COLUMN-PARTITIONED MATRICES OVER RINGS WITHOUT INVERTIBLE TRANSVERSAL SUBMATRICES

STEPHAN FOLDES AND ERKKO LEHTONEN

ABSTRACT. Let the columns of a $p \times q$ matrix M over any ring be partitioned into n blocks, $M = [M_1, \ldots, M_n]$. If no $p \times p$ submatrix of M with columns from distinct blocks M_i is invertible, then there is an invertible $p \times p$ matrix Q and a positive integer $m \leq p$ such that $QM = [QM_1, \ldots, QM_n]$ is in reduced echelon form and in all but at most $m - 1$ blocks QM_i the last m entries of each column are either all zero or they include a non-zero non-unit.

1. COLUMN-PARTITIONED MATRICES AND TRANSVERSAL SUBMATRICES

Generalizing the concept of row-reduced form of matrices over fields, we shall say that a matrix M over any ring with identity is in *reduced echelon form* if among all matrices QM where Q is invertible it has the maximum possible number of distinct standard unit vectors appearing as columns. (A $p \times p$ matrix Q over a ring R with identity is called *invertible* if it is an invertible member of the ring of all $p \times p$ matrices over R, i.e., if there is a $p \times p$ matrix Q' over R such that $QQ' = Q'Q$ is the $p \times p$ identity matrix. The ring R is not assumed to be commutative, but even when it is commutative, the ring of $p \times p$ matrices will generally not be commutative. With the above definition of reduced echelon form, while for any matrix M there is a square matrix Q such that QM is in reduced echelon form, this Q and QM are clearly not unique. There is an essential uniqueness of the reduced echelon form for matrices over fields, well known from classical linear algebra, which does not extend to matrices over arbitrary rings. For the general theory of non-commutative rings, see, e.g., [3].)

Theorem. *Let R be any ring with identity, possibly non-commutative. Let the columns of a $p \times q$ matrix M with entries in R be partitioned into n blocks, $M = [M_1, \ldots, M_n]$. Suppose that no $p \times p$ submatrix extracted from M with columns from distinct blocks M_i is invertible. Then there is an invertible $p \times p$ matrix Q and a positive integer $m \leq p$ such that $QM = [QM_1, \ldots, QM_n]$ is in reduced echelon form and in all but at most $m - 1$ blocks QM_i the last m entries of each column are either all zero or they include a non-zero non-unit.*

ARS COMBINATORIA 97(2010), pp. 33-39

Remark 1. If $n = q$ and R is a field, then the Theorem is an obvious consequence of a rank-deficient matrix over a field having a null row in its reduced row-echelon form.

Remark 2. If $n < p$, then the Theorem trivially holds with the identity matrix as Q and $m = p$.

Remark 3. Excluding the trivial case mentioned in Remark 2, we have $m \leq p \leq n \leq q$.

In order to prove the Theorem, it will be convenient to recast it in a somewhat more general form, using the following definitions and notation for purposes of precision and simplicity in the proof.

An $A \times B$ *matrix* with entries in a ring R is any map $M : A \times B \to R$, where A, B are finite sets of positive integers. The matrix product MN of $M : A \times B \to R$ and $N : B \times C \to R$ is a map $A \times C \to R$ whose value on $(a, c) \in A \times C$ is defined by the usual convolution formula. For $A' \subseteq A$, $B' \subseteq B$, we denote by $M[A', B']$ the restriction of M (as a map) to $A' \times B'$; thus $M = M[A, B]$. If any of A' or B' is a singleton $\{a\}$, then we may omit the set braces and write a for $\{a\}$.

Whenever we refer to *elementary row operations* on an $A \times B$ matrix M, we mean left multiplication of M by an $A \times A$ matrix E of one of the following two types:

(1) a diagonal matrix all whose diagonal entries are units (scaling of rows by units),

(2) the sum of the identity matrix and a matrix with a single non-zero entry in an off-diagonal position (adding a multiple of a row to another row).

All such matrices E are invertible.

Remark 4. Rows i and j can be transposed by a composition of elementary row operations as follows: add row i multiplied by -1 to row j, add row j to row i, scale row j by -1, add row i to row j. The transposition of rows is not included here as an elementary row operation, because in the current setting, the order of rows is irrelevant.

For any set B, a *partition* is a set Π of nonempty pairwise disjoint subsets of B the union of which is B. A *partial transversal* of Π is a subset J of B intersecting every partition class $K \in \Pi$ in at most one element.

