Superlinear Bounds on Matrix Searching

Maria M. Klawe*

The technique of matrix searching in totally monotone matrices and their generalizations is steadily finding ever more applications in a wide variety of areas of computer science, especially computational geometry and dynamic programming problems (see [AKMSW87], [AK87], [AP88], [AS87], [AS89], [EGG88], [KK88], [W88]). Although an asymptotically optimal linear time algorithm is known for the most basic problem of finding row minima and maxima in totally monotone matrices [AKMSW87], for most of the generalizations of totally monotone matrices, only superlinear algorithms are known, though until now no superlinear lower bounds have been proved. This paper gives the first superlinear bound for matrix searching in two types of totally monotone partial matrices. We also give a matching upper bound for a subclass of one of them, though unfortunately the proof of the lower bound does not apply to this subclass. These types of matrices, which we refer to as v-matrices and h-matrices, respectively, were introduced by Aggarwal and Suri [AS89] who used them to find the farthest visible pair in a simple polygon. In addition, these matrix classes are natural extensions of staircase matrices which have applications in computational geometry and dynamic programming problems.

The precise results of this paper are as follows. We show that any algorithm for finding row maxima or minima in totally monotone partial $2n \times n$ matrices with the property that the non-blank entries in each column form a contiguous segment, can be forced to evaluate $\Omega(n\alpha(n))$ entries of the matrix in order to find the row maxima or minima, where $\alpha(n)$ denotes the very slowly growing inverse of Ackermann's function. A similar result is obtained for $n \times 2n$ matrices with contiguous non-blank segments in each row.

We also give an $O(m\alpha(n)+n)$ time algorithm to find row maxima and minima in totally monotone partial $n \times m$ matrices with the property that the non-blank entries in each column form a contiguous segment ending at the bottom row. This upper bound comes from extending the Klawe-Kleitman algorithm [KK88] for matrix searching in staircase matrices. The lower bounds are proved by introducing the concept of an independence set in a partial matrix and showing that any matrix searching algorithm for these types of partial matrices can be forced to evaluate every element in the independence set. Wiernik's $\Omega(n\alpha(n))$ lower bound on the lower envelope of n line segments in the plane ([W86]) is then used to construct an independence set of size $\Omega(n\alpha(n))$ in the matrices of size $2n \times n$ and $n \times 2n$.

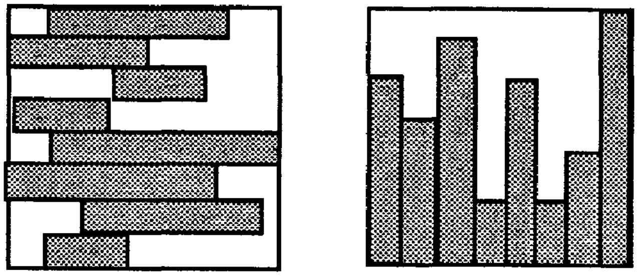
1. Introduction

A **partial matrix** is a matrix in which entries are either real numbers or are blank. A partial matrix $M = (M_{ij})$ is called **totally monotone** if for every $i < i', j < j'$ such that all entries of the 2×2 submatrix, $M_{ij}, M_{ij'}, M_{i'j}$, and $M_{i'j'}$, are non-blank, whenever $M_{ij} \leq M_{ij'}$ we have $M_{i'j} \leq M_{i'j'}$. A **totally monotone matrix** is a totally monotone partial matrix with no blank entries.

We will call a totally monotone partial matrix a **v-matrix** (vertical matrix) if the set of non-blank entries in each column forms a contiguous interval. Similarly, an **h-matrix** (horizontal matrix) is a totally monotone partial matrix such that the set of non-blank entries in each row forms a contiguous interval. Finally, a **skyline** matrix is a v-matrix such that every column's non-blank segment ends at the bottom row. A partial matrix is

* Dept. of Computer Science, UBC, Vancouver, BC V6T 1W5

a staircase matrix if it is both a v-matrix and an h-matrix (this definition is slightly more general than the one given in [AK87] and [KK88], but the algorithms of those papers can be trivially extended to handle this definition of staircase matrix). Examples of an h-matrix and skyline matrix are shown in Figure 1, where the grey areas indicate the regions containing non-blank entries.



(a) h-matrix (b) skyline matrix

Figure 1

Totally monotone matrices were introduced by Aggarwal, Klawe, Moran, Shor and Wilber in [AKMSW87], who showed that several problems in computational geometry could be reduced to finding the maximum or minimum value in each row of a totally monotone matrix. We will use the term **matrix searching** to refer to the task of finding row minima or maxima in a matrix. Aggarwal et al gave a linear time algorithm (which we will refer to as the SMAWK algorithm) for matrix searching in totally monotone matrices, yielding faster algorithms for a broad collection of problems. Wilber [W88] used the SMAWK algorithm to get a linear time algorithm for a dynamic programming problem known as the concave least weight subsequence problem. Aggarwal and Klawe [AK87] generalized totally monotone matrices to staircase matrices, and showed that additional problems of computational geometry could be reduced to matrix searching in staircase matrices. Aggarwal and Klawe [AK87] also gave an $O(m \log\log n)$ time algorithm for searching staircase matrices of size $n \times m$, again yielding faster algorithms for several problems in computational geometry. Klawe and Kleitman [KK88] gave an $O(m\alpha(n) + n)$ time algorithm for matrix searching in staircase matrices, and extended this algorithm to handle a class of dynamic programming problems satisfying convex quadrangle inequalities introduced by Eppstein, Galil and Giancarlo [EGG88]. In [AS89], Aggarwal and Suri introduced v-matrices and h-matrices, and used matrix searching in these matrices to give a faster algorithm for computing the farthest visible vertex pair in a simple polygon.

This paper has two main contributions. The first is a superlinear lower bound for matrix searching in v-matrices and h-matrices. This is the first superlinear lower bound for matrix searching in totally monotone matrices. The problem of extending this lower bound to staircase matrices remains open, and requires at least one more idea since we can show that our current techniques will not suffice. The second contribution is the extension of the Klawe-Kleitman matrix searching algorithm for totally monotone staircase matrices to skyline matrices. The question of extending the algorithm to either v-matrices or h-matrices remains open. In the next section we outline the proofs of the lower bound. Section 3 contains a sketch of the extension of the Klawe-Kleitman algorithm to skyline matrices, and the final section describes remaining open problems.

