Submatrix Maximum Queries in Monge Matrices are Equivalent to Predecessor Search

Paweł Gawrychowski\textsuperscript{1}, Shay Mozes\textsuperscript{2⋆}, and Oren Weimann\textsuperscript{3⋆}

\textsuperscript{1} University of Warsaw, gawry@mimuw.edu.pl
\textsuperscript{2} IDC Herzliya, snozes@idc.ac.il
\textsuperscript{3} University of Haifa, oren@cs.haifa.ac.il

Abstract. We present an optimal data structure for submatrix maximum queries in $n \times n$ Monge matrices. Our result is a two-way reduction showing that the problem is equivalent to the classical predecessor problem in a universe of polynomial size. This gives a data structure of $O(n)$ space that answers submatrix maximum queries in $O(\log \log n)$ time, as well as a matching lower bound, showing that $O(\log \log n)$ query-time is optimal for any data structure of size $O(n \text{polylog}(n))$. Our result settles the problem, improving on the $O(\log^2 n)$ query-time in SODA’12, and on the $O(\log n)$ query-time in ICALP’14.

In addition, we show that partial Monge matrices can be handled in the same bounds as full Monge matrices. In both previous results, partial Monge matrices incurred additional inverse-Ackerman factors.

1 Introduction

Data structures for range queries and for predecessor queries are among the most studied data structures in computer science. Given an $n \times n$ matrix $M$, a range maximum (also called submatrix maximum) data structure can report the maximum entry in any query submatrix (a set of consecutive rows and a set of consecutive columns) of $M$. Given a set $S \subseteq [0,U)$ of $n$ integers from a polynomial universe $U$, a predecessor data structure can report the predecessor (and successor) in $S$ of any query integer $x \in [0,U)$. In this paper, we prove that these two seemingly unrelated problems are in fact equivalent when the matrix $M$ is a Monge matrix.

Range maximum queries. A long line of research over the last three decades including [4,11,12,15,23] achieved range maximum data structures of $O(n^2)$ space and $O(1)$ query time\textsuperscript{1}, culminating with the $O(n^2)$-space $O(1)$-query data structure of Yuan and Atallah [23]. In general matrices, this is optimal since representing the input matrix already requires $\Theta(n^2)$ space. In fact, reducing the additional space to $O(n^2/c)$ is known to incur an $O(c)$ query-time [6] and such tradeoffs can indeed be achieved for any value of $c$ [5,6].

However, in many applications, the matrix $M$ is not stored explicitly but any entry of $M$ can be computed when needed in $O(1)$ time. One such case is when the matrix $M$ is sparse, or simply has $N = o(n^2)$ nonzero entries. In this case the problem is known in computational geometry as the orthogonal range searching problem on the $n \times n$ grid. In this case as well, various data structures with $O(N)$-space and $O(1)$-query appear in a long history of results including [3,8,10,13,15]. For a survey on orthogonal range searching see [21]. Another case where the additional space can be made $o(n^2)$ (and in fact even $O(n)$) is when the matrix is a Monge matrix.

Range maximum queries in Monge matrices. A matrix $M$ is Monge if for any pair of rows $i < j$ and columns $k < \ell$ we have that $M[i,k] + M[j,\ell] \geq M[i,\ell] + M[j,k]$.\textsuperscript{5} Submatrix maximum queries on Monge matrices have various important applications in combinatorial optimization and computational geometry such as problems involving distances in the plane, and in problems on convex $n$-gons. See [7] for a survey on Monge matrices and their uses in combinatorial optimization. Submatrix maximum queries on Monge matrices are used in algorithms that efficiently find the largest empty rectangle containing a query point, in

\textsuperscript{1} Mozes and Weimann supported in part by Israel Science Foundation grant 794/13.
\textsuperscript{2} The $O(\cdot)$ notation hides polylogarithmic factors in $n$.
\textsuperscript{5} Monge matrices are often defined with $a \leq$ (rather than $\geq$) in the condition. Our results apply to both definitions, and to minimum (rather than maximum) queries.
dynamic distance oracles for planar graphs, and in algorithms for maximum flow in planar graphs. See [17] for more details on the history of this problem and its applications.

Given an $n \times n$ Monge matrix $M$ it is possible to obtain compact data structures of only $\tilde{O}(n)$ space that can answer submatrix maximum queries in $O(1)$ time. The first such data structure was given by Kaplan, Mozes, Nussbaum and Sharir [17]. They presented an $O(n \log n)$-space data structure with $O(\log^2 n)$ query time. This was improved in [16] to $O(n)$ space and $O(\log n)$ query time.

**Breakpoints and Partial Monge matrices.** Given an $m \times n$ Monge matrix $M$, let $r(c)$ be the row containing the maximum element in the $c$-th column of $M$. It is easy to verify that the $r(\cdot)$ values are monotone, i.e., $r(1) \leq r(2) \leq \ldots \leq r(n)$. Columns $c$ such that $r(c-1) < r(c)$ are called the breakpoints of $M$. A Monge matrix consisting of $m < n$ rows has $O(m)$ breakpoints, which can be found in $O(n)$ time using the SMAWK algorithm [2].

