T} b$ can be found in strongly polynomial time.*

Proof. By reduction to Corollary 12.2e. ∎

Although these results suggest a symmetry between matchings and vertex covers, we mention here that the nonbipartite version of Theorem 21.9 holds true (Section 32.4), but that finding a maximum-size stable set in a nonbipartite graph is NP-complete (see Section 64.2).

21.6. Transportation

The minimum-weight perfect b-matching problem is close to the classical transportation problem. Given a bipartite graph $G = (V, E)$ and a vector $b \in \mathbb{R}^V_+$, a *b-transportation* is a vector $x \in \mathbb{R}^E_+$ with

(21.17) $$x(\delta(v)) = b_v$$

for each $v \in V$. So a b-transportation is a fractional version of a perfect b-matching. Integer b-transportations are exactly the perfect b-matchings.

The following characterization of the existence of a b-transportation was shown (in a much more general form) by Rado [1943] — compare Corollary 21.1b:

Theorem 21.11. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{R}^V_+$. Then there exists a b-transportation if and only if $b(C) \geq \frac{1}{2} b(V)$ for each vertex cover C.*

Proof. Since the inequalities $b(C) \geq \frac{1}{2}b(V)$ (for vertex covers C), define a rational polyhedral cone, we can assume that b is rational, and hence, by scaling, that b is integer. Then the theorem follows from Corollary 21.1b. ∎

Note that, trivially, there exists a b-transportation if and only if b belongs to the convex cone in $\mathbb{R}^V$ generated by the incidence vectors of the edges of G. So Theorem 21.11 characterizes this cone.

A negative cycle criterion follows directly from the corresponding criterion for transshipments. For any b-transportation x in a bipartite graph $G = (V, E)$ and any cost function $c : E \to \mathbb{R}$, make the directed graph $D_x = (V, A)$ as follows. Let U and W be the colour classes of G. For each edge $e = uv$ of G, with $u \in U$ and $v \in W$, let A have an arc (u, v) of cost c_e, and, if $x_e > 0$, an arc (v, u).of cost $-c_e$. Then (Tolstoĭ [1930]):

Theorem 21.12. *x is a minimum-cost b-transportation if and only if D_x has no negative-cost directed circuits.*

Proof. Directly from Theorem 12.3. ∎

Transportations in a complete bipartite graph can be formulated in terms of matrices. Fixing vectors $a \in \mathbb{R}^m_+$ and $b \in \mathbb{R}^n_+$, an $m \times n$ matrix $X = (x_{i,j})$ is called a *transportation* if

(21.18)
$$\begin{aligned}
&\text{(i)} && x_{i,j} \geq 0 && i = 1, \ldots, m; j = 1, \ldots, n,\\
&\text{(ii)} && \sum_{j=1}^{n} x_{i,j} = a_i && i = 1, \ldots, m,\\
&\text{(iii)} && \sum_{i=1}^{m} x_{i,j} = b_j && j = 1, \ldots, n.
\end{aligned}$$

Clearly, a transportation exists if and only if $\sum_i a_i = \sum_j b_j$.

Given an $m \times n$ 'cost' matrix $C = (c_{i,j})$, the cost of a transportation $X = (x_{i,j})$ is defined as $\sum_{i,j} c_{i,j}x_{i,j}$. Then the *transportation problem* (also called the *Hitchcock-Koopmans transportation problem*) is:

(21.19) given: vectors $a \in \mathbb{Q}^m_+$, $b \in \mathbb{Q}^n_+$ and an $m \times n$ 'cost' matrix $C = (c_{i,j})$,
find: a minimum-cost transportation.

So it is equivalent to solving the LP problem of minimizing $\sum_{i,j} c_{i,j}x_{i,j}$ over (21.18). The transportation problem formed a major impulse to introduce linear programming. Hitchcock [1941] and Dantzig [1951a] showed that the simplex method applies to the transportation problem.

The transportation problem is also a special case of the minimum-cost b-transshipment problem, and hence can be solved with the methods of Chapter 12. In particular, it is solvable in strongly polynomial time.

Linear programming also yields a min-max relation, originally due to Hitchcock [1941] (also implicit in Kantorovich [1939]):

Theorem 21.13 (Hitchcock's theorem)**.** *The minimum cost of a transportation is equal to the maximum value of $y^{\mathsf{T}}a + z^{\mathsf{T}}b$, where $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ such that $y_i + z_j \leq c_{i,j}$ for all i, j.*

Proof. This is LP-duality. ∎

(Hitchcock [1941] gave a direct proof.)

The transportation problem differs from the minimum-weight perfect b-matching problem in having a complete bipartite graph $K_{m,n}$ as underlying bipartite graph and in not requiring integrality of the output. This last however is not a restriction, as Dantzig [1951a] showed:

Theorem 21.14. *If a and b are integer, the transportation problem has an integer optimum solution x.*

Proof. Directly from the total unimodularity of the matrix underlying the system (21.18), which is the incidence matrix of the complete bipartite graph $K_{m,n}$. ∎

For a different proof, see the proof of Corollary 21.15a below.

Notes. Ford and Fulkerson [1955,1957b], Gleyzal [1955], Munkres [1957], and Egerváry [1958] described primal-dual methods for the transportation problem, and Ford and Fulkerson [1956a,1957a] extended it to the capacitated version.

If the a_i and b_j are small integers, the transportation problem can be reduced to the assignment problem, by 'splitting' each i into a_i or b_i copies. (This observation is due to Egerváry [1958], and in a different context to Tutte [1954b].)

21.6a. Reduction of transshipment to transportation

It is direct to transform a transportation problem to a transshipment problem. Orden [1955] observed a reverse reduction (similar to the reduction described in Section 16.7c). Indeed, let input $D = (V, A)$, $b \in \mathbb{R}^V$ and $k \in \mathbb{R}^A$ for the transshipment problem be given. Split each vertex v into two vertices v', v'' and replace each arc (u, v) by an arc (u', v''), with cost $k(u, v)$. Moreover, add arcs (v', v''), each with cost 0. Let $N := \sum_{v \in V} |b(v)|$. Define $b'(v') := -N$ and $b'(v'') := b(v) + N$. Then a minimum-cost b'-transshipment in the new structure gives a minimum-cost b-transshipment in the original structure. Since the new graph is bipartite with all edges oriented from one colour class to the other, we have a reduction to the transportation problem.

(Orden [1955] also gave an alternative reduction of the transshipment problem to the transportation problem. Let A' be the set of pairs (u, v) with $b_u < 0$ and $b_v > 0$ and with v is reachable in D from U. For each $(u, v) \in A'$, let $k'(u, v)$ be the length of a shortest $u - v$ path in D, taking k as length function. Then the (bipartite) transshipment problem for $D' := (V, A')$, b, and k' is equivalent to the original transshipment problem.)

Fulkerson [1960] gave the following reduction of the *capacitated* transshipment problem to the uncapacitated transportation problem. Let be given directed graph

$D = (V, A)$, $b \in \mathbb{R}^V$, a 'capacity' function $c \in \mathbb{R}^A$, and a 'cost' function $k \in \mathbb{R}^A$. Define $V' := V \cup A$ and $E' := \{\{a, v\} \mid a = (v, u) \text{ or } a = (u, v)\}$. Define $w(\{a, v\}) := k(a)$ if v is head of a, and $:= 0$ if v is tail of a. Let $b'(a) := c(a)$ and $b'(v) := b(v) + c(\delta^{\text{out}}(v))$. Then a minimum-cost b-transshipment subject to c corresponds to a minimum-cost b'-transportation. (More can be found in Wagner [1958].)

21.6b. The transportation polytope

Given $a \in \mathbb{R}^m_+$ and $b \in \mathbb{R}^n_+$, the *transportation polytope* is the set of all matrices $X = (x_{i,j})$ in $\mathbb{R}^{m \times n}$ satisfying (21.18). The transportation polytope was first studied by Hitchcock [1941]. The following result is due to Dantzig [1951a].

Theorem 21.15. *Let $X = (x_{i,j})$ belong to the transportation polytope. Then X is a vertex of the transportation polytope if and only if the set $F := \{ij \mid x_{i,j} > 0\}$ forms a forest in the complete bipartite graph $K_{m,n}$.*

Proof. If F contains a circuit $C = (i_0, j_1, i_1, j_2, i_2, \ldots, j_k, i_k)$, with $i_k = i_0$, define $Y = (y_{i,j})$ by: $y_{i,j} := 1$ if $(i, j) = (i_h, j_h)$ for some $h = 1, \ldots, k$, $y_{i,j} := -1$ if $(i, j) = (i_{h-1}, j_h)$ for some $h = 1, \ldots, k$, and $y_{i,j} := 0$ for all other (i, j). Then $X + \varepsilon Y$ belongs to the transportation polytope for any ε close enough to 0 (positive or negative), and hence X is not a vertex of the transportation polytope.

Conversely, if X is not a vertex of the transportation polytope, there exists a nonzero matrix $Y = (y_{i,j})$ such that $X + \varepsilon Y$ is in the transportation polytope for any ε close enough to 0 (positive or negative). Then Y satisfies $\sum_{j=1}^n y_{i,j} = 0$ for $i = 1, \ldots, m$ and $\sum_{i=1}^m y_{i,j} = 0$ for $j = 1, \ldots, n$. Since Y is nonzero, the set $F' := \{ij \mid y_{i,j} \neq 0\}$ contains a circuit. Since $F' \subseteq F$, it implies that F contains a circuit. ∎

This gives:

Corollary 21.15a. *If a and b are integer vectors, the transportation polytope is an integer polyhedron.*

Proof. By Theorem 21.15, for any vertex $X = (x_{i,j})$ of the transportation polytope, the set of pairs (i, j) with $x_{i,j}$ not an integer is a forest. Hence, if it is nonempty, this forest has an end edge, say (i, j). Assume without loss of generality that i has degree 1 in this forest. Then $x_{i,j}$ is equal to a_i minus $\sum_{j' \neq j} x_{i,j'}$, which is an integer as a_i and each of the $x_{i,j'}$ $(j' \neq j)$ is an integer. ∎

The dimension of the transportation polytope is easy to determine (Koopmans and Reiter [1951], Dulmage and Mendelsohn [1962], Klee and Witzgall [1968]):

Theorem 21.16. *If $a > 0$ and $b > 0$, the dimension of the transportation polytope is equal to $(m - 1)(n - 1)$.*

Proof. Let $X = (x_{i,j})$ be a vector in the relative interior of the transportation polytope. So $x_{i,j} > 0$ for all i, j. For each (i, j) with $i \in \{1, \ldots, m - 1\}$ and $j \in \{1, \ldots, n - 1\}$, we can correct any small perturbation of $x_{i,j}$ by a unique change of the $x_{i,n}$ and $x_{m,j}$. So the dimension of the transportation polytope is $(m-1)(n-1)$. ∎

Notes. Balinski [1974] (cf. Balinski and Rispoli [1993]) showed the Hirsch conjecture for some classes of transportation polytopes. For counting and estimating the number of vertices of transportation polytopes, see Simonnard and Hadley [1959], Demuth [1961], Wintgen [1964], Szwarc and Wintgen [1965], Klee and Witzgall [1968], Bolker [1972], and Ahrens [1981]. For counting facets, see Klee and Witzgall [1968].

Given $C = (c_{i,j}) \in \mathbb{R}^{m\times n}$, the *dual transportation polyhedron* is the set of all vectors $(u;v) \in \mathbb{R}^m \times \mathbb{R}^n$ satisfying[38]:

(21.20)
$$\begin{aligned} &u_1 = 0 \\ &u_i + v_j \geq c_{i,j} \qquad i = 1,\ldots,m; j = 1,\ldots,n. \end{aligned}$$

(The condition $u_1 = 0$ is added for normalization.) It is easy to see that the dimension of the dual transportation polyhedron is $m + n - 1$, and that $(u;v)$ satisfying (21.20) is a vertex of the dual transportation polyhedron if and only if the graph with vertex set $\{p_1,\ldots,p_m,q_1,\ldots,q_n\}$ and edge set $\{\{p_i,q_j\} \mid u_i + v_j = c_{i,j}\}$ is connected.

Balinski [1984] showed with the 'signature method' that the diameter of the dual transportation polyhedron is at most $(m-1)(n-1)$, thus proving the Hirsch conjecture for this class of polyhedra.

Balinski and Russakoff [1984] characterized vertices and higher-dimensional faces of dual transportation polyhedra. More can be found in Zhu [1963], Balinski [1983], and Kleinschmidt, Lee, and Schannath [1987].

21.7. b-edge covers and w-stable sets

Exchanging $\leq$ and $\geq$ appropriately in the definitions of b-matchings and w-vertex covers gives the b-edge covers and the w-stable sets. These concepts again turn out to be each others dual.

Let $G = (V,E)$ be a graph, with $V \times E$ incidence matrix A. For $b : V \to \mathbb{Z}_+$, a *b-edge cover* is a function $x : E \to \mathbb{Z}_+$ such that for each vertex v of G:

(21.21) $\quad x(\delta(v)) \geq b_v.$

In other words, x is a b-edge cover if and only if x is an integer vector satisfying $x \geq \mathbf{0}$, $Ax \geq b$. So if $b = \mathbf{1}$, then $\{0,1\}$-valued b-edge covers are precisely the incidence vectors of edge covers.

For $w : E \to \mathbb{Z}_+$, a *w-stable set* is a function $y : V \to \mathbb{Z}_+$ such that for each edge $e = uv$ of G:

(21.22) $\quad y_u + y_v \leq w_e.$

[38] We write $(u;v)$ for $\begin{pmatrix} u \\ v \end{pmatrix}$.

In other words, y is a w-stable set if and only if y is an integer vector satisfying $y \geq \mathbf{0}$, $y^\mathsf{T} A \leq w^\mathsf{T}$. So if $w = \mathbf{1}$, then $\{0,1\}$-valued w-stable sets are precisely the incidence vectors of stable sets.

In this case, b-edge covers and w-stable sets are related by the following LP-duality equation:

(21.23) $$\min\{w^\mathsf{T} x \mid x \geq \mathbf{0}, Ax \geq b\} = \max\{y^\mathsf{T} b \mid y \geq \mathbf{0}, y^\mathsf{T} A \leq w^\mathsf{T}\}.$$

Since A is totally unimodular (Theorem 18.2), both optima are attained by integer vectors. This gives (where the *w-weight* of a vector x equals $w^\mathsf{T} x$ and the *b-weight* of a vector y equals $y^\mathsf{T} b$):

Theorem 21.17. *Let $G = (V, E)$ be a bipartite graph and let $b : V \to \mathbb{Z}_+$ and $w : E \to \mathbb{Z}_+$. Then the minimum w-weight $w^\mathsf{T} x$ of a b-edge cover x is equal to the maximum b-weight of a w-stable set.*

Proof. See above. ∎

Taking $b = \mathbf{1}$, we obtain Corollary 19.5a. For $w = \mathbf{1}$, we get a min-max relation for minimum-size b-edge cover:

Corollary 21.17a. *Let $G = (V, E)$ be a bipartite graph and let $b : V \to \mathbb{Z}_+$. Then the minimum size of a b-edge cover is equal to the maximum b-weight of a stable set.*

Proof. This is the special case $w = \mathbf{1}$ of Theorem 21.17. ∎

Again, an alternative way of proving this is by derivation from the Kőnig-Rado edge cover theorem (Theorem 19.4): Split each vertex v into b_v copies, replace each edge uv by $b_u b_v$ edges connecting the b_u copies of u with the b_v copies of v.

21.8. The b-edge cover and the w-stable set polyhedron

The total unimodularity of the incidence matrix of a bipartite graph also gives descriptions of the corresponding polyhedra.

The *b-edge cover polyhedron* is the convex hull of the b-edge covers. For bipartite graphs it is determined by:

(21.24)
(i) $x_e \geq 0$ for each $e \in E$,
(ii) $x(\delta(v)) \geq b_v$ for each $v \in V$.

Theorem 21.18. *The b-edge cover polyhedron of a bipartite graph $G = (V, E)$ is determined by* (21.24).

Proof. Directly from the facts that system (21.24) amounts to $x \geq \mathbf{0}, Ax \geq b$ and that A is totally unimodular. ∎

This extends Theorem 19.6 on the edge cover polytope.

Similarly, the *w-stable set polyhedron*, being the convex hull of the w-stable sets, is, for bipartite graphs, determined by:

(21.25) (i) $y_v \geq 0$ for each $v \in V$,
(ii) $y_u + y_v \leq w_e$ for each $e = uv \in E$.

Theorem 21.19. *The w-stable set polyhedron of a bipartite graph is determined by* (21.25).

Proof. Directly from the total unimodularity of the incidence matrix of a bipartite graph. ∎

This generalizes the necessity part of Theorem 19.7.

21.9. Simple b-edge covers

Again, call a vector x *simple* if it is $\{0,1\}$-valued. Then a simple b-edge cover corresponds to a set F of edges with $\deg_F(v) \geq b_v$ for each $v \in V$. We will identify the vector and the set. Note that a simple b-edge cover can exist only if $b_v \leq \deg(v)$ for each vertex v.

It is easy to derive the following min-max relation for simple b-edge covers from Theorem 21.4 on the maximum size of a simple b-matching ($E[X]$ denote the set of edges spanned by X):

Theorem 21.20. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{Z}_+^V$ with $b_v \leq \deg(v)$ for each vertex v. Then the minimum size of a simple b-edge cover in G is equal to the maximum value of $b(X) - |E[X]|$ taken over $X \subseteq V$.*

Proof. Define $b'(v) := \deg(v) - b(v)$ for each vertex v. Then a subset F of E is a simple b-edge cover if and only if $E \setminus F$ is a simple b'-matching. By Theorem 21.4, the maximum size of a simple b'-matching is equal to the minimum value of $b'(V \setminus X) + |E[X]|$ taken over $X \subseteq V$. Hence the minimum size of a simple b-edge cover is equal to the maximum value of

$$\begin{aligned}(21.26)\qquad & |E| - b'(V \setminus X) - |E[X]| = |E| - \sum_{v \in V \setminus X} (\deg(v) - b(v)) - |E[X]| \\ & = |E| - 2|E[V \setminus X]| - |\delta(X)| + b(V \setminus X) - |E[X]| \\ & = b(V \setminus X) - |E[V \setminus X]|,\end{aligned}$$

taken over $X \subseteq V$. ∎

Alternatively, the theorem follows from the fact that both optima in the LP-duality equation (where A is the $V \times E$ incidence matrix of G):

$$\begin{aligned}(21.27)\qquad & \min\{\mathbf{1}^{\mathsf{T}} x \mid \mathbf{0} \leq x \leq \mathbf{1}, Ax \geq b\} \\ & = \max\{y^{\mathsf{T}} b - z^{\mathsf{T}} \mathbf{1} \mid y \geq \mathbf{0}, z \geq \mathbf{0}, y^{\mathsf{T}} A - z^{\mathsf{T}} \leq \mathbf{1}^{\mathsf{T}}\}\end{aligned}$$

have integer optimum solutions, since A is totally unimodular.

If b is equal to a constant k, Theorem 21.20 amounts to (with the help of the edge cover variant of Kőnig's edge-colouring theorem (Theorem 20.5)):

Corollary 21.20a. *Let $G = (V, E)$ be a bipartite graph and let $k \in \mathbb{Z}_+$. Then the minimum size of the union of k disjoint edge covers is equal to the maximum value of $k|X| - |E[X]|$ taken over $X \subseteq V$.*

Proof. Apply Theorem 21.20 to $b_v := k$ for all $v \in V$. We obtain a formula for the maximum size of a subset F of E with $\deg_F(v) \geq k$ for all $v \in V$. By Theorem 20.5, F is the union of k disjoint edge covers. ∎

By the total unimodularity of the incidence matrix of bipartite graphs, the *simple b-edge cover polytope* (the convex hull of the simple b-edge covers) of a bipartite graph $G = (V, E)$ is determined by:

(21.28)
$$\begin{aligned} 0 \leq x_e \leq 1 \quad & \text{for each } e \in E, \\ x(\delta(v)) \geq b_v \quad & \text{for each } v \in V. \end{aligned}$$

LP-duality also gives a min-max formula for the minimum weight of simple b-edge covers:

Theorem 21.21. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{Z}_+^V$ and $w \in \mathbb{Z}_+^E$. Then the minimum weight $w^\mathsf{T} x$ of a simple b-edge cover x is equal to the maximum value of*

(21.29)
$$\sum_{v \in V} y_v b_v - \sum_{e \in E} z_e$$

where $y \in \mathbb{Z}_+^V$ and $z \in \mathbb{Z}_+^E$ with $y_u + y_v - z_e \leq w_e$ for each edge $e = uv$.

Proof. Directly from the LP-duality equation

(21.30)
$$\begin{aligned} & \min\{w^\mathsf{T} x \mid \mathbf{0} \leq x \leq \mathbf{1}, Ax \geq b\} \\ & = \max\{y^\mathsf{T} b - z^\mathsf{T} \mathbf{1} \mid y \geq \mathbf{0}, z \geq \mathbf{0}, y^\mathsf{T} A - z^\mathsf{T} \leq w^\mathsf{T}\} \end{aligned}$$

(where A is the $V \times E$ incidence matrix of G), using the total unimodularity of A. ∎

21.10. Capacitated b-edge covers

If we require that a b-edge cover x satisfies $x \leq c$ for some 'capacity' function $c : E \to \mathbb{Z}_+$, we speak of a *capacitated* b-edge cover. So simple b-edge covers correspond to capacitated b-edge covers with $c = \mathbf{1}$.

Theorem 21.22. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{Z}_+^V$ and $c \in \mathbb{Z}_+^E$ with $c(\delta(v)) \geq b_v$ for each $v \in V$. Then the minimum size of a b-edge cover $x \leq c$ is equal to*

(21.31) $$\max_{X \subseteq V} b(X) - c(E[X]).$$

Proof. The proof is similar to that of Theorem 21.20. ∎

Alternatively, we can reduce this theorem to Theorem 21.20, by replacing each edge e by c_e parallel edges, or we can use total unimodularity similarly to (21.27).

Theorem 21.23. *Let $G = (V, E)$ be a bipartite graph and let $b \in \mathbb{Z}_+^V$ and $c, w \in \mathbb{Z}_+^E$. Then the minimum weight $w^\mathsf{T} x$ of a b-edge cover $x \leq c$ is equal to the maximum value of*

(21.32) $$\sum_{v \in V} y_v b_v - \sum_{e \in E} z_e c_e$$

where $y \in \mathbb{Z}_+^V$ and $z \in \mathbb{Z}_+^E$ with $y_u + y_v - z_e \leq w_e$ for each edge $e = uv$.

Proof. Directly from the LP-duality equation

(21.33) $$\begin{aligned} &\min\{w^\mathsf{T} x \mid \mathbf{0} \leq x \leq c, Ax \geq b\} \\ &= \max\{y^\mathsf{T} b - z^\mathsf{T} c \mid y \geq \mathbf{0}, z \geq \mathbf{0}, y^\mathsf{T} A - z^\mathsf{T} \leq w^\mathsf{T}\} \end{aligned}$$

(where A is the $V \times E$ incidence matrix of G), using the total unimodularity of A. ∎

By the total unimodularity of the incidence matrix of G, the convex hull of b-edge covers $x \leq c$ of a bipartite graph G is determined by the inequalities

(21.34)
(i) $0 \leq x_e \leq c_e$ for each $e \in E$,
(ii) $x(\delta(v)) \geq b_v$ for each $v \in V$.