2. The Lower Bound

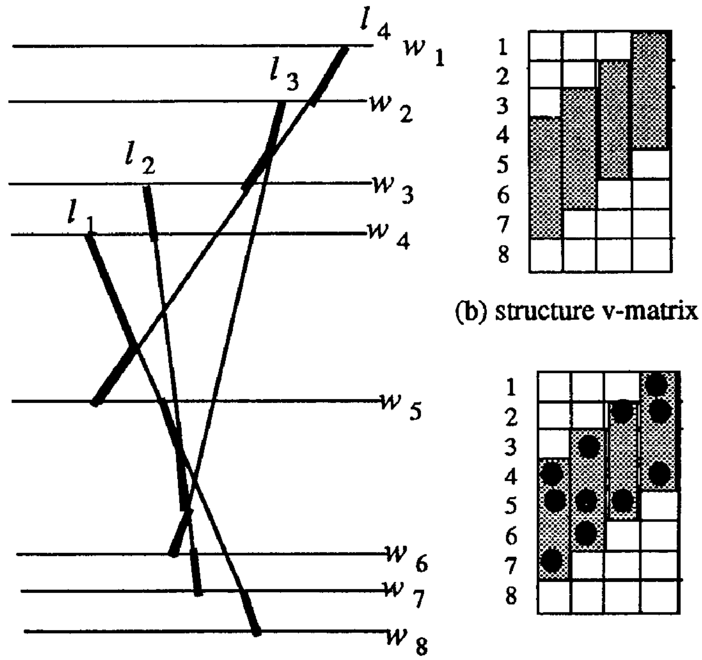
We assume that an algorithm for matrix searching in a partial matrix is given as input the pattern of non-blank entries in the matrix. For a v-matrix this simply the positions of the top and bottom non-blank entry in each column. We will refer to this pattern matrix indicating the positions of non-blanks as the **structure matrix** (or structure v-matrix or h-matrix as appropriate) of the partial matrix. The algorithm may query the value of any entry in the matrix at any time, and at the end must report the position of the maximum [minimum] value in each row. We will prove a lower bound on the number of entries that must be evaluated in the worst-case.

Our strategy to prove the lower bound is as follows. Given a fixed structure matrix, we define the concept of an independence set for that structure matrix. Next we show, using Wiernik's $\Omega(n\alpha(n))$ lower bound on the lower envelope of n line segments in the plane ([W86]), that there is a structure matrix of size $2n \times n$ possessing the column interval property which has an independence set of size $\Omega(n\alpha(n))$. Transposing this matrix gives a structure h-matrix of size $2n \times n$ with an independence set of size $\Omega(n\alpha(n))$. The final step is to exhibit an adversary which can respond to queries in such a way that that the matrix created is totally monotone, and such that any element of the independence set which has not yet been queried is still a candidate, but not a certainty, for the maximum in its row. In the remainder of this section we give the definition of independence set and show how Wiernik's result gives a structure matrix with the desired size of independence set. For the v-matrix case we construct the adversary directly from Wiernik's result, but this does not seem to work for the h-matrix case.

Let A be an $n \times m$ structure matrix. A subset $S \subset \{1, \dots, n\} \times \{1, \dots, m\}$ is said to be **independent for** A if every (i,j) in S, the entry A_{ij} is non-blank and there exists some $j' \neq j$ such that (i,j') is also in S. Moreover, for every $i < i'$ and $j < j'$ such that both (i,j') and (i',j) are in S, we have that either A_{ij} is blank or $A_{i'j'}$ is blank. For any matrix M we will call the ordered pair (i,j) the **index** of the entry M_{ij}.

Given a set of line segments $l_1, \dots, l_n$ in the plane, we define their **left envelope** to be the set of points $\{z : z \in l_i$ for some i, and z is the leftmost point in the intersection of $\cup_{j=1}^{n} l_i$ with the horizontal line through $z\}$. Figure 2(a) shows a set of line segments and their left envelope. It is easy to see that the left envelope is always the union of a finite set of line segments. Wiernik [W86] gives a construction of n line segments $l_1, \dots, l_n$ in the plane such that their left envelope has $\Omega(n\alpha(n))$ segments. For each i let (x_1^i, y_1^i) and (x_2^i, y_2^i) be the top and bottom endpoints of l_i respectively, and let L_i be the infinite line extending l_i. Suppose the line segments are ordered so that whenever $i < j$, as y goes to ∞ the line L_i is eventually to the left of L_j. We use the l_i to define a $2n \times n$ structure matrix, A as follows. Let $\{w_1, \dots, w_{2n}\} = \{y_j^i : j = 1, 2, i = 1, \dots, n\}$ arranged in decreasing order. Without loss of generality we may assume that the $\{w_i\}$ are all distinct. The i-th row of the structure matrix corresponds to w_i and the j-th column corresponds to the line segment l_j. More precisely, the top non-blank entry in the j-th column of A is in the row i such that $w_i = y_1^j$ and the bottom non-blank entry is in the row $i' - 1$ where $w_{i'} = y_2^j$. A is obviously a structure v-matrix. We now show that A has an independence set of size $\Omega(n\alpha(n))$. Figure 2(b) shows the structure v-matrix corresponding to the line segments in Figure 2(a).

We start with a set T that is almost an independent set. The only way in which it may fail is that there may be some rows in which T only has one entry. Let $T = \{(i,j) :$ there is some y_0 with $w_i \geq y_0 > w_{i+1}$ such that the line segment forming the left envelope at $y = y_0$ is $l_j\}$. It is easy to see that A must be non-blank at every (i,j) in T. Suppose $i < i'$ and $j < j'$ such that both (i,j') and (i',j) are in T, and suppose both (i,j) and (i',j') are non-blank in A. Let z be the y-coordinate of the intersection of L_j and $L_{j'}$. Because of the ordering of the line-segments and the fact that $(i,j') \in T$, it is not hard to see that we must have $z > w_{i+1}$ and hence $z > w_{i'}$. Since (i',j') is non-blank, it is impossible that $(i',j) \in T$ since $l_{j'}$ lies to the left of l_j for the entire interval between $w_{i'}$ and $w_{i'+1}$. Thus at least one of (i,j) and (i',j') must be blank. Figure 2(c) shows the set T for the line segments in Figure 2(a).



