

MATROIDS AND THE GREEDY ALGORITHM *

Jack EDMONDS

National Bureau of Standards, Washington, D.C., U.S.A.

and

University of Waterloo, Ontario, Canada

Linear-algebra rank is the solution to an especially tractable optimization problem. This tractability is viewed abstractly, and extended to certain more general optimization problems which are linear programs relative to certain derived polyhedra.

(0) Many discrete programming algorithms are being proposed. One thing most of them have in common is that they do not work very well. Solving problems that are a priori finite, astronomically finite, is preferably a matter of finding algorithms that are somehow better than finite. These considerations prompt looking for good algorithms and trying to understand how and why at least a few combinatorial problems have them.

(1) Let H be a finite (for convenience) set of real-valued vectors, $x = [x_j], j \in E$. Often all the members of H will be integer-valued; often they will all be $\{0, 1\}$ -valued. The index-set, E , is any finite set of elements. Let $c = [c_j], j \in E$, be any real vector on E , called the *objective* or *weighting* of E . The problem of finding a member of H which maximizes (or minimizes) $cx = \sum_j c_j x_j, j \in E$, we call a *loco* problem or *loco* programming. “Loco” stands for “linear-objective combinatorial”.

(2) In order for a loco problem to be a completely defined problem, the way H is given must of course be specified. There are various ways to describe implicitly very large sets H so that it is relatively easy to determine whether or not any particular vector is a member of H . One well-known way is in linear programming, where H is the set of extreme points of the solution-set of a given finite system, L , of linear equations and \leq type linear inequalities in the variables x_j (briefly, a

* This paper was presented at the Princeton Symposium on Mathematical Programming, 1967, and intended for its Proceedings. With permission of the author it appears here instead.

given finite “linear system, L , in x ”) for which cx is bounded. Another well-known way is in integer linear programming, where H is the set of integer-valued solutions of a given finite linear system having a bounded solution-set. (Strictly speaking, for example, in order to be able to call a general integer l.p. problem a loco problem, we should allow H to be merely “discrete” rather than finite. However, for present purposes this is not important.)

(3) A purpose of this paper is to treat a certain loco problem which is not directly an l.p., and which probably cannot be reduced to an integer l.p. except by introducing some tremendously large linear system. On the other hand, a main point of the paper is that we can, by introducing a tremendously large linear system, make the loco problem into an l.p., and it *is* worth doing.

(4) For the present type of loco problem, called *matroidal*, or an m.l. problem, there is an algorithm, called the *greedy algorithm*, which is particularly simple and efficient. (If you have ever taught the optimum assignment problem, probably some students have proposed it to you for that.) Linear programming concepts, though evidently not the simplex method, provide a useful way of viewing it.

(5) Any matrix A , whose column-set we denote by E , determines the H of an m.l. problem, though quite differently than the way the constraint matrices of l.p.’s or integer l.p.’s determine their H ’s. In fact, this matrix A can be over any field.

(6) A vector $x = [x_j], j \in E$, of zeroes and ones is called *the (incidence) vector* of the subset of j ’s such that $x_j = 1$. For any family K of subsets of E , and for any weighting $c = [c_j]$ of E , to find a $B \in K$ such that its weight $c(B) = \sum_{j \in B} c_j$, $j \in B$, is maximum is clearly a loco problem — a “{0, 1}-loco problem” — where the members of H are the incidence vectors of the members of K .

An m.l. problem having “constraint matrix”, A , is a {0, 1}-loco problem where K is the family of subsets of the column-set E of A which are column-bases of A (or, to take a slight variant, where K is the family of linearly independent subsets of E .) For any given A and any weighting $c = [c_j], j \in E$, the m.l. problem is then to find a maximum (or minimum) weight column-basis of A (or a maximum weight linearly independent subset of E .)

(7) The *greedy algorithm* for the {0, 1}-loco problem, maximize cx over the incidence vectors of the members of K , is: In each step, choose any largest weight member of E , not already chosen, which together with the members already chosen forms a subset of some member of K . Stop when the chosen members of E comprise a member of K .

(8) Of course, the algorithm presumes the use of a subroutine which will decide for any given $J \subseteq E$, whether or not J is contained in a member of K .