21.11. Relations between *b*-matchings and *b*-edge covers

Like for matchings and edge covers, there is also a close relation between maximum-size b-matchings and minimum-size b-edge covers, as was shown by Gallai [1959a]. This gives a connection between Corollaries 21.1a and 21.17a.

Let $G = (V, E)$ be an undirected graph without isolated vertices, and let $b \in \mathbb{Z}_+^V$. Define:

(21.35) $$\begin{aligned} \nu_b(G) &:= \text{the maximum size of a } b\text{-matching}, \\ \rho_b(G) &:= \text{the minimum size of a } b\text{-edge cover.} \end{aligned}$$

Theorem 21.24. *Let $G = (V, E)$ be an undirected graph without isolated vertices, and let $b \in \mathbb{Z}_+^V$. Then*

(21.36) $$\nu_b(G) + \rho_b(G) = b(V).$$

Proof. This can be reduced to Gallai's theorem (Theorem 19.1), by splitting each vertex v into b_v copies, and replacing each edge $e = uv$ by $b_u b_v$ edges connecting the b_u copies of u with the b_v copies of v. ∎

A direct proof of the previous theorem is given in the proof of the following theorem, also due to Gallai [1959a]:

Theorem 21.25. *Let $G = (V, E)$ be an undirected graph and let $b \in \mathbb{Z}_+^V$. Then for each maximum-size b-matching x there is a minimum-size b-edge cover y with $x \leq y$. Conversely, for each minimum-size b-edge cover y there is a maximum-size b-matching x with $x \leq y$.*

Proof. Let x be a maximum-size b-matching. For each vertex v of G, increase the value of x on some edge incident with v, by $b_v - x(\delta(v))$. We obtain a b-edge cover y satisfying

$$(21.37) \qquad y(E) = x(E) + \sum_{v \in V} (b_v - x(\delta(v))) = b(V) - x(E).$$

Conversely, let y be a minimum-size b-edge cover. For each vertex v of G, decrease the value of y on edges incident with v, by a total amount of $y(\delta(v)) - b_v$ (as long as $y \geq \mathbf{0}$). We obtain a b-matching x satisfying

$$(21.38) \qquad x(E) \geq y(E) - \sum_{v \in V} (y(\delta(v)) - b_v) = b(V) - y(E).$$

(21.37) and (21.38) imply that the y (x, respectively) obtained from x (y, respectively) is optimum, thus showing the theorem, and also showing (21.36). ∎

In a bipartite graph, a minimum-size b-edge cover and a maximum-weight stable set can be found in strongly polynomial time, by reduction to Theorem 21.9:

Corollary 21.25a. *Given a bipartite graph $G = (V, E)$ and $b \in \mathbb{Z}_+^V$, a minimum-size b-edge cover and a maximum b-weight stable set can be found in strongly polynomial time.*

Proof. Since stable sets are exactly the complements of vertex covers, finding a maximum b-weight stable sets is directly reduced to finding a minimum b-weight vertex cover. The construction given in the proof of Theorem 21.25 implies that a maximum-size b-matching gives a minimum-size b-edge cover in polynomial time. So Theorem 21.9 gives the present corollary. ∎

Moreover, for the weighted case:

Theorem 21.26. *A minimum-weight capacitated b-edge cover in a bipartite graph can be found in strongly polynomial time.*

Proof. Directly from Corollary 12.2d, by orienting the edges from one colour class to the other. ∎

21.12. Upper and lower bounds

We finally consider upper *and* lower bounds. That is, for a graph $G = (V, E)$ and $a, b \in \mathbb{R}^V$ and $d, c \in \mathbb{R}^E$, we consider vectors $x \in \mathbb{R}^E$ satisfying:

(21.39) (i) $d_e \leq x_e \leq c_e$ for each $e \in E$,
(ii) $a_v \leq x(\delta(v)) \leq b_v$ for each $v \in V$,

If integer, x is both a b-matching and an a-edge cover.

The optimization problem can be reduced again to minimum-cost circulation, and hence:

Theorem 21.27. *Given $w : E \to \mathbb{Q}$, an integer vector x maximizing $w^{\mathsf{T}}x$ over* (21.39) *can be found in strongly polynomial time.*

Proof. This is a special case of Corollary 12.2d, by orienting the edges of G from one colour class to the other. ∎

Corresponding min-max and polyhedral characterizations directly follow from LP-duality and the total unimodularity of the incidence matrix of G. We formulate them for existence and optimum size of solutions of (21.39).

The following was formulated by Kellerer [1964]:

Theorem 21.28. *Let $G = (V, E)$ be a bipartite graph and let $a, b \in \mathbb{Z}^V$ and $d, c \in \mathbb{Z}^E$ with $a \leq b$ and $d \leq c$. Then there exists an $x \in \mathbb{Z}^E$ satisfying* (21.39) *if and only if for each $X \subseteq V$ one has*

(21.40) $$c(E[X]) - d(E[V \setminus X]) \geq \max\{a(S \cap X) - b(T \setminus X), a(T \cap X) - b(S \setminus X)\},$$

where S and T are the colour classes of G.

Proof. From Corollary 11.2i, by orienting all edges from S to T and taking $U := (S \setminus X) \cup (T \cap X)$. ∎

This theorem has several special cases. For $d = \mathbf{0}$ it implies the following result due to Fulkerson [1959a] (a generalization of Theorem 16.8):

Corollary 21.28a. *Let $G = (V, E)$ be a bipartite graph with colour classes S and T, let $a, b \in \mathbb{Z}_+^V$ with $a \leq b$, and let $c \in \mathbb{Z}_+^E$. Then there is a vector $x \leq c$ that is both a b-matching and an a-edge cover if and only if there exist $y \in \mathbb{Z}_+^E$ and $z \in \mathbb{Z}_+^E$ with $y \leq c$ and $z \leq c$, such that*

(21.41) *$y(\delta(v)) \leq b_v$ and $z(\delta(v)) \geq a_v$ for each $v \in S$ and*
$y(\delta(v)) \geq a_v$ and $z(\delta(v)) \leq b_v$ for each $v \in T$.

Proof. Note that (21.40) can be decomposed into two inequalities, one involving $a|S$ and $b|T$ only, the other involving $a|T$ and $b|S$ only[39]. This gives the present corollary. ∎

The special case $d = \mathbf{0}$, $c = \mathbf{1}$ is:

Corollary 21.28b. *Let $G = (V, E)$ be a bipartite graph with colour classes S and T and let $a, b \in \mathbb{Z}_+^V$ with $a \leq b$. Then E has a subset F that is both a b-matching and an a-edge cover if and only if E has subsets F' and F'' such that F' contains at least a_v edges covering v if $v \in S$ and at most b_v edges covering v if $v \in T$, and F'' contains at least a_v edges covering v if $v \in T$ and at most b_v edges covering v if $v \in S$.*

Proof. Directly from Corollary 21.28a by taking $c = \mathbf{1}$. ∎

A min-max relation for such vectors can be derived from Hoffman's circulation theorem (Theorem 11.2):

Theorem 21.29. *Let $G = (V, E)$ be a bipartite graph and let $a, b \in \mathbb{Z}^V$ and $d, c \in \mathbb{Z}^E$, such that there exists an $x \in \mathbb{Z}^E$ satisfying* (21.39). *Then the minimum size of such a vector x is equal to*

(21.42) $$\max_{Z \subseteq V}(a(Z) - c(E[Z]) + d(E[V \setminus Z])),$$

while the maximum size of such a vector x is equal to

(21.43) $$\min_{Z \subseteq V}(c(E[V \setminus Z]) - d(E[Z]) + b(Z)).$$

For each integer value τ between (21.42) *and* (21.43) *there exists such a vector x of size τ.*

Proof. Choose $\tau \in \mathbb{Z}$. Make a directed graph $D = (V, A)$ as follows.

Let S and T be the colour classes of G. Orient each edge of G from S to T. Add new vertices s and t. For each $v \in S$, make an arc from s to v, with $d(s, v) := a_v$ and $c(s, v) := b_v$. For each $v \in T$, make an arc from v to t, with $d(v, t) := a_v$ and $c(v, t) := b_v$. Finally, make an arc from t to s with $d(t, s) := c(t, s) := \tau$.

It suffices to show that D has a circulation x satisfying $d \leq x \leq c$ if and only if τ is between (21.42) and (21.43). We do this by using Hoffman's circulation theorem. Choose a subset X of the vertex set of D. Consider Hoffman's condition:

(21.44) $$d(\delta^{\text{in}}(X)) \leq c(\delta^{\text{out}}(X)).$$

Since by assumption some vector x satisfying (21.39) exists, (21.44) holds if $s, t \in X$ or $s, t \notin X$ (as ignoring the bounds on (t, s) there is a circulation).

If $s \in X$ and $t \notin X$, we have

[39] $f|X$ denotes the restriction of a function f to a set X.

(21.45) $d(\delta^{\text{in}}(X)) = \tau + d(E[(S \setminus X) \cup (T \cap X)])$

and

(21.46) $c(\delta^{\text{out}}(X)) = b(S \setminus X) + c(E[(S \cap X) \cup (T \setminus X)]) + b(T \cap X)$.

Hence (21.44) for such X is equivalent to

(21.47) $\tau \leq b(Z) + c(E[V \setminus Z]) - d(E[Z])$

for all $Z \subseteq V$ (take $Z = (S \setminus X) \cup (T \cap X)$). That is, to τ being at most (21.43).

If $t \in X$ and $s \notin X$, we have

(21.48) $d(\delta^{\text{in}}(X)) = a(S \cap X) + d(E[(S \setminus X) \cup (T \cap X)]) + a(T \setminus X)$

and

(21.49) $c(\delta^{\text{out}}(X)) = \tau + c(E[(S \cap X) \cup (T \setminus X)])$.

Hence (21.44) for such X is equivalent to

(21.50) $\tau \geq a(Z) - c(E[Z]) + d(E[V \setminus Z])$

for all $Z \subseteq V$ (take $Z = (S \cap X) \cup (T \setminus X)$). That is, to τ being at least (21.42). ∎

A special case is the following theorem of Folkman and Fulkerson [1969]:

Corollary 21.29a. *Let $G = (V, E)$ be a bipartite graph, let $a, b \in \mathbb{Z}_+^V$, and let $\tau \in \mathbb{Z}_+$. Then E has a subset F with $a_v \leq \deg_F(v) \leq b_v$ for each $v \in V$ and with $|F| = \tau$ if and only if*

(21.51) $|E[Z]| \geq \max\{a(Z) - \tau, \tau - b(V \setminus Z), a(S \cap Z) - b(T \setminus Z), a(T \cap Z) - b(S \setminus Z)\}$

for each $Z \subseteq V$, where S and T are the colour classes of G.

Proof. Directly from Theorems 21.28 and 21.29. ∎

21.13. Further results and notes

21.13a. Complexity survey on weighted bipartite *b*-matching and transportation

Complexity survey for weighted b-matching in bipartite graphs ($*$ indicates an asymptotically best bound in the table):

$O(n^4 B)$	Munkres [1957]
$O(\beta \cdot \text{MF}(n, m, B))$	Ford and Fulkerson [1955,1957b]

≫

continued

	$O(n^2 mB)$	Iri [1960]
*	$O(\beta \cdot \mathrm{SP}_+(n,m,W))$	Edmonds and Karp [1970]
	$O(nW \cdot \mathrm{MF}(n,m,B))$	Edmonds and Karp [1972]
*	$O(m \log B \cdot \mathrm{SP}_+(n,m,W))$	Edmonds and Karp [1972]
	$O(nm \log(nB))$	Dinits [1973a]
	$O(n \log \beta \cdot \mathrm{SP}_+(n,m,W))$	Lawler [1976b]
	$O(n \log W \cdot \mathrm{MF}(n,m,B))$	Röck [1980]
	$O(m^2 \log n \cdot \mathrm{MF}(n,m,B))$	Tardos [1985a]
*	$O(\beta^{3/4} m \log W)$	Gabow [1985b]
*	$O(\beta^{1/2} n^{1/3} m \log W)$	Gabow [1985b] *for simple graphs*
	$O(n^2 \log n \cdot \mathrm{SP}_+(n,m,W))$	Galil and Tardos [1986,1988]
	$O(nm \log(n^2/m) \log(nW))$	Goldberg and Tarjan [1987,1990]
	$O(n \log n(m + n \log n))$	Orlin [1988,1993]
*	$O((\beta^{1/2} m + \beta \log \beta) \log(nW))$	Gabow and Tarjan [1989]
*	$O(n_1 m + n_1^3 \log(n_1 W))$	Ahuja, Orlin, Stein, and Tarjan [1994]
*	$O(n_1 m \log(2 + \frac{n_1^2}{m} \log(n_1 W)))$	Ahuja, Orlin, Stein, and Tarjan [1994]
*	$O(n \log n(m + n_1 \log n_1))$	Kleinschmidt and Schannath [1995]

Here $B := \|b\|_\infty$, $\beta := \|b\|_1$, $W := \|w\|_\infty$ (assumed to be integer), and $n_1 := \min\{|S|, |T|\}$, where S and T are the colour classes of the bipartite graph. By $\mathrm{SP}_+(n,m,W)$ we denote the time required for solving a shortest path problem in a digraph with n vertices, m arcs, and nonnegative integer length function l with $\|l\|_\infty \leq W$. $\mathrm{MF}(n,m,B)$ denotes the time required to solve a maximum flow problem in a digraph with n vertices, m arcs, and integer capacity function c with $\|c\|_\infty \leq B$.

Complexity survey for the uncapacitated transportation problem:

	$O(n^4 B)$	Munkres [1957]
	$O(\beta \cdot \mathrm{MF}(n, n^2, B))$	Ford and Fulkerson [1955,1957b]
*	$O(n^3 \log(nB))$	Edmonds and Karp [1972], Dinits [1973a]
	$O(n^4 W)$	Edmonds and Karp [1972]
*	$O(\beta^{3/4} n^2 \log W)$	Gabow [1985b]
*	$O(\beta^{1/2} n^{7/3} \log W)$	Gabow [1985b]
	$O(n^4 \log n \cdot \mathrm{MF}(n, n^2, W))$	Tardos [1985a]
	$O(n^4 \log n)$	Galil and Tardos [1986,1988]
	$O(n^3 \log(nW))$	Goldberg and Tarjan [1987,1990]

≫

continued

	$O(n^3 \log n)$	Orlin [1988,1993]
*	$O((\beta^{1/2}n^2 + \beta \log \beta) \log(nW))$	Gabow and Tarjan [1989]
*	$O(n_1 n^2 + n_1^3 \log(n_1 W))$	Ahuja, Orlin, Stein, and Tarjan [1994]
*	$O(n_1 n^2 \log(2 + \frac{n_1^2}{n^2} \log(n_1 W)))$	Ahuja, Orlin, Stein, and Tarjan [1994]
*	$O(n_1^2 n \log^2 n)$	Tokuyama and Nakano [1992,1995]
*	$O(n_1 n^2 \log n)$	Kleinschmidt and Schannath [1995]

Complexity survey for weighted capacitated b-matching in bipartite graphs:

*	$O(n \max\{B, C\} \cdot \mathrm{SP}_+(n, m, W))$	Edmonds and Karp [1970]
	$O(nW \cdot \mathrm{MF}(n, m, \max\{B, C\}))$	Edmonds and Karp [1972]
*	$O(n \log \beta \cdot \mathrm{SP}_+(n, m, W))$	Lawler [1976b]
	$O(n \log W \cdot \mathrm{MF}(n, m, \max\{B, C\}))$	Röck [1980]
	$O(m^2 \log n \cdot \mathrm{MF}(n, m, \max\{B, C\}))$	Tardos [1985a]
	$O(\beta^{3/4} mC \log W)$	Gabow [1985b]
	$O(\beta^{1/2} n^{1/3} mC \log W)$	Gabow [1985b]
	$O(n^2 \log n \cdot \mathrm{SP}_+(n, m, W))$	Galil and Tardos [1986,1988]
*	$O(nm \log(n^2/m) \log(nW))$	Goldberg and Tarjan [1987,1990]
*	$O(m \log n \cdot \mathrm{SP}_+(n, m, W))$	Orlin [1988,1993]
*	$O(\beta^{1/2} mC \log(nW))$	Gabow and Tarjan [1988b,1989]
*	$O(n^{2/3} mC^{4/3} \log(nW))$	Gabow and Tarjan [1989]
*	$O((\beta^{1/2} m + \beta \log \beta) \log(nW))$	Gabow and Tarjan [1989]
*	$O((nm + \beta \log \beta) \log(nW))$	Gabow and Tarjan [1989]
*	$O(n_1 m + n_1^3 \log(n_1 W))$	Ahuja, Orlin, Stein, and Tarjan [1994]
*	$O(n_1 m \log(2 + \frac{n_1^2}{m} \log(n_1 W)))$	Ahuja, Orlin, Stein, and Tarjan [1994]

Here $C := \|c\|_\infty$.

Complexity survey for the capacitated transportation problem:

	$O(n^4 W)$	Edmonds and Karp [1972]
	$O(n^3 \log \max\{B, C\})$	Edmonds and Karp [1972]
	$O(n^4 \log W)$	Röck [1980]
	$O(n^4 \log n \cdot \mathrm{MF}(n, n^2, \max\{B, C\}))$	Tardos [1985a]
*	$O(n^2 B)$	Gabow [1985b]

$\gg$

continued

	$O(\beta^{3/4}n^2C\log W)$	Gabow [1985b]
	$O(\beta^{1/2}n^{7/3}C\log W)$	Gabow [1985b]
*	$O(n^3\log(nW))$	Goldberg and Tarjan [1987,1990]
*	$O(n^4\log n)$	Galil and Tardos [1986,1988], Orlin [1988,1993]
*	$O(n^2\beta^{1/2}C\log(nW))$	Gabow and Tarjan [1988b,1989]
*	$O(n^{8/3}C^{4/3}\log(nW))$	Gabow and Tarjan [1989]
*	$O((\beta^{1/2}n^2+\beta\log\beta)\log(nW))$	Gabow and Tarjan [1989]
*	$O(n_1n^2+n_1^3\log(n_1W))$	Ahuja, Orlin, Stein, and Tarjan [1994]
*	$O(n_1n^2\log(2+\frac{n_1^2}{n^2}\log(n_1W)))$	Ahuja, Orlin, Stein, and Tarjan [1994]

Let $G=(V,E)$ be a bipartite graph, with colour classes S and T say. The existence of a perfect (capacitated) b-matching can be reduced quite directly to the problem of finding a maximum $s-t$ flow in the digraph obtained from G by adding two new vertices s and t, orienting each edge from S to T, and adding an arc (s,s') for each $s'\in S$, and adding an arc (t',t) for each $t'\in T$. Similarly, a maximum (capacitated) b-matching can be found.

It implies that if $\mathrm{MF}(n,m,C)$ is the running time of a maximum flow algorithm for inputs with n vertices, m arcs, and integer capacity function c with $\|c\|_\infty\le C$, then a maximum-size (capacitated) b-matching can be found in time $O(\mathrm{MF}(n,m,C))$, for bipartite graphs with n vertices, m edges and $b\in\mathbb{Z}^V$ satisfying $\|b\|_\infty\le C$ (and capacity function $c\in\mathbb{Z}^E$ satisfying $\|c\|_\infty\le C$).

In some cases, one can obtain better bounds, in particular if one of the colour classes is considerably smaller than the other. To this end, let $n_1:=\min\{|S|,|T|\}$. Implementing the shortest augmenting path rule described in Section 10.5, then gives an $O(n_1m^2)$ running time, since a shortest $s-t$ path has length at most $2n_1+1=O(n_1)$, implying that the number of iterations is bounded by n_1m.

Similarly, the blocking flow method of Dinits [1970] described in Section 10.6 can be performed in $O(n_1^2m)$ time, since the bound in Theorem 10.6 becomes $O(n_1m)$, while there are $O(n_1)$ blocking flow iterations. The method of Karzanov [1974] can be sharpened to $O(n_1^2n)$, as was shown by Gusfield, Martel, and Fernández-Baca [1987]. Ahuja, Orlin, Stein, and Tarjan [1994] gave a method taking the minimum of $O(n_1m+n_1^3)$, $O(n_1m+n_1^2\sqrt{m})$, $O(n_1m+n_1^2\sqrt{\log C})$, and $O(n_1m\log(2+\frac{n_1^2}{m}))$ time.

For the special case where $b_u=1$ for each u in the smaller colour class, Adel'son-Vel'skiĭ, Dinits, and Karzanov [1975] gave an $O(n_1^{5/3}n)$ algorithm for finding a b-factor.

21.13b. The matchable set polytope

Let $G = (V, E)$ be a graph. A subset X of V is called *matchable*, if G has a matching M with $\bigcup M = X$; that is, if the subgraph $G[X]$ of G induced by X has a perfect matching.

The *matchable set polytope* of G is the convex hull of the incidence vectors of matchable sets. Theorem 21.11 implies a characterization of the matchable set polytope in case G is bipartite.

For any graph, each vector in the matchable set polytope trivially satisfies:

(21.52) (i) $0 \leq x_v \leq 1$ for each $v \in V$,
(ii) $x(C) \leq \frac{1}{2}x(V)$ for each stable set C.

If G is bipartite, this set of inequalities determines the matchable set polytope, a result of Balas and Pulleyblank [1983]:

Theorem 21.30. *If G is bipartite, the matchable set polytope is determined by* (21.52).

Proof. Let x satisfy (21.52). By Theorem 21.11, there exists an x-transportation $y \in \mathbb{R}_+^E$. That is, $x = Ay$, where A is the $V \times E$ incidence matrix of G.

As x satisfies (21.52)(i), y satisfies $y \geq \mathbf{0}$, $Ay \leq \mathbf{1}$. So, by Corollary 18.1b, y belongs to the matching polytope of G. So y is a convex combination of vectors χ^M, where M ranges over the matchings in G. Then x is a convex combination of the vectors χ^S, where S is matchable (that is, the set of vertices covered by some matching M). This follows from the fact that $A\chi^M = \chi^S$ if M is a matching and S is the set of vertices covered by M.

So x belongs to the matchable set polytope. ∎

It is easy to check that only for bipartite graphs the matchable set polytope is determined by (21.52).

Note that for bipartite graphs $G = (V, E)$, by Theorem 21.11, condition (21.52)(ii) is equivalent to x belonging to the convex cone generated by the incidence vectors (in $\mathbb{R}^V$) of edges, considered as subsets of V.

Qi [1987] gave an algorithm for the separation problem for the matchable set polytope of a bipartite graph. For more on the matchable set polytope, see Balas and Pulleyblank [1983] and Section 25.5d.

21.13c. Existence of matrices

If the bipartite graph is a complete bipartite graph, theorems on the existence of b-matchings and b-edge covers amount to theorems on the existence of matrices obeying prescribed bounds on the row and column sums. This gives the following theorem of Gale [1956,1957] and Ryser [1957]:

Theorem 21.31 (Gale-Ryser theorem)**.** *Let $a, b \in \mathbb{Z}_+^m$ and $a', b' \in \mathbb{Z}_+^n$ with $a \leq b$ and $a' \leq b'$ and satisfying $a_1 \geq a_2 \geq \cdots \geq a_m$ and $a'_1 \geq a'_2 \geq \cdots \geq a'_n$. Then there exists a $\{0,1\}$-valued $m \times n$ matrix with ith row sum between a_i and b_i ($i = 1, \ldots, m$) and jth column sum between a'_j and b'_j ($j = 1, \ldots, n$) if and only if*

(21.53) (i) $\displaystyle\sum_{i=1}^{k} a_i \leq \sum_{j=1}^{n} \min\{k, b'_j\}$ *for all* $k = 1, \ldots, m$,

(ii) $\displaystyle\sum_{j=1}^{k} a'_j \leq \sum_{i=1}^{m} \min\{k, b_i\}$ *for all* $k = 1, \ldots, n$.