Some applications involve partial Monge matrices rather than full Monge matrices. A partial Monge matrix is a Monge matrix where some of the entries are undefined, but the defined entries in each row and in each column are contiguous. The total number of breakpoints in a partial Monge matrix is still $O(m)$ [16], and they can be found in $O(n \cdot \alpha(n))$ time$^6$ using an algorithm of Klawe and Kleitman [18]. This was used in [16,17] to extend their solutions to partial Monge matrices at the cost of an additional $\alpha(n)$ factor to the query time.$^7$

**Our results.** In this paper, we fully resolve the submatrix maximum query problem in $n \times n$ Monge matrices by presenting a data structure of $O(n)$ space and $O(\log \log n)$ query time. Consequently, we obtain an improved query time for other applications such as finding the largest empty rectangle containing a query point. We complement our upper bound with a matching lower bound, showing that $O(\log \log n)$ query-time is optimal for any data structure of size $O(n \cdot \text{polylog}(n))$. In fact, implicit in our upper and lower bound is an equivalence between the predecessor problem in a universe of polynomial size and the range maximum query problem in Monge matrices. The upper bound essentially reduces a submatrix query to a predecessor problem, and vice versa, the lower bound reduces the predecessor problem to a submatrix query problem.

Finally, we extend our result to partial Monge matrices with the exact same bounds (i.e., $O(n)$ space and $O(\log \log n)$ query time). Our result is the first to achieve such extension with no overhead.

**Techniques.** Let $M$ be an $n \times n$ Monge matrix.$^8$ Consider a full binary tree $T$ whose leaves are the rows of $M$. Let $M_u$ be the submatrix of $M$ composed of all rows (i.e., leaves) in the subtree of a node $u$ in $T$. Both existing data structures for submatrix maximum queries [16,17] store, for each node $u$ in $T$ a data structure $D_u$. The goal of $D_u$ is to answer submatrix maximum queries for queries that include an arbitrary interval of columns and exactly all rows of $M_u$. This way, an arbitrary query is covered in [16,17] by querying the $D_u$ structures of $O(\log n)$ canonical nodes of $T$. An $O(\log n)$ bound is thus inherent for any solution that examines the canonical nodes. We overcome this obstacle by designing a stronger data structure $D_u$. Namely, one that supports queries that include an arbitrary interval of columns and a prefix of rows or a suffix of rows of $M_u$. This way, an arbitrary query can be covered by just two $D_u$s. The idea behind the new design is to efficiently encode the changes in column maxima as we add rows to $M_u$ one by one. Retrieving this information is done using weighted ancestor search and range maximum queries on trees. This is a novel use of these techniques.

For our lower bound, we show that for any set of $n$ integers $S \subseteq [0, n^2]$ there exists an $n \times n$ Monge matrix $M$ such that the predecessor of $x$ in $S$ can be found with submatrix minimum queries on $M$. The predecessor lower bound of Pătraşcu and Thorup [22] then implies that $O(n \cdot \text{polylog}(n))$ space requires $O(\log \log n)$ query time. We overcome two technical difficulty here: First, $M$ should be Monge. Second, there must be an $O(n \cdot \text{polylog}(n))$-size representation of $M$ which can retrieve any entry $M[i,j]$ in $O(1)$ time.

Finally, for handling partial Monge matrices, and unlike previous solutions for this case, we do not directly adapt the solution for the full Monge case to partial Monge matrices. Instead we decompose the partial Monge matrix into many full Monge matrices, that can be preprocessed to be queried cumulatively in an efficient way. This requires significant technical work and careful use of the structure of the decomposition.

---

$^6$ Here $\alpha(n)$ is the inverse-Ackerman function.

$^7$ In [17], there was also an additional $\log n$ factor to the space.

$^8$ We consider $m \times n$ matrices, but for simplicity we sometimes state the results for $n \times n$ matrices.
Roadmap. In Section 2 we present an $O(n \log n)$-space data structure for Monge matrices that answers submatrix maximum queries in $O(\log \log n)$ time. In Section 3 we reduce the space to $O(n)$. Our lower bound is given in Section 4, and the extension to partial Monge matrices in Section 5.

## 2 Data structure for Monge matrices

Our goal in this section is to construct, for a given $m \times n$ Monge matrix $M$, a data structure of size $O(m \log n)$ that answers submatrix maximum queries in $O(\log \log n)$ time. In Section 3 we show how to reduce the space from $O(n \log n)$ to $O(n)$ when $m = n$. We will actually show a stronger result, namely the structure allows us to reduce in $O(1)$ time a submatrix maximum query into $O(1)$ predecessor queries on a set consisting of $n$ integers from a polynomial universe.

We denote by $\text{pred}(m, n)$ the complexity of a predecessor query on a set of $m$ integers from a universe $\{0, \ldots, n - 1\}$. It is well known that there are $O(m)$ data structures achieving $\text{pred}(m, n) = \min\{O(\log m), O(\log \log n)\}$.

Recall that a submatrix maximum query returns the maximum $M[i, j]$ over all $i \in [i_0, i_1]$ and $j \in [j_0, j_1]$ for a given $i_0 \leq i_1$ and $j_0 \leq j_1$. We start by answering the easier subcolumn maximum queries within these space and time bounds. That is, finding the maximum $M[i, j]$ over all $i \in [i_0, i_1]$ for a given $i_0 \leq i_1$ and $j$.