(a) the left envelope

(b) structure v-matrix

(c) the set T

Figure 2

We complete the construction of the independent set, S, by removing all points from T which are the unique point in their row. We claim that S has size $\Omega(n\alpha(n))$. Since we removed at most $2n$ points, it suffices to show that the size of T is $\Omega(n\alpha(n))$. This follows immediately from the observation that in any interval in which no line segment begins or ends, each l_j can occur in the left envelope at most once.

We now turn to the problem of constructing an adversary which will force a row-maxima finding algorithm to evaluate every entry whose index is in the independent set. We first define a v-matrix, M, whose structure matrix is A. We then prove that M is totally monotone. Next we will define a set of v-matrices M^f, such that each M^f has structure matrix A and agrees with M on all entries outside the independence set. We then prove that each M^f is totally monotone. Finally we construct an adversary for the searching algorithm such that the positions of the row-maxima cannot be known until each element whose index is in the independence set has been queried, and such that the final matrix will be M^f for some f.

Let M be the v-matrix with structure matrix A defined by M_{ij} = the maximum number of lines lying to the right of l_j at any point strictly between w_i and w_{i+1}

whenever A_{ij} is non-blank. Note that M_{ij} assumes a maximal value in row i if and only if part of l_j is in the left envelope between w_i and w_{i+1} . The next lemma proves that M is totally monotone.

Lemma 2.1. M is totally monotone.

Proof. Suppose $i < i', j < j'$ such that all entries of the 2×2 submatrix, $M_{ij}, M_{ij'}, M_{i'j}$, and $M_{i'j'}$, are non-blank, and $M_{ij} \le M_{ij'}$. We must show that $M_{i'j} \le M_{i'j'}$. Let z be the y-coordinate of the intersection of L_j and $L_{j'}$. Since $j < j'$ we know L_j lies to the left of $L_{j'}$ as y goes to ∞. Since $M_{ij} \le M_{ij'}$ we must have l_j lying to the right of $l_{j'}$ at some point strictly between w_i and w_{i+1}. Thus we must have $z > w_{i+1}$ and hence $z > w_{i'}$ since $w_{i+1} \ge w_{i'}$. This shows that l_j is to the right of $l_{j'}$ at every point between $w_{i'}$ and $w_{i'+1}$, and hence $M_{i'j} \le M_{i'j'}$. ∎

For each function f from S to the non-negative real numbers, we define M^f to be the v-matrix such that $M^f_{ij} = M_{ij} + f(i,j)$ for $(i,j) \in S$ and $M^f_{ij} = M_{ij}$ otherwise. The next lemma shows that M^f is totally monotone.

Lemma 2.2. For any function f from S to the non-negative real numbers, the v-matrix M^f is totally monotone.

Proof. Suppose $i < i', j < j'$ such that all entries of the 2×2 submatrix, $M^f_{ij}, M^f_{ij'}, M^f_{i'j}$, and $M^f_{i'j'}$, are non-blank, and $M^f_{ij} \le M^f_{ij'}$. We must show that $M^f_{i'j} \le M^f_{i'j'}$. If none of the indices are in S this follows from Lemma 2.1, so we may assume that at least one of the indices is in S. Since S is an independence set, we cannot have both (i,j') and (i',j) in S. Also $M^f_{ij} \le M^f_{ij'}$ implies that if (i,j) is in S then (i,j') is also. Moreover, if (i',j') is in S then either (i',j) is also, or $M^f_{i'j} \le M^f_{i'j'}$. Thus it suffices to consider the cases $(i,j') \in S$ and $(i',j) \in S$. Suppose we have $(i,j') \in S$. This implies that $M_{ij} \le M_{ij'}$, and hence $M_{i'j} < M_{i'j'}$ by Lemma 2.1. In addition, $M^f_{i'j} = M_{i'j}$ since $(i',j) \notin S$, and $M_{i'j'} \le M^f_{i'j'}$ since f only assumes non-negative values. Combining this gives $M^f_{i'j} \le M^f_{i'j'}$ as desired. Now suppose $(i',j) \in S$. Let z be the y-coordinate of the intersection of L_j and $L_{j'}$. Since $(i',j) \in S$ we must have $w_{i'} > z$, and hence l_j lies to left of $l_{j'}$ at every point between w_i and w_{i+1}, contradicting the assumption $M^f_{ij} \le M^f_{ij'}$, and completing the proof. ∎

We are now ready to define the behaviour of an adversary for any row-maxima finding algorithm on v-matrices with structure matrix A. When the algorithm queries the entry with index (i,j), the adversary will respond with M_{ij} for $(i,j) \notin S$ and $M_{ij} + k + 1$ for $(i,j) \in S$, where k is the number of entries with indices in S that the algorithm has queried so far. By Lemma 2.2 the matrix produced by the adversary is totally monotone. Moreover, if (i,j) is the last index in S to be queried by the algorithm, the adversary could answer M_{ij} instead of $M_{ij} + |S|$ and still produce a totally monotone matrix. Since S has at least two indices in row i, the question of whether M_{ij} is a row-maxima cannot be answered without evaluating it. This shows that the algorithm must evaluate $|S| = O(n\alpha(n))$ entries of the v-matrix in order to determine the positions of the row-maxima.

We now turn to the proof of the lower bound for the h-matrix case. Let A, T be the tranposed versions of the structure matrix and "pre-independence set" from the proof for the v-matrix case. Clearly A is a structure h-matrix. As before let S be the set obtained by deleting any element of T which is the unique element of T in its row. It is easy to check that S is an independence set for A from the definition of independence set.