Proof. *Necessity.* Consider any inequality in (21.53)(i). The number of 1's in rows $1, \ldots, k$ is at least the left-hand side and at most the right-hand side. This proves necessity of the inequality. Necessity of the inequalities (ii) is shown similarly.

Sufficiency. This follows from Theorem 21.28 applied to the complete bipartite graph $G = K_{m,n}$. Then we must show that for each $I \subseteq \{1, \ldots, m\}$ and $J \subseteq \{1, \ldots, n\}$ one has:

(21.54) $|I| \cdot |J| \geq \max\{a(I) - b'(\overline{J}), a'(J) - b(\overline{I})\}$,

where $\overline{I} := \{1, \ldots, m\} \setminus I$ and $\overline{J} := \{1, \ldots, n\} \setminus J$. By symmetry, it suffices to show

(21.55) $|I| \cdot |J| \geq a(I) - b'(\overline{J})$.

This follows from (21.53)(i), since

(21.56) $\displaystyle a(I) \leq \sum_{i=1}^{|I|} \leq \sum_{j=1}^{n} \min\{|I|, b'_j\} \leq |J| \cdot |I| + b'(\overline{J})$

for any $J \subseteq \{1, \ldots, n\}$. ∎

(Gale [1956,1957] proved this theorem for $a = \mathbf{0}$ and $b' = \infty$, and Ryser [1957] for $a = b$ and $a' = b'$.)

Corollary 21.28a due to Fulkerson [1959a], is equivalent to the following result extending the Gale-Ryser theorem:

Theorem 21.32. *Let* $(c_{i,j})$ *be a nonnegative* $m \times n$ *matrix and let* $a, b \in \mathbb{Z}_+^m$ *and* $a', b' \in \mathbb{Z}_+^n$ *with* $a \leq b$ *and* $a' \leq b'$. *Then there exists an integer* $m \times n$ *matrix* $(x_{i,j})$ *satisfying*

(21.57) (i) $0 \leq x_{i,j} \leq c_{i,j}$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$,

(ii) $\displaystyle a_i \leq \sum_{j=1}^{n} x_{i,j} \leq b_i$ for all $i = 1, \ldots, m$,

(iii) $\displaystyle a'_j \leq \sum_{i=1}^{m} x_{i,j} \leq b'_j$ for all $j = 1, \ldots, n$,

if and only if there exist an $m \times n$ *matrix* $(x'_{i,j})$ *satisfying*

(21.58) (i) $0 \leq x'_{i,j} \leq c_{i,j}$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$,

(ii) $\displaystyle\sum_{j=1}^{n} x'_{i,j} \leq b_i$ for all $i = 1, \ldots, m$,

(iii) $\displaystyle a'_j \leq \sum_{i=1}^{m} x'_{i,j}$ for all $j = 1, \ldots, n$,

and an $m \times n$ *matrix* $(x''_{i,j})$ *satisfying*

$$
\text{(21.59)} \quad
\begin{array}{lll}
\text{(i)} & 0 \le x''_{i,j} \le c_{i,j} & \text{for all } i = 1, \ldots, m \text{ and } j = 1, \ldots, n, \\
\text{(ii)} & a_i \le \sum_{j=1}^{n} x''_{i,j} & \text{for all } i = 1, \ldots, m, \\
\text{(iii)} & \sum_{i=1}^{m} x''_{i,j} \le b'_j & \text{for all } j = 1, \ldots, n.
\end{array}
$$

Proof. This is equivalent to Corollary 21.28a. ∎

21.13d. Further notes

Corollary 11.2c implies the following result of Hoffman [1956a]. Let $G = (V, E)$ be a bipartite graph and let $0 < \alpha < 1$. Then E has a subset F such that

$$
\text{(21.60)} \quad \lfloor \tfrac{\deg_E(v)}{\alpha} \rfloor \le \deg_F(v) \le \lceil \tfrac{\deg_E(v)}{\alpha} \rceil
$$

for each vertex v.

Ikura and Nemhauser [1982] gave a strongly polynomial-time primal simplex algorithm for the maximum-weight stable set problem in bipartite graphs (the number of pivot steps is at most n^2; the method corresponds to a strongly polynomial-time dual simplex algorithm for the minimum-size b-edge cover problem, which is a special case of a minimum-flow problem). (An improvement was given by Armstrong and Jin [1996].) An interior-point method was described by Mizuno and Masuzawa [1989]. For more on capacitated b-matchings (in terms of matrices), see Anstee [1983].

We refer for further notes on algorithmic aspects of the transportation problem to Section 12.5d on the equivalent transshipment problem.

Heller [1963,1964] gave necessary and sufficient conditions for a linear program to be equivalent to a transportation problem. Katerinis [1987] and Enomoto, Ota, and Kano [1988] gave sufficient conditions for bipartite graphs to have a k-factor.

Goodman, Hedetniemi, and Tarjan [1976] gave a linear-time algorithm finding a maximum-weight simple b-matching in a tree.

Faster algorithms for transportation problems where the cost satisfies a quadrangle inequality where given by Karp and Li [1975] and Aggarwal, Bar-Noy, Khuller, Kravets, and Schieber [1995].

Variants of the transportation problem (minimax, bottleneck) were investigated by Szwarc [1966,1971], Hammer [1969,1971], Garfinkel and Rao [1971], Srinivasan and Thompson [1972a,1972b,1976], Derigs and Zimmermann [1979], Derigs [1982], Russell, Klingman, and Partow-Navid [1983], and Ahuja [1986]. Prager [1957b] and Kellerer [1961] gave a generalization.

Prager [1955] gave an extension to quadratic cost functions, i.e. given $b \in \mathbb{R}^m$, $d \in \mathbb{R}^n$, and $c_{i,j} \ge 0$, $q_{i,j} \ge 0$ ($i = 1, \ldots, m$; $j = 1, \ldots, n$):

$$
\text{(21.61)} \quad
\begin{array}{lll}
\text{minimize} & \sum_{i=1}^{m} \sum_{j=1}^{n} (c_{i,j} x_{i,j} + q_{i,j} x_{i,j}^2), & \\
\text{subject to} & \sum_{j=1}^{n} x_{i,j} = b_i & \text{for } i = 1, \ldots, m, \\
 & \sum_{i=1}^{m} x_{i,j} = d_j & \text{for } j = 1, \ldots, n, \\
 & x_{i,j} \ge 0 & \text{for } i = 1, \ldots, m; j = 1, \ldots, n.
\end{array}
$$

Among the books surveying transportation are Ford and Fulkerson [1962], Dantzig [1963], Murty [1976,1983], Bazaraa and Jarvis [1977], Papadimitriou and Steiglitz [1982], Gondran and Minoux [1984], Derigs [1988a], Nemhauser and Wolsey [1988], and Bazaraa, Jarvis, and Sherali [1990].

21.13e. Historical notes on the transportation and transshipment problems

Transportation can be considered as the special case of transshipment where all arcs are oriented from a source to a sink. By the techniques described in Section 21.6a, transshipment problems can be reduced conversely to transportation problems. This makes the history of the two problems intertwined. We should notice also that the transshipment problems studied by Kantorovich and Koopmans were in fact transportation problems, due to the fact that their cost functions are metrics.

Tolstoĭ

The first to study the transportation problem mathematically seems to be A.N. Tolstoĭ. In the collection *Transportation Planning, Volume I* of the National Commissariat of Transportation of the Soviet Union, Tolstoĭ [1930] published an article called *Methods of finding the minimal total kilometrage in cargo-transportation planning in space*. In it, Tolstoĭ described a number of approaches to solve the transportation problem, illuminated by applications to the transportation of salt, cement, and other cargo between sources and destination points along the railway network of the Soviet Union. He seems to be the first to give a negative cycle criterion for optimality. Moreover, a for that time large-scale instance of the transportation problem was solved to optimality.

First, Tolstoĭ considered the problem for the case where there are two sources. He observed that in that case one can order the destination points by the difference between the distances to the two sources. In that case, one source can provide the destinations starting from the beginning of the list, until the supply of that source has been used up. The other source supplies the remaining demands. Tolstoĭ observed that the list is independent of the supplies and demands, and hence

> such table is applicable for the whole life-time of factories, or sources of production.
> Using this table, one can immediately compose an optimal transportation plan every year, given quantities of output produced by these two factories and demands of the destination points.

Next, Tolstoĭ studied the transportation problem for the case where all sources and destinations are along one circular railway line. In this case, considering the negative cycle criterion yields directly the optimum solution. He calls this phenomenon 'circle dependency'.

Finally, Tolstoĭ combined the two methods into a heuristic to solve a concrete transportation problem coming from cargo transportation along the Soviet railway network. The problem has 10 sources and 68 sinks, and 155 links between sources and sinks (all other distances are taken infinite):

	Arkhangelsk	Yaroslavl'	Murom	Balakhonikha	Dzerzhinsk	Kishert'	Sverdlovsk	Artemovsk	Iledzhk	Dekonskaya	*demand:*
Agryz				709	1064	693					2
Aleksandrov					397			1180			4
Almaznaya								81		65	1.5
Alchevskaya								106		114	4
Baku								1554		1563	10
Barybino								985		968	2
Berendeevo		135			430						10
Bilimbai						200	59				1
Bobrinskaya								655		663	10
Bologoe		389						1398			1
Verkhov'e								678		661	1
Volovo								757		740	3
Vologda	634					1236					2
Voskresensk				427				1022		1005	1
V.Volochek		434						1353		1343	5
Galich	815	224				1056					0.5
Goroblagodatskaya						434	196				0.5
Zhlobin								882		890	8
Zverevo								227		235	5
Ivanovo					259						6
Inza				380	735					1272	2
Kagan								2445	2379		0.5
Kasimov			0								1
Kinel'				752		1208			454	1447	2
Kovylkino				355						1213	2
Kyshtym						421	159				3
Leningrad	1237	709						1667		1675	55
Likino			223		328						15
Liski								443		426	1
Lyuberdzhy			268		411					1074	1
Magnitogorskaya						932	678		818		1
Mauk						398	136				5
Moskva			288	378	405			1030		1022	141
Navashino			12	78							2
Nizhegol'								333		316	1
Nerekhta		50			349						5
Nechaevskaya			92								0.5
N.-Novgorod					32						25
Omsk						1159	904		1746		5
Orenburg									76		1.5
Penza				411				1040	883	1023	7
Perm'	1749					121					1
Petrozavodsk	1394										1
Poltoradzhk								1739	3085	1748	4
Pskov								1497		1505	10
Rostov/Don								287		296	20
Rostov/Yarosl		56			454						2
Rtishchevo								880		863	1
Savelovo		325						1206		1196	5
Samara				711					495	1406	7
San-Donato						416	157				1
Saratov								1072		1055	15
Sasovo				504				1096		1079	1
Slavyanoserbsk								119		115	1.1
Sonkovo		193						1337			0.5
Stalingrad								624		607	15.4
St.Russa		558						1507		1515	5
Tambov								783		766	4
Tashkent								3051	1775		3
Tula								840		848	8
Tyumen'						584	329				6
Khar'kov								251		259	60
Chelyabinsk						511	257		949		2
Chishmy				1123		773			889		0.5
Shchigry								566		549	4
Yudino				403	757	999					0.5
Yama								44		52	5
Yasinovataya								85		93	6
supply:	5	11.5	8.5	12	100	12	15	314	10	55	543

Table of distances (in kilometers) between sources and destinations, and of supplies and demands (in kilotons).
Tolstoĭ gave no distance for Kasimov. We have inserted a distance 0 to Murom, since from Tolstoĭ's solution it appears that Kasimov is connected only to Murom (by a waterway). Hence the distance is irrelevant.

Tolstoĭ's heuristic also makes use of insight into the geography of the Soviet Union. He goes along all sources (starting with the most remote source), where, for each source X, he lists those sinks for which X is the closest source or the second closest source. Based on the difference of the distances to the closest and second closest sources, he assigns cargo from X to the sinks, until the supply of X has been used up. In case Tolstoĭ foresees circle dependency, he deviates from this rule to avoid that a negative-length circuit would arise. No backtracking occurs.

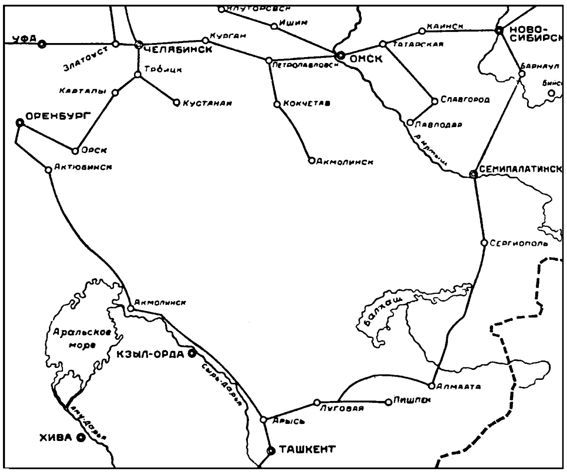


Figure 21.1
Figure from Tolstoĭ [1930] to illustrate a negative cycle.

In the following quotation, Tolstoĭ considers the cycles Dzerzhinsk-Rostov-Yaroslavl'-Leningrad-Artemovsk-Moscow-Dzerzhinsk and Dzerzhinsk-Nerekhta-Yaroslavl'-Leningrad-Artemovsk-Moscow-Dzerzhinsk. It is the sixth step in his method, after the transports from the factories in Iletsk, Sverdlovsk, Kishert', Balakhonikha, and Murom have been set:

> 6. The Dzerzhinsk factory produces 100,000 tons. It can forward its production only in the Northeastern direction, where it sets its boundaries in interdependency with the Yaroslavl' and Artemovsk (or Dekonskaya) factories.

	From Dzerzhinsk	From Yaroslavl'	Difference to Dzerzhinsk
Berendeevo	430 km	135 km	−295 km
Nerekhta	349 ,,	50 ,,	−299 ,,
Rostov	454 ,,	56 ,,	−398 ,,

	From Dzerzhinsk	From Artemovsk	Difference to Dzerzhinsk
Aleksandrov	397 km	1,180 km	+783 km
Moscow	405 ,,	1,030 ,,	+625 ,,

> The method of differences does not help to determine the boundary between the Dzerzhinsk and Yaroslavl' factories. Only the circle dependency, specified to be

an interdependency between the Dzerzhinsk, Yaroslavl' and Artemovsk factories, enables us to exactly determine how far the production of the Dzerzhinsk factory should be advanced in the Yaroslavl' direction.
Suppose we attach point Rostov to the Dzerzhinsk factory; then, by the circle dependency, we get:

Dzerzhinsk-Rostov	454 km	−398 km	Nerekhta	349 km	−299 km
Yaroslavl'- ,,	56 ,,		,,	50 ,,	
Yaroslavl'-Leningrad	709 ,,	+958 ,,	These points remain		
Artemovsk- ,,	1,667 ,,		unchanged because only the		
Artemovsk-Moscow	1,030 ,,	−625 ,,	quantity of production sent		
Dzerzhinsk- ,,	405 ,,		by each factory changes		
Total		−65 km			+34 km

Therefore, the attachment of Rostov to the Dzerzhinsk factory causes over-run in 65 km, and only Nerekhta gives a positive sum of differences and hence it is the last point supplied by the Dzerzhinsk factory in this direction.
As a result, the following points are attached to the Dzerzhinsk factory:

N. Novgorod	25,000 tons	
Ivanova	6,000 ,,	
Nerekhta	5,000 ,,	
Aleksandrov	4,000 ,,	
Berendeevo	10,000 ,,	
Likino	15,000 ,,	
Moscow	35,000 ,,	(remainder of factory's production)
Total	100,000 tons	

After 10 steps, when the transports from all 10 factories have been set, Tolstoĭ 'verifies' the solution by considering a number of cycles in the network, and he concludes that his solution is optimum:

> Thus, by use of successive applications of the method of differences, followed by a verification of the results by the circle dependency, we managed to compose the transportation plan which results in the minimum total kilometrage.

The objective value of Tolstoĭ's solution is 395,052 kiloton-kilometers. Solving the problem with modern linear programming tools (CPLEX) shows that Tolstoĭ's solution indeed is optimum. But it is unclear how sure Tolstoĭ could have been about his claim that his solution is optimum. Geographical insight probably has helped him in growing convinced of the optimality of his solution. On the other hand, it can be checked that there exist feasible solutions that have none of the negative-cost cycles considered by Tolstoĭ in their residual graph, but that are yet not optimum[40].

In the September 1939 issue of *Sotsialisticheskiĭ Transport*, Tolstoĭ [1939] published an article *Methods of removing irrational transportations in planning*, in which he again described his method of 'circle dependency', and applied it to the planning of driving empty cars and transporting heavy cargoes on the U.S.S.R. railway network. In this paper, Tolstoĭ restricted himself to sources and sinks arranged along a circular railway line, for which he gave his 'circle dependency' method:

[40] The maximum objective value of a feasible solution, whose residual graph contains no nonnegative-cost cycle of length 4, and none of the seven longer nonnegative-length cycles considered by Tolstoĭ (of lengths 6 and 8), is equal to 397,226.

> Before counting distances from cargo-senders to points of destination which form a circle dependency, it is necessary to attach points of destination to cargo-senders with complete distribution of waggons. In case of circle dependency determined by geographical location it can be done without special calculations. Then, by calculation of km in circle dependency, the initial attachment can be verified and if not correct, then it can be improved.

Tolstoĭ illustrated the method by the circuit Smolensk - Vitebsk - Velikiye-Luki - Zemtsy - Rzhev - Vyazma - Smolensk of the U.S.S.R. network. A negative-length directed circuit in the auxiliary directed graph gives an improvement, as in the following Table given by Tolstoĭ [1939]:

Source of cargoes	Amount km	Difference of distance	Amount of carriages
Vyazma-Smolensk	176	-37	$4-3=1$
Vitebsk ,,	139		$0+3=4$
Vitebsk-V. Luki	156	-37	$3-3=0$
Zemtsy ,,	119		$2+3=5$
Zemtsy-Rzhev	123	$+7$	$5-3=2$
Vyazma ,,	130		$1+3=4$

Altogether . . . -67

Tolstoĭ then remarked:

> The negative total difference shows that the distribution was wrong and that there is an over-run of 67 km for every waggon which goes from upper cargo-senders.

According to Kantorovich [1987], there were some attempts to introduce Tolstoĭ's work by the appropriate department of the People's Commissariat of Transport. Tolstoĭ's method was also explained in the book *Planning Goods Transportation* by Pariĭskaya, Tolstoĭ, and Mots [1947].

Kantorovich

Apparently unaware (by that time) of the work of Tolstoĭ, L.V. Kantorovich studied a general class of problems, that includes the transportation problem. It formed a major impulse to the study of linear programming. In his memoirs, Kantorovich [1987] writes:

> Once some engineers from the veneer trust laboratory came to me for consultation with a quite skilful presentation of their problems. Different productivity is obtained for veneer-cutting machines for different types of materials; linked to this the output of production of this group of machines depended, it would seem, on the chance factor of which group of raw materials to which machine was assigned. How could this fact be used rationally?
> This question interested me, but nevertheless appeared to be quite particular and elementary, so I did not begin to study it by giving up everything else. I put this question for discussion at a meeting of the mathematics department, where there were such great specialists as Gyunter, Smirnov himself, Kuz'min, and Tartakovskii. Everyone listened but no one proposed a solution; they had already turned to someone earlier in individual order, apparently to Kuz'min. However, this question nevertheless kept me in suspense. This was the year of my marriage, so I was also distracted by this. In the summer or after the vacation

> concrete, to some extent similar, economic, engineering, and managerial situations started to come into my head, that also required the solving of a maximization problem in the presence of a series of linear constraints.
> In the simplest case of one or two variables such problems are easily solved—by going through all the possible extreme points and choosing the best. But, let us say in the veneer trust problem for five machines and eight types of materials such a search would already have required solving about a billion systems of linear equations and it was evident that this was not a realistic method. I constructed particular devices and was probably the first to report on this problem in 1938 at the October scientific session of the Herzen Institute, where in the main a number of problems were posed with some ideas for their solution.
> The universality of this class of problems, in conjunction with their difficulty, made me study them seriously and bring in my mathematical knowledge, in particular, some ideas from functional analysis.

In a footnote, Kantorovich's son V.L. Kantorovich adds:

> In L.V. Kantorovich's archives a manuscript from 1938 is preserved on "Some mathematical problems of the economics of industry, agriculture, and transport" that in content, apparently, corresponds to this report and where, in essence, the simplex method for the machine problem is described.

L.V. Kantorovich recalled that he created in January 1939 'a method of Lagrange (resolving) multipliers'.

> What became clear was both the solubility of these problems and the fact that they were widespread, so representatives of industry were invited to a discussion of my report at the university.

This meeting took place on 13 May 1939 at the Mathematical Section of the Institute of Mathematics and Mechanics of the Leningrad State University. A second meeting, which was devoted specifically to problems connected with construction, was held on 26 May 1939 at the Leningrad Institute for Engineers of Industrial Construction. These meetings provided the basis of the monograph *Mathematical Methods in the Organization and Planning of Production* (Kantorovich [1939]).

According to the Foreword by A.R. Marchenko to this monograph, Kantorovich's work was highly praised by mathematicians, and, in addition, at the special meeting industrial workers unanimously evinced great interest in the work.

The relevance was described by Kantorovich as follows:

> I want to emphasize again that the greater part of the problems of which I shall speak, relating to the organization and planning of production, are connected specifically with the Soviet system of economy and in the majority of cases do not arise in the economy of a capitalist society. There the choice of output is determined not by the plan but by the interests and profits of individual capitalists. The owner of the enterprise chooses for production those goods which at a given moment have the highest price, can most easily be sold, and therefore give the largest profit. The raw material used is not that of which there are huge supplies in the country, but that which the entrepreneur can buy most cheaply. The question of the maximum utilization of equipment is not raised; in any case, the majority of enterprises work at half capacity.
> In the USSR the situation is different. Everything is subordinated not to the interests and advantage of the individual enterprise, but to the task of fulfilling the state plan. The basic task of an enterprise is the fulfillment and overfulfillment of its plan, which is a part of the general state plan. Moreover, this not only means fulfillment of the plan in aggregate terms (i.e. total value of output, total tonnage, and so on), but the certain fulfillment of the plan for all kinds of output; that is, the fulfillment of the assortment plan (the fulfillment of the plan for each kind of output, the completeness of individual items of output, and so on).