(9) Methods are well-known [5, 6] for finding an optimum edge-weight-sum spanning tree in an edge-weighted (connected) graph, G , which are the greedy algorithm together with elaborations for ensuring that at each step the set of edges so-far-chosen is the edge-set of some forest in G (a subset of the edges of some spanning tree of G) – or, alternatively, is an edge-set contained in the complement of some spanning tree.

A graph G may be regarded as a matrix A of zeroes and ones, mod 2, which has exactly two ones in every column. The columns of A are the edges of the graph and the rows of A are the nodes of the graph. An edge and a node “meet” if there is a one located in that row and that column. Assuming the graph is “connected”, i.e., the rows of A cannot be partitioned into two non-empty sets such that every column has both of its ones in the same set, the column-bases of A are precisely the edge-sets of the “spanning-trees” of G and the linearly independent sets of columns are the edge-sets of “forests” in G . Thus, the optimum spanning tree problem is an m.l. problem relative to a matrix, A , of the type just described.

(10) A *basis* of any subset $S \subseteq E$ of the columns of matrix A may be defined as a maximal linearly independent subset J of S . *Maximal* here means that there is no linearly independent subset of S which properly contains J .

(11) The fact that the greedy algorithm will always yield a maximum weight basis of the set E of columns of a matrix A for any $\{0, 1\}$ -valued weighting $c = [c_j]$, $j \in E$, is one of the best known theorems of loco programming – indeed, of all mathematics. It is precisely the fact that:

(12) For any $S \subseteq E$, every basis J of S has the same cardinality, $|J|$, called the *rank* $r(S)$ of S .

(Take S to be $\{j \in E : c_j = 1\}$.)

(13) An *independence system* $M = (E, F)$ on E is defined to be a set E and a non-empty family, F , of so-called *independent* (or *M-independent*) subsets of E , such that every subset of an independent set is independent.

(14) For any independence system M on a set E , and for any $S \subseteq E$, a *basis* (or *M-basis*) of S is defined as in (10), replacing “linearly independent” by “*M*-independent”. The *M*-bases of E are also called the *bases* of M .

(15) A *clutter* K on a set E is a family of subsets of E such that no member of K is contained in another. Clearly the family of bases of any independence system on E form a clutter on E , and conversely for any clutter K on E there is a unique independence system on E whose family of bases is K .

(16) A *matroid*, $M = (E, F)$, is an independence system having property (12).

(17) In other words, a matroid is an independence system $M = (E, F)$ whose clutter, K , of bases (the M -bases of E) is such that, for any $\{0, 1\}$ -valued weighting of E , the greedy algorithm always yields a maximum weight member of K .

(18) We will see immediately that: *For any matroid M on E , and for any weighting of E , the greedy algorithm always gives a maximum weight member of the family K of bases of M .*

(19) This is equivalent to the “if” part of the statement that:

(20) *For any matroid M on E , and for any weighting, $c = [c_j]$, $j \in E$, a basis $B \in K$ of M has maximum weight, $c(B) = \sum c_i$, $i \in B$, if and only if,*

(21) *for every $j \in B$, the set $\{i \in B : c_i > c_j\}$ is an M -basis of the set $\{i \in E : c_i > c_j\}$.*

(22) To see the equivalence asserted in (19), observe that a basis $B \in K$ has property (21) and is arranged in an order such that the c_i 's, $i \in B$, are non-increasing if and only if it is the output of an application of the greedy algorithm, arranged in the order in which its members were chosen.

(23) For any matroid M on the set E and for any weighting, $c = [c_j]$, $j \in E$, let B_0, B_1 , and B_2 be bases of M , each arranged in any order such that the weights of its members are non-increasing. By (12), we have that $|B_0| = |B_1| = |B_2|$.

(24) Suppose that, for some k , the k th member, call it j , of B_1 has smaller weight than the k th member of B_2 . Then by (12), $J = \{i \in B_1 : c_i > c_j\}$ is not a basis of $S = \{i \in E : c_i > c_j\}$, since $|J| < k$ and since the first k members of B_2 are contained in a basis of S . Thus, B_1 does not satisfy (21).

(25) If $c(B_1) < c(B_2)$ then certainly the hypothesis of (24) is satisfied. Thus the “if” part of (20)–(21) is proved.