We first define $S_i = \{j : (i,j) \in S\}$. Similar to the proof in the v-matrix case we will construct an h-matrix M with structure matrix A such that for each function f from S to the non-negative reals, the matrix M^f defined by $M^f_{ij} = M_{ij}$ for $(i,j) \notin S$ and $M^f_{ij} = |S_i| + f(i,j)$ is totally monotone. Given M, the adversary which forces a row-maxima finding algorithm to evaluate each entry with an index in S is completely analogous to the v-matrix case. The construction of M takes a bit more work in this case than in the v-matrix case. For each i let $A_i = \{(i,j) : A_{ij} \text{ is non-blank}\}$. We begin by defining a partial order on each A_i.

Let $j, j' \in A_i$. We define a relation $\propto_i$ on A_i by $j \propto_i j'$ if any of the following hold:

(i) $(i,j) \notin S$ and $(i,j') \in S$.

(ii) Neither (i,j) nor (i,j') are in S, $j < j'$ and for some $h < i$ we have $(h,j') \in S$ and A_{hj} non-blank.

(iii) Neither (i,j) nor (i,j') are in S, $j' < j$ and for some $i' > i$ we have $(i',j') \in S$ and $A_{i'j}$ non-blank.

Let $\prec_i$ be the transitive closure of $\propto_i$, i.e. $j \prec_i j'$ if for any $k \ge 1$ there exist $j_0, j_1, \ldots, j_k \in A_i$ with $j = j_0 \propto_i j_1 \propto_i \ldots \propto_i j_k = j'$.

Remark 2.3. Whenever $p, q \in A_i$ with $p \prec_i q$ we have $(i,p) \notin S$.

Proof. This follows immediately from the observation that whenever $p, q \in A_i$ with $p \propto_i q$ we have $(i,p) \notin S$. ∎

Lemma 2.4. Suppose A is a structure h-matrix. Then $\prec_i$ is a partial order on A_i.

Proof. Since $\prec_i$ is obviously transitive, it suffices to show that we cannot have $j \prec_i j$ for any j in A_i. Suppose the contrary. Let $k \geq 1$ and $j_0, j_1, \ldots, j_k \in A_i$ such that $j = j_0 \propto_i j_1 \propto_i \ldots \propto_i j_k = j$, and suppose that j and k are chosen so that k is minimal, i.e. whenever $j'_0 \propto_i j'_1 \propto_i \ldots \propto_i j'_{k'} = j'_0$ we have $k' \geq k$. It is easy to check from the definition of $\propto_i$ that we never have $j' \propto_i j'$ for any $j' \in A_i$. It is also not hard to see that we cannot have $j \propto_i j' \propto_i j$ for any pair j, j' in A_i. To see this, suppose without loss of generality that $j < j'$. Then in order to have $j \propto_i j' \propto_i j$, we must have some $h < i$ such that $(h, j') \in S$ and A_{hj} non-blank, and some $i' > i$ such that $(i', j) \in S$ and $Ai'j'$ non-blank. However, this contradicts the independence of S since we have $h < i', j < j'$ with both (h, j') and (i', j) in S and both Ahj and $A_{i'j'}$ non-blank. Thus we may assume that $k \geq 3$, and that $j_0 < j_s$ for $s = 1, ..., k-1$. Also, note that (i, j_s) is not in S for $0 \leq s \leq k$. This is obvious for $0 \leq s \leq k-1$ by Remark 2.3, and also for $s = k$ since $j_k = j_0$. Thus whenever $j_s < j_{s+1}$ there is some $h_s < i$ such that $(h_s, j_{s+1}) \in S$ and $A_{h_s j_s}$ is non-blank, and whenever $j_s > j_{s+1}$ there is some $i_s > i$ such that $(i_s, j_{s+1}) \in S$ and $A_{i_s j_s}$ is non-blank.

Choose r such that $|j_r - j_{r+1}|$ is maximal. Without loss of generality we assume that $j_r < j_{r+1}$ (the proof for the other case is symmetric). Let t such that $j_t > j_s$ for $s \neq t$. It is not hard to prove that for any q with $j_0 < q \leq j_t$, there is some s and some s' such that $j_s < q \leq j_{s+1}$ and $j_{s'+1} < q \leq j_{s'}$. For example, taking s to be maximal such that $j_w < q$ for each $w \leq s$, and s' to be minimal such that $j_x < q$ for each $x > s'$ will do. Thus there is some y such that $j_{y+1} < j_{r+1} \leq j_y$. Now since $|j_r - j_{r+1}|$ is maximal, we must have $j_r \leq j_{y+1}$. We have $h_r < i < i_y$ and both (h_r, j_{r+1}) and (i_y, j_{y+1}) in S and both $A_{h_r j_r}$ and $A_{i_y j_y}$ are non-blank. Moreover, since A has the row interval property and $j_r \leq j_{y+1} < j_{r+1} \leq j_y$, we must have that $A_{h_r j_{y+1}}$ and $A_{i_y j_{r+1}}$ are non-blank. Now this contradicts the independence of S, completing the proof. ∎

If for $i = 1, \ldots, n$ we have a linear order $<_i$ on each A_i, we define the **canonical partial matrix** generated by the $\{<_i\}$ to be the matrix M with M_{ij} = the position of j in the $<_i$ ordering of A_i if A_{ij} is non-blank, and blank otherwise. We will say that a set $\{\prec_i : \prec_i$ is a partial order on $A_i\}$ is **consistent** if whenever $j, j' \in A_i$ with $j < j'$ and $j \prec_i j'$, for every i' with $i < i'$ and $j, j' \in A_{i'}$ we have $j \prec_{i'} j'$.

Remark 2.5. If the linear orderings $\{<_i\}$ are consistent then the canonical partial matrix generated by the $\{<_i\}$ is totally monotone.

Proof. This follows immediately from the definition of total monotonicity. ∎

Suppose $\prec_i$ is a partial order on A_i and j, j' are incomparable elements of A_i with $j < j'$. We define the partial order $\prec_i^+ (j, j')$ on A_i to be the extension of $\prec_i$ obtained by adding the relation $j \prec_i^+ j'$, and similarly define the partial order $\prec_i^- (j, j')$ on A_i to be the extension of $\prec_i$ obtained by adding the relation $j' \prec_i^- j$. If $P = \{\prec_s : s = 1, \ldots, n\}$ we use $P_i^+(j, j')$ and $P_i^-(j, j')$ to denote the sets obtained by replacing $\prec_i$ in P by $\prec_i^+ (j, j')$ and $\prec_i^- (j, j')$ respectively.