In the monograph, Kantorovich outlined a new method to maximize a linear function under given linear constraints. One of the problems studied was a rudimentary form of a transportation problem:

(21.62) given: an $m \times n$ matrix $(a_{i,j})$;
find: an $m \times n$ matrix $(x_{i,j})$ such that:

$$\begin{array}{lll} \text{(i)} & x_{i,j} \geq 0 & \text{for all } i,j; \\ \text{(ii)} & \sum_{i=1}^{m} x_{i,j} = 1 & \text{for each } j = 1,\ldots,n; \\ \text{(iii)} & \sum_{j=1}^{n} a_{i,j} x_{i,j} \text{ is independent of } i \text{ and is maximized.} & \end{array}$$

Another problem studied by Kantorovich was 'Problem C' which can be stated as follows:

(21.63)

$$\begin{array}{lll} \text{maximize} & \lambda & \\ \text{subject to} & \sum_{i=1}^{m} x_{i,j} = 1 & (j = 1,\ldots,n) \\ & \sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j,k} x_{i,j} = \lambda & (k = 1,\ldots,t) \\ & x_{i,j} \geq 0 & (i = 1,\ldots,m; j = 1,\ldots,n). \end{array}$$

The interpretation is: let there be n machines, which can do m jobs. Let there be one final product consisting of t parts. When machine i does job j, $a_{i,j,k}$ units of part k are produced ($k = 1,\ldots,t$). Now $x_{i,j}$ is the fraction of time machine i does job j. The number λ is the amount of the final product produced. 'Problem C' was later seen (by H.E. Scarf, upon a suggestion by Kantorovich — see Koopmans [1959]) to be equivalent to the general linear programming problem.

Kantorovich's method consists of determining dual variables ('resolving multipliers') and finding the corresponding primal solution. If the primal solution is not feasible, the dual solution is modified following prescribed rules. Kantorovich also indicated the role of the dual variables in sensitivity analysis, and he showed that a feasible primal solution for Problem C can be shown to be optimal by specifying optimal dual variables.

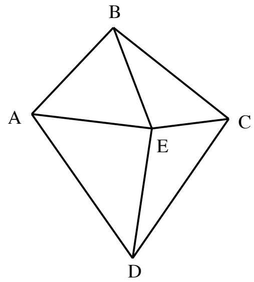
Kantorovich gave a wealth of practical applications of his methods, which he based mainly in the Soviet plan economy:

> Here are included, for instance, such questions as the distribution of work among individual machines of the enterprise or among mechanisms, the correct distribution of orders among enterprises, the correct distribution of different kinds of raw materials, fuel, and other factors. Both are clearly mentioned in the resolutions of the 18th Party Congress.

He described the applications to transportation:

> Let us first examine the following question. A number of freights (oil, grain, machines and so on) can be transported from one point to another by various methods; by railroads, by steamship; there can be mixed methods, in part by railroad, in part by automobile transportation, and so on. Moreover, depending on the kind of freight, the method of loading, the suitability of the transportation, and the efficiency of the different kinds of transportation is different. For example, it is particularly advantageous to carry oil by water transportation if oil tankers

are available, and so on. The solution of the problem of the distribution of a given freight flow over kinds of transportation, in order to complete the haulage plan in the shortest time, or within a given period with the least expenditure of fuel, is possible by our methods and leads to Problems A or C.
Let us mention still another problem of different character which, although it does not lead directly to questions A, B, and C, can still be solved by our methods. That is the choice of transportation routes.



Let there be several points A, B, C, D, E (Fig. 1) which are connected to one another by a railroad network. It is possible to make the shipments from B to D by the shortest route BED, but it is also possible to use other routes as well: namely, BCD, BAD. Let there also be given a schedule of freight shipments; that is, it is necessary to ship from A to B a certain number of carloads, from D to C a certain number, and so on. The problem consists of the following. There is given a maximum capacity for each route under the given conditions (it can of course change under new methods of operation in transportation). It is necessary to distribute the freight flows among the different routes in such a way as to complete the necessary shipments with a minimum expenditure of fuel, under the condition of minimizing the empty runs of freight cars and taking account of the maximum capacity of the routes. As was already shown, this problem can also be solved by our methods.

Kantorovich [1987] wrote in his memoirs:

The university immediately published my pamphlet, and it was sent to fifty People's Commissariats. It was distributed only in the Soviet Union, since in the days just before the start of the World War it came out in an edition of one thousand copies in all.
The number of responses was not very large. There was quite an interesting reference from the People's Commissariat of Transportation in which some optimization problems directed at decreasing the mileage of wagons was considered, and a good review of the pamphlet appeared in the journal *The Timber Industry*.
At the beginning of 1940 I published a purely mathematical version of this work in Doklady Akad. Nauk [76], expressed in terms of functional analysis and algebra. However, I did not even put in it a reference to my published pamphlet—taking into account the circumstances I did not want my practical work to be used outside the country.

In the spring of 1939 I gave some more reports—at the Polytechnic Institute and the House of Scientists, but several times met with the objection that the work used mathematical methods, and in the West the mathematical school in economics was an anti-Marxist school and mathematics in economics was a means for apologists of capitalism. This forced me when writing a pamphlet to avoid the term "economic" as much as possible and talk about the organization and planning of production; the role and meaning of the Lagrange multipliers had to be given somewhere in the outskirts of the second appendix and in the semi Aesopian language.

(Here reference [76] is Kantorovich [1940].) Kantorovich mentioned that the new area opened by his work played a definite role in forming the Leningrad Branch of the Mathematical Institute (LOMI), where he worked with M.K. Gavurin on this area. The problem that they studied occurred to them by itself, but they soon found out that railway workers were already studying the problem of planning haulage on railways, applied to questions of driving empty cars and transport of heavy cargoes.

Kantorovich and Gavurin wrote their method (the method of 'potentials') in a paper *Application of mathematical methods in questions of analysis of freight traffic* (Kantorovich and Gavurin [1949]), which was presented in January 1941 to the mathematics section of the Leningrad House of Scientists, but according to Kantorovich [1987]:

> The publication of this paper met with many difficulties. It had already been submitted to the journal *Railway Transport* in 1940, but because of the dread of mathematics already mentioned it was not printed then either in this or in any other journal, despite the support of Academicians A.N. Kolmogorov and V.N. Obraztsov, a well-known transport specialist and first-rank railway General.

Kantorovich [1987] said that he fortunately made an abstract version of the problem, Kantorovich [1942], in which he considered the following generalization of the transportation problem.

Let R be a compact metric space, with two measures μ and μ'. Let $\mathcal{B}$ be the collection of measurable sets in R. A *translocation (of masses)* is a function $\Psi : \mathcal{B} \times \mathcal{B} \to \mathbb{R}_+$ such that for each $X \in \mathcal{B}$ the functions $\Psi(X, .)$ and $\Psi(., X)$ are measures and such that

(21.64) $$\Psi(X, R) = \mu(X) \text{ and } \Psi(R, X) = \mu'(X)$$

for each $X \in \mathcal{B}$.

Let a continuous function $r : R \times R \to \mathbb{R}_+$ be given. (The value $r(x, y)$ represents the work needed to transfer a unit mass from x to y.) Then the *work* of a *translocation* Ψ is by definition:

(21.65) $$\int_R \int_R r(x, y)\Psi(d\mu, d\mu').$$

Kantorovich argued that, if there exists a translocation, then there exists a *minimal* translocation, that is, a translocation Ψ minimizing (21.65).

He calls a translocation Ψ *potential* if there exists a function $p : R \to \mathbb{R}$ such that for all $x, y \in R$:

(21.66)
(i) $|p(x) - p(y)| \leq r(x, y)$;
(ii) $p(y) - p(x) = r(x, y)$ if $\Psi(U_x, U_y) > 0$ for any neighbourhoods U_x of x and U_y of y.

Kantorovich showed:

Theorem 21.33. *A translocation Ψ is minimal if and only if it is potential.*

This framework applies to the transportation problem (when $m = n$), by taking for R the space $\{1, \ldots, n\}$, with the discrete topology.

Kantorovich's proof of Theorem 21.33 is by a construction of a potential, that however only is correct if r satisfies the triangle inequality. Kantorovich remarked that his method is algorithmic:

> The theorem just demonstrated makes it easy for one to prove that a given mass translocation is or is not minimal. He has only to try and construct the potential in the way outlined above. If this construction turns out to be impossible, i.e. the given translocation is not minimal, he at least will find himself in the possession of the method how to lower the translocation work and eventually come to the minimal translocation.

Beside to a problem of leveling a land area, Kantorovich gave as application:

> Problem 1. *Location of consumption stations with respect to production stations.* Stations $A_1, A_2, \cdots, A_m$, attached to a network of railways deliver goods to an extent of $a_1, a_2, \cdots, a_m$ carriages per day respectively. These goods are consumed at stations $B_1, B_2, \cdots, B_n$ of the same network at a rate of $b_1, b_2, \cdots, b_n$ carriages per day respectively ($\sum a_i = \sum b_k$). Given the costs $r_{i,k}$ involved in moving one carriage from station A_i to station B_k, assign the consumption stations such places with respect to the production stations as would reduce the total transport expenses to a minimum.

As mentioned, Kantorovich's results remained unnoticed for some time by Western researchers. In a note introducing a reprint of the article of Kantorovich [1942], in *Management Science* in 1958, the following reassurance is given:

> It is to be noted, however, that the problem of determining an effective method of actually acquiring the solution to a specific problem is *not* solved in this paper. In the category of development of such methods we seem to be, currently, ahead of the Russians.

Kantorovich's method was elaborated by Kantorovich and Gavurin [1949], where moreover single- and multicommodity transportation models are studied, with applications to the railway network of the U.S.S.R.

Hitchcock

Independently, Hitchcock [1941] studied the transportation problem:

(21.67) given: an $m \times n$ matrix $C = (c_{i,j})$ and vectors $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$;
find: an $m \times n$ matrix $X = (x_{i,j})$ such that:
(i) $x_{i,j} \geq 0$ for all i, j;
(ii) $\displaystyle\sum_{j=1}^{n} x_{i,j} = a_i$ for each $i = 1, \ldots, m$;
(iii) $\displaystyle\sum_{i=1}^{m} x_{i,j} = b_j$ for each $j = 1, \ldots, n$;
(iv) $\displaystyle\sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} x_{i,j}$ is as small as possible.

The interpretation of the problem is, in Hitchcock's words:

> When several factories supply a product to a number of cities we desire the least costly manner of distribution. Due to freight rates and other matters the cost of a ton of product to a particular city will vary according to which factory supplies it, and will also vary from city to city.

Hitchcock showed that the minimum is attained at a vertex of the feasible region, and he outlined a scheme for solving the transportation problem which has much in common with the simplex method for linear programming. It includes pivoting (eliminating and introducing basic variables) and the fact that nonnegativity of certain dual variables implies optimality. He showed that the *complementary slackness* conditions characterize optimality: $(x^*_{i,j})$ is an optimum vertex if and only if there exists a combination $\sum_{i,j} \lambda_{i,j} x_{i,j}$ of the left-hand sides of the constraints (ii) and (iii) such that $\lambda_{i,j} \geq c_{i,j}$ for all i, j and such that $\lambda_{i,j} = c_{i,j}$ if $x^*_{i,j} > 0$.

Hitchcock however seemed to have overlooked the possibility of cycling of his method, although he pointed at an example in which some dual variables are negative while yet the primal solution is optimum.

Hitchcock also gave a method to find an initial basic solution, now known as the *north-west rule*: set $x_{1,1} := \min\{a_1, b_1\}$; if the minimum is attained by a_1, reset $b_1 := b_1 - a_1$ and recursively find a basic solution $x_{i,j}$ satisfying $\sum_{j=1}^{n} x_{i,j} = a_i$ for each $i = 2, \ldots, m$ and $\sum_{i=2}^{m} x_{i,j} = b_j$ for each $j = 1, \ldots, n$; if the minimum is attained by b_1, proceed symmetrically. (The north-west rule was also described by Salvemini [1939] and Fréchet [1951] in a statistical context, namely in order to complete correlation tables given the marginal distributions.)

Koopmans

Also independently, Koopmans investigated transportation problems. In March 1942, Koopmans was appointed as a statistician on the staff of the British Merchant Shipping Mission, and later the Combined Shipping Adjustment Board (CSAB), a British-American agency dealing with merchant shipping problems during the Second World War (as they should go in convoys, under military protection). Influenced by his teacher J. Tinbergen (cf. Tinbergen [1934]) he was interested in tanker freights and capacities (cf. Koopmans [1939]). According to Koopmans' personal diary, in August 1942 while the Board was being organized, there was not much work for the statisticians,

> and I had a fairly good time working out exchange ratio's between cargoes for various routes, figuring how much could be carried monthly from one route if monthly shipments on another route were reduced by one unit.

At the Board he studied the assignment of ships to convoys so as to accomplish prescribed deliveries, while minimizing empty voyages (cf. Dorfman [1984]). According to the memoirs of his wife (Wanningen Koopmans [1995]), when Koopmans was with the Board,

> he had been appalled by the way the ships were routed. There was a lot of redundancy, no intensive planning. Often a ship returned home in ballast, when with a little effort it could have been rerouted to pick up a load elsewhere.

In his autobiography (published posthumously), Koopmans [1992] described how he came to the problem:

> My direct assignment was to help fit information about losses, deliveries from new construction, and employment of British-controlled and U.S-controlled ships into a unified statement. Even in this humble role I learned a great deal about the difficulties of organizing a large-scale effort under dual control—or rather in this case four-way control, military and civilian cutting across U.S. and U.K.

> controls. I did my study of optimal routing and the associated shadow costs of transportation on the various routes, expressed in ship days, in August 1942 when an impending redrawing of the lines of administrative control left me temporarily without urgent duties. My memorandum, cited below, was well received in a meeting of the Combined Shipping Adjustment Board (that I did not attend) as an explanation of the "paradoxes of shipping" which were always difficult to explain to higher authority. However, I have no knowledge of any systematic use of my ideas in the combined U.K.-U.S. shipping problems thereafter.

In the memorandum to the Board, Koopmans [1942] analyzed the sensitivity of the optimum shipments for small changes in the demands. In this memorandum, Koopmans did not give a method to find an optimum shipment. Further study led him to a 'local search' method for the transportation problem, stating that it leads to an optimum solution. According to Dorfman [1984], Koopmans found these results in 1943, but, due to wartime restrictions, published them only after the war (Koopmans [1948], Koopmans and Reiter [1949a,1949b,1951]). Koopmans [1948] wrote:

> Let us now for the purpose of argument (since no figures of war experience are available) assume that one particular organization is charged with carrying out a world dry-cargo transportation program corresponding to the actual cargo flows of 1925. How would that organization solve the problem of moving the empty ships economically from where they become available to where they are needed? It seems appropriate to apply a procedure of trial and error whereby one draws tentative lines on the map that link up the surplus areas with the deficit areas, trying to lay out flows of empty ships along these lines in such a way that a minimum of shipping is at any time tied up in empty movements.

The 'trial and error' method mentioned is one of local improvements, corresponding to finding a negative-cost directed circuit in the residual digraph. Koopmans' *first theorem* is that it leads to an optimum solution:

> If, under the assumptions that have been stated, no improvement in the use of shipping is possible by small variations such as have been illustrated, then there is no—however thoroughgoing—rearrangement in the routing of empty ships that can achieve a greater economy of tonnage.

He illustrated the method by giving an optimum solution for a 3×12 transportation problem, with the following supplies and demands:

Net receipt of dry cargo in overseas trade, 1925

Unit: Millions of metric tons per annum

Harbour	Received	Dispatched	Net receipts
New York	23.5	32.7	−9.2
San Francisco	7.2	9.7	−2.5
St. Thomas	10.3	11.5	−1.2
Buenos Aires	7.0	9.6	−2.6
Antofagasta	1.4	4.6	−3.2
Rotterdam	126.4	130.5	− 4.1
Lisbon	37.5	17.0	20.5
Athens	28.3	14.4	13.9
Odessa	0.5	4.7	−4.2
Lagos	2.0	2.4	−0.4
Durban	2.1	4.3	−2.2
Bombay	5.0	8.9	−3.9
Singapore	3.6	6.8	−3.2
Yokohama	9.2	3.0	6.2
Sydney	2.8	6.7	−3.9
Total	266.8	266.8	0.0

Koopmans [1948] moreover claimed that there exist *potentials* $p_1, \ldots, p_n$ and $q_1, \ldots, q_m$ such that $c_{i,j} \geq p_i - q_j$ for all i, j and such that $c_{i,j} = p_i - q_j$ for each i, j for which $x_{i,j} > 0$.

The potentials give the *marginal costs* when modifying the input data. That is, if both a_i and b_j increase by 1, then the minimum cost increases by at least $p_i - q_j$. This is Koopmans' *second theorem.*

In the proof, Koopmans assumed that the cost function is symmetric and satisfies the triangle inequality. Moreover, he assumed that the graph of arcs having a positive transshipment value is weakly connected. The latter restriction was removed in a later paper by Koopmans and Reiter [1951]. In this paper, they investigated the economic implications of the model and the method:

> For the sake of definiteness we shall speak in terms of the transportation of cargoes on ocean-going ships. In considering only shipping we do not lose generality of application since ships may be "translated" into trucks, aircraft, or, in first approximation, trains, and ports into the various sorts of terminals. Such translation is possible because all the above examples involve particular types of movable transportation equipment.

They use the graph model, and in a footnote they remark:

> The cultural lag of economic thought in the application of mathematical methods is strikingly illustrated by the fact that linear graphs are making their entrance into transportation theory just about a century after they were first studied in relation to electrical networks, although organized transportation systems are much older than the study of electricity.

(For a review of Koopmans' research, see Scarf [1992].)

Robinson, 1950

Robinson [1950] might be the earliest reference stating clearly and generally that the absence of a negative-cost directed circuit in the residual digraph is necessary and sufficient for optimality. She mentioned that it can be 'verified directly', and observed that it gives an algorithm to find an optimum transportation. She concluded with:

> The number of steps in the iterative procedure depends on the "goodness" of the initial choice of X_0. The method does not seem to lend itself to machine calculation but may be efficient for hand computation with matrices of small order.

Linear programming and the simplex method

The breakthrough of general linear programming came at the end of the 1940s. In 1947, Dantzig formulated the linear programming problem and designed the *simplex method* for the linear programming problem, published in Dantzig [1951b]. The success of the method was enlarged by a simple tableau-form and a simple pivoting rule, and by the efficiency in practice. In another paper, Dantzig [1951a] described a direct implementation of the simplex method to the transportation problem (including an anti-cycling rule based on perturbation; variants were given by Charnes and Cooper [1954] and Eisemann [1956]).

The simplex method for transportation was described in terms of graphs by Koopmans and Reiter [1951], and Flood [1952,1953] aimed at giving a purely mathematical description of it. A continuous model of transportation was studied by Beckmann [1952].

Votaw and Orden [1952] reported on early computational results (on the SEAC), and claimed (without proof) that the simplex method is polynomial-time for the transportation problem (a statement refuted by Zadeh [1973a]):

> As to computation time, it should be noted that for moderate size problems, say $m \times n$ up to 500, the time of computation is of the same order of magnitude as the time required to type the initial data. The computation time on a sample computation in which m and n were both 10 was 3 minutes. The time of computation can be shown by study of the computing method and the code to be proportional to $(m + n)^3$.

Application to practice

The new ideas of applying linear programming to the transportation problem were quickly disseminated. Applications to routing empty boxcars over the U.S. railroads were given by Fox [1952] and Nerlove [1953]. Dantzig and Fulkerson [1954b,1954a] studied a rudimentary form of a minimum-cost circulation problem in order to determine the minimum number of tankers to meet a fixed schedule. Similarly, Bartlett [1957] and Bartlett and Charnes [1957] studied methods to determine the minimum railway stock to run a given schedule.

Applicability of linear programming to transportation to practice was also met with scepticism. At a Conference on Linear Programming in May 1954 in London, Land [1954] presented a study of applying linear programming to the problem of transporting coal for the British Coke Industry:

> The real crux of this piece of research is whether the saving in transport cost exceeds the cost of using linear programming.

In the discussion which followed, T. Whitwell of Powers Samas Accounting Machines Ltd remarked

> that in practice one could have one's ideas of a solution confirmed or, much more frequently, completely upset by taking a couple of managers out to lunch.

Gleyzal's primal-dual method for the transportation problem

Gleyzal [1955] published the following primal-dual method for the transportation problem (with integer data). Let $x_{i,j}$ be a feasible solution of the transportation problem. Transform $x_{i,j}$ such that the set $\{u_i v_j \mid x_{i,j} > 0\}$ contains no circuit, and transform $c_{i,j}$ such that $c_{i,j} = 0$ if $x_{i,j} > 0$. (These are easy by first cancelling circuits, and next redefining $c_{i,j}$.)

If $c_{i,j} \geq 0$ for all i, j we are done. Suppose that $c_{i_0,j_0} < 0$ for some i_0, j_0. Let $A := \{(u_i, v_j) \mid c_{i,j} \leq 0\} \cup \{(v_j, u_i) \mid x_{i,j} > 0\}$. If u_{i_0} is reachable in A from v_{j_0}, A contains a directed circuit C containing (u_{i_0}, v_{j_0}). Then we can reset $x_{i,j} := x_{i,j} - 1$ if (v_j, u_i) is in C and $x_{i,j} := x_{i,j} + 1$ if (u_i, v_j) is in C. This decreases $c^{\mathsf{T}} x$.

If u_{i_0} is not reachable in A from v_{j_0}, then for any vertex v let $r(v) := 1$ if v is reachable in A from v_{j_0} and $r(v) := 0$ otherwise. Reset $c_{i,j} := c_{i,j} - r(u_i) + r(v_j)$. This increases $\sum(c_{i,j} \mid c_{i,j} < 0)$, and hence the method terminates.

Munkres on the transportation problem

Munkres [1957] extended his variant of the Hungarian method for the assignment problem to the transportation problem. In graph terms, it amounts to the following.

Let $G = (V, E)$ be a complete bipartite graph, with colour classes U and W of size n, and let be given a weight function $w : E \to \mathbb{Z}_+$ and a function $b : V \to \mathbb{Z}_+$ with $b(U) = b(W)$. We must find a function $x : E \to \mathbb{Q}_+$ such that $\sum_{e\in\delta(v)} x_e = b_v$ for each vertex v and such that $\sum_e w_e x_e$ is minimized.

Let F be the set of edges e with $w_e = 0$ and let $H = (V, F)$. Suppose that we have found an $x : E \to \mathbb{Q}_+$ such that $x_e = 0$ if $e \notin F$ and such that $\sum_{e\in\delta(v)} x_e \leq b_v$ for each $v \in V$. Let U' and W' be the sets of vertices v in U and W for which strict inequality holds. If U', and hence W', are empty, x is an optimum solution. Otherwise, perform the following iteratively.

Orient each edge of H from U to W, and orient each edge e of H with $x_e > 0$ also from W to U (so they are two-way). Now determine the set R_M of vertices reachable by a directed path from U'.

Case 1: $R_M \cap W' \neq \emptyset$. Then D has a $U'-W'$ path, on which we can alternatingly increase and decrease the value of x_e, so as to make $\sum_e x_e$ larger.

Case 2: $R_M \cap W' = \emptyset$. So $w(uv) > 0$ for each $u \in U \cap R_M$ and $v \in W \setminus R_M$. Let h be the minimum of these $w(uv)$. Decrease $w(uv)$ by h if $u \in U \cap R_M, v \in W \setminus R_M$, and increase $w(uv)$ by h if $u \in U \setminus R_M, v \in W \cap R_M$.

This describes the iteration. Note that between any two occurrences of Case 1, only n times Case 2 can occur, since at each such iteration the set $R_M \cap W$ increases. Moreover, after Case 2 we can continue the previous search for R_M. So between any two Case 1-iterations, the Case 2-iterations take $O(n^2)$ time altogether.

Now Case 1 can occur at most $\sum_{v\in U} b_v$ times. So the algorithm is finite, and has running time $O(n^4 B)$ where $B := \max\{b_v \mid v \in V\}$. This specializes to the Hungarian method if $b_v = 1$ for all $v \in V$.