(26) M does have a basis, say B_0 , which satisfies (21) since, as noted in (22), the greedy algorithm produces one. By (24), for any basis, say B_1 , of M , there is no k such that the k th member of B_0 has smaller weight than the k th member of B_1 .

(27) Therefore if B_1 is a maximum weight basis of M then there is no k such that the k th member of B_1 has a weight different from the weight of the k th member of B_0 .

(28) Suppose B_1 is a maximum weight basis of M , but does not satisfy (21). That is, for some $j \in B_1$, $J = \{i \in B_1 : c_i > c_j\}$ is not a basis of $S = \{i \in E : c_i > c_j\}$. Then for some $i \in S - J$, the set $J \cup \{i\}$ is independent and thus contained in some basis, say B_2 , of M . Since $c_i > c_j$, there is a k such that the k th member of B_2 , has larger weight than the k th member of B_1 . Thus, by (27), the k th member of B_2 has larger weight than the k th member of B_0 , which contradicts (26). Thus the “only if” part of (20)–(21) is proved.

(29) Where $M = (E, F)$ is any independence system, we say that an element $j \in E$ *M*-depends on an independent set $J \in F$ when either $j \in J$ or else $(J \cup \{j\}) \notin F$. More generally we say that an element $j \in E$ *M*-depends on a set $S \subseteq E$ when j *M*-depends on some independent subset of S .

(30) It is easy to prove that for a basis B of M condition (21) is equivalent to the following condition:

(31) For every $j \in E$, j *M*-depends on $\{i \in B : c_i \geq c_j\}$.

(32) Thus, by (20)–(21), where M is a matroid, (30) is another n. and s. condition for a basis B to be of maximum weight.

(33) In a sense, any loco problem, maximize cx over a finite set H of vectors, can be regarded as the linear programming problem, maximize cx over the extreme points (*vertices*) of the solution-set (*polyhedron*) P of a linear system L , by taking L to be such that P is the convex hull of H .

(34) It is well-known that there exists such an L , that the vertices of its polyhedron P are all members of H , that $H \subset P$, and that any linear function cx of vectors $x \in P$ can be maximized over P (using linear programming) by a vertex x^0 of P . It follows immediately from these facts that x^0 maximizes cx over H .

(35) In fact, it can be shown that if H is all $\{0, 1\}$ -valued vectors then the vertex-set of P is precisely H . (This is rather beside the point.)

(36) An apparent difficulty of the above approach is, of course, that L will generally be astronomically large, astronomically degenerate, and not known in any practical way, even when H is rather small, and that for loco problems of interest, H itself will be astronomically large.

(37) There is a “negative principle”, which, with some justification, seems to have had wide acceptance in recent years, to the effect that

L 's, unless they happen to be small, are a futile aspect of loco problems otherwise prescribed. To some extent at least, the principle is wrong. I have shown this in solving two other loco problems in two other publications [1, 3]. The matroidal loco problem provides an especially simple context for illustrating again the same basic idea. We have already shown here that we can easily do without it for solving the m.l. problem itself. However, I hope you find the polyhedral approach to be interesting, and it does seem to be essential in solving more complicated loco problems which I will mention later.

(38) Contrary to the "negative principle", it is plausible that if H is describable in some combinatorially pleasant way, i.e., so that it is easy to recognize whether or not any given vector is a member of H , then an L which determines the convex hull of H might be describable in some combinatorially pleasant way, i.e., so that it is easy to recognize whether or not any given linear constraint is a member of L . If this is possible then the l.p. duality theorem, applied to L and any cx , will provide a useful criterion for confirming that a given $x^0 \in H$ is one which maximizes cx over H . Conversely, one might expect a good algorithm for maximizing any cx over H to reveal, by its termination criteria, a good description of an L .

(39) For any matroid M on a set E , the set V of vertices of the solution-set P of linear system [(40), (41)] is precisely the set H' of incidence vectors of independent sets of M .

(40) $x_j \geq 0$ for every $j \in E$.

(41) $\sum_j x_j \leq r(A)$, $j \in A$, for every $A \subseteq E$, where $r(A)$ is the M -rank of A .