Reformulation of the Theorem. *Let R be any ring with identity, possibly non-commutative, and let M be an $A \times B$ matrix with entries in R. Consider a partition Π of B into n classes, $\Pi = \{B_1, \ldots, B_n\}$. Suppose that for every partial transversal J of Π with $|J| = |A|$, the submatrix $M[A, J]$ is not invertible. Then there is an invertible $A \times A$ matrix Q and a nonempty subset $A' \subseteq A$ such that QM is in reduced echelon form and at most $|A'| - 1$*

of the matrices $(QM)[A', B_i]$, $1 \leq i \leq n$, can have a column containing a unit entry but no non-zero non-units.

2. Proof of the reformulation

If $n < |A|$, then the statement clearly holds with $A' = A$. Therefore we can assume that $n \geq |A|$.

If there is no subset $P \subseteq B$ with $|P| = |A|$ such that $M[A, P]$ is invertible, then let $t < |A|$ be the largest positive integer such that there is some invertible $A \times A$ matrix Q and t distinct standard unit vectors that appear as columns of QM. (In case there is no such positive t, then obviously no entry of M is a unit and the claimed result holds with any singleton A'.) Clearly QM has exactly $m = |A| - t > 0$ rows without units, and the Theorem easily follows.

Suppose therefore that there are subsets $P \subseteq B$, $|P| = |A|$ (but none with P being a partial transversal of Π) such that $M[A, P]$ is invertible. Call such subsets P *admissible sets.*

Define the *spread* of an admissible set P as the set $\{i \in \{1, \dots, n\} : P \cap B_i \neq \emptyset\}$. The *weight* (with respect to P) of a block B_i is defined as $w_i = |P \cap B_i|$. The *profile* of P is the multiset of the weights w_i where i is in the spread of P. The *profile sequence* of P is the monotone increasing ordered profile of P. Denote the inverse of $M[A, P]$ by Q. For each $1 \leq i \leq n$, define the set

$$A_i = \{r \in A : (QM)[r, c] = 1,\, c \in P \cap B_i\}.$$

For an admissible set P, denote

$$\hat{A} = \bigcup_{w_i \leq 2} A_i, \qquad \hat{B} = \bigcup_{w_i \leq 1} B_i,$$

and let $D_0 = (\hat{A} \cup \{0\}) \times \hat{B} \times \hat{A}$. The elements $(s, c, t) \in D_0$ will be considered as the *arrows* of a directed graph G_0 with vertex set $A \cup \{0\}$, where the source of (s, c, t) is s and its target is t, and the element c distinguishes between parallel arrows; we say that (s, c, t) is an arrow from row s to row t through column c. For any subset $D' \subseteq D_0$, we shall mean by "the graph D'" the subgraph of G_0 which contains all vertices (i.e., $A \cup \{0\}$) but only those arrows that are in D'. Let D be the set of arrows $(s, c, t) \in D_0$ that satisfy the following conditions: $(QM)[t, c]$ is a unit; $s \neq t$; if c belongs to a block B_k of weight 1 then s is the unique member of A_k, else $s = 0$. A directed path $(\alpha_1, \dots, \alpha_l) = ((s_1, c_1, t_1), \dots, (s_l, c_l, t_l))$ in the graph D is said to be *clear* if for all $1 \leq i < l$, denoting by T_i the set of targets of the arrows α_j, $j > i$, we have $QM[T_i, c_i] = 0$.

Define pairwise disjoint subsets $D_1, D_2, \dots, D_k, \dots$ of D inductively as follows, denoting $\bigcup_{w_i=2} A_i$ by T. The members of D_1 are the arrows of the graph D with target in T. The members of D_{k+1} are the arrows

(s,c,t) of the graph D whose target is the source of an arrow in the graph D_k and for which there is a clear path in the graph $\bigcup_{i=1}^{k} D_i \cup \{(s,c,t)\}$ starting at (s,c,t) and ending with an arrow with target in T. Let D_P be the union of all D_k, $k \geq 1$, and call the graph D_P the *connection graph* of P, and denote it by G_P. We will need to distinguish two cases. If there is no directed path from 0 to a member of T in G_P, then we say that P is of the *first kind.* Otherwise we say that P is of the *second kind* and the length of the shortest directed path from 0 to a member of T in G_P is called the *connection distance* for P.

Let $\mathcal{P}$ be the set of all admissible sets of maximum spread (i.e., meeting as many blocks of Π as possible). This set is quasi-ordered by the majorization relation between profile sequences. (Recall that a monotone sequence $a_1 \leq a_2 \leq \cdots \leq a_s$ is said to *majorize* a sequence $b_1 \leq b_2 \leq \cdots \leq b_s$ when for all $1 \leq i \leq s$, $a_1 + \cdots + a_i \geq b_1 + \cdots + b_i$, with equality for $i = s$. Majorization is a partial order on the set of finite monotone increasing sequences of integers.) Let $\mathcal{P}_1$ be the set of maximal members of $\mathcal{P}$ (i.e., the members of $\mathcal{P}$ whose profile sequence is not strictly majorized by the profile sequence of another member of $\mathcal{P}$). Let P be an admissible set in $\mathcal{P}_1$ of the first kind if such exists, otherwise let P be a member of $\mathcal{P}_1$ (necessarily of the second kind) whose connection distance is as small as possible.