Further early methods

Also Ford and Fulkerson [1955,1957b] (cf. Ford and Fulkerson [1956c,1956d]) extended the Hungarian method to general transportation problems. Their method is essentially the same as that of Munkres [1957], except that successive occurrences of Case 1 iterations are combined to a maximum flow computation. A similar primal-dual method for the transportation problem was described by Egerváry [1958].

Ford and Fulkerson [1956a,1957a] extended the method of Ford and Fulkerson [1955,1957b] for the uncapacitated transportation problem to the *capacitated* transportation problem.

Orden [1955] showed the equivalence of the transshipment problem and the transportation problem. He also noted that the class of transportation problems covers the majority of the applications of linear programming which are in practical use or under active development. Also Prager [1957a] studied the transshipment problem by reduction to a transportation problem and by methods of elastostatics (cf. Kuhn [1957]).

Gallai [1957,1958a,1958b] studied the minimum-cost and the maximum-profit circulation problem, for which he gave min-max relations (see Section 12.5b). He also considered vertex capacities and demands. Beside combinatorial proofs based on potentials, Gallai gave proofs based on linear programming duality and total unimodularity.

A minimum-cost flow algorithm (in disguised form) was given by Ford and Fulkerson [1958b], to solve the 'dynamic flow' problem described in Section 12.5c. They described a method which essentially consists of repeatedly finding a zero-length $r - s$ path in the residual graph, making lengths nonnegative by translating the cost with the help of the current potential p. If no zero-length path exists, the potential is updated. (This is Routine I of Ford and Fulkerson [1958b].) The complexity of this was studied by Fulkerson [1958].

Yakovleva [1959] gave some implementations of the method of Kantorovich and Gavurin [1949]. The paper considers three cases of the problem in a digraph with demands (positive, negative, and zero) of vertices and costs of arcs: (i) noncapacitated case, (ii) capacitated case, and (iii) bipartite case (without zero demands). Two methods are developed for finding feasible potentials or improving the current flow. Time bounds are not indicated.

Among the other early algorithms for minimum-cost flow are successive shortest paths methods (Busacker and Gowen [1960], Iri [1960]), *out-of-kilter* methods (Minty [1960], Fulkerson [1961]), cycle-cancelling (Klein [1967]), and successive shortest paths maintaining potentials (Tomizawa [1971], Edmonds and Karp [1972]). An alternative method, which transforms the transportation problem to a nonlinear programming problem, with computational results, was given by Gerstenhaber [1958,1960].

Polynomial-time algorithms

Edmonds and Karp [1972] gave the first polynomial-time algorithm for the minimum-cost flow problem, based on capacity-scaling. They realized that in fact the method is only *weakly* polynomial; that is, the number of steps depends also on the size of the numbers in the input:

> Although it is comforting to know that the minimum-cost flow algorithm terminates, the bounds on the number of augmentations are most unfavorable. The scaling method of the next two sections is a variant of this algorithm in which the bound depends logarithmically, rather than linearly, on the capacities. A challenging open problem is to emulate the results of Section 1.2 for the maximum-value flow problem by giving a method for the minimum-cost flow problem having a bound on computation which is a polynomial in the number of nodes, and is independent of both costs and capacities.

Tarjan [1983] wrote: 'There is still much to be learned about the minimum cost flow problem'. Soon after, Edmonds and Karp's question was resolved by Tardos [1985a], by giving a strongly polynomial-time minimum-cost circulation algorithm. Her work has inspired a stream of further developments, part of which was discussed in Chapter 12.

Chapter 22

Transversals

The study of transversals of a family of sets is close to that of matchings in a bipartite graph, but with a shift in focus. While matchings are subsets of the edge set, transversals are subsets of one of the colour classes. This gives rise to a number of optimization and polyhedral problems and results that deserve special attention.
In this chapter we study transversals of one family of sets, while in the next chapter we go over to *common* transversals of two families of sets.

22.1. Transversals

Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of sets. A set T is called a *transversal* of $\mathcal{A}$ if there exist distinct elements $a_1 \in A_1, \ldots, a_n \in A_n$ such that $T = \{a_1, \ldots, a_n\}$. So T is an *unordered* set with $|T| = n$. (Instead of 'transversal' one uses also the term *system of distinct representatives* or *SDR*.)

Transversals are closely related to matchings in bipartite graphs. In particular, the basic result on the existence of a transversal (Hall [1935]), is a consequence of Kőnig's matching theorem. This can be seen with the following basic construction of a bipartite graph $G = (V, E)$ associated with a family $\mathcal{A} = (A_1, \ldots, A_n)$ of subsets of a set S:

(22.1) $$V := \{1, \ldots, n\} \cup S,$$
$$E := \{\{i, s\} \mid i = 1, \ldots, n; s \in A_i\},$$

assuming that S is disjoint from $\{1, \ldots, n\}$ (which for our purposes can be done without loss of generality). So G has colour classes $\{1, \ldots, n\}$ and S. (This construction was given by Skolem [1917].)

Then trivially

(22.2) a set T is a transversal of $\mathcal{A}$ if and only if G has a matching M of size n such that T is the set of vertices in S covered by M.

So the existence of a transversal of $\mathcal{A}$ can be reduced to the existence of a matching in G of size n. Hence Kőnig's matching theorem applies to the existence of transversals.

It is convenient to introduce the following notation, for any family $(A_1, \ldots, A_n)$ of sets and any $I \subseteq \{1, \ldots, n\}$:

(22.3) $$A_I := \bigcup_{i \in I} A_i.$$

Theorem 22.1 (Hall's marriage theorem). *A family $\mathcal{A} = (A_1, \ldots, A_n)$ of sets has a transversal if and only if*

(22.4) $$|A_I| \geq |I|$$

for each subset I of $\{1, \ldots, n\}$.

Proof. Necessity of the condition being easy, we prove sufficiency. Let G be the graph associated to $\mathcal{A}$ (as in (22.1)). Now the theorem is equivalent to Theorem 16.7 (taking $U := \{1, \ldots, n\}$). ∎

Condition (22.4) is called *Hall's condition.* The name 'marriage theorem' is due to Weyl [1949].

The polynomial-time algorithm given in Section 16.3 for finding a maximum matching in a bipartite graph directly yields a polynomial-time algorithm for finding a transversal of a family $(A_1, \ldots, A_n)$ of sets. In fact, Theorem 16.5 implies an $O(\sqrt{n}\, m)$ algorithm, where $m := \sum_i |A_i|$.

22.1a. Alternative proofs of Hall's marriage theorem

We give two alternative, direct proofs of the sufficiency of Hall's condition (22.4) for the existence of a transversal. Call a subset I of $\{1, \ldots, n\}$ *tight* if equality holds in (22.4).

If there is a $y \in A_n$ such that $A_1 \setminus \{y\}, \ldots, A_{n-1} \setminus \{y\}$ has a transversal, then we are done. Hence, we may assume that for each $y \in A_n$ there is a tight $I \subseteq \{1, \ldots, n-1\}$ with $y \in A_I$ (using induction).

The proof given by Easterfield [1946] (also by M. Hall [1948], Halmos and Vaughan [1950], and Mann and Ryser [1953]) continues as follows. Choose any such tight subset I. Without loss of generality, $I = \{1, \ldots, k\}$. By induction, $(A_1, \ldots, A_k)$ has a transversal, which must be $T := A_I$. Moreover, $(A_{k+1} \setminus T, \ldots, A_n \setminus T)$ has a transversal, Z say. This follows inductively, since for each $J \subseteq \{k+1, \ldots, n\}$,

(22.5) $$\Big|\bigcup_{i \in J}(A_i \setminus T)\Big| = \Big|\bigcup_{i \in I \cup J} A_i\Big| - |T| \geq |I| + |J| - |T| = |J|.$$

Then $T \cup Z$ is a transversal of $(A_1, \ldots, A_k, A_{k+1}, \ldots, A_n)$.

The proof due to Everett and Whaples [1949] continues slightly different. They noted that the collection of tight subsets of $\{1, \ldots, n\}$ is closed under taking intersections and unions. That is, if I and J are tight, then also $I \cap J$ and $I \cup J$ are tight, since

(22.6) $$|I| + |J| = |A_I| + |A_J| \geq |A_{I \cap J}| + |A_{I \cup J}| \geq |I \cap J| + |I \cup J| = |I| + |J|,$$

giving equality throughout. (In (22.6), the first inequality holds as $A_{I \cap J} \subseteq A_I \cap A_J$ and $A_{I \cup J} = A_I \cup A_J$.)

Since for each $y \in A_n$ there is a tight subset I of $\{1, \ldots, n-1\}$ with $y \in A_I$, it follows, by taking the union of them, that there is a tight subset I of $\{1, \ldots, n-1\}$ with $A_n \subseteq A_I$. For $J := I \cup \{n\}$ this gives the contradiction $|A_J| = |A_I| = |I| < |J|$.

The the closedness of tight subsets under intersections and unions was also noticed by Maak [1936] and Weyl [1949], who gave alternative proofs of a theorem of Rado and Hall's marriage theorem, respectively.

Edmonds [1967b] gave a linear-algebraic proof of Hall's marriage theorem (cf. Section 16.2b). Ford and Fulkerson [1958c] derived Hall's marriage theorem from the max-flow min-cut theorem.

22.2. Partial transversals

Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of sets. A set T is called a *partial transversal* if it is a transversal of some subfamily $(A_{i_1}, \ldots, A_{i_k})$ of $(A_1, \ldots, A_n)$. (Instead of 'partial transversal' one uses also the term *partial system of distinct representatives* or *partial SDR.*)

Again, by the construction (22.1), we can study partial transversals with the help of bipartite matching theory. In particular, if G is the graph associated to a family $\mathcal{A}$ of subsets of a set S,

(22.7) a set T is a partial transversal of $\mathcal{A}$ if and only if G has a matching M such that T is the set of vertices in S covered by M.

This yields the following so-called defect form of Hall's marriage theorem, which is equivalent to Kőnig's matching theorem (cf. Ore [1955]):

Theorem 22.2 (defect form of Hall's marriage theorem)**.** *Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of subsets of a set S. Then the maximum size of a partial transversal of $\mathcal{A}$ is equal to the minimum value of*

(22.8) $$|S \setminus X| + |\{i \mid A_i \cap X \neq \emptyset\}|,$$

where X ranges over all subsets of S.

Proof. Let G be the graph constructed in (22.1). The maximum size of a partial transversal of $\mathcal{A}$ is equal to the maximum size of a matching in G. By Kőnig's matching theorem, this is equal to the minimum size of a vertex cover of G. This minimum is attained by a vertex cover of form $(S \setminus X) \cup \{i \mid A_i \cap X \neq \emptyset\}$, which shows the theorem. ∎

An equivalent way of characterizing the maximum size of a partial transversal is:

Corollary 22.2a. *The maximum size of a partial transversal of $\mathcal{A}$ is equal to the minimum value of*

(22.9) $$\Big|\bigcup_{i \in I} A_i\Big| + n - |I|,$$

taken over $I \subseteq \{1, \ldots, n\}$.

Proof. Directly from Theorem 22.2, since we can assume that $S \setminus X = A_I$ where $I := \{i \mid A_i \cap X = \emptyset\}$. ∎

Note that it needs an argument to state that each partial transversal is a subset of a transversal, if a transversal exists. This was shown by Hoffman and Kuhn [1956b] (solving a problem of Mann and Ryser [1953]):

Theorem 22.3. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a system of sets having a transversal. Then each partial transversal is contained in a transversal.*

Proof. Directly from Theorem 16.8, using construction (22.1) (taking $R := \{1, \ldots, n\} \cup T$, where T is a partial transversal). ∎

One can generalize this to the case where the family need not have a transversal:

Theorem 22.4. *Let $\mathcal{A}$ be a family of sets. Then each partial transversal is contained in a maximum-size partial transversal.*

Proof. Again directly from Theorem 16.8, using construction (22.1). ∎

In other words, each inclusionwise maximal partial transversal is a maximum-size partial transversal. This is the basis of the fact that partial transversals form the independent sets of a matroid — see Chapter 39. It is equivalent to:

Corollary 22.4a (exchange property of transversals)**.** *Let $\mathcal{A}$ be a family of sets and let T and T' be partial transversals of $\mathcal{A}$, with $|T| < |T'|$. Then there exists an $s \in T' \setminus T$ such that $T \cup \{s\}$ is a partial transversal.*

Proof. To prove this, we can assume that each set in $\mathcal{A}$ is contained in $T \cup T'$. This implies that, if no s as required exists, T is an inclusionwise maximal partial transversal. However, as $|T'| > |T|$, this contradicts Theorem 22.4. ∎

Brualdi and Scrimger [1968] (extending a result of Mirsky and Perfect [1967]) observed:

Theorem 22.5. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of sets, let k be the maximum size of a partial transversal, and let $\mathcal{A}' = (A_1, \ldots, A_k)$ have a transversal. Then each maximum-size partial transversal of $\mathcal{A}$ is a transversal of $\mathcal{A}'$.*

Proof. Via construction (22.1) this follows from Corollary 16.8b. ∎

So when studying the collection of partial transversals of a certain collection $\mathcal{A}$ of sets, we can assume that $\mathcal{A}$ has a transversal.

22.3. Weighted transversals

Consider the problem of finding a minimum-weight transversal: given a family $\mathcal{A} = (A_1, \ldots, A_n)$ of subsets of a set S and a weight function $w : S \to \mathbb{Q}$, find a transversal T of $\mathcal{A}$ minimizing $w(T)$. This problem can be easily reduced to a minimum-weight perfect matching problem, implying that a minimum-weight transversal can be found in strongly polynomial time. In fact:

Theorem 22.6. *A minimum-weight transversal can be found in time $O(nm)$ where n is the number of sets and $m := \sum_i |A_i|$.*

Proof. Make the graph G as in (22.1) and define $w(\{i, s\}) := w(s)$ for each edge $\{i, s\}$ of G. Denote $R := \{1, \ldots, n\}$. Starting with $M = \emptyset$, we can apply the Hungarian method, to obtain an extreme matching of size n. The elements of S covered by M form a maximum-weight transversal. As each iteration of the Hungarian method takes $O(m)$ time, this gives the theorem. ∎

Note that in this algorithm, we grow a partial transversal until it is a (complete) transversal. In this respect it is a 'greedy method': we never backtrack. Again, this is a preview of the fact that transversals form a 'matroid' — see Chapter 39.

The method similarly solves the problem of finding a maximum-weight partial transversal:

Theorem 22.7. *A maximum-weight partial transversal can be found in time $O(rm)$, where r is the maximum size of a partial transversal and where $m := \sum_i |A_i|$.*

Proof. As above. ∎

22.4. Min-max relations for weighted transversals

We can also obtain a min-max relation for the minimum weight of a transversal:

Theorem 22.8. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of subsets of a set S having a transversal and let $w : S \to \mathbb{Z}$ be a weight function. Then the minimum weight of a transversal of $\mathcal{A}$ is equal to the maximum value of*

$$y(S) + \sum_{i=1}^{n} \min_{s \in A_i} (w(s) - y(s)) \tag{22.10}$$

taken over $y : S \to \mathbb{Z}_+$.

Proof. Let $t := |S|$. For $i = n+1, \ldots, t$, let $A_i := S$. Consider the bipartite graph $G = (V, E)$ defined by (22.1), for the family $(A_1, \ldots, A_t)$. Define a length function l on the edges of G as follows. For any edge $e = is$ of G, with $s \in A_i$, define $l_e := w(s)$ if $i \leq n$ and $l_e := 0$ otherwise. Then the minimum weight of a transversal of $(A_1, \ldots, A_n)$ is equal to the minimum length of a perfect matching in G. By Theorem 17.5 (a variant of Egerváry's theorem), the latter value is equal to the maximum value of $y(V)$ where $y \in \mathbb{Q}^V$ with $y(s) + y(i) \leq l(is)$ for each $i = 1, \ldots, t$ and $s \in A_i$. We can assume that the minimum of $y(s)$ over $s \in S$ is equal to 0 (since subtracting a constant to $y(s)$ for any $s \in S$ and adding it to $y(i)$ for any $i \in \{1, \ldots, t\}$ maintains the properties required for y). Then $y(i) = \min_{s \in A_i}(w(s) - y(s))$ if $i \leq n$ and $y(i) = 0$ if $i > n$. So $y(V)$ is equal to the value of (22.10). ∎

A min-max relation for the maximum weight of a partial transversal follows similarly:

Theorem 22.9. *Let $\mathcal{A} = (A_1, \ldots, A_k)$ be a family of subsets of a set S and let $w : S \to \mathbb{Z}_+$ be a weight function. Then the maximum weight of a partial transversal of $\mathcal{A}$ is equal to the minimum value of*

$$(22.11) \qquad y(S) + \sum_{i=1}^{k} \max\{0, \max_{s \in A_i}(w(s) - y(s))\}$$

over functions $y : S \to \mathbb{Z}_+$.

Proof. Directly from Egerváry's theorem (Theorem 17.1), using construction (22.1). ∎

22.5. The transversal polytope

Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of subsets of a set S. The *partial transversal polytope* $P_{\text{partial transversal}}(\mathcal{A})$ of $\mathcal{A}$ is the convex hull of the incidence vectors (in $\mathbb{R}^S$) of the partial transversals of $\mathcal{A}$. That is,

$$(22.12) \qquad P_{\text{partial transversal}}(\mathcal{A}) = \text{conv.hull}\{\chi^T \mid T \text{ is a partial transversal of } \mathcal{A}\}.$$

It is easy to see that each vector x in the partial transversal polytope satisfies:

$$(22.13) \qquad \begin{array}{lll} \text{(i)} & 0 \leq x_s \leq 1 & \text{for each } s \in S, \\ \text{(ii)} & x(S \setminus A_I) \leq n - |I| & \text{for each } I \subseteq \{1, \ldots, n\}. \end{array}$$

Corollary 22.9a. *System* (22.13) *determines the partial transversal polytope and is TDI.*

Proof. Consider a weight function $w : S \to \mathbb{Z}_+$. Let ω be the maximum weight of a partial transversal. By Theorem 22.9, there exists a function $y : S \to \mathbb{Z}_+$ such that

$$\omega = y(S) + \sum_{i=1}^{n} \max\{0, \max_{s \in A_i}(w(s) - y(s))\}. \tag{22.14}$$

For each $j \in \mathbb{Z}_+$, let I_j be the set of $i \in \{1, \ldots, n\}$ with

$$\max_{s \in A_i}(w(s) - y(s)) \leq j. \tag{22.15}$$

So $I_j = \{1, \ldots, n\}$ for j large enough.

Then

$$w - y \leq \sum_{j=0}^{\infty} \chi^{S \setminus A_{I_j}}, \tag{22.16}$$

since for $k := w(s) - y(s)$, we have for each $j < k$ there is no $i \in I_j$ with $s \in A_i$. Hence $s \in S \setminus A_{I_j}$ for all $j < k$. So y and the I_j give an integer feasible dual solution.

The fact that they are optimum follows from:

$$\begin{aligned} y(S) + \sum_{j=0}^{\infty}(n - |I_j|) &= y(S) + \sum_{j=0}^{\infty} \sum_{\substack{i=1 \\ \max_{s \in A_i}(w(s)-y(s)) > j}}^{n} 1 \\ &= y(S) + \sum_{i=1}^{n} \sum_{\substack{j=0 \\ \max_{s \in A_i}(w(s)-y(s)) > j}}^{\infty} 1 \\ &= y(S) + \sum_{i=1}^{n} \max\{0, \max_{s \in A_i}(w(s) - y(s))\} = \omega, \end{aligned} \tag{22.17}$$

by (22.14). ∎

Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of subsets of a set S. The *transversal polytope* $P_{\text{transversal}}(\mathcal{A})$ of $\mathcal{A}$ is the convex hull of the incidence vectors (in $\mathbb{R}^S$) of the transversals of $\mathcal{A}$. That is,

$$P_{\text{transversal}}(\mathcal{A}) = \text{conv.hull}\{\chi^T \mid T \text{ is a transversal of } \mathcal{A}\}. \tag{22.18}$$

It is easy to see that each vector x in the transversal polytope satisfies:

$$\begin{array}{lll} \text{(i)} & 0 \leq x_s \leq 1 & \text{for each } s \in S, \\ \text{(ii)} & x(A_I) \geq |I| & \text{for each } I \subseteq \{1, \ldots, n\}, \\ \text{(iii)} & x(S) = n. & \end{array} \tag{22.19}$$

Corollary 22.9b. *System* (22.19) *determines the transversal polytope and is TDI.*

Proof. The transversal polytope is the facet of the partial transversal polytope determined by the equality $x(S) = n$. This is constraint (22.13)(ii) for $I = \emptyset$, set to equality. Now each inequality in (22.19) is a nonnegative integer combination of the inequalities in (22.13) and of $-x(S) \leq -n$ (since $-x(A_I) = x(S \setminus A_I) - x(S) \leq (n - |I|) - n = -|I|$). So using Theorem 5.25, the corollary follows. ∎

One may note that the number of facets of the matching polytope of a bipartite graph $G = (V, E)$ is at most $|V| + |E|$, while the number of facets of the closely related partial transversal polytope can be exponential in the size of the input (the family $\mathcal{A}$). In fact, the partial transversal polytope is a projection of the matching polytope of the corresponding graph. Thus we have an illustration of the phenomenon that projection can increase the number of facets dramatically, while this has no negative effect on the complexity of the corresponding optimization problem.

22.6. Packing and covering of transversals

The following min-max relation for the maximum number of disjoint transversals is an easy consequence of Hall's marriage theorem:

Theorem 22.10. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of sets and let k be a natural number. Then $\mathcal{A}$ has k disjoint transversals if and only if*

(22.20) $$|A_I| \geq k|I|$$

for each subset I of $\{1, \ldots, n\}$.

Proof. Replace each set A_i by k copies, yielding the family $\mathcal{A}'$. Then by Hall's marriage theorem and (22.20), $\mathcal{A}'$ has a transversal. This can be split into k transversals of $\mathcal{A}$. ∎

A generalization to disjoint partial transversals of prescribed sizes was given by Higgins [1959] (cf. Mirsky [1966], Mirsky and Perfect [1966]):

Theorem 22.11. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of sets and let $d_1, \ldots, d_k \in \{1, \ldots, n\}$. Then $\mathcal{A}$ has k disjoint partial transversals of sizes $d_1, \ldots, d_k$ respectively if and only if*

(22.21) $$|A_I| \geq \sum_{j=1}^{k} \max\{0, |I| - n + d_j\}$$

for each $I \subseteq \{1, \ldots, n\}$.

Proof. Necessity follows from the fact that if $T_1, \ldots, T_k$ are partial transversals as required, then

$$|A_I| \geq \sum_{j=1}^{k} |A_I \cap T_j| \geq \sum_{j=1}^{k} \max\{0, |I| - n + d_j\} \tag{22.22}$$

for each $I \subseteq \{1,\ldots,n\}$, since $|A_I \cap T_j| + (n - d_j) \geq |I|$.