(42) Since the bases of M are the independent sets J such that $|J| = r(E)$, it follows immediately from (39) that:

(43) The set of vertices of the solution-set of [(40), (41), (44)] is precisely the set H of incidence vectors of bases of M .

(44) $\sum_j x_j = r(E)$, $j \in E$.

(45) For any independent set J of M , the incidence vector x^0 of J satisfies (40) since it is all 0's and 1's. It satisfies (41) for any $A \subseteq E$, since $A \cap J$ is an independent subset of A and since the value of the left side of (41) is $|A \cap J|$.

(46) A vertex of the solution-set (polyhedron) of a finite linear system L may be defined as the unique solution of some linear system L' obtained from L by replacing certain \leq 's of L by $=$'s. The x^0 of (45) is the unique solution of the relations, $x_j = 0$ for $j \notin J$, and $x_j = r(\{j\})$ for $j \in J$, which are obtained from certain relations of [(40), (41)] by replacing inequality signs.

- (47) Therefore, by (45) and (46), $H' \subseteq V$.
- (48) The harder part of (39) is to show that $V \subseteq H'$.
- (49) This would follow immediately by showing that every vertex of P is integer-valued. In view of the easy ways known to prove, using the above definition of vertex, that the vertices of the polyhedron of an integer transportation problem are integer valued, the technique we use to do (48) may seem backwards. Actually, however, I think the technique is especially to the point – the point being the intimate relationship between having a good algorithm for the loco problems given by a class of H 's and having a good description of L 's for these H 's.
- (50) A *vertex* of a convex polyhedron P (i.e. solution-set of some finite linear system) can alternatively be defined as an $x^0 \in P$ such that some linear function cx is maximized over P by x^0 and *only* by x^0 .
- (51) Using the greedy algorithm (a mild variant of (7)), we will obtain, for any weighting $c = [c_j], j \in E$, the vector x^0 of an independent set J of M . We will show, using the weak l.p. duality principle, that x^0 maximizes cx over all solutions of [(40), (41)], and thus also over all members of H' . In view of (50), this will immediately imply $V \subseteq H'$. It will also immediately imply, for any matroid M and for any weighting c , that the greedy algorithm always yields an x^0 which maximizes cx over H' .

The variant of the greedy algorithm spoken of in (51) is:

- (52) Consider the set $E' = \{j \in E : c_j \geq 0\}$ in any order, $j(1), j(2), \dots, j(m)$, such that $c_{j(1)} \geq c_{j(2)} \geq \dots \geq c_{j(m)} \geq 0$. For each $k = 1, 2, \dots, m$, let $A_k = \{j(1), \dots, j(k)\}$. Let $x^0 = [x_j^0], j \in E$, be the vector such that $x_{j(1)}^0 = r(A_1)$, $x_{j(k)}^0 = r(A_k) - r(A_{k-1})$ for $k = 2, \dots, m$, and $x_j^0 = 0$ for every other $j \in E$.
- (53) It is easy to verify, by induction on k , and using the matroid-properties of M , that x^0 is the incidence vector of a set $J \subseteq E$ which is an M -basis of E' obtained by application of the greedy algorithm, (7), to E' .
- (54) The dual of the l.p., maximize cx subject to [(40), (41)], is
- (55) Minimize $ry = \sum_A r(A) \cdot y(A), A \subseteq E$, subject to
- (56) $y(A) \geq 0$ for every $A \subseteq E$; and
- (57) $\sum_A y(A) \geq c_j, j \in A$, for every $j \in E$.
- (58) The weak l.p. duality principle says that for every solution x of [(40), (41)] and for every solution $y = [y(A)], A \subseteq E$, of [(56), (57)]:
- (59) $ry - cx = \sum_{j \in A \subseteq E} x_j [\sum_A y(A) - c_j] + \sum_A y(A) [r(A) - \sum_{j \in A} x_j] \geq 0$, where
- (60) Let $y^0 = [y^0(A)], A \subseteq E$, be $y^0(A_k) = c_{j(k)} - c_{j(k+1)}$ for $k = 1, \dots, m - 1; y^0(A_m) = c_{j(m)}$; and $y^0(A) = 0$ for every other $A \subseteq E$.