We construct a full binary tree $T$ over the rows of $M$. Every leaf of the tree corresponds to a single row of $M$, and every inner node corresponds to the range of rows in its subtree. To find the maximum $M[i, j]$ over all $i \in [i_0, i_1]$ and $j$, we first locate the lowest common ancestor (lca) $u$ of the leaves corresponding to $i_0$ and $i_1$ in the tree. Then we decompose the query into two parts: one fully within the range of rows $M_r$ of the left child of $u$, and one fully within the range of rows $M_l$ of the right child of $u$. The former ends at the last row of $M_l$ and the latter starts at the first row of $M_r$. We equip every node with two data structures allowing us to answer such simpler subcolumn maximum queries. Because of symmetry (if $M$ is Monge, so is $M'$, where $M'[i, j] = M[n + 1 - i, n + 1 - j]$) it is enough to show how to answer subcolumn maximum queries starting at the first row.

**Lemma 1.** Given an $m \times n$ Monge matrix $M$, a data structure of size $O(m)$ can be constructed in $O(m \log n)$ time to answer in $O(\text{pred}(m, n))$ time subcolumn maximum queries starting at the first row of $M$.

**Proof.** Consider queries spanning an entire column $c$ of $M$. To answer such a query, we only need to find the corresponding $r(c)$. If we store the breakpoints of $M$ in a predecessor structure, where every breakpoint $c$ links to its corresponding value of $r(c)$, a query can be answered with a single predecessor search. More precisely, to determine the maximum in the $c$-th column of $M$, we locate the largest breakpoint $c' \leq c$, and then set $r(c) = r(c')$. Hence we can construct a data structure of size $O(m)$ to answer entire column maximum queries in $O(\text{pred}(m, n))$ time.

Let $M_i$ be a Monge matrix consisting of the first $i$ rows of $M$. By applying the above reasoning to every $M_i$, separately, we immediately get a structure of size $O(m^2)$ answering subcolumn maximum queries starting at the first row of $M$ in $O(\text{pred}(m, n))$ time. We want to improve on this by utilizing the dependency of the structures constructed for different $i$’s. Namely it can be observed that the list of breakpoints of $M_{i+1}$ is a prefix of the list of breakpoints of $M_i$ to which we append at most one new element. In other words, if the breakpoints of $M_i$ are stored on a stack, we need to pop zero or more elements and push at most one new element to represent the breakpoints of $M_{i+1}$. Consequently, instead of storing a separate list for every $M_i$, we can succinctly describe the content of all stacks with a single tree $T$ on at most $m + 1$ nodes. For every $i$, we store a pointer to a node $s(i) \in T$, such that the ancestors of $s(i)$ (except for the root) are exactly the breakpoints of $M_i$. Whenever we pop an element from the current stack, we move to the parent of the current node, and whenever we push an element, we create a new node and make it a child of the current node. Initially, the tree consists of just the root. Every node is labelled with a column number and by construction these numbers are strictly increasing on any path starting at the root (the root is labelled with $-\infty$). Therefore, a predecessor search for $j$ among the breakpoints of $M_i$ reduces to finding the leafmost ancestor of $s(i)$ whose label is at most $j$. This is known as the weighted ancestor problem.
Weighted ancestor queries on a tree of size $O(m)$ are equivalent to predecessor searching on a number of sets of $O(m)$ total size $[9]$, achieving the claimed space and query time bounds.

To finish the proof, we need to bound the construction time. The bottleneck is constructing the tree $T$. Let $c_1 < c_2 < \ldots < c_k$ for some $k \leq i$ be the breakpoints of $M_i$. As long as $M[i+1, c_k] \geq M[r(c_k), c_k]$ we decrease $k$ by one, i.e., remove the last breakpoint. This process is repeated $O(m)$ times in total. If $k = 0$ we create a new breakpoint $c_1 = 1$. If $k \geq 1$ and $M[i+1, c_k] < M[r(c_k), c_k]$, we check if $M[i+1, n] \geq M[r(c_k), n]$. If so, we need to create a new breakpoint. To this end, we need to find the smallest $j$ such that $M[i+1, j] \geq M[r(c_k), j]$. This can be done in $O(\log n)$ using binary search. Consequently, $T$ can be constructed in $O(m \log n)$ time. Then augmenting it with a weighted ancestor structure takes $O(m)$ time.

We apply Lemma 1 twice to every node of the full version tree $T$. Once for subcolumn maximum queries starting at the first row and once for queries ending at the last row. Since the total size of all structures at the same level of the tree is $O(m)$, the total size of our subcolumn maximum data structure becomes $O(m \log m)$, and it can be constructed in $O(m \log m \log n)$ time to answer queries in $O(\text{pred}(m, n))$ time. Hence we have proved the following.

**Theorem 1.** Given an $m \times n$ Monge matrix $M$, a data structure of size $O(m \log m)$ can be constructed in $O(m \log m \log n)$ time to answer subcolumn maximum queries in $O(\text{pred}(m, n))$ time.

By symmetry (a transpose of a Monge matrix is Monge) we can answer subrow maximum queries (where the query is a single row and a range of columns) in $O(\text{pred}(n, m))$ time. We are now ready to tackle general submatrix maximum queries.