To see sufficiency, let $B_1,\ldots,B_k$ be disjoint sets, disjoint also from all A_i, with $|B_j| = n - d_j$ for $j = 1,\ldots,k$. Define $A_{i,j} := A_i \cup B_j$ for $i = 1,\ldots,n$ and $j = 1,\ldots,k$. Then $\mathcal{A}$ has k disjoint partial transversals as required, if $(A_{i,j} \mid i = 1,\ldots,n; j = 1,\ldots,k)$ has a transversal. So it suffices to check Hall's condition (22.4) for the latter family. Take $K \subseteq \{1,\ldots,n\} \times \{1,\ldots,k\}$. Let $I := \{i \mid \exists j : (i,j) \in K\}$ and $J := \{j \mid \exists i : (i,j) \in K\}$. Then

$$\begin{aligned} |\bigcup_{(i,j)\in K} A_{i,j}| &= |\bigcup_{i\in I} A_i| + |\bigcup_{j\in J} B_j| \\ &\geq \sum_{j=1}^{k} \max\{0, |I| - n + d_j\} + \sum_{j\in J}(n - d_j) \geq \sum_{j\in J} |I| = |I| \cdot |J| \\ &\geq |K|. \end{aligned} \tag{22.23}$$

∎

(A proof based on total unimodularity was given by Hoffman [1976b].)

As to covering by partial transversals, Mirsky [1971b] (p. 51) mentioned that R. Rado proved in 1965:

Theorem 22.12. *Let $\mathcal{A} = (A_1,\ldots,A_n)$ be a family of subsets of a set S and let k be a natural number. Then S can be covered by k partial transversals if and only if*

$$k \cdot |\{i \mid A_i \cap X \neq \emptyset\}| \geq |X| \tag{22.24}$$

for each subset X of S.

Proof. Let $\mathcal{A}'$ be the family obtained from $\mathcal{A}$ by taking each set k times. Then S can be covered by k partial transversals if and only if S is a partial transversal of $\mathcal{A}'$. By the defect form of Hall's marriage theorem (Theorem 22.2), this last is equivalent to the condition that

$$|S \setminus X| + k \cdot |\{i \mid A_i \cap X \neq \emptyset\}| \geq |S| \tag{22.25}$$

for each $X \subseteq S$. This is equivalent to (22.24). ∎

For covering by partial transversals of prescribed size, there is the following easy consequence of the exchange property of transversals (Corollary 22.4a):

Theorem 22.13. *Let $\mathcal{A}$ be a family of subsets of a set S and let $k \in \mathbb{Z}_+$. If S can be covered by k partial transversals, it can be covered by k partial transversals each of size $\lfloor |S|/k \rfloor$ or $\lceil |S|/k \rceil$.*

Proof. Let $T_1,\ldots,T_k$ be partial transversals partitioning S. If $|T_i| \geq |T_j| + 2$ for some i,j, we can replace T_i and T_j by $T_i \setminus \{s\}$ and $T_j \cup \{s\}$ for some

$s \in T_i$. Repeating this, we finally achieve that $\big||T_i| - |T_j|\big| \leq 1$ for all i, j. Hence $\lfloor |S|/k \rfloor \leq |T_i| \leq \lceil |S|/k \rceil$ for all i. ∎

22.7. Further results and notes

22.7a. The capacitated case

Capacitated versions of the theorems on transversals can be derived straightforwardly from the previous results. First, Halmos and Vaughan [1950] showed the following generalized (but straightforwardly equivalent) version of Hall's marriage theorem:

Theorem 22.14. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of sets and let $b \in \mathbb{Z}_+^n$. Then there exist disjoint subsets $B_1, \ldots, B_n$ of $A_1, \ldots, A_n$ respectively with $|B_i| = b_i$ for $i = 1, \ldots, n$ if and only if*

(22.26) $\qquad |A_I| \geq b(I)$

for each $I \subseteq \{1, \ldots, n\}$.

Proof. Let $\mathcal{A}'$ be the family of sets obtained from $\mathcal{A}$ by repeating any A_i b_i times. Then the existence of the B_i is equivalent to the existence of a transversal of $\mathcal{A}'$. Moreover, (22.26) is equivalent to Hall's condition for $\mathcal{A}'$. ∎

This theorem concerns taking multiplicities on the sets in $\mathcal{A}$. If we put multiplicities on the elements of S, there is the following observation of R. Rado (as reported by Mirsky and Perfect [1966]):

Theorem 22.15. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of sets and let $r \in \mathbb{Z}_+$. Then there exist $x_i \in A_i$ $(i = 1, \ldots, n)$ such that no element occurs more than r times among the s_i if and only if*

(22.27) $\qquad |A_I| \geq |I|/r$

for each $I \subseteq \{1, \ldots, n\}$.

Proof. Let $\mathcal{A}'$ be the family of sets obtained from $\mathcal{A}$ by replacing any A_i by $A_i \times \{1, \ldots, r\}$. Then the existence of the required s_i is equivalent to the existence of a transversal of $\mathcal{A}'$. Moreover, (22.27) is equivalent to Hall's condition for $\mathcal{A}'$. ∎

These theorems are in fact direct consequences of the general Theorem 21.28. This theorem moreover gives the following result of Vogel [1961], which puts multiplicities both on the sets in $\mathcal{A}$ and on the elements of the underlying set S:

Theorem 22.16. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a family of subsets of a set S. Let $a \in \mathbb{Z}_+^n$ and $b \in \mathbb{Z}_+^S$. Then there exist subsets $B_1, \ldots, B_n$ of $A_1, \ldots, A_n$ respectively such that $|B_i| = a_i$ for $i = 1, \ldots, n$ and such that each $s \in S$ occurs in at most $b(s)$ of the B_i if and only if*

(22.28) $\qquad b(X) + \sum_{i \in I} |A_i \setminus X| \geq a(I)$

for each $X \subseteq S$ *and each* $I \subseteq \{1,\dots,n\}$.

Proof. Consider the system

$$\text{(22.29)} \qquad \begin{array}{ll} 0 \le x(i,s) \le 1 & \text{for } i \in \{1,\dots,n\} \text{ and } s \in A_i, \\ a_i \le x(\delta(i)) \le a_i & \text{for } i \in \{1,\dots,n\}, \\ 0 \le x(\delta(s)) \le b_s & \text{for } s \in S, \end{array}$$

and apply Theorem 21.28. ∎

This has as special case (Vogel [1961]):

Corollary 22.16a. *Let* $\mathcal{A} = (A_1,\dots,A_n)$ *be a family of subsets of a set* S *and let* $r, s \in \mathbb{Z}_+$. *Then there exist subsets* $B_1,\dots,B_n$ *of* $A_1,\dots,A_n$ *respectively such that* $|B_i| = s$ *for each* i *and such that each element belongs to at most* r *of the* B_i *if and only if*

$$\text{(22.30)} \qquad r|X| + \sum_{i \in I} |A_i \setminus X| \ge s|I|$$

for each $I \subseteq \{1,\dots,n\}$ *and each* $X \subseteq S$.

Proof. This is a special case of Theorem 22.16. ∎

Similar methods apply to systems of *restricted* representatives, considered by Ford and Fulkerson [1958c]. Let $\mathcal{A} = (A_1,\dots,A_n)$ be a collection of subsets of a set S and let $a, b \in \mathbb{Z}_+^S$ with $a \le b$. A *system of restricted representatives* (or *SRR*) of $\mathcal{A}$ (with respect to a and b) is a sequence $(s_1,\dots,s_n)$ such that

(22.31) (i) $s_i \in A_i$ for $i = 1,\dots,n$;
(ii) $a(s) \le |\{i \mid s_i = s\}| \le b(s)$ for $s \in S$.

Ford and Fulkerson [1958c] showed:

Theorem 22.17. $\mathcal{A}$ *has a system of restricted representatives if and only if*

$$\text{(22.32)} \qquad a(S - \bigcup_{i \notin I} A_i) \le |I| \le b(\bigcup_{i \in I} A_i)$$

for each $I \subseteq \{1,\dots,n\}$.

Proof. Consider the system

$$\text{(22.33)} \qquad \begin{array}{ll} 0 \le x(i,s) \le \infty & \text{for } i \in \{1,\dots,n\}, s \in A_i, \\ x(\delta(i)) = 1 & \text{for } i \in \{1,\dots,n\}, \\ a_s \le x(\delta(s)) \le b_s & \text{for } s \in S, \end{array}$$

and apply Theorem 21.28. ∎

(For an alternative proof, see Mirsky [1968a].)

Considering both upper and lower bounds, the following theorem of Hoffman and Kuhn [1956a] follows from Hoffman's circulation theorem:

Theorem 22.18. *Let* $\mathcal{A} = (A_1,\dots,A_n)$ *be a collection of subsets of a set* S, *let* $\mathcal{P} = (P_1,\dots,P_m)$ *be a partition of* S, *and let* $a, b \in \mathbb{Z}_+^m$ *with* $a \le b$. *Then* $\mathcal{A}$ *has a transversal* T *satisfying* $a_i \le |T \cap P_i| \le b_i$ *for each* $i = 1,\dots,m$ *if and only if*

(22.34) $$|P_I \cap A_J| \geq \max\{|J| - b(\overline{I}), |J| - n + a(I)\}$$

for all $I \subseteq \{1, \ldots, m\}$ *and* $J \subseteq \{1, \ldots, n\}$*, where* $\overline{I} := \{1, \ldots, n\} \setminus I$*.*

Proof. Make a directed graph as follows. Its vertex set is $\{r\} \cup \{u_1, \ldots, u_n\} \cup S \cup \{p_1, \ldots, p_m\} \cup \{t\}$, and there are arcs

(22.35)
(r, u_i) for $i = 1, \ldots, n$,
(u_i, s) for $i = 1, \ldots, n$ and $s \in A_i$,
(s, p_j) for $j = 1, \ldots, m$ and $s \in P_j$,
(p_j, t) for $j = 1, \ldots, m$.

Put lower bound a_j and capacity b_j on each arc (p_j, t). On any other arc, put lower bound 0 and capacity 1. Then a transversal as required exists if and only if there is an integer $r-t$ flow of value n satisfying the lower bounds and capacities. Applying Corollary 11.2e gives the present theorem. ∎

(The proof of Hoffman and Kuhn [1956a] is based on the duality theorem of linear programming. Gale [1956,1957] and Fulkerson [1959a] derived the theorem from network flow theory. For further extensions, see Mirsky [1968b].)

22.7b. A theorem of Rado

Rado [1938] proved the following generalization (but also consequence) of Hall's marriage theorem:

Theorem 22.19. *Let* $A_1, \ldots, A_n$*,* $B_1, \ldots, B_n$ *be sets. Then there exists an injection* $f : A_1 \cup \cdots \cup A_n \to B_1 \cup \cdots \cup B_n$ *such that* $f[A_i] \subseteq B_i$ *for* $i = 1, \ldots, n$ *if and only if each set obtained by intersections and unions of sets from* $A_1, \ldots, A_n$ *has size at most the size of the result of the same operations applied to* $B_1, \ldots, B_n$*.*

Proof. Let $A := A_1 \cup \cdots \cup A_n$. For each $s \in A$, define

(22.36) $$C_s := \bigcap_{\substack{i \\ s \in A_i}} B_i.$$

Then for each subset S of A one has

(22.37) $$\Big|\bigcup_{s \in S} C_s\Big| = \Big|\bigcup_{s \in S} \bigcap_{\substack{i \\ s \in A_i}} B_i\Big| \geq \Big|\bigcup_{s \in S} \bigcap_{\substack{i \\ s \in A_i}} A_i\Big| \geq |S|.$$

Hence, by Hall's marriage theorem, $(C_s \mid s \in A)$ has a transversal. This gives an injection $f : A \to B_1 \cup \cdots \cup B_n$ with $f(s) \in C_s$ for $s \in A$. This is as required. ∎

22.7c. Further notes

Shmushkovich [1939], de Bruijn [1943], Hall [1948], Henkin [1953], Tutte [1953], Mirsky [1967], Rado [1967a] (with H.A. Jung), Brualdi and Scrimger [1968], Folkman [1970], McCarthy [1973], Damerell and Milner [1974], Steffens [1974], Podewski and Steffens [1976], Nash-Williams [1978], Aharoni [1983c], and Aharoni, Nash-Williams, and Shelah [1983] considered extensions of Hall's marriage theorem to the infinite case. Perfect [1968] gave proofs of theorems on transversals with Menger's theorem.

For a 'very general theorem' see Brualdi [1969a]. For counting transversals, see Hall [1948], Rado [1967b], and Ostrand [1970].

Gale [1968] showed that for any family $\mathcal{A}$ of subsets of a finite set S and any total order $<$ on S, there is a transversal T of $\mathcal{A}$ such that for each transversal T' of $\mathcal{A}$ there exists a one-to-one function $\phi : T' \to T$ with $\phi(s) \geq s$ for each $s \in T'$. (Gale showed that this in fact characterizes matroids.)

The standard work on transversal theory is Mirsky [1971b]. Also Brualdi [1975] and Welsh [1976] provide surveys. Surveys on the relations between the theorems of Hall, Kőnig, Menger, and Dilworth were given by Jacobs [1969] and Reichmeider [1984].

22.7d. Historical notes on transversals

Results on transversals go back to the papers by Miller [1910] and Chapman [1912], who showed that if H is a subgroup of a finite group G, then the partitions of G into left cosets and into right cosets have a common transversal. This is an easy result, due to the fact that each component of the intersection graph of left and right cosets is a complete bipartite graph. This implies that any common partial transversal can be extended to a common (full) transversal (Chapman [1912]).

This result was extended by Scorza [1927] to: if H and K are subgroups of a finite group G, with $|H| = |K|$, then there exist $x_1, \ldots, x_m \in G$ with $x_1 H \cup \cdots \cup x_m H = G = K x_1 \cup \cdots \cup K x_m$ and $m = |G|/|H|$. (Again this can be derived easily from the fact that each component of the intersection graph of left cosets of H and right cosets of K is a complete bipartite graph.)

As an extension of these results, in October 1926, van der Waerden [1927] presented the following theorem at the *Mathematisches Seminar* in Hamburg:

> *Es seien zwei Klasseneinteilungen einer endlichen Menge $\mathcal{M}$ gegeben. Die eine soll die Menge in μ zueinander fremde Klassen $\mathcal{A}_1, \ldots, \mathcal{A}_\mu$ zu je n Elementen zerlegen, die andere ebenfalls in μ fremde Klassen $\mathcal{B}_1, \ldots, \mathcal{B}_\mu$ zu je n Elementen. Dann gibt es ein System von Elementen $x_1, \ldots, x_\mu$, derart, daß jede A-Klasse und ebenso jede B-Klasse under den x_i durch ein Element vertreten wird.*[41]

The proof of van der Waerden is based on an augmenting path argument. Moreover, van der Waerden remarked that E. Artin had communicated orally to him that the result can be sharpened to the existence of n disjoint such common transversals.

In a note added in proof, van der Waerden observed that his theorem follows from Kőnig's theorem on the existence of a perfect matching in a regular bipartite graph:

> **Zusatz bei der Korrektur.** Ich bemerke jetzt, daß der hier bewiesene Satz mit einem Satz von DÉNES KŐNIG über reguläre Graphen äquivalent ist.[42]

[41] Let be given two partitions of a finite set $\mathcal{M}$. One of them should decompose the set into μ mutually disjoint classes $\mathcal{A}_1, \ldots, \mathcal{A}_\mu$ each of n elements, the other likewise in μ disjoint classes $\mathcal{B}_1, \ldots, \mathcal{B}_\mu$ each of n elements. Then there exists a system of elements $x_1, \ldots, x_\mu$ such that each $\mathcal{A}$-class and likewise each $\mathcal{B}$-class is represented by one element among the x_i.

[42] **Note added in proof.** I now notice that the theorem proved here is equivalent to a theorem of DÉNES KŐNIG on regular graphs.

Van der Waerden's article is followed by an article of Sperner [1927] (presented at the *Mathematisches Seminar* in Januari 1927) that gives a 'simple proof' of van der Waerden's result. We quote the full article (containing page references to van der Waerden's paper):

> Der auf S. 185 ff. bewiesene Satz gestattet auch folgenden einfachen Beweis.
> Der Satz lautete:
> Zwei beliebige Klasseneinteilungen von $m \cdot n$ Elementen in m Klassen zu je n Elementen haben immer ein gemeinsames Repräsentantensystem (vgl. S. 185).
> Der Satz ist evident für die Klassenzahl 1. Wir nehmen an, er sei bewiesen für die Klassenzahl m (und beliebiges n). Dan folgt für dieses m:
> 1. Die beiden Klasseneinteilungen haben sogar n verschiedene und zueinander fremde Repräsentantensysteme.
> Beweis wie auf S. 187 oben.
> 2. Streicht man daher in beiden Einteilungen dieselben k Elemente, wo $0 \leq k \leq n-1$, dan werden höchstens $n-1$ Repräsentantensysteme verletzt und wenigstens eins bleibt erhalten. Da man auch umgekehrt $m \cdot n - k$ Elemente durch k neue ergänzen kann, um diese nachher wieder zu streichen, so gilt:
> Zwei beliebige Klasseneinteilungen von $m \cdot n - k$ Elementen in m Klassen zu je höchstens n Elementen, wo $0 \leq k \leq n - 1$, haben immer ein gemeinsames Repräsentantensystem.
> Nunmehr wenden wir vollständige Induktion an. Es seien zwei Klasseneinteilungen von $(m+1) \cdot n$ Elementen in $m+1$ Klassen zu je n Elementen gegeben. Dann greifen wir aus beiden Einteilungen je eine Klasse heraus, etwa die Klassen $\mathcal{A}$ und $\mathcal{B}$, die aber wenigstens 1 Element gemeinsam haben sollen, etwa A. Streichen wir dann in beiden Einteilungen die in $\mathcal{A}$ und $\mathcal{B}$ vorkommenden Elemente (also höchstens $2n - 1$, aber wenigstens n Elemente), so bleiben zwei Klasseneinteilungen von $m \cdot n - k$ Elementen in je m Klassen zu je höchstens n Elementen übrig, wo $0 \leq k \leq n - 1$. Zwei solche Einteilungen haben aber nach 2. ein gemeinsames Repräsentantensystem, das man sofort durch Hinzufügen von A zu einem gemeinsamen Repräsentantensysteme der beiden Einteilungen von $(m + 1)n$ Elementen erweitert.[43]

[43] The theorem proved on p. 185 and following pages allows also the following simple proof. The theorem reads:

Two arbitrary partitions of $m \cdot n$ elements into m classes of n elements each, always have a common system of representatives (cf. p. 185).

The theorem is evident for class number 1. We assume that it be proved for class number m (and arbitrary n). Then the following follows for this m:

1. Both partitions even have n different and disjoint systems of representatives.

Proof like on p. 187 above.

2. Therefore, if one cancels in both partitions the same k elements, where $0 \leq k \leq n - 1$, then at most $n - 1$ systems of representatives are injured and at least one is preserved. As one can also, reversely, complete $m \cdot n - k$ elements by k new ones, to cancel them after it again, the following therefore holds:

Two arbitrary partitions of $m \cdot n - k$ elements into m classes of at most n elements each, where $0 \leq k \leq n - 1$, always have a common system of representatives.

Now we apply complete induction. Let be given two partitions of $(m+1) \cdot n$ elements into $m+1$ classes of n elements each. Then we select from each of the two partitions one class, say the classes $\mathcal{A}$ and $\mathcal{B}$, that however should have at least 1 element in common, say A. If we then cancel in both partitions the elements occurring in $\mathcal{A}$ and $\mathcal{B}$ (so at most $2n-1$, but at least n elements), two partitions of $m \cdot n - k$ elements into m classes of at most n elements each thus remain, where $0 \leq k \leq n - 1$. Two such partitions have however, according to 2., a common system of representatives, that one extends, by adding A, to a common system of representatives of both partitions of $(m + 1)n$ elements.

Hall

After having mentioned Kőnig's result on the existence of a common transversal for two partitions of a set where all classes have the same size, Hall [1935] said that he is 'concerned with a slightly different problem': to find a transversal

> for a finite collection of (arbitrarily overlapping) subsets of any given set of things. The solution, Theorem 1, is very simple.

Calling a transversal a 'C.D.R. (= complete system of distinct representatives)' and denoting a finite system $T_1, \ldots, T_m$ of subsets of a set S by '(1)', Hall formulated his theorem as follows:

> *In order that a C.D.R. of* (1) *shall exist, it is sufficient that for each* $k = 1, 2, \ldots, m$ *any selection of* k *of the sets* (1) *shall contain between them at least* k *elements of* S.

This result now is known as 'Hall's marriage theorem'.

In order to prove this theorem, Hall first showed the following lemma. Let $(A_1, \ldots, A_n)$ be a system of sets with at least one transversal and let R be the intersection of all transversals. Then there is an $I \subseteq \{1, \ldots, n\}$ with $A_I = R$ and $|I| = |R|$.

Hall proved this with the help of an alternating path argument. Having the lemma, the theorem is easy, by induction on n: we may assume that $(A_1, \ldots, A_{n-1})$ has a transversal; let R' be the intersection of all these transversals. So by the lemma, $R' = A_{I'}$ for some $I' \subseteq \{1, \ldots, n-1\}$ with $|I'| = |R'|$. Hence $A_n \not\subseteq R'$, since otherwise for $I := I' \cup \{n\}$ one has $\left|\bigcup_{i \in I} A_i\right| = |R'| < |I|$. Therefore, $(A_1, \ldots, A_{n-1})$ has a transversal not containing A_n as a subset, implying that $(A_1, \ldots, A_n)$ has a transversal.

Hall derived as a consequence that if $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ are two partitions of a finite set S, then the two partitions have a common transversal if and only if for each subset I of $\{1, \ldots, n\}$, the set $\bigcup_{i \in I} A_i$ intersects at least $|I|$ sets among $B_1, \ldots, B_n$. Hall remarked that the theorem of Kőnig [1916] on the existence of a perfect matching in a regular bipartite graph follows as an immediate corollary, and that also a theorem of Rado [1933] can be derived (the Kőnig-Rado edge cover theorem — Theorem 19.4), but he did not observe that Hall's marriage theorem is equivalent to a theorem of Kőnig [1931] (Kőnig's matching theorem — Theorem 16.2).

As for *common* transversals, Maak [1936] showed that if $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ are partitions of a finite set S, then $\mathcal{A}$ and $\mathcal{B}$ have a common transversal if and only if for each $I \subseteq \{1, \ldots, n\}$, the set $\bigcup_{i \in I} A_i$ contains at most $|I|$ of the sets B_i as a subset. This can be derived from Frobenius' theorem (Frobenius [1917]).

The basic characterization of common transversals of two arbitrary families of sets was given by Ford and Fulkerson [1958c] — see Section 23.1.

Shmushkovich [1939] and de Bruijn [1943] extended the results to the infinite case. Weyl [1949] introduced the name 'marriage theorem' for Hall's marriage theorem. Maak [1952] gave some historical notes.

Chapter 23

Common transversals

We consider sets that are transversals of two families of sets simultaneously. Again we denote, for any family $(A_1, \ldots, A_n)$ of sets and any $I \subseteq \{1, \ldots, n\}$,

$$A_I := \bigcup_{i \in I} A_i.$$

23.1. Common transversals

Let $\mathcal{A}$ and $\mathcal{B}$ be families of sets. A set T is called a *common transversal* of $\mathcal{A}$ and $\mathcal{B}$ if T is a transversal of both $\mathcal{A}$ and $\mathcal{B}$. Similarly, T is called a *common partial transversal* of $\mathcal{A}$ and $\mathcal{B}$ if T is a partial transversal of both $\mathcal{A}$ and $\mathcal{B}$.

When considering two families $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_m)$ of subsets of a set S, it is helpful to construct the following directed graph $D = (V, A)$:

(23.1) $V := \{a_1, \ldots, a_n\} \cup S \cup \{b_1, \ldots, b_m\}$,
$A := \{(a_i, s) \mid i = 1, \ldots, n; s \in A_i\} \cup \{(s, b_i) \mid i = 1, \ldots, m; s \in B_i\}$,

where $a_1, \ldots, a_n, b_1, \ldots, b_m$ are distinct new elements, not in S.