Lemma (Gap Condition). *There are no units in $(QM)[A_i, B_j]$ whenever $w_i \geq w_j + 2$.*

Proof. Suppose on the contrary that $(QM)[r,c]$ is a unit for some $r \in A_i$, $c \in B_j$ with $w_i \geq w_j + 2$. Then there is a $c' \in A_i \cap P$ such that $(QM)[r,c'] = 1$, and we can make column c into a standard unit vector with elementary row operations that do not affect the columns indexed by $P \setminus \{c'\}$. Thus the set $P' = P \cup \{c\} \setminus \{c'\}$ is admissible, but it either has a larger spread than P (if $w_j = 0$) or it has the same spread as P (if $w_j > 0$) but its profile sequence majorizes that of P', a contradiction. □

We now continue the proof of the Theorem. Since P is not a partial transversal of Π, there must be a block of weight at least 2, and there is of course a block of weight 0. Assume first that there is no block of weight 2. In this case, let $I = \{i : w_i > 2\}$, and the claimed result holds by choosing $A' = \bigcup_{i \in I} A_i$, because by the Gap Condition, $(QM)[A', B_i]$ does not contain a unit for any $i \notin I$, and $|I| < |A'|$.

We can thus assume that there is a block of weight 2. If P is of the first kind, let S be the set of indices $i \in \{1, \ldots, n\}$ such that $w_i = 1$ and there is no arrow (s,c,t) with $c \in B_i$ on any path in G_P terminating in T. In this case we obtain the result, if we let $I = \{1, \ldots, n\} \setminus S$ and choose $A' = \bigcup_{i \in I} A_i$. For, if $i \notin I$ and $(QM)[r,c]$ is a unit for some $r \in A'$, $c \in B_i$, then there is an $r' \in T$ such that $(QM)[r',c]$ is a non-zero non-unit.

(Such an r is necessarily in an A_k with $w_k = 1$: this follows from the Gap Condition for blocks of weight 0; and if B_i is a block of weight 1 and $r \in A_k$ with $w_k = 2$, then there would be an arrow from the single element of A_i to an element of T through c, and so $i \in I$, a contradiction.)

If P is of the second kind, it is clear that the connection distance is at least 2. In G_P, take a shortest directed path $((s_1, c_1, t_1), \ldots, (s_l, c_l, t_l))$ from 0 to a vertex in T. For the last arrow (s_l, c_l, t_l) in this path, we have $t_l \in A_k$ for some k with $w_k = 2$ and c belongs to a block B_j of weight 1. There is a $c \in B_k \cap P$ such that $(QM)[t_l, c_l]$ is a unit and $(QM)[t_l, c] = 1$, and we can do elementary row transformations and make B_k into a block of weight 1 and B_j into a block of weight 2 with respect to a new admissible set $P' = P \cup \{c_l\} \setminus \{c\}$. These row transformations do not affect the columns indexed by $\{c_1, \ldots, c_{l-1}\} \cup P \setminus \{c\}$. Therefore $((s_1, c_1, t_1), \ldots, (s_{l-1}, c_{l-1}, t_{l-1})$ is a clear path in $G_{P'}$ from 0 to t_{l-1} and t_{l-1} now belongs to the set A_j of rows corresponding to a block B_j of weight 2 with respect to P'. The set P' has the same spread, the same profile, and the same profile sequence as P, it is still of the second kind, but its connection distance is smaller than that of P, a contradiction exhausting the last possible case. This completes the proof of the Theorem. □

Remark 5. The Gap Condition in the above proof shows that the matrix QM will indeed have some rows in which some blocks are completely free of units.

3. Matrices over fields

The Theorem above applies to any ring R, whether commutative or not. In the special case that R is a field, the Theorem overlaps as we shall show below with Rado's [7] matroid-theoretical generalization of Hall's [2] theorem on systems of distinct representatives as reformulated and extended by Perfect [5, 6]. (See also Welsh [8] for an exposition of these results based on the submodularity of the rank function.) However, the Rado–Perfect results do not apply to matrices over arbitrary rings, as the columns of such matrices do not have the combinatorial properties stipulated by matroid theory's abstract generalization of linear independence.