At a high level, the idea is identical to the one used for subcolumn maximum queries: we construct a full binary tree $T$ over the rows of $M$, where every node corresponds to a range of rows. To find maximum $M[i, j]$ over all $i \in [i_0, i_1]$ and $j \in [j_0, j_1]$ for a given $i_0 \leq i_1$ and $j_0 \leq j_1$, we locate the lowest common ancestor of the leaves corresponding to $i_0$ and $i_1$ and decompose the query into two parts, the former ending at the last row of $M_i$ and the latter starting at the first row of $M_j$. Every node is equipped with two data structures allowing us to answer submatrix maximum queries starting at the first row or ending at the last row. As before, it is enough to show how to answer submatrix maximum queries starting at the first row.

**Lemma 2.** Given an $m \times n$ Monge matrix $M$, and a data structure that answers subrow maximum queries on $M$ in $O(\text{pred}(m, n))$ time, one can construct in $O(m \log m)$ time a data structure consuming $O(m)$ additional space, that answers submatrix maximum queries starting at the first row of $M$ in $O(\text{pred}(m, n) + \text{pred}(n, m))$ time.

**Proof.** We extend the proof of Lemma 1. Let $c_1 < c_2 < \ldots < c_k$ be the breakpoints of $M$ stored in a predecessor structure. For every $i \geq 2$ we precompute and store the value

$$m_i = \max_{j \in [c_{i-1}, c_i]} M[r(c_{i-1}), j].$$

These values are augmented with a (one dimensional) range maximum query data structure. To begin with, consider a submatrix maximum query starting at the first row of $M$ and ending at the last row of $M$, i.e., we need to calculate the maximum $M[i, j]$ over all $i \in [1, m]$ and $j \in [j_0, j_1]$. We find in $O(\text{pred}(m, n))$ the successor of $j_0$, denoted $c_i$, and the predecessor of $j_1$, denoted $c_{i'}$. There are three possibilities:

1. The maximum is reached for $j \in [j_0, c_i)$,
2. The maximum is reached for $j \in [c_i, c_{i'})$,
3. The maximum is reached for $j \in [c_{i'}, j_1]$.

The first and the third possibilities can be calculated with subrow maximum queries in $O(\text{pred}(n, m))$, because both ranges span an interval of columns and a single row. The second possibility can be calculated with a range maximum query on the range $(i, i')$. Consequently, we can construct a data structure of size $O(m)$ to answer such submatrix maximum queries in $O(\text{pred}(m, n) + \text{pred}(n, m))$ time.

---

9 Technically, the reduction adds $O(\log^* m)$ to the query time, but this can be avoided.
The above solution can be generalized to queries that start at the first row of $M$ but do not necessarily end at the last row of $M$. This is done by considering the Monge matrices $M_i$ consisting of the first $i$ rows of $M$. For every such matrix, we need a predecessor structure storing all of its breakpoints, and additionally a range maximum structure over their associated values. Hence now we need to construct a similar tree $T$ as in Lemma 1 on $O(m)$ nodes, but now every node has both a weight and a value. The weight of a node is the column number of the corresponding breakpoint $c_k$, and the value is its $m_k$ (or undefined if $k = 1$). As in Lemma 1, the breakpoints of $M_i$ are exactly the ancestors of the node $s(i)$. Note that every $m_k$ is defined in terms of $c_{k-1}$ and $c_k$, but this is not a problem because the predecessor of a breakpoint does not change during the whole construction. We maintain a weighted ancestor structure using the weights (in order to find $c_i$ and $c_j$ in $O(\text{pred}(m,n))$ time), and a generalized range maximum structure using the values. A generalized range maximum structure of a tree $T$, given two query nodes $u$ and $v$, returns the maximum value on the unique $u$-to-$v$ path in $T$. It can be implemented in $O(m)$ space and $O(1)$ query time after $O(m \log m)$ preprocessing [12] once we have the values. The values can be computed with subrow maximum queries in $O(m \cdot \text{pred}(n,m)) = O(m \log m)$ total time.

By applying Lemma 2 twice to every node of the full binary tree $T$, we construct in $O(m \log^2 m)$ time a data structure of size $O(m \log m)$ to answer submatrix maximum queries in $O(\text{pred}(m,n) + \text{pred}(n,m))$ time. In order to apply Lemma 2 to a node of $T$ we need a subrow maximum query data structure for the corresponding rows of the matrix $M$. Note, however, that a single subrow maximum query data structure for $M$ can be used for all nodes of $T$.

**Theorem 2.** Given an $m \times n$ Monge matrix $M$, and a data structure answering subrow maximum queries on $M$ in $O(\text{pred}(n,m))$ time, one can construct in $O(m \log^2 m)$ time a data structure taking $O(m \log m)$ additional space, that answers submatrix maximum queries on $M$ in $O(\text{pred}(m,n) + \text{pred}(n,m))$ time.

By combining Theorem 1 with Theorem 2, given an $n \times n$ Monge matrix $M$, a data structure of size $O(n \log n)$ can be constructed in $O(n \log^2 n)$ time to answer submatrix maximum queries in $O(\text{pred}(n,n))$ time.