Lemma 2.6. If $P = \{\prec_s : s = 1, \ldots, n\}$ is consistent and j, j' are incomparable elements of A_i with $j < j'$, then at least one of $P_i^+(j, j')$ and $P_i^-(j, j')$ must be consistent.

Proof. Suppose not. Since $P_i^+(j, j')$ is not consistent, there is some $i' > i$ with $j, j' \in A_{i'}$ and $j' \prec_{i'} j$. Similarly, since $P_i^-(j, j')$ is not consistent, there is some $h < i$ with $j, j' \in A_h$ and $j \prec_h j'$, but this contradicts the consistency of P. ∎

Corollary 2.7. If $P = \{\prec_s : s = 1, \ldots, n\}$ is consistent then there is a consistent set P' of linear orderings extending P.

Proof. This follows immediately from Lemma 2.6. ∎

Lemma 2.8. Suppose A is a structure h-matrix, and $j_1, \ldots, j_k \in A_i$ with $j_1 \propto_i \ldots \propto_i j_k$. Then if $(j_s - j_{s-1})(j_k - j_{k-1}) < 0$ for each $s = 2, \ldots, k-1$, then j_k cannot lie between j_{s-1} and j_s for $s = 2, \ldots, k-1$.

Proof. First suppose $j_k < j_{k-1}$. This implies that $j_{s-1} < j_s$ for each $s = 2, \ldots, k-1$. Thus it suffices to show $j_k < j_{s-1}$ for each $s = 2, \ldots, k$. The proof is by backwards induction on s. This holds for $s = k$ since we assumed $j_k < j_{k-1}$, so suppose $3 \leq s \leq k$ and $j_k < j_{s-1}$. Since $j_{s-2} \propto_i j_{s-1}$ there is some $h < i$ such that $j_{s-2}, j_{s-1} \in A_h$ and $(h, j_{s-1}) \in S$. Similarly as $j_{k-1} \propto_i j_k$ there is some $i' > i$ such that $j_{k-1}, j_k \in A_{i'}$ and $(i', j_k) \in S$. If $j_k \geq j_{s-2}$ then $j_k \in A_h$ because A has the row interval property and $j_{s-2} \leq j_k < j_{s-1}$. This contradicts the independence of S. The argument for the case $j_k > j_{k-1}$ is symmetric. ∎

Lemma 2.9. Suppose A is a structure h-matrix, and $j_1, \ldots, j_k \in A_i$ with $j_1 \propto_i \ldots \propto_i j_k$, where $k \geq 2$. Let p be minimal such that $(j_s - j_{s-1})(j_k - j_{k-1}) > 0$ for all

s with $p \le s \le k$. Then j_s lies between j_{p-1} and j_p for $s = 1, \dots, p$.

Proof. The proof is by induction on k. It is obviously true for $k = 2$ so assume $k > 2$ and that the hypothesis holds for $k-1$. Let q be minimal such that $(j_s - j_{s-1})(j_{k-1} - j_{k-2}) > 0$ for all s with $q \le s \le k-1$. Without loss of generality suppose $j_k < j_{k-1}$. If $p \le k-1$ it is easy to see that statement holds since clearly $p = q$. Thus suppose $p = k$. This implies that $(j_s - j_{s-1})(j_k - j_{k-1}) < 0$ for $s = q, \dots, k-1$. Now by Lemma 2.8, we have that $j_k < j_{s-1}$ for $s = q, \dots, k$ and the interval $[j_k, j_{k-1}]$ contains the interval $[j_{s-1}, j_s]$ for $s = q, \dots, k-1$. This completes the proof as the interval $[j_{q-1}, j_q]$ contains all the j_s for $s = 1, \dots, q$ by the induction hypothesis.∎

Corollary 2.10. Suppose A is a structure h-matrix, $j_1, \dots, j_k \in A_i$ with $j_1 \propto_i \dots \propto_i j_k$, where $k \ge 2$. Then j_k is either the maximum or minimum of $\{j_s : 1 \le s \le k\}$.

Proof. Let p be as in Lemma 2.9. Without loss of generality suppose $j_p < j_{p+1} < \dots < j_{k-1} < j_k$. If $p = 1$ we are done so assume $p > 1$. Then $j_p < j_{p-1}$ so by Lemma 2.9 it suffices to show that $j_{p-1} < j_k$. If $j_k < j_{p-1}$ then j_{p+1} lies between j_p and j_{p-1} but this is impossible by Lemma 2.8 since $(j_p - j_{p-1})(j_{p+1} - j_p) < 0$.∎

Lemma 2.11. Suppose $x, y \ge 2, a_1 < \dots a_x, b_y < \dots < b_1, a_1 < b_y$ and $a_x < b_1$. Then there exist u, v with $2 \le u \le x, 2 \le v \le y$ such that $a_{u-1} \le b_v < a_u \le b_{v-1}$.

Proof. Choose $u, v \ge 2$ such that $b_v < a_u$ and such that $a_u - b_v$ is minimal. It is always possible to do this since $b_y < a_x$ and $x, y \ge 2$, and clearly by the minimality of $a_u - b_v$ we have $a_{u-1} \le b_v < a_u \le b_{v-1}$.∎

Theorem 2.12. Suppose A is a structure h-matrix. Then the set $\{\prec_i\}$ is consistent.