Perfect's version of Rado's theorem, specifically as in Theorem 2 of [6], states the following, when applied to any $A \times B$ matrix M over a field, any partition Π of B into n classes, and any positive integer k:

> *There is a partial transversal P of Π of size k such that $M[A, P]$ has rank k if and only if for all $\Theta \subseteq \Pi$ the rank of $M[A, \bigcup \Theta]$ is at least $k + |\Theta| - n$ (where $\bigcup \Theta$ denotes the union of the partition blocks in Θ).*

The "only if" part is obvious here, while the "if" part states in particular that if $M[A, P]$ is non-invertible for all partial transversals P of Π of size

$|A|$ then for some $\Theta \subseteq \Pi$ the rank ρ of $M[A, \bigcup \Theta]$ is less than $|A|+|\Theta|-n = |A|-(|\Pi|-|\Theta|)$, in other words $|A|-\rho > |\Pi \setminus \Theta|$. Gaussian elimination then yields an invertible $A \times A$ matrix Q and a set $A' \subseteq A$ of size $|A|-\rho$ such that the matrix $(QM)[A', \bigcup \Theta]$ is identically null, i.e., at most $|\Pi \setminus \Theta| \leq |A'|-1$ of the matrices $(QM)[A', B']$, $B' \in \Pi$, can have a non-null entry.

Thus the Rado–Perfect result on independent partial transversals in matroids and our Theorem above overlap in the following Corollary, where the implication (2) $\Rightarrow$ (1) is obvious.

Corollary. *Let the columns of a $p \times q$ matrix M with entries in any field be partitioned into n blocks, $M = [M_1, \ldots, M_n]$. The following are equivalent.*

(1) All $p \times p$ submatrices extracted from M with columns from distinct blocks M_i are noninvertible.

(2) There is an invertible $p \times p$ matrix Q and a positive integer $m \leq p$ such that in $QM = [QM_1, \ldots, QM_n]$ the last m rows are null in all but at most $m-1$ blocks QM_i.

4. Concluding remarks

An application arises in the algebraic theory of n-ary operations on any set A, i.e., maps $A^n \rightarrow A$. Let C be a fixed set of operations on A, possibly of different arities. For operations f and g on A of arities n and m, respectively, we denote $f \leq_C g$ if and only if $f = g(h_1, \ldots, h_m)$ where $h_1, \ldots, h_m \in C$ are all n-ary. The relation $\leq_C$ is a quasi-order (a reflexive and transitive relation) on the set of all operations on A if and only if C is a clone on A (a set of operations that contains all projection maps and is closed under functional composition). The Corollary as applied to matrices with entries in the two-element field is used in [4] to establish the descending chain condition in the quasi-order $\leq_C$ in the particular case where C is the clone of projections and quasi-linear functions of Burle [1] on a finite set of $k \geq 3$ elements. An operation f on A is *quasi-linear* if it has the form $f = g(h_1(x_1) \oplus \cdots \oplus h_n(x_n))$, where $h_1, \ldots, h_n : A \rightarrow \{0,1\}$, $g : \{0,1\} \rightarrow A$ and $\oplus$ denotes addition modulo 2.

As noted before, the Theorem in its full generality relating to matrices over arbitrary rings does not seem to fit within matroid theory. However, it may be possible to develop some relaxation of matroid transversal theory or of submodular function theory, possibly in the spirit of Welsh [8, Chapter 7], which could shed additional light on why the Theorem works over arbitrary rings.

Acknowledgements

The authors thank Miguel Couceiro and Jenő Szigeti for useful discussions of the subject.

REFERENCES

[1] G. A. BURLE, The classes of k-valued logics containing all one-variable functions, *Diskretnyi Analiz* **10** (1967) 3–7 (in Russian).

[2] P. HALL, On representatives of subsets, *J. London Math. Soc.* **10** (1935) 26–30.

[3] T. Y. LAM, *A First Course in Noncommutative Rings,* 2nd edition, Springer, 2001.

[4] E. LEHTONEN, Subfunction relations defined by the clones containing all unary operations, arXiv:math.CO/0703867.

[5] H. PERFECT, Independence spaces and combinatorial problems, *Proc. London Math. Soc. (3)* **19** (1969) 17–30.

[6] H. PERFECT, A generalization of Rado's theorem on independent transversals, *Proc. Camb. Phil. Soc.* **66** (1969) 513–515.

[7] R. RADO, A theorem on independence relations, *Quart. J. Math. (Oxford)* **13** (1942) 83–89.

[8] D. J. A. WELSH, *Matroid Theory,* Academic Press, 1976.

(S. Foldes) INSTITUTE OF MATHEMATICS, TAMPERE UNIVERSITY OF TECHNOLOGY, P.O. BOX 553, FI-33101 TAMPERE, FINLAND
E-mail address: stephan.foldes@tut.fi

(E. Lehtonen) INSTITUTE OF MATHEMATICS, TAMPERE UNIVERSITY OF TECHNOLOGY, P.O. BOX 553, FI-33101 TAMPERE, FINLAND
E-mail address: erkko.lehtonen@tut.fi