### 3 Obtaining linear space

In this section we show how to decrease the space of the data structure presented in Section 2 to be linear. We extend the idea developed in our previous paper [16]. The previous linear space solution was based on partitioning the matrix $M$ into $n/x$ matrices $M_1, M_2, \ldots, M_{n/x}$, where each $M_i$ is a slice of $M$ consisting of $x = \log n$ consecutive rows. Then, instead of working with the matrix $M$, we worked with the $(n/x) \times n$ matrix $M'$, where $M'[i,j]$ is the maximum entry in the $j$-th column of $M_i$.

**Subcolumn queries.** Consider a subcolumn query. Suppose the query is entirely contained in some $M_i$. This means it spans less than $x = \log n$ rows. In [16], since the desired query time was $O(\log n)$, a query simply inspected all elements of the subcolumn. In this case however, since the desired query time is only $O(\log \log n)$, we apply the above partitioning scheme twice. We explain this now.

We start with the following lemma, that provides an efficient data structure for queries consisting of a single column and all rows in rectangular matrices. The statement of the lemma was taken almost verbatim from the previous solution [16]. Its query time was originally stated in terms of query to a predecessor structure, but here we prefer to directly plug in the bounds implied by atomic heaps [14] (which support predecessor searches in constant time provided $x$ is $O(\log n)$). This requires only an additional $O(n)$ time and space preprocessing.

**Lemma 3 ([16]).** Given an $x \times n$ Monge matrix, a data structure of size $O(x)$ can be constructed in $O(x \log n)$ time to answer entire-column maximum queries in $O(1)$ time, if $x = O(\log n)$.

Our new subcolumn data structure is summarized in the following theorem. It uses the above lemma and two applications of the partitioning scheme.
**Theorem 3.** Given an $m \times n$ Monge matrix $M$, a data structure of size $O(m)$ can be constructed in $O(m \log n)$ time to answer subcolumn maximum queries in $O(\log \log (n + m))$ time.

**Proof.** We first partition $M$ into $n/x$ matrices $M_1, M_2, \ldots, M_{n/x}$, where $x = \log m$. Every $M_i$ is a slice of $M$ consisting of $x$ consecutive rows. Next, we partition every $M_i$ into $x/x'$ matrices $M_{i,1}, M_{i,2}, \ldots, M_{i,x'}$, where $x' = \log \log m$. Every $M_{i,j}$ is a slice of $M_i$ consisting of $x'$ consecutive rows (without loss of generality, assume that $x$ divides $m$ and $x'$ divides $x$). Now we define a new $(m/x) \times n$ matrix $M'$, where $M'[i,j]$ is the maximum entry in the $j$-th column of $M_i$. Similarly, for every $M_i$ we define a new $(x/x') \times n$ matrix $M'_i$, where $M'_i[j,k]$ is the maximum entry in the $k$-th column of $M_{i,j}$.

We apply Lemma 3 on every $M_i$ and $M_{i,j}$ in $O(m \log n)$ total time and $O(m)$ total space, so that any $M'[i,j]$ or $M'_i[j,k]$ can be retrieved $O(1)$ time. Furthermore, it can be easily verified that $M'$ and all $M'_i$'s are also Monge. To prove this, it is enough to argue that if $N$ is an $4 \times 2$ Monge matrix, the $2 \times 2$ matrix $N'$ created by partitioning $N$ into two slices consisting of two rows and computing the maximum in every column of every slice is also Monge.

To this end, we need to compare:

$$N'[1,1] + N'[2,2] = \max(N[1,1], N[2,1]) + \max(N[3,2], N[4,2])$$

and

$$N'[1,2] + N'[2,1] = \max(N[3,1], N[4,1]) + \max(N[1,2], N[2,1]).$$

Let $\max(N[1,2], N[2,1]) = N[i, 2]$, where $i \in \{1, 2\}$, and similarly $\max(N[3,1], N[4,1]) = N[i', 1]$, where $i' \in \{3, 4\}$. Then

$$(N'[1,1] + N'[2,2]) - (N'[1,2] + N'[2,1]) \geq (N[i', 1] + N[i, 2]) - (N[i', 1] + N[i, 2])$$

which is at least 0 because of $N$ being Monge.

Therefore, because $M'$ and all $M'_i$ are all Monge, and by Lemma 3 their entries can be accessed in $O(1)$ time, we can apply Theorem 1 on $M'$ and every $M'_i$. The total construction time is $O((m/x) \log(m/x) \log n + (m/x)(x/x') \log(x/x') \log n) = O(m \log n)$, and the total size of all structures constructed so far is $O((m/x) \log(m/x) + (m/x)(x/x') \log(x/x')) = O(m)$.

Now consider a subcolumn maximum query. If the range of rows is fully within a single $M_{i,j}$, the query can be answered naively in $O(x'') = O(\log \log m)$ time. Otherwise, if the range of rows is fully within a single $M_i$, the query can be decomposed into a prefix fully within some $M_{i,j}$, an infix corresponding to a range of rows in $M'_i$, and a suffix fully within some $M_{i,j}$. The maximum in the prefix and the suffix can be computed naively in $O(x'') = O(\log \log m)$ time, and the maximum in the infix can be computed in $O(\log \log n)$ using the structure constructed for $M'_i$. Finally, if the range of rows starts inside some $M_i$ and ends inside another $M_j$, the query can be decomposed into two queries fully within $M_i$ and $M_j$, respectively, which can be processed in $O(\log \log n)$ time as explained before, and an infix corresponding to a range of rows of $M'$.