(61) It is straight forward to verify that y^0 satisfies [(56), (57)], and that $ry^0 = cx^0$. Thus, by (58)–(59), y^0 minimizes ry subject to [(56), (57)], and x^0 maximizes cx subject to [(40), (41)].

(62) Since c was arbitrary and $x^0 \in H'$, it follows immediately from (61) and (50) that $V \subseteq H'$.

(63) Since $x^0 \in H' \subseteq P$, it also follows immediately from (61) that, for the r.h.l. problem, maximize cx over H' , the greedy algorithm, (52), always works.

(64) The P of (39), we call *the polyhedron* of matroid M .

(65) I have extended the present material in several directions, some to appear under the titles “Matroid Intersections” and “Submodular functions, matroids, and certain polyhedra”. One result is the following:

(66) Let $v(P)$ denote the set of vertices of a polyhedron P . Where M_1 and M_2 are any two matroids on the same set E , and where P_1 and P_2 are, respectively, their polyhedra, $v(P_1 \cap P_2) = v(P_1) \cap v(P_2)$.

(67) In other words, $v(P_1 \cap P_2)$ consists entirely of the family $H'_1 \cap H'_2$ of incidence vectors of sets $J \subseteq E$ which are both M_1 -independent and M_2 -independent. In other words, where L_1 and L_2 are linear systems which, respectively, determine P_1 and P_2 , the “2-matroid” loco problem, maximize cx over $x \in H'_1 \cap H'_2$, is the l.p. problem, maximize cx by a vertex of the solution-set of $L_1 \cup L_2$.

(68) There is a short proof of (66). Even better, there is a long proof of (66) which is at the same time a very good algorithm, considerably more complicated but not much less efficient than the greedy algorithm, for the 2-matroid loco problem. Like the greedy algorithm, its efficiency is of course modulo the efficiency of being able to recognize for the vector x of any given $J \subseteq E$ whether or not $x \in H'_1 \cap H'_2$.

(69) There is no known constructively good representation for general matroids. They are not all representable by linear independence in matrices. Though, the systems L by which we have here described matroids are very redundant, the irredundant subsystems are generally exponentially large relative to $|E|$.

(70) So-called transversal matroids are a combinatorially interesting class. In [4] I show that M is transversal matroid if and only if its independent sets are the linearly independent sets of columns in a matrix A whose entries are all zeroes and distinct algebraic indeterminates. A point made in the paper is that though there does exist a good well-known combinatorial algorithm which will recognise when a set of columns of such an A is independent, and thus which when combined

with the greedy algorithm will provide a good algorithm for an m.l. problem (which is, in fact, an optimum assignment problem, where man j is worth either nothing or c_j on each job), Gauss elimination is a finite but very bad algorithm for recognizing when a set of columns of such an A is independent. Where, for example, A is any matrix whose entries are all zeroes and not-necessarily-distinct algebraic indeterminates, there is no good algorithm known for recognizing when a set of the columns is linearly independent.

(71) An instance of the 2-matroid loco problem is presented in [3]. Let G be any connected directed graph having $n + 1$ nodes. Let E be the edge-set of G . Let a set $J \subseteq E$ be M_1 -independent if it is the edge-set of a forest in G . Let $J \subseteq E$ be M_2 -independent if $|J| \leq n$ and, for each node v , at most one member of J is directed toward v . M_1 and M_2 are matroids. A set is a basis of both M_1 and M_2 if and only if it is the edge-set of a "directed spanning tree" of G . The resulting 2-matroid loco-problem (by adding a large constant to each edge-weight) is to find an optimum edge-weight directed spanning tree of G . This problem does not seem to be reducible to any previously solved problems. An algorithm is given for it in [3] which is considerably simpler than the general 2-matroid loco algorithm.

(72) Let $J \subseteq E$ be M_3 -independent if $|J| \leq n$ and, for each node v , at most one member of J is directed away from v . A set is a basis of M_1 , M_2 , and M_3 , if and only if it is the edge-set of an open traveling-salesman tour (Hamiltonian path) of G .

(73) If we knew how to well-solve "3-matroid" loco problems we could well-solve the traveling salesman problem. One might say the latter is "two-thirds solved".

(74) Unfortunately, usually

$$v(P_1 \cap P_2 \cap P_3) \not\subseteq v(P_1) \cap v(P_2) \cap v(P_3)$$

where P_1 , P_2 , and P_3 are the polyhedra of three matroids on E .