Then one has, if $m = n$:

(23.2) a subset T of S is a common transversal of $\mathcal{A}$ and $\mathcal{B}$ if and only if D has n vertex-disjoint paths from $\{a_1, \ldots, a_n\}$ to $\{b_1, \ldots, b_n\}$ such that T is the set of vertices in S traversed by these paths.

A similar statement can be formulated with respect to common partial transversals.

With Menger's theorem, it yields the following characterization of the existence of a common transversal, due to Ford and Fulkerson [1958c]:

Theorem 23.1. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of sets. Then $\mathcal{A}$ and $\mathcal{B}$ have a common transversal if and only if*

(23.3) $|A_I \cap B_J| \geq |I| + |J| - n$

for all $I, J \subseteq \{1, \ldots, n\}$.

Proof. To see necessity, let T be a common transversal. To prove (23.3), we can assume that $A_i \subseteq T$ and $B_j \subseteq T$ for all i, j. Then

(23.4) $$|A_I \cap B_J| = |A_I| + |B_J| - |A_I \cup B_J| \geq |I| + |J| - |T| \geq |I| + |J| - n$$

for all $I, J \subseteq \{1, \ldots, n\}$.

To see sufficiency, make the digraph D associated to $\mathcal{A}, \mathcal{B}$ as in (23.1). Let $U := \{a_1, \ldots, a_n\}$ and $W := \{b_1, \ldots, b_n\}$. Then by (23.2), $\mathcal{A}$ and $\mathcal{B}$ have a common transversal if D has n disjoint $U - W$ paths. By Menger's theorem, these paths exist if $|C| \geq n$ for each $C \subseteq U \cup S \cup W$ intersecting each $U - W$ path. To check this condition, let $I := \{i \mid a_i \notin C\}$ and $J := \{j \mid b_j \notin C\}$. Then

(23.5) $$C \cap S \supseteq A_I \cap B_J,$$

since $A_I \cap B_J$ is equal to the set of vertices in S that are on a $U - W$ path not intersected by $C \cap (U \cup W)$. So (23.3) implies

(23.6) $$|C \cap S| \geq |A_I \cap B_J| \geq |I| + |J| - n = (n - |C \cap U|) + (n - |C \cap W|) - n,$$

giving $|C| \geq n$. ∎

(For a direct derivation of this theorem from Hall's marriage theorem, see Perfect [1969c]. For a derivation from the Kőnig-Rado edge cover theorem, see Perfect [1980].)

This construction also implies, with Theorem 9.8, that a common transversal of two collections of n subsets of S can be found in time $O(n^{3/2}|S|)$ (cf. Adel'son-Vel'skiĭ, Dinits, and Karzanov [1975]).

Perfect [1968] (cf. McDiarmid [1973]) strengthened Theorem 23.1 to a min-max relation for the maximum size of a common partial transversal:

Corollary 23.1a. *Let* $\mathcal{A} = (A_1, \ldots, A_n)$ *and* $\mathcal{B} = (B_1, \ldots, B_m)$ *be families of sets and let* $k \in \mathbb{Z}_+$. *Then* $\mathcal{A}$ *and* $\mathcal{B}$ *have a common partial transversal of size* k *if and only if*

(23.7) $$|A_I \cap B_J| \geq |I| + |J| - n - m + k$$

for all $I \subseteq \{1, \ldots, n\}$ *and* $J \subseteq \{1, \ldots, m\}$.

Proof. We may assume that $m = n$ (if, say, $n < m$, add $m - n$ copies of $\emptyset$ to $\mathcal{A}$). Let X be a set disjoint from all A_i and B_i with $|X| = n - k$. Replace each A_i by $A'_i := A_i \cup X$ and each B_i by $B'_i := B_i \cup X$. Then $\mathcal{A}$ and $\mathcal{B}$ have a common partial transversal of size k if and only if $\mathcal{A}' = (A'_1, \ldots, A'_n)$ and $\mathcal{B}' = (B'_1, \ldots, B'_n)$ have a common transversal. Applying Theorem 23.1 to $\mathcal{A}'$ and $\mathcal{B}'$ gives this corollary. ∎

Generally, a common partial transversal of families $\mathcal{A}$ and $\mathcal{B}$ need not be contained in a common transversal, even not if a common transversal exists:

let $\mathcal{A} := (\{a\}, \{b, c\})$ and $\mathcal{B} := (\{b\}, \{a, c\})$. Then $\{c\}$ is a common partial transversal, while $\{a, b\}$ is the only common transversal.

The following result of Perfect [1968] and Welsh [1968] characterizes subsets contained in common transversals. It is a special case of a theorem of Ford and Fulkerson [1958c] (cf. Theorem 23.14), and will be derived from Theorem 23.1 with a method of Mirsky and Perfect [1968].

Corollary 23.1b. *Let* $\mathcal{A} = (A_1, \ldots, A_n)$ *and* $\mathcal{B} = (B_1, \ldots, B_n)$ *be families of subsets of a set* S *and let* $X \subseteq S$. *Then* $\mathcal{A}$ *and* $\mathcal{B}$ *have a common transversal containing* X *if and only if*

$$(23.8) \qquad |A_I \cap B_J| \geq |I| + |J| - n + |X \setminus (A_I \cup B_J)|$$

for all $I, J \subseteq \{1, \ldots, n\}$.

Proof. To see necessity, we can assume that there is a common transversal T containing each A_i, each B_j, and X. Then for all $I, J \subseteq \{1, \ldots, n\}$:

$$(23.9) \qquad |A_I \cap B_J| = |A_I| + |B_J| - |A_I \cup B_J| \geq |I| + |J| + |X \setminus (A_I \cup B_J)| - n$$

since $|A_I \cup B_J| + |X \setminus (A_I \cup B_J)| \leq |T| = n$.

To see sufficiency, let $X = \{x_1, \ldots, x_k\}$ and let $x'_1, \ldots, x'_k$ be new elements. For each $i = 1, \ldots, n$, let A'_i be the set obtained from A_i by replacing any occurrence of x_j by x'_j. Then $\mathcal{A}$ and $\mathcal{B}$ have a common transversal containing X if the families

$$(23.10) \qquad \begin{aligned} &\mathcal{A}' := (A'_1, \ldots, A'_n, \{x_1\}, \ldots, \{x_k\}) \text{ and} \\ &\mathcal{B}' := (B_1, \ldots, B_n, \{x'_1\}, \ldots, \{x'_k\}) \end{aligned}$$

have a common transversal. So by Theorem 23.1 we must check condition (23.3) for $\mathcal{A}'$ and $\mathcal{B}'$. Let $I, J \subseteq \{1, \ldots, n\}$ and $I', J' \subseteq \{1, \ldots, k\}$. Define $Y := \{x_i \mid i \in I'\}$ and $Z := \{x_i \mid i \in J'\}$. Then

$$(23.11) \qquad \begin{aligned} &\Big|\Big(\bigcup_{i \in I} A'_i \cup \bigcup_{i \in I'} \{x_i\}\Big) \cap \Big(\bigcup_{j \in J} B_j \cup \bigcup_{j \in J'} \{x'_j\}\Big)\Big| \\ &= |(A_I \cap B_J) \setminus X| + |A_I \cap Z| + |B_J \cap Y| \\ &= |(A_I \cap B_J) \setminus X| + |Z| - |Z \setminus A_I| + |Y| - |Y \setminus B_J| \\ &\geq |(A_I \cap B_J) \setminus X| + |Z| - |X \setminus A_I| + |Y| - |X \setminus B_J| \\ &= |A_I \cap B_J| - |X \setminus (A_I \cup B_J)| + |Y| + |Z| - |X| \\ &\geq |I| + |J| + |Y| + |Z| - |X| - n = |I| + |I'| + |J| + |J'| - n - k \end{aligned}$$

(the last inequality follows from (23.8)). ∎

23.2. Weighted common transversals

Consider the problem of finding a minimum-weight common transversal: given families $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ of subsets of a set S and a weight function $w : S \to \mathbb{Q}$, find a common transversal T of $\mathcal{A}$ and $\mathcal{B}$

minimizing $w(T)$. This problem can easily be solved by solving an associated minimum-cost flow problem.

Alternatively, it can be solved with the Hungarian method, as follows. For $s \in S$, introduce a copy s' of s. Let $S' := \{s' \mid s \in S\}$. Let $a_1, \ldots, a_n, b_1, \ldots, b_n$ be vertices. Make a bipartite graph G with colour classes $\{a_1, \ldots, a_n\} \cup S$ and $\{b_1, \ldots, b_n\} \cup S'$. Vertex a_i is connected with vertex $s' \in S'$ if $s \in A_i$. Vertex b_i is connected with vertex $s \in S$ if $s \in B_i$. Moreover, each $s \in S$ is connected with its copy $s' \in S'$. This describes all edges of G.

For any perfect matching M in G, the set of $s \in S$ with $\{s, s'\} \notin M$ is a common transversal of $\mathcal{A}$ and $\mathcal{B}$. Conversely, each common transversal can be obtained in this way from a perfect matching in G.

Therefore, a minimum-weight common transversal of $\mathcal{A}$ and $\mathcal{B}$ can be found by determining a maximum-weight perfect matching in G, taking weight $w(s)$ on any edge $\{s, s'\}$ and weight 0 on any other edge of G. So by Theorem 17.3 we can find a minimum-weight common transversal in time $O(k(m + k \log k))$, where

$$(23.12) \qquad k := n + |S| \text{ and } m := \sum_{i=1}^{n} (|A_i| + |B_i|).$$

Due to the special structure of G and its weight function one can sharpen this to:

Theorem 23.2. *A minimum-weight common transversal can be found in time $O(n(m + k \log k))$, with m and k as in (23.12).*

Proof. We may assume that $w(s) \geq 0$ for each $s \in S$ (we can add a constant to all weights). Then we can start the Hungarian method with the matching M consisting of all edges $\{s, s'\}$ with $s \in S$. This matching is extreme (that is, has maximum weight among all matchings of size $|M|$), and the Hungarian method requires only n iterations to obtain a maximum-weight perfect matching. ∎

Note that, unlike what happened in finding a minimum-weight transversal for *one* family of sets, in the algorithm above we do not grow a common partial transversal — we do backtrack.

We can also obtain a min-max relation for the minimum weight of a common transversal:

Theorem 23.3. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of subsets of a set S and let $w : S \to \mathbb{Z}$ be a weight function. Then the minimum weight of a common transversal of $\mathcal{A}$ and $\mathcal{B}$ is equal to the maximum value of*

$$(23.13) \qquad \sum_{i=1}^{n} (\min_{s \in A_i} w_1(s) + \min_{s \in B_i} w_2(s)) + (w(S) - w_1(S) - w_2(S))$$

taken over $w_1, w_2 \in \mathbb{Z}^S$ *satisfying* $w_1 + w_2 \geq w$.

Proof. Consider the graph G above. By Theorem 17.5 (or by total unimodularity), the maximum weight of a perfect matching in G is equal to the minimum value of

$$(23.14) \qquad \sum_{i=1}^{n}(\lambda_i + \mu_i) + \sum_{s\in S}(w_1(s) + w_2(s))$$

taken over $\lambda, \mu \in \mathbb{Z}^n$ and $w_1, w_2 \in \mathbb{Z}^S$ satisfying

$$(23.15) \qquad \begin{array}{ll} \lambda_i + w_1(s) \geq 0 & \text{for } i = 1,\ldots,n \text{ and } s \in A_i, \\ w_1(s) + w_2(s) \geq w(s) & \text{for } s \in S, \\ \mu_i + w_2(s) \geq 0 & \text{for } i = 1,\ldots,n \text{ and } s \in B_i. \end{array}$$

We can assume that $\lambda_i = \max\{-w_1(s) \mid s \in A_i\}$ and $\mu_i = \max\{-w_2(s) \mid s \in B_i\}$ for each $i = 1,\ldots,n$.

Now the minimum weight of a common transversal is equal to $w(S)$ minus the maximum weight of a perfect matching in G. So it is equal to the maximum value of

$$(23.16) \qquad w(S) - \sum_{s\in S}(w_1(s) + w_2(s)) + \sum_{i=1}^{n}(\min_{s\in A_i} w_1(s) + \min_{s\in B_i} w_2(s)),$$

where $w_1, w_2 \in \mathbb{Z}^S$ satisfy $w_1 + w_2 \geq w$. This is equal to (23.13). ∎

23.3. Weighted common partial transversals

A maximum-weight common partial transversal can be found with the Hungarian method, like described at the beginning of Section 23.2. At any stage of the Hungarian method the current matching M is extreme (that is, it has optimum weight among all matchings of size $|M|$). So we can also apply it (like in Theorem 23.2) to find a maximum-weight common partial transversal of two families $\mathcal{A} = (A_1,\ldots,A_k)$ and $\mathcal{B} = (B_1,\ldots,B_l)$ of subsets of a set S. Taking

$$(23.17) \qquad n := k + l + |S| \text{ and } m := \sum_{i=1}^{k}|A_i| + \sum_{i=1}^{l}|B_i|,$$

we have:

Theorem 23.4. *A maximum-weight common partial transversal can be found in time* $O(\min\{k,l\}(m + n\log n))$.

Proof. As above. ∎

Note that, even if all weights are positive, a maximum-weight common partial transversal need not be a common transversal (a statement that is true if we delete 'common'). To see this, let $\mathcal{A} = (\{a\}, \{b, c\})$, $\mathcal{B} = (\{b\}, \{a, c\})$, and $w(a) = w(b) = 1$, $w(c) = 3$. Then $\{c\}$ is the only maximum-weight common partial transversal, while $\{a, b\}$ is the only common transversal.

A min-max relation for the maximum weight of a common partial transversal can be derived from a min-max relation for the maximum weight of a matching in a bipartite graph, or from linear programming duality using total unimodularity, as we do in the proof below:

Theorem 23.5. *Let $\mathcal{A} = (A_1, \ldots, A_k)$ and $\mathcal{B} = (B_1, \ldots, B_l)$ be families of subsets of a set S and let $w : S \to \mathbb{Z}_+$ be a weight function. Then the maximum weight of a common partial transversal of $\mathcal{A}$ and $\mathcal{B}$ is equal to the minimum value of*

$$\text{(23.18)} \qquad \sum_{i=1}^{k} \max_{s \in A_i} w_1(s) + \sum_{i=1}^{l} \max_{s \in B_i} w_2(s) + (w - w_1 - w_2)(S)$$

where $w_1, w_2 \in \mathbb{Z}_+^S$ with $w_1 + w_2 \leq w$.

Proof. The maximum weight of a common partial transversal is equal to the maximum of $w^{\mathsf{T}} x$ where $x \in \mathbb{Z}^S$ such that there exist $y_1(i, s) \in \mathbb{Z}_+$ ($i = 1, \ldots, k$; $s \in A_i$) and $y_2(i, s) \in \mathbb{Z}_+$ ($i = 1, \ldots, l$; $s \in B_i$) satisfying

$$\text{(23.19)} \qquad \begin{array}{ll} \displaystyle\sum_{s \in A_i} y_1(i, s) \leq 1 & \text{for } i = 1, \ldots, k, \\ \displaystyle\sum_{s \in B_i} y_2(i, s) \leq 1 & \text{for } i = 1, \ldots, l, \\ x_s = \displaystyle\sum_{i, s \in A_i} y_1(i, s) & \text{for } s \in S, \\ x_s = \displaystyle\sum_{i, s \in B_i} y_2(i, s) & \text{for } s \in S, \\ 0 \leq x_s \leq 1 & \text{for } s \in S. \end{array}$$

By linear programming duality and the total unimodularity of the constraint matrix in (23.19), the maximum value is equal to the minimum value of

$$\text{(23.20)} \qquad \sum_{i=1}^{k} z_1(i) + \sum_{i=1}^{l} z_2(i) + \sum_{s \in S} u(s),$$

where $z_1, z_2 \in \mathbb{Z}_+^k$ and $u \in \mathbb{Z}_+^S$ satisfy

$$\text{(23.21)} \qquad \begin{array}{ll} z_1(i) \geq w_1(s) & \text{for } i = 1, \ldots, k \text{ and } s \in A_i, \\ z_2(i) \geq w_2(s) & \text{for } i = 1, \ldots, l \text{ and } s \in B_i, \\ w_1(s) + w_2(s) + u(s) \geq w(s) & \text{for } s \in S, \end{array}$$

for some $w_1, w_2 \in \mathbb{Z}^E$. We may assume that $w_1, w_2 \geq \mathbf{0}$, since replacing any negative $w_j(s)$ by 0 does not violate (23.21). We may assume that $w_1 + w_2 + u = w$, since $w \geq \mathbf{0}$, and hence we can decrease $w_1(s)$, $w_2(s)$ or $u(s)$ if $w_1(s) + w_2(s) + u(s) > w(s)$. This gives the theorem. ∎

By specializing w to the all-one function, Theorem 23.5 reduces to Corollary 23.1a on the maximum size of a common partial transversal. We can also derive an alternative min-max relation for the maximum weight of a common partial transversal, expressed in

(23.22) $m(\mathcal{C}, w) :=$ maximum weight of a partial transversal of $\mathcal{C}$

for any family $\mathcal{C}$ and weight function w (so we can plug in a min-max relation for $m(\mathcal{C}, w)$ to obtain a genuine min-max relation):

Corollary 23.5a. *Let $\mathcal{A} = (A_1, \ldots, A_k)$ and $\mathcal{B} = (B_1, \ldots, B_l)$ be families of subsets of a set S and let $w : S \to \mathbb{Z}_+$ be a weight function. Then the maximum weight of a common partial transversal of $\mathcal{A}$ and $\mathcal{B}$ is equal to the minimum value of $m(\mathcal{A}, w_1) + m(\mathcal{B}, w_2)$, taken over $w_1, w_2 \in \mathbb{Z}_+^S$ with $w_1 + w_2 = w$.*

Proof. Clearly, the maximum value here cannot be larger than the minimum value, since $w(T) = w_1(T) + w_2(T) \leq m(\mathcal{A}, w_1) + m(\mathcal{B}, w_2)$ for any maximum-weight common partial transversal T.

To see equality, consider w_1 and w_2 of Theorem 23.5, and let $w_2' := w - w_1$. Then for any partial transversal T_1 of $\mathcal{A}$ one has

(23.23) $$w_1(T_1) \leq \sum_{i=1}^{k} \max_{s \in A_i} w_1(s).$$

Moreover, for any partial transversal T_2 of $\mathcal{B}$ one has

(23.24) $$\begin{aligned} w_2'(T_2) &= w_2(T_2) + (w - w_1 - w_2)(T_2) \\ &\leq \sum_{i=1}^{k} \max_{s \in B_i} w_2(s) + (w - w_1 - w_2)(S). \end{aligned}$$

So by Theorem 23.5 we have that $m(\mathcal{A}, w_1) + m(\mathcal{B}, w_2')$ is not more than the maximum w-weight of a common partial transversal. ∎

The obvious generalization to common partial transversals of *three* families is not true: take

(23.25) $\mathcal{A} = (\{a\}, \{b, c\})$, $\mathcal{B} = (\{b\}, \{a, c\})$, and $\mathcal{C} = (\{c\}, \{a, b\})$,

and $w(a) = w(b) = w(c) = 1$. Then the maximum weight of a common partial transversal is 1, but one cannot decompose w as $w = w_1 + w_2 + w_3$ with $m(\mathcal{A}, w_1) + m(\mathcal{B}, w_2) + m(\mathcal{C}, w_3) = 1$.

23.4. The common partial transversal polytope

Let $\mathcal{A} = (A_1, \ldots, A_k)$ and $\mathcal{B} = (B_1, \ldots, B_l)$ be families of subsets of a set S. The *common partial transversal polytope* $P_{\text{common partial transversal}}(\mathcal{A}, \mathcal{B})$ of

$\mathcal{A}$ and $\mathcal{B}$ is the convex hull of the incidence vectors (in $\mathbb{R}^S$) of the common partial transversals of $\mathcal{A}$ and $\mathcal{B}$. That is,

(23.26) $P_{\text{common partial transversal}}(\mathcal{A},\mathcal{B}) = \text{conv.hull}\{\chi^T \mid T$ is a common partial transversal of $\mathcal{A}$ and $\mathcal{B}\}$.

It is easy to see that each vector x in the common partial transversal polytope satisfies:

(23.27)
(i) $0 \leq x_s \leq 1$ for $s \in S$,
(ii) $x(S \setminus A_I) \leq k - |I|$ for $I \subseteq \{1,\ldots,k\}$,
(iii) $x(S \setminus B_I) \leq l - |I|$ for $I \subseteq \{1,\ldots,l\}$.

In fact, this fully determines the common partial transversal polytope:

Theorem 23.6. *The common partial transversal polytope is determined by* (23.27).

Proof. We must show that for any weight function $w \in \mathbb{Z}_+^S$, the maximum value of $w^{\mathsf{T}}x$ over (23.27) is equal to the maximum weight μ of any common partial transversal. By Corollary 23.5a, there exist weight functions $w_1, w_2 \in \mathbb{Z}^S$ with $w = w_1 + w_2$ and $\mu = m(\mathcal{A}, w_1) + m(\mathcal{B}, w_2)$. Now any x satisfying (23.27) belongs to the partial transversal polytopes of $\mathcal{A}$ and $\mathcal{B}$. So $w_1^{\mathsf{T}}x \leq m(\mathcal{A}, w_1)$ and $w_2^{\mathsf{T}}x \leq m(\mathcal{B}, w_2)$. Hence $w^{\mathsf{T}}x \leq \mu$. ∎

Since (23.27) is the union of the systems that determine the partial transversal polytope of $\mathcal{A}$ and of $\mathcal{B}$, we have:

Corollary 23.6a. *Let $\mathcal{A}$ and $\mathcal{B}$ be families of subsets of a set S. Then*

(23.28) $P_{\text{common partial transversal}}(\mathcal{A},\mathcal{B})$
$= P_{\text{partial transversal}}(\mathcal{A}) \cap P_{\text{partial transversal}}(\mathcal{B})$.

Proof. Directly from Theorem 23.6 and Corollary 22.9a. ∎

Also:

Theorem 23.7. *System* (23.27) *is TDI.*

Proof. Directly from Corollaries 23.5a and 22.9a. ∎

Again one cannot make the obvious extension to three families of sets, by considering the families (23.25). In that case, the vector $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ belongs to the intersection of the three partial transversal polytopes, but does not belong to the common partial transversal polytope.

23.5. The common transversal polytope

Similar results hold for the common transversal polytope. Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of subsets of a set S. The *common transversal polytope* $P_{\text{common transversal}}(\mathcal{A})$ of $\mathcal{A}$ and $\mathcal{B}$ is the convex hull of the incidence vectors (in $\mathbb{R}^S$) of the common transversals of $\mathcal{A}$ and $\mathcal{B}$. That is,

(23.29) $P_{\text{common transversal}}(\mathcal{A}, \mathcal{B}) = \text{conv.hull}\{\chi^T \mid T$ is a common transversal of $\mathcal{A}$ and $\mathcal{B}\}$.

It is easy to see that each vector x in the common transversal polytope satisfies:

(23.30)
(i) $0 \leq x_s \leq 1$ for $s \in S$,
(ii) $x(A_I) \geq |I|$ for $I \subseteq \{1, \ldots, n\}$,
(iii) $x(B_I) \geq |I|$ for $I \subseteq \{1, \ldots, n\}$,
(iv) $x(S) = n$.