Proof. Suppose there exist $i < i', j < j'$ such that $j, j' \in A_i \cap A_{i'}$ and $j \prec_i j', j' \prec_{i'} j$. Then there exist $k, k', j_1, \dots, j_k, j'_1, \dots, j'_{k'}$ such that $j = j_1 \propto_i \dots \propto_i j_k = j'$ and $j' = j'_1 \propto_{i'} \dots \propto_{i'} j'_{k'} = j$. Moreover, since $j < j'$ by Corollary 2.10 we have $j' > j_s$ for $s = 1, \dots, k-1$ and $j < j'_s$ for $s = 2, \dots, k'$. Let p be minimal such that $(j_s - j_{s-1})(j_k - j_{k-1}) > 0$ for all s with $p \le s \le k$, and let p' be minimal such that $(j'_s - j'_{s-1})(j'_{k'} - j'_{k'-1}) > 0$ for all s with $p' \le s \le k'$. By Lemma 2.9 we have $j_{p-1} \le j, \quad j_{p-1} < j_p < j_{p+1} < \dots < j_{k-1} < j_k = j', \quad j' \le j'_{p'-1}$ and $j = j'_{k'} < j'_{k'-1} < \dots < j'_{p'} < j'_{p'-1}$. Now by Lemma 2.11 there exist $u, v \ge 2$ such that $j_{u-1} \le j'_v < j_u \le j'_{v-1}$. Now since $j_{u-1} < j_u$ and $j_{u-1} \propto_i j_u$, there is some $h < i$ such that $j_{u-1}, j_u \in A_h$ and $(h, j_u) \in S$. Moreover, as A is a structure h-matrix, j'_v must be in A_h also. Likewise, as $j'_{v-1} > j'_v$ and $j'_{v-1} \propto_{i'} j'_v$, there is some $r > i'$ such that $j'_{v-1}, j'_v \in A_r$ and $(r, j'_v) \in S$. Finally, as A is a structure h-matrix, j_u must be in A_r also. Combining all this we have $h < i < i' < r, \quad j'_v < j_u, \quad (h, j_u) \in S, \quad (r, j'_v) \in S$ and both (h, j'_v) and (r, j_u) non-blank, contradicting the independence of S.∎

By Corollary 2.7, there is a consistent set, $\{<_i\}$, of linear orderings which extend the set $\{\prec_i\}$ of partial orders. Let $M = (M_{ij})$ be the canonical partial matrix generated by the $\{<_i\}$. By Remark 2.5, M is a totally monotone. Thus M is an h-matrix with A as its structure matrix. Recall that for any function f from S to the positive reals, we define the matrix M^f by $M^f_{ij} = M_{ij}$ for $(i,j) \notin S$ and $M^f_{ij} = |S_i| + f(i,j)$.

Theorem 2.13. The partial matrix M^f is totally monotone.

Proof. Suppose $i < i', j < j'$ such that all entries of the 2×2 submatrix, $M^f_{ij}, M^f_{ij'}, M^f_{i'j}$, and $M^f_{i'j'}$, are non-blank, and $M^f_{ij} \le M^f_{ij'}$. We must show that $M^f_{i'j} \le M^f_{i'j'}$. There is nothing to prove if none of the indices are in S since M is totally monotone. Also note that if exactly one of indices in a row is in S then the relationship between the two entries in that row is the same in M and in M^f. It is not hard to see that this implies we may restrict our attention to the two cases $(i,j), (i,j') \in S$ and $(i',j), (i',j') \notin S$; and $(i,j), (i,j') \notin S$ and $(i',j), (i',j') \in S$. Suppose $(i,j), (i,j') \in S$ and $(i',j), (i',j') \notin S$. Then by the definition of $\alpha_{i'}$ we have $j \alpha_{i'} j'$ and hence $M_{i'j} < M_{i'j'}$. The argument for the other case is analogous. ∎

As we noted at the beginning of the proof for the h-matrix lower bound, given Theorem 2.13 the remainder of the proof of the lower bound for h-matrices can now be completed along entirely analogous lines as the proof for the v-matrix case.

3. The Matrix Searching Algorithm for Skyline Matrices

In this section we extend the almost linear time matrix searching algorithm of [KK89] to a more general type of partial totally monotone matrices, skyline matrices. Recall that a **skyline** matrix is a v-matrix in which every non-blank column segment ends at the bottom row. The extension follows from the following observation. Suppose that, given some particular type of partial totally monotone matrix, we choose parameters n, m and t and let $q_c(t,n,m)$ denote the worst-case number of comparisions needed to find the row-minima of any of this type of partial matrix with those parameter values. Then if this function $q_c(t,n,m)$ satisfies the three key propositions in [KK89], namely Lemma 2.1, Corollary 2.4 and Theorem 2.6, it will also satisfy Theorem 2.9, i.e. $q_c(n,n,m) = O(m\alpha(n)+n)$.

For simplicity and completeness we restate the three key propositions from [KK89]. As in [KK89] we define the functions $L_i(n)$ for $i = -1, 0, 1, 2, \ldots$ recursively as follows. $L_{-1}(n) = n/2$, and for $i \geq 0$, $L_i(n) = \min_s\{L_{i-1}^s(n) \leq 1\}$. Thus $L_0(n) = \lceil \log n \rceil, L_1(n)$ is essentially $\log^*(n), L_2(n)$ is essentially $\log^{**}(n)$ etc. We now define $\alpha(n) = \min\{s : L_s(n) \leq s\}$.

Proposition 3.1 (Lemma 2.1 in [KK89]). For any positive integer a we have

$$q_c(n,n,m) \leq q_c(n/a, n, m) + O(am+n).$$

Proposition 3.2 (Corollary 2.4 in [KK89]). For any positive integer a we have

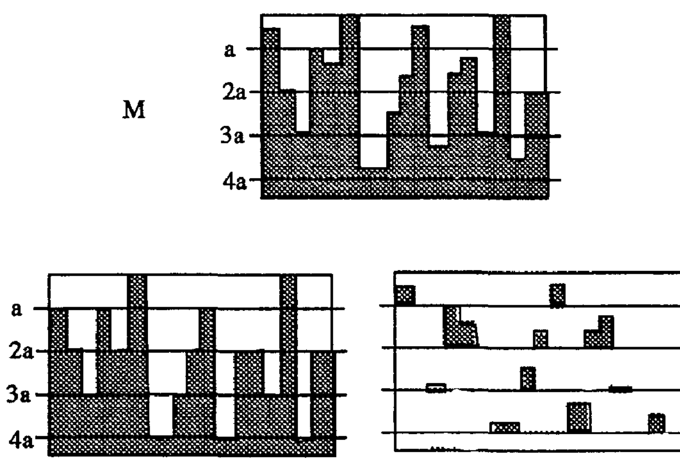
$$q_c(n,n,m) \leq q_c(n/a, n/a, m) + O(am+n).$$

Proposition 3.3 (Theorem 2.6 in [KK89]). There is a constant c_1 such that for $s \geq 0$ we have $q_c(n,n,m) \leq c_1(m + nL_s(n)) + \max\{\sum_{i=1}^k q_c(n_i/L_{s-1}(n_i), n_i, m_i) : \sum_{i=1}^k n_i \leq nL_s(n) \text{ and } \sum_{i=1}^k m_i \leq m + nL_s(n)\}$.