(75) Another useful result about the l.p. of (67) is that: *If c is integer-valued then there is an optimum solution to the dual of this l.p. which is integer-valued.*

(76) Thus, for instance, taking c to be all ones, and applying the l.p. duality theorem, and sub-additivity of each of the rank-functions, $r_1(A)$ for M_1 , and $r_2(A)$ for M_2 , we have that:

(77) *For any two matroids $M_1 = (E, F_1)$ and $M_2 = (E, F_2)$, $\max |J|, J \in F_1 \cap F_2$, equals $\min [r_1(S_1) + r_2(S_2)], S_1 \cup S_2 = E$.*

(78) This result, and a good algorithm for finding a maximum cardinality $J \in F_1 \cap F_2$, also follow from results in [2].

(79) Whitney [8] showed that if $M = (E, F)$ is a matroid with rank function $r(A)$, then where $F^* = \{J \subseteq E: r(E-J) = r(E)\}$, $M^* = (E, F^*)$ is a matroid, called the dual of M . The bases of M^* are the complements in E of the bases of M . Clearly, any good method for recognizing M -independent sets will provide a good method for recognizing M^* -independent sets.

(80) In [2] there is an algorithm which, for any two matroids $M_1 = (E, F_1)$ and $M_2^* = (E, F_2^*)$, will produce a set $J_1 \in F_1$ and a set $J_2^* \in F_2^*$ such that $|J_1 \cup J_2^*|$ is maximum. It is a good one modulo methods for recognizing M_1 -independence and M_2 -independence. Having obtained J_1 and J_2^* , extend J_2^* to a basis B_2^* of M_2^* . Clearly, $J_2^* \subseteq B_2^* \subseteq J_1 \cup J_2^*$, since otherwise $|J_1 \cup B_2^*| > |J_1 \cup J_2^*|$. Thus, $J = (J_1 \cup J_2^*) - B_2^* \subseteq J_1$ is $(M_1$ and $M_2)$ -independent, i.e., $J \in F_1 \cap F_2$. Furthermore, there is no larger $J' \in F_1 \cap F_2$, for if there were, $E-J'$ would contain a basis B' of M_2^* , and we would have $|J' \cup B'| > |J \cup B_2^*| = |J_1 \cup J_2^*|$.

(81) Another instance of the 2-matroid loco problem is the optimum assignment problem. Let $c = [c_{hk}]$ be real-valued, $[a_h]$ and $[b_k]$ be integer-valued and such that $\sum a_h = \sum b_k$. Let E be the set of ordered pairs (h, k) . Let $J \subseteq E$ be M_1 -independent when, for every h , at most a_h members of J contain h in the first component. Let $J \subseteq E$ be M_2 -independent when, for every k , at most b_k members of J contain k in the second component. Clearly, M_1 and M_2 are matroids. They are especially simple matroids and for them the linear systems we have described here are especially redundant.

References

- [1] Jack Edmonds, "Maximum matching and a polyhedron with $\{0, 1\}$ -vertices," *Journal of Research of the National Bureau of Standards* 69B (1965) 125-130.
- [2] Jack Edmonds, "Matroid partition, Math. of the Decision Sciences," *American Mathematical Society Lectures in Applied Mathematics* 11 (1968) pp. 335-345.
- [3] Jack Edmonds, "Optimum branchings," *Journal of Research of the National Bureau of Standards* 71B (1967) 233-240; reprinted with [2], pp. 346-361.
- [4] Jack Edmonds, "Systems of distinct representatives and linear algebra," *Journal of Research National Bureau of Standards* 71B (1967) 241-245.
- [5] J.B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem, *Proceedings American Mathematical Society* 7 (1956) 48-50.
- [6] P. Rosenstiehl, "L'arbre minimum d'un graphe," *Proc. Int'l. Symp. on the Theory of Graphs in Rome 1966* (Dunod/Gordon-Breach, 1967).
- [7] R. Rado, "Note on independence functions," *Proceedings London Mathematical Society* 7 (1957) 300-320.
- [8] H. Whitney, "On the abstract properties of linear dependence," *American Journal of Mathematics* 57 (1935) 509-533.