The maximum in the infix can also be computed in $O(\log \log n)$ time using the structure constructed for $M'$. $\square$

**Submatrix queries.** We are ready to present the final version of our data structure. It is based on two applications of the partitioning scheme, and an additional trick of transposing the matrix.

**Theorem 4.** Given an $n \times n$ Monge matrix $M$, a data structure of size $O(n)$ can be constructed in $O(n \log n)$ time to answer submatrix maximum queries in $O(\log \log n)$ time.

**Proof.** We partition $M$ as described in the proof of Theorem 3, i.e., $M$ is partitioned into $n/x$ matrices $M_1, M_2, \ldots, M_{n/x}$, where $x = \log n$, and every $M_i$ is then partitioned into $x/x'$ matrices $M_{i,1}, M_{i,2}, \ldots, M_{i,x'}$, where $x' = \log \log n$. Then we define smaller Monge matrices $M'$ and $M'_i$, and provide $O(1)$ time access to their entries with Lemma 3. We apply Theorem 3 to the transpose of $M'$ to get a subrow maximum query data structure for $M'$. This takes $O(n)$ space and $O(n \log n)$ time. With this data structure we can apply Theorem 2 on $M'$, which takes an additional $O\left(\frac{n}{\log n} \log \frac{n}{\log n}\right) = O(n)$ space and $O(n \log n)$ time. We would have liked to apply Theorem 3 to the transpose of all $M'_i$ as well, but this would require $O(n)$ space for each matrix, which we cannot afford. Since we do not have subrow maximum query data structure for the $M'_i$'s, we cannot
apply Theorem 2 to them directly. However, note that the subrow maximum query data structure is used in Theorem 2 in two ways (see the proof of Lemma 2). The first use is in directly finding the subrow maximum in cases 1 and 3 in the proof of Lemma 2. In the absence of the subrow structure, we can still report the two rows containing the candidate maximum, although not the maximum itself. The second use is in computing the values for the generalized range maximum structure required to handle case 2 in that proof. In this case, we do not really need the fast query of the data structure of Theorem 3, and can use instead the slower linear space data structure from [16, Lemma 2] to compute the values in \( O(n \log n) \) time. Thus, we can apply Theorem 2 to each \( M'_i \), and get at most two candidate rows of \( M'_i \) (from cases 1 and 3), and one candidate entry of \( M'_i \) (from case 2), with the guarantee that the submatrix maximum is among these candidates.

We repeat the above preprocessing on the transpose of \( M \). Now consider a submatrix maximum query. If the range of rows starts inside some \( M_i \) and ends inside another \( M_{i'} \), the query can be decomposed into two queries fully within \( M_i \) and \( M_{i'} \), respectively, and an infix corresponding to a range of rows of \( M' \). The maximum in the infix can be computed in \( O(\log \log n) \) time using the structure constructed for \( M' \). Consequently, it is enough to show how to answer a query in \( O(\log \log n) \) time when the range of rows is fully within a single \( M_i \). In such a case, if the range of rows starts inside some \( M_{i,j} \) and ends inside another \( M_{i,j'} \), the query can be decomposed into a prefix fully within \( M_{i,j} \), an infix corresponding to a range of rows in \( M'_i \) and a suffix fully within some \( M_{i,j'} \). As we explained above, even though we cannot locate the maximum in the infix exactly, we can isolate at most 2 rows (plus a single entry) of \( M'_i \), such that the maximum lies in one of these rows. Each row of \( M'_i \) corresponds to a range of rows fully inside some \( M_{i,j} \). Consequently, we reduced the query in \( O(\log \log n) \) time to a constant number of queries such that the range of rows in each query is fully within a single \( M_{i,j} \). Since each \( M_{i,j} \) consists of \( O(\log \log n) \) rows of \( M \), we have identified, in \( O(\log \log n) \) time, a set of \( O(\log \log n) \) rows of \( M \) that contain the desired submatrix maximum.

Now we repeat the same procedure on the transpose of \( M \) to identify a set of \( O(\log \log n) \) columns of \( M \) that contain the desired submatrix maximum. Since a submatrix of a Monge matrix is also Monge, the submatrix of \( M \) corresponding to these sets of candidate rows and columns is an \( O(\log \log n) \times O(\log \log n) \) Monge matrix. By running the SMAWK algorithm [2] in \( O(\log \log n) \) time on this small Monge matrix, we can finally determine the answer.

\[ \square \]

4 Lower Bound

A predecessor structure stores a set of \( n \) integers \( S \subseteq [0, U) \), so that given \( x \) we can determine the largest \( y \in S \) such that \( y \leq x \). As shown by Pătraşcu and Thorup [22], for \( U = n^2 \) any predecessor structure consisting of \( O(n \text{ polylog}(n)) \) words needs \( \Omega(\log \log n) \) time to answer queries, assuming that the word size is \( \Theta(\log n) \). We will use their result to prove that our structure is in fact optimal.