For M a skyline matrix, we will say that row i is a **top row** if it is the top row of some column's non-blank segment. A skyline matrix M is said to be of shape (t,n,m) if it has at most t top rows, at most n rows and at most m columns. We will denote the worst case number of comparisons needed to find the row-minima of a skyline matrix of this shape by $q_c(t,n,m)$. We will now prove that the three propositions above hold for this definition of the function q_c, thus providing an $O(m\alpha(n)+n)$ time algorithm for finding row-minima in skyline matrices. The proofs of Propositions 3.1 and 3.3 are very similar to those in [KK89] for staircase matrices, so we only sketch the proof of 3.1 to indicate how the arguments must be modified for skyline matrices. We give a complete proof of Proposition 3.2, since it is somewhat more subtle than the proof of the corresponding result for staircases matrices.

In order to translate the proofs for staircase matrices to skyline matrices we need to give the appropriate definition of stepsize approximation and border matrices in this setting.

For each column j of M let $t(j)$ be the top row of the non-blank segment in column j. For $i \leq k$ and $j \leq l$, we will use the notation $M[i,k;j,l]$ to denote the skyline matrix obtained by taking the intersection of rows $i, \ldots, k$ of M with columns $j, \ldots, l$ of M. For any positive integer a, we define the **stepsize a approximation** of M to be the submatrix of M obtained by, for each j with $t(j) \leq a\lfloor n/a \rfloor$, truncating the non-blank entries in column j so that the non-blank segment begins at row $a\lceil t(j)/a \rceil$, and for each j with $t(j) > a\lfloor n/a \rfloor$ replacing all the entries in column j with blanks. We denote the stepsize a approximation of M by M_a. An example is illustrated in Figure 3. Clearly M_a is a skyline matrix of shape $(\lfloor n/a \rfloor, n, m)$. We define the **$a$-border matrices of M**, which we denote by $M(a,i)$ for $i = 1, \ldots, \lceil n/a \rceil$, by $M(a,i)$ is the skyline matrix with at most $a-1$ rows whose non-blank entries are the non-blank entries of M in rows $(i-1)a+1, \ldots, \min(ia-1, n)$ which are blank in M_a. An example of a-border matrices is also illustrated in Figure 3. We will denote the set of a-border matrices of M by $\Gamma(M,a)$. It is easy to see that the skyline matrices $M_a, M(a,1), \ldots, M(a, \lceil n/a \rceil)$ disjointly cover the non-blank entries of M.



Stepsice a approximation M_a a-border matrices

Figure 3

Given these definitions, the proof of the first proposition is identical to the proof of Lemma 2.1 in [KK89].

Proposition 1. For any positive integer a we have

$$q_c(n,n,m) \leq q_c(n/a, n, m) + O(am+n).$$

Proof. Let M be a skyline matrix of shape (n,n,m). Since the stepsize a approximation M_a is of shape $(n/a,n,m)$, it can be processed in $q_c(n/a,n,m)$ time. Since the total number of non-blank entries in the a-border matrices $M(a,i)$ for $i=1,\ldots,\lfloor n/a \rfloor$, is less than am we can process these border matrices in $O(am)$ time. Finally in $O(n)$ comparisons we can compare the row-minima found in M_a with the row-minima found in the a-border matrices, and hence determine the row-minima of M. ∎

The proof of the second proposition requires a little more care than the corresponding proposition for staircase matrices.

Proposition 2. For any positive integer a we have

$$q_c(n,n,m) \le q_c(n/a,n/a,m) + O(am+n).$$

Proof. Let N be a skyline matrix of shape (n,n,m). By the proof of Proposition 1 it suffices to show that finding the row-minima of the stepsize approximation matrix, N_a, can be reduced to finding the row-minima of a skyline matrix of shape $(n/a,n/a,m)$ in $O(m+n)$ time. Let $M=N_a$, and for each j let $t(j)$ be the row such that the non-blank segment of column j of M begins at row $t(j)$. Let S be the $\lfloor n/a \rfloor \times m$ skyline matrix where $S_{i,j}=M_{ai,j}$ if $t(j) \le (i-1)a$ and is blank otherwise. Let $s(i)$ be the column containing the minimum value in row i of S, and let $d(i)$ such that $M_{ai,d(i)}$ is minimal among the $M_{ai,j}$ such that $t(j)=ai$. Finally let $j(i)$ be the column containing the minimum value in row ai of N. Note that $j(i)$ must be either $s(i)$ or $d(i)$. Note that S is a skyline matrix of shape $(n/a,n/a,m)$. Thus it suffices to show that given the $\{s(i)\}$, the row minima of the rows of M_a can be found in $O(m+n)$ time.

We first note that we can find the $d(i)$ in at most $O(m)$ time since we only need to look at one entry in each column. Let $r=\lfloor n/a \rfloor$. For each $i=1,\ldots,r+1$ let $J(i)$ be the set of columns $\{j : s(i) \le j \le j(i-1), t(j) \le a(i-1)$, and $j \le s(k)$ for each $k<i$ such that $t(j) \le a(k-1)\}$, where we adopt the convention that $j(0)=m$ and $s(r+1)=1$. Let $A(i)$ be the matrix consisting of the intersection of rows $a(i-1)+1,\ldots,\min(n,ai-1)$ of M with the columns in $J(i)$.