Corollary 23.7a. *The common transversal polytope is determined by* (23.30).

Proof. The common transversal polytope is the facet of the common partial transversal polytope determined by the equality $x(S) = n$. So we must show that (23.30) implies (23.27), which is trivial, since if x satisfies (23.30), then $x(S \setminus A_I) = x(S) - x(A_I) \leq n - |I|$ and $x(S \setminus B_I) = x(S) - x(B_I) \leq n - |I|$ for any $I \subseteq \{1, \ldots, n\}$. ∎

Again this implies:

Corollary 23.7b. *Let $\mathcal{A}$ and $\mathcal{B}$ be families of subsets of a set S. Then*

(23.31) $P_{\text{common transversal}}(\mathcal{A}, \mathcal{B}) = P_{\text{transversal}}(\mathcal{A}) \cap P_{\text{transversal}}(\mathcal{B})$.

Proof. Directly from Corollaries 23.7a and 22.9b. ∎

In fact:

Theorem 23.8. *System* (23.30) *is TDI.*

Proof. This follows from Theorem 23.7, using Theorem 5.25. ∎

Weinberger [1976] proved the following conjecture of Fulkerson [1971a], which generalizes Theorem 18.8. Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of subsets of a set S. Then the up hull $P^{\uparrow}_{\text{common transversal}}(\mathcal{A}, \mathcal{B})$ of the common transversal polytope is determined by:

(23.32) $x(U) \geq n-$ maximum size of a common partial transversal contained in $S \setminus U$, for $U \subseteq S$. This will follow from Theorem 46.3 on polymatroids.

23.6. Packing and covering of common transversals

Fulkerson [1971b] and de Sousa [1971] detected that results on bipartite edge-colouring (or related results) imply characterizations of packings of common transversals. It was noticed by Brualdi [1971b] that the methods in fact yield more general results.

Basic is the following exchange property given by de Sousa [1971]:

Theorem 23.9. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_m)$ be families of subsets of a set S and let $k \in \mathbb{Z}_+$. Suppose that S can be covered by k partial transversals of $\mathcal{A}$ and that S can also be covered by k partial transversals of $\mathcal{B}$. Then S can be covered by k common partial transversals of $\mathcal{A}$ and $\mathcal{B}$.*

Proof. Let $T_1, \ldots, T_k$ be a partition of S into k partial transversals of $\mathcal{A}$. Since each T_i is a partial transversal of $\mathcal{A}$, it follows that each A_i has a subset A'_i such that $|A'_i| \leq k$ and such that $A'_1, \ldots, A'_n$ partition S. We can assume that $A'_i = A_i$ for each i, and hence that $\mathcal{A}$ is a partition of S into classes of size at most k.

Similarly, we can assume that $\mathcal{B}$ is a partition of S into classes of size at most k.

Now make a bipartite graph G, with colour classes $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_m\}$, connecting a_i and b_j by $|A_i \cap B_j|$ parallel edges. So G has maximum degree k, and hence, by Kőnig's edge-colouring theorem, the edges of G can be coloured with k colours. It implies that S can be partitioned as required. ∎

A consequence is a min-max formula for the minimum number of common partial transversals needed to cover S, stated by Brualdi [1971b]:

Corollary 23.9a. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of subsets of a set S, each with union S. Then the minimum number of common partial transversals of $\mathcal{A}$ and $\mathcal{B}$ needed to cover S is equal to*

$$\left\lceil \max_{\substack{X \subseteq S \\ X \neq \emptyset}} \max\left\{\frac{|X|}{|\{i \mid A_i \cap X \neq \emptyset\}|}, \frac{|X|}{|\{i \mid B_i \cap X \neq \emptyset\}|}\right\}\right\rceil. \tag{23.33}$$

Proof. From Theorem 23.9, using Theorem 22.12. ∎

Theorem 23.9 also gives a variant of the exchange property (de Sousa [1971]):

Corollary 23.9b. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of subsets of a set S and let $k \in \mathbb{Z}_+$. Suppose that S can be partitioned into k*

transversals of $\mathcal{A}$, and also can be partitioned into k transversals of $\mathcal{B}$. Then S can be partitioned into k common transversals of $\mathcal{A}$ and $\mathcal{B}$.

Proof. Directly from Theorem 23.9, since $|S| = nk$. ∎

This implies another variant (de Sousa [1971]):

Corollary 23.9c. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of subsets of a set S and let $k \in \mathbb{Z}_+$. Suppose that S has a partition $(S_1, \ldots, S_n)$ with $|S_i| = k$ and $S_i \subseteq A_i$ for $i = 1, \ldots, n$. Suppose moreover that S has a partition $(Z_1, \ldots, Z_n)$ with $|Z_i| = k$ and $Z_i \subseteq B_i$ for $i = 1, \ldots, n$. Then S can be partitioned into common transversals of $\mathcal{A}$ and $\mathcal{B}$.*

Proof. Note that if S has a partition $(S_1, \ldots, S_n)$ with $|S_i| = k$ and $S_i \subseteq A_i$ for $i = 1, \ldots, n$, then S can be partitioned into k transversals of $\mathcal{A}$. Similarly for $\mathcal{B}$. So the present corollary follows from Corollary 23.9b. ∎

This gives the following basic min-max relation for the maximum number of disjoint common transversals, given by Fulkerson [1971b] and de Sousa [1971]:

Corollary 23.9d. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of sets and let k be a natural number. Then $\mathcal{A}$ and $\mathcal{B}$ have k disjoint common transversals if and only if*

(23.34) $$|A_I \cap B_J| \geq k(|I| + |J| - n).$$

for all $I, J \subseteq \{1, \ldots, n\}$.

Proof. Necessity of (23.34) being easy, we show sufficiency.

Let $\mathcal{A}'$ arise by taking k copies of $\mathcal{A}$ and let $\mathcal{B}'$ arise from taking k copies of $\mathcal{B}$. Condition (23.34) implies that $\mathcal{A}'$ and $\mathcal{B}'$ have a common transversal, S say (by Theorem 23.1). Then we can partition S into subsets $A'_1, \ldots, A'_n$, with $A'_i \subseteq A_i$ and $|A'_i| = k$. Similarly, we can partition S into subsets $B'_1, \ldots, B'_n$, with $B'_i \subseteq B_i$ and $|B'_i| = k$. Then by Corollary 23.9c, S has a partition into k common transversals of $\mathcal{A}$ and $\mathcal{B}$. ∎

(Note that if $\mathcal{A}$ and $\mathcal{B}$ are partitions of a set, this corollary reduces to Corollary 20.9a.)

The following open problem, dealing with packing common transversals, was mentioned by Fulkerson [1971b]: Let $\mathcal{A}$ and $\mathcal{B}$ be families of subsets of a set S and let $c \in \mathbb{Z}_+^S$. What is the maximum number k of common transversals $T_1, \ldots, T_k$ such that

(23.35) $$\chi^{T_1} + \cdots + \chi^{T_k} \leq c?$$

More generally than Corollary 23.9d, one has for disjoint common *partial* transversals of prescribed size (Fulkerson [1971b]):

Theorem 23.10. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_m)$ be families of sets and let $k, p \in \mathbb{Z}_+$. Then there exist k disjoint common partial transversals of size p if and only if*

$$(23.36) \qquad |A_I \cap B_J| \geq k(|I| + |J| + p - n - m)$$

for all $I \subseteq \{1, \ldots, n\}$ and $J \subseteq \{1, \ldots, m\}$.

Proof. Construct $\mathcal{A}'$ and $\mathcal{B}'$ as in Corollary 23.9d. By Corollary 23.1a, (23.36) implies that $\mathcal{A}'$ and $\mathcal{B}'$ have a common partial transversal, T say, of size pk. Then each A_i has a subset A'_i such that $|A'_i| \leq k$ and such that $A'_1, \ldots, A'_n$ partition T. We can assume that $A'_i = A_i$ for each i, and hence that $\mathcal{A}$ is a partition of T into classes of size at most k.

Similarly, we can assume that $\mathcal{B}$ is a partition of T into classes of size at most k.

Now make a bipartite graph G, with colour classes $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_m\}$, connecting a_i and b_j by $|A_i \cap B_j|$ parallel edges. So G has kp edges and maximum degree k, and hence, by Theorem 20.8, the edges of G can be coloured with k colours, each of size p. It implies that T can be partitioned into common partial transversals of $\mathcal{A}$ and $\mathcal{B}$ of size p. ∎

Similarly to Theorem 23.9 one can prove the following exchange property:

Theorem 23.11. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of subsets of a set S and let $k \in \mathbb{Z}_+$. Suppose that $\mathcal{A}$ has k disjoint transversals and that also $\mathcal{B}$ has k disjoint transversals. Then S has k disjoint subsets $S_1, \ldots, S_k$ such that each S_i contains a transversal of $\mathcal{A}$ and contains a transversal of $\mathcal{B}$.*

Proof. As $\mathcal{A}$ has k disjoint transversals, there exist disjoint sets $A'_1, \ldots, A'_n$ with $A'_i \subseteq A_i$ and $|A'_i| = k$ for $i = 1, \ldots, k$. For our purposes, we can assume that $A'_i = A_i$. Let Y be the union of the A_i. Similarly, we can assume that $B_1, \ldots, B_n$ have size k each and partition some set Z.

Again, make a bipartite graph G, with colour classes $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$, connecting a_i and b_j by $|A_i \cap B_j|$ parallel edges. Then G has maximum degree at most k, and hence, by Kőnig's edge-colouring theorem (Theorem 20.1), G is k-edge-colourable. It gives a partition of $Y \cap Z$ into k classes each intersecting any A_i and B_i in at most one element. We can extend this partition to a partition of $Y \cup Z$ into classes each intersecting any A_i and any B_i in exactly one element. This is a partition as required. ∎

This implies another min-max relation:

Corollary 23.11a. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of subsets of a set S. Then the maximum number k for which there exist disjoint subsets $S_1, \ldots, S_k$ each containing a transversal of $\mathcal{A}$ and a transversal of $\mathcal{B}$ is equal to*

$$\lfloor \min_{\emptyset \neq I \subseteq \{1,\ldots,n\}} \min\{\tfrac{|A_I|}{|I|}, \tfrac{|B_I|}{|I|}\} \rfloor. \tag{23.37}$$

Proof. Directly from Theorems 22.10 and 23.11. ∎

An analogue of Corollary 23.9d for covering by common transversals is:

Theorem 23.12. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of subsets of S and let $X \subseteq S$. Then X can be covered by k common transversals if and only if*

$$k|A_I \cap B_J| \geq k(|I| + |J| - n) + |X \setminus (A_I \cup B_J)| \tag{23.38}$$

for all $I, J \subseteq \{1, \ldots, n\}$.

Proof. To see necessity, let $T_1, \ldots, T_k$ be common transversals covering X and let $I, J \subseteq \{1, \ldots, n\}$. Then

$$\begin{aligned} k|A_I \cap B_J| &\geq \sum_{j=1}^{k} |A_I \cap B_J \cap T_j| \\ &= \sum_{j=1}^{k} (|A_I \cap T_j| + |B_J \cap T_j| - |T_j \cap (A_I \cup B_J)|) \\ &\geq \sum_{j=1}^{k} (|I| + |J| - |T_j \cap (A_I \cup B_J)|) \\ &= k(|I| + |J| - n) + \sum_{j=1}^{k} |T_j \setminus (A_I \cup B_J)| \\ &\geq k(|I| + |J| - n) + |X \setminus (A_I \cup B_J)|. \end{aligned} \tag{23.39}$$

To see sufficiency, make a directed graph D, with vertex set

$$\{r\} \cup \{a_1, \ldots, a_n\} \cup S \cup S' \cup \{b_1, \ldots, b_n\}, \tag{23.40}$$

where S' is a set consisting of, for each $s \in S$, a (new) copy s' of s, and with arcs, with demands and capacities, as follows:

(23.41)
(r, a_i) with demand k and capacity k, for $i = 1, \ldots, n$,
(a_i, s) with demand 0 and capacity ∞ for $i = 1, \ldots, n$ and $s \in A_i$,
(s, s') with demand 1 (if $s \in X$) or 0 (if $s \notin X$) and capacity k, for $s \in S$,
(s', b_i) with demand 0 and capacity ∞, for $i = 1, \ldots, n$ and $s \in B_i$,
(b_i, r) with demand k and capacity k, for $i = 1, \ldots, n$.

Then by Hoffman's circulation theorem (Theorem 11.2), (23.38) implies the existence of a circulation f obeying the demands and capacities. Indeed, consider any set U of vertices of D. Let $I := \{i \mid a_i \in U\}$, $J := \{j \mid b_j \notin U\}$, $Y := U \cap S$ and $Z := \{s \in S \mid s' \notin U\}$. We can assume that the capacity of the arcs leaving U is finite, and hence, if $i \in I$, then $A_i \subseteq Y$ and if $j \in J$, then $B_j \subseteq Z$. That is, $A_I \subseteq Y$ and $B_J \subseteq Z$.

If $r \notin U$, then the total demand of the arcs entering U is equal to

$$k|I| + |X \setminus (Y \cup Z)| \tag{23.42}$$

and the total capacity of the arcs leaving U is equal to

$$k|Y \cap Z| + k(n - |J|). \tag{23.43}$$

Since $A_I \subseteq Y$ and $B_J \subseteq Z$, (23.38) implies that (23.42) is at most (23.43).

If $r \in U$, then the total demand of the arcs entering U is equal to

$$k|J| + |X \setminus (Y \cup Z)| \tag{23.44}$$

and the total capacity of the arcs leaving U is equal to

$$k|Y \cap Z| + k(n - |I|). \tag{23.45}$$

Since $A_I \subseteq Y$ and $B_J \subseteq Z$, (23.38) implies that (23.44) is at most (23.45).

So Hoffman's condition is satisfied, and hence there exists a circulation f.

Now f is at most k on any arc. Hence, by Corollary 11.2b, f is the sum of k $\{0,1\}$-valued circulations $f_1, \ldots, f_k$. For each circulation f_i, the set T_i of $s \in S$ with $f_i(s, s') = 1$ is a common transversal of $\mathcal{A}$ and $\mathcal{B}$. Moreover, since $f(s, s') \geq 1$ for each $s \in X$, these common transversals cover X. ∎

A covering theorem different from Theorem 23.12 is due to Brualdi [1971b]:

Theorem 23.13. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_m)$ be families of subsets of a set S. Suppose that S can be covered by k common partial transversals of $\mathcal{A}$ and $\mathcal{B}$. Then S can be covered by k common partial transversals each of size $\lfloor |S|/k \rfloor$ or $\lceil |S|/k \rceil$.*

Proof. The assumption implies that each A_i contains a subset A'_i with $|A'_i| \leq k$, such that the A'_i partition S. For our purposes, we can assume that $A'_i = A_i$ for each i. Similarly, we can assume that $\mathcal{B}$ is a partition of S into classes of size at most k.

Again, make a bipartite graph G, with colour classes $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_m\}$, connecting a_i and b_j by $|A_i \cap B_j|$ parallel edges. Then G has maximum degree at most k, and hence, by Theorem 20.8, G is k-edge-colourable, where each colour has size $\lfloor |S|/k \rfloor$ or $\lceil |S|/k \rceil$. This yields a partition of S into k common partial transversals as required. ∎

23.7. Further results and notes

23.7a. Capacitated common transversals

Recall the definition of system of restricted representatives: Let $\mathcal{A} = (A_1, \ldots, A_n)$ be a collection of subsets of a set S and let $a, b \in \mathbb{Z}_+^S$ with $a \leq b$. A *system of restricted representatives* (or *SRR*) of $\mathcal{A}$ (with respect to a and b) is a sequence $(s_1, \ldots, s_n)$ such that

(23.46) (i) $s_i \in A_i$ for $i = 1, \ldots, n$;
(ii) $a(s) \leq |\{i \mid s_i = s\}| \leq b(s)$ for $s \in S$.

Ford and Fulkerson [1958c] derived the following characterization of the existence of a common system of restricted representatives from the max-flow min-cut theorem (we give the derivation from Corollary 23.1b based on splitting elements, due to Mirsky and Perfect [1968]):

Theorem 23.14. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ be families of subsets of a set S and let $a, b \in \mathbb{Z}_+^S$ with $a \leq b$. Then $\mathcal{A}$ and $\mathcal{B}$ have a common system of restricted representatives if and only if*

(23.47) $$b(A_I \cap B_J) \geq |I| + |J| - n + a(S \setminus (A_I \cup B_J))$$

for all $I, J \subseteq \{1, \ldots, n\}$.

Proof. Let for any $s \in S$, Z_s be a set of $b(s)$ (new) elements. Replace in each A_i and B_j, any occurrence of any $s \in S$ by the elements of Z_s. Choose from each Z_s, $a(s)$ elements, forming the set X. Then $\mathcal{A}$ and $\mathcal{B}$ have a common system of restricted representatives if and only if the new families have a common transversal containing X. Trivially, condition (23.47) is equivalent to condition (23.8) for the new families, and hence the theorem follows from Corollary 23.1b. ∎

(More can be found in Mirsky [1968b].)

23.7b. Exchange properties

Mirsky [1968a] showed the following exchange property of common transversals:

Theorem 23.15. *Let $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_m)$ be families of sets. Let $I', I'' \subseteq \{1, \ldots, n\}$ and $J', J'' \subseteq \{1, \ldots, m\}$. Suppose that $(A_i \mid i \in I')$ and $(B_j \mid j \in J')$ have a common transversal, and also that $(A_i \mid i \in I'')$ and $(B_j \mid j \in J'')$ have a common transversal. Then there exist I and J with $I' \subseteq I \subseteq I' \cup I''$ and $J'' \subseteq J \subseteq J' \cup J''$ such that $(A_i \mid i \in I)$ and $(B_j \mid j \in J)$ have a common transversal.*

Proof. Directly from Corollary 9.12a applied to the digraph defined in (23.1). ∎

This implies (Mirsky [1968a]):

Corollary 23.15a. *Let $\mathcal{A}$ and $\mathcal{B}$ be families of sets and let $\mathcal{A}'$ and $\mathcal{B}'$ be subfamilies of $\mathcal{A}$ and $\mathcal{B}$ respectively. Then there exist subfamilies $\mathcal{A}_0$ and $\mathcal{B}_0$ of $\mathcal{A}$ and $\mathcal{B}$*

respectively satisfying $\mathcal{A}' \subseteq \mathcal{A}_0$ and $\mathcal{B}' \subseteq \mathcal{B}_0$ and having a common transversal if and only if (i) *$\mathcal{A}'$ and some subfamily of $\mathcal{B}$ have a common transversal and* (ii) *$\mathcal{B}'$ and some subfamily of $\mathcal{A}$ have a common transversal.*

Proof. Directly from Theorem 23.15. ∎

23.7c. Common transversals of three families

It is NP-complete to test if *three* families of sets have a common transversal, even if each of the three families is a partition of S (E.L. Lawler — cf. Karp [1972b]).

Theorem 23.16. *Testing if three partitions have a common transversal is* NP-*complete.*

Proof. I. It suffices to show the NP-completeness of the following problem:

(23.48) given disjoint sets X, Y, Z with $|X| = |Y| \geq |Z|$ and a collection $\mathcal{C}$ of subsets U of $W := X \cup Y \cup Z$ with $|U \cap X| = |U \cap Y| = 1$ and $|U \cap Z| \leq 1$, decide if $\mathcal{C}$ contains a partition of W as subcollection.

To see this, first observe that we can assume that $|X| = |Y| = |Z|$. Indeed, we can extend Z by a set R of size $|X| - |Z|$ and replace each doubleton $\{x, y\}$ in $\mathcal{C}$ by all sets $\{x, y, w\}$ with $w \in R$. Then the new collection contains a partition if and only if the original collection contains one.

So we can assume that $|X| = |Y| = |Z|$. For $w \in W$, define $\mathcal{C}_w := \{C \in \mathcal{C} \mid w \in C\}$. Then the collection $\{\mathcal{C}_w \mid w \in W\}$ is the union of three partitions of $\mathcal{C}$. Moreover, these three partitions have a common transversal if and only if $\mathcal{C}$ contains a partition of W. So this reduces problem (23.48) to the problem of finding a common transversal of three partitions of a set.

II. So it suffices to show the NP-completeness of (23.48). We derive this from the NP-completeness of the (more general) *partition problem*: decide if a given collection $\mathcal{B}$ of subsets of a set Z contains a partition of Z as a subcollection (Corollary 4.1b).

Let $V := \{(B, z) \mid z \in B \in \mathcal{B}\}$. Make, for each $B \in \mathcal{B}$, an (arbitrary) directed circuit on $\{(B, z) \mid z \in B\}$. This makes the directed graph D on V (consisting of vertex-disjoint directed circuits). Define $X := V \times \{1\}$ and $Y := V \times \{2\}$. Let $\mathcal{C}$ be the collection of

(23.49) all triples $\{(B, z, 1), (B, z, 2), z\}$ for all $B \in \mathcal{B}$ and $z \in B$, and all pairs $\{(B, z, 1), (B, z', 2)\}$, for all $B \in \mathcal{B}$ and $z, z' \in B$ such that D contains an arc from z to z'.

So each element of $X \cup Y$ is in precisely two sets in $\mathcal{C}$: a triple and a pair. Any partition $\mathcal{P} \subseteq \mathcal{C}$ of $X \cup Y \cup Z$ contains, for any $B \in \mathcal{B}$, either all triples containing B or all pairs containing B. (Here containing B means: containing (B, z, i) for some z, i.)

This implies that $\mathcal{C}$ contains a partition of $X \cup Y \cup Z$ if and only if $\mathcal{B}$ contains a partition of Z. ∎

As indicated in this proof, the problem of finding a common transversal of three partitions is equivalent to the 3-*dimensional matching problem*: given a partition

U, V, W of a finite set S and a collection $\mathcal{C}$ of subsets X of S satisfying $|X \cap U| = |X \cap V| = |X \cap W| = 1$, does $\mathcal{C}$ have a subcollection that partitions S?

The following *necessary* condition for the existence of a common transversal of three families $\mathcal{A} = (A_1, \ldots, A_n)$, $\mathcal{B} = (B_1, \ldots, B_n)$, and $\mathcal{C} = (C_1, \ldots, C_n)$ of sets is not sufficient: for all $I, J, K \subseteq \{1, \ldots, n\}$

$$(23.50) \qquad |A_I \cap B_J \cap C_K| \geq |I| + |J| + |K| - 2n.$$

(This would generalize condition (23.3).) To see this, consider $\mathcal{A} = (\{a\}, \{b, c\})$, $\mathcal{B} = (\{b\}, \{a, c\})$, $\mathcal{C} = (\{c\}, \{a, b\})$.

More on common transversals of more than two families is given by Brown [1976,1984], Dacić [1977,1979], Longyear [1977], and Zaverdinos [1981]. Woodall [1982] studied fractional transversals, and described a good characterization for the existence of a common fractional transversal for more than two families, based on linear programming.

23.7d. Further notes

Weinberger [1974b] observed that if the families $\mathcal{A} = (A_1, \ldots, A_n)$ and $\mathcal{B} = (B_1, \ldots, B_n)$ of subsets of a set S are uniform (that is, all sets have the same size) and regular (that is, each $s \in S$ is in the same number of sets), then $\mathcal{A}$ and $\mathcal{B}$ have a common transversal.

Further work on common transversals (including extensions to the infinite case) is reported by Perfect [1969b], Brualdi [1970b,1971a], and Davies and McDiarmid [1976].