Given a set of \( n \) integers \( S \subseteq [0, n^2) \) we want to construct \( n \times n \) Monge matrix \( M \) such that the predecessor of any \( x \in S \) can be found using one submatrix minimum query on \( M \) and \( O(1) \) additional time (to decide which query to ask and then return the final answer). Then, assuming that for any \( n \times n \) Monge matrix there exists a data structure of size \( O(n \text{ polylog}(n)) \) answering submatrix minimum queries in \( o(\log \log n) \) time, we can construct a predecessor structure of size \( O(n \text{ polylog}(n)) \) answering queries in \( o(\log \log n) \) time, which is not possible. The technical difficulty here is twofolds. First, \( M \) should be Monge. Second, we are working in the indexing model, i.e., the data structure for submatrix minimum queries can access the matrix. Therefore, for the lower bound to carry over, \( M \) should have the following property: there is a data structure of size \( O(n \text{ polylog}(n)) \) which retrieves any \( M[i,j] \) in \( O(1) \) time. Guaranteeing that both properties hold simultaneously is not trivial.

Before we proceed, let us comment on the condition \( S \subseteq [0, n^2) \). While quadratic universe is enough to invoke the \( \Omega(\log \log n) \) lower bound for structures of size \( O(n \text{ polylog}(n)) \), our reduction actually implies that even for larger polynomially bounded universes, i.e., \( S \subseteq [0, n^c) \), for any fixed \( c \), it is possible to construct \( n \times n \) Monge matrix \( M \) such that the predecessor of \( x \) in \( S \) can be found with \( O(1) \) submatrix minimum queries on \( M \) and \( O(1) \) additional time (and, as previously, any \( M[i,j] \) can be retrieved in \( O(1) \) time with a structure of size \( O(n) \)). This is a consequence of the following lemma.
Lemma 4. For any $c$, predecessor queries on a set of $n$ integers $S \subseteq [0,n^c]$ can be reduced in $O(1)$ time to $O(1)$ predecessor queries on a set of $n$ integers $S' \subseteq [0,n^2)$ with a structure of size $O(n)$.

Proof. We explain the reduction for $c = 4$. Larger $c$ are processed similarly.

Let $S = \{x_1, x_2, \ldots, x_n\}$. We represent every $x_i$ in base $n^2$ as $x_i = y_i \cdot n^2 + z_i$, where $y_i, z_i \in [0,n^2)$. We create a new set $S' \subseteq [0, n^2)$ storing all $y_i$s and $z_i$s. Let $\text{rank}(x)$ denote the rank of $x$ in $S$. We create another set $S'' \subseteq [0, n^2)$ storing elements of the form $\text{rank}(y_i) \cdot n + \text{rank}(z_i)$. It can be seen that finding the predecessor of $x$ in $S$ can be solved by first representing $x = y \cdot n^2 + z$, computing $\text{rank}(y)$ and $\text{rank}(z)$, and finally locating the predecessor of $\text{rank}(y) \cdot n + \text{rank}(z)$ in $S''$. Consequently, a predecessor query on $S$ can be reduced into two predecessor queries on $S'$ and $S''$, respectively. $S'$ and $S''$ can be combined into a single set $S''' \subseteq [0, 2n^2)$, such that predecessor queries in either of them can be answered with predecessor queries on $S'''$, by simply shifting every element of $S''$ by $n^2$. Finally, the size of $S'''$, which is up to $3n$ right now, can be reduced to $n$ by storing every third element. Knowing the predecessor of $x$ among these chosen elements allows us to find the true predecessor in $O(1)$ time by inspecting at most three elements stored explicitly for every element of the reduced $S'''$.

The following propositions are easy to verify:

Proposition 1. A matrix $M$ is Monge iff $M[i, j] + M[i + 1, j + 1] \leq M[i + 1, j] + M[i, j + 1]$ for all $i, j$ such that all these entries are defined.

Proposition 2. If a matrix $M$ is Monge, then for any vector $H$ the matrix $M'$, where $M'[i, j] = M[i, j] + H[j]$ for all $i, j$, is also Monge.

Theorem 5. For any set of $n$ integers $S \subseteq [0, n^2)$, there exists a data structure of size $O(n)$ returning any $M[i, j]$ in $O(1)$ time, where $M$ is a Monge matrix such that the predecessor of $x$ can be found using $O(1)$ time and one submatrix minimum query on $M$.

Proof. We partition the universe $[0, n^2)$ into $n$ parts $[0, n), [n, 2n), \ldots$. The $i$-th part $[i \cdot n, (i + 1) \cdot n)$ defines a Monge matrix $M_i$ consisting of $|S \cap [i \cdot n, (i + 1) \cdot n)|$ rows and $n$ columns. The idea is to encode the predecessor of $x \in [0, n^2)$ by the minimum element in the $(x \mod n + 1)$-th column of $M_{x/n}$. We first describe how these matrices are defined, and then show how to stack them together.