We claim that each $A(i)$ is a totally monotone matrix containing all the row minima in rows $a(i-1)+1,\ldots,\min(n,ai-1)$ of M, and $\sum_{i=1}^{r+1}|J(i)| \le m+r+1$. We now prove this claim. It is easy to see that each $A(i)$ is a totally monotone matrix since j in $J(i)$ implies that column j has no blank entries in rows $a(i-1)+1,\ldots,\min(n,ai-1)$. We now show that $A(i)$ contains all the row minima in rows $a(i-1)+1,\ldots,\min(n,ai-1)$ of M. In other words, for each j not in $J(i)$, and $a(i-1)+1 \le h \le \min(n,ai-1)$, we must show that $M_{h,j}$ is not a row minima. If $t(j) > a(i-1)$ then $t(j) \ge ai$ so $M_{h,j}$ is blank, and hence not a row minima. Thus we may assume $t(j) \le a(i-1)$. Now by total monotonicity, if $j < s(i)$ then we must have $M_{h,j} > M_{h,s(i)}$ and if $j > j(i-1)$ then we must have $M_{h,j(i-1)} \le M_{h,j}$, so in either case $M_{h,j}$ is not a (leftmost) row minima. Finally suppose $j > s(k)$ for some $k < i$ such that $t(j) \le a(k-1)$. Again by total monotonicity we have $M_{h,j} \ge M_{h,s(k)}$ so again $M_{h,j}$ is not a row minima.

We now show that $\sum_{i=1}^{r+1}|J(i)| \le m+r+1$. It suffices to show that for $i<i'$, the sets $J(i)\backslash\{s(i)\}$ and $J(i')$ are disjoint. Suppose j belongs to both $J(i)$ and $J(i')$. Then since $i<i'$ and $t(j) \le a(i-1)$ because j is in $J(i)$, we must have $j \le s(i)$ because j is in $J(i')$. However j in $J(i)$ implies $j \ge s(i)$ and hence $j=s(i)$.

The proof is completed by observing that finding the row minima of the $A(i)$ requires at most $O(n+\sum_{i=1}^{r+1}|J(i)|)=O(m+n)$ time. ∎

In order to indicate how the proof of Proposition 3 is translated from the proof for staircase matrices in [KK89], we need to define the concept of **i-th slice** for skyline matrices. For each i, the i-th slice of M is the set of columns $\{j : t(j)=i\}$. Given this definition it is fairly straightforward to translate the proof of Theorem 2.6 to handle skyline matrices.

4. Open Problems

There are many interesting problems in matrix searching which remain open (see [AP88] for example). In this section we restrict ourselves to problems related to upper and lower bounds for matrix searching in partial matrices. The first obvious group of problems concerns closing the gap between the current upper and lower bounds for the partial matrices discussed in this paper. Specifically for staircase and skyline matrices we have $O(m\alpha(n)+n)$ upper bounds ([KK88] and this paper respectively) for searching matrices of size $n \times m$ but only linear lower bounds. For v-matrices and h-matrices there are fairly straightforward $O(m\log n+n)$ upper bounds [AS89] and lower bounds of $\Omega(n\alpha(n))$ (this paper). It would be interesting to improve these lower bounds to $\Omega(m\alpha(n))$ for the case $m>n$. Another problem which seems to be difficult is to find a better upper bound for horizontal skyline matrices, i.e. h-matrices in which each row's non-blank segment starts in the first column.

A completely different direction involves Monge matrices [AP88]. These are matrices which satisfy the condition, for every $i < i', j < j'$ such that all entries of the 2×2 submatrix, $M_{ij}, M_{ij'}, M_{i'j}$, and $M_{i'j'}$, are non-blank, we have $M_{ij} + M_{i'j'} \geq M_{ij'} + M_{i'j}$. It is easy to see that Monge implies totally monotone but the reverse is not true. In most applications of totally monotone matrices, the matrix in question is actually Monge, so it would be worthwhile to get a superlinear lower bound for matrix searching of Monge matrices.

Finally, there are a number of open problems concerning the techniques used to prove the lower bound for h-matrices. First, is it possible to find a simpler construction of the matrix M directly from the line segments and their left envelope as was done in the v-matrix case? Next, for any structure matrix A with an independent set S, one can define the relations $\{\propto_i\}$ as was done in section 2. It is easy to find structure matrices in which some of the $\{\prec_i\}$ are not partial orders. It seems natural to try to characterize the family of structure matrices for which $\{\prec_i\}$ is a consistent set of partial orders. This paper proves that structure h-matrices have this property, and we believe that a similar proof can be given for structure v-matrices though it seems to be slightly more difficult. We conjecture that in fact this will hold for any structure matrix in which for every non-blank entry, the set of non-blank entries in either its row or column form a contiguous segment.

References

[AKMSW87] A. Aggarwal, M. Klawe, S. Moran, P. Shor, and R. Wilber, Geometric applications of a matrix searching algorithm, Algorithmica 2(1987), pp. 195-208.

[AK87] A. Aggarwal and M. Klawe, Applications of generalized matrix searching to geometric algorithms, IBM Research Division Technical Report 1987, to appear in Discrete Applied Math.

[AP88] A. Aggarwal and J. Park, Notes on searching in multidimensional arrays, Proc. 29th Ann. IEEE Symposium on Found. Comp. Sci. (1988), pp.497-512.

[AS87] A. Aggarwal and S. Suri, Fast algorithms for computing the largest empty rectangle, Proc. 3rd Ann. Symp. Comp. Geom.(1987), pp.278-290.

[AS89] A. Aggarwal and S. Suri, Computing the farthest visible pair in a simple polygon, preprint.

[EGG88] D. Eppstein, Z. Galil and R Giancarlo, Speeding up dynamic programming, Proc. 29th Ann. IEEE Symposium on Found. Comp. Sci. (1988), pp.488-496.

[HL87] D.S. Hirschberg and L.L. Larmore, The least weight subsequence problem,SIAM J. Computing 16, 1987, pp. 628-638.

[KK88] M. Klawe and D.j. Kleitman, An almost linear time algorithm for generalized matrix searching, IBM Research Division Technical Report 1988, to appear in SIAM J. Discrete Math., February 1990.

[W86] A. Wiernik, Planar realizations of nonlinear Davenport-Schinzel sequences by segments, Proc. 27th Ann. IEEE Symposium on Found. Comp. Sci. (1986), pp. 97-106.

[W88] R. Wilber, The concave least weight subsequence revisited, J. Algorithms 9 (1988), pp.418-425.