Consider any $0 \leq i < n$. Every element in $S \cap [i \cdot n, (i + 1) \cdot n) = \{a_1, a_2, \ldots, a_k\}$ has a unique corresponding row in $M_i$. Let $a_j = i \cdot n + a'_j$, so that $a'_1 < a'_2 < \ldots < a'_k$ and $a'_j \in [0, n)$ for all $j$, and also define $a'_{j+1} = n$. We describe an incremental construction of $M_i$. For technical reasons, we start with an artificial top row containing $1, 2, 3, \ldots, n$. Then we add the rows corresponding to $a'_1, a'_2, \ldots, a'_k$. The row corresponding to $a'_j$ consists of three parts. The middle part starts at the $(a'_j + 1)$-th column, ends at the $a'_{j+1}$-th column, and contains only 1’s. The elements in the left part decrease by 1 and end with 2 at the $a'_j$-th column, similarly the elements in the right part (if any) start with 2 at the $(a'_{j+1} + 1)$-th column and increase by 1. Formally, the $k$-th element of the $(j + 1)$-th row, denoted $M_i[j + 1, k]$, is defined as follows.

$$
M_i[j + 1, k] = \begin{cases} 
  a'_j - k + 2 & \text{if } k \in [1, a'_j] \\
  1 & \text{if } k \in [a'_j + 1, a'_{j+1}] \\
  k - a'_{j+1} + 1 & \text{if } k \in [a'_{j+1} + 1, n] 
\end{cases}
$$

Finally, we end with an artificial bottom row containing $n, n-1, \ldots, 1$. See the upper part of Figure 1 for an example. We need to argue that every $M_i$ is Monge. By Proposition 1, it is enough to consider every pair of adjacent rows $r_1, r_2$ there. Define $r'_1[j] = r_1[j] - r_1[j - 1]$ and similarly $r'_2[j] = r_2[j] - r_2[j - 1]$. To prove that $M_i$ is Monge, it is enough to argue that $r'_2[j] \geq r'_1[j]$ for all $j \geq 2$. By construction, both $r'_1$ and $r'_2$ are of the form $-1, -1, \ldots, -1, 0, 0, \ldots, 0, 1, 1, \ldots, 1$, and all 0’s in $r'_2$ are on the right of all 0’s in $r'_1$. Therefore, $M_i$ is Monge.

Now one can observe that the predecessor of $x \in [0, n^2)$ can be found by looking at the $(x \mod n + 1)$-th column of $M_{x/n}$. We check if $x < a_1$, and if so return the predecessor of $a_1$ in the whole $S$. This can be done in $O(1)$ time and $O(n)$ additional space by explicitly storing $a_1$ and its predecessor for every $i$.  

8
Initially, we set a subcolumn minimum query. requires only storing only need to find then we can return subcolumn minimum queries on the final all elements in the same column of \( M_j \) is also Reduction for \( 1 \leq j,k \leq M \), we just directly use Equation 1, which requires only storing \( a_1',a_2',\ldots,a_n \) in \( O(n) \) total space.

We cannot simply concatenate all \( M_i \)'s to form a larger Monge matrix. We use Proposition 2 instead. Initially, we set \( M = M_0 \). Then we consider every other \( M_i \) one-by-one maintaining invariant that the current \( M \) is Monge and its last row is \( n,n-1,\ldots,1 \). In every step we add the vector \( H = [-n+1,-n+3,\ldots,n-1] \) to the current matrix \( M_i \), obtaining a matrix \( M' \) whose last row is \( 1,2,\ldots,n \). By Proposition 2, \( M' \) is Monge. Then we can construct the new \( M \) by appending \( M_i \) without its first row to \( M' \). Because the first row of \( M_i \) is also \( 1,2,\ldots,n \), the new \( M \) is also Monge. Furthermore, because we add the same value to all elements in the same column of \( M_i \), answering subcolumn minimum queries on \( M_i \) can be done with subcolumn minimum queries on the final \( M \). The lower part of Figure 1 depicts the final Monge matrix \( M \).

We need to argue that elements of \( M \) can be accessed in \( O(1) \) using a data structure of size \( O(1) \). To retrieve \( M[j,k] \), first we lookup in \( O(1) \) time the appropriate \( M_{i} \) from which it originates. This can be preprocessed and stored for every \( j \) in \( O(n) \) total space and allows us to reduce the question to retrieving \( M_i[j',k] \). Because Proposition 2 is applied exactly \( n-1-i \) times after appending \( M_i \) to the current \( M \), then we can return \( M_i[j',k] + (n-1-i)H[k] \). To find \( M_i[j',k] \), we just directly use Equation 1, which requires only storing \( a_1',a_2',\ldots,a_n \) in \( O(n) \) total space.

\[
M_0 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 3 & 2 & 1 & 1 & 1 & 1 & 1 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}
\]

\[
M_1 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 3 & 2 & 1 & 1 & 1 & 1 & 1 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}
\]

\[
M_2 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 3 & 2 & 1 & 1 & 1 & 1 & 1 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}
\]

\[
M_3 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 3 & 2 & 1 & 1 & 1 & 1 & 1 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}
\]

\[
M_4 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 3 & 2 & 1 & 1 & 1 & 1 & 1 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}
\]

\[
M_5 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 3 & 2 & 1 & 1 & 1 & 1 & 1 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}
\]

\[
M_6 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 3 & 2 & 1 & 1 & 1 & 1 & 1 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}
\]

\[
M_7 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 3 & 2 & 1 & 1 & 1 & 1 & 1 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}
\]

\[
\begin{bmatrix} 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ -48 & -33 & -18 & -3 & 12 & 27 & 42 & 57 \\ -45 & -32 & -19 & -6 & 8 & 22 & 36 & 50 \\ -41 & -28 & -15 & -2 & 11 & 24 & 37 & 50 \\ -34 & -23 & -12 & -1 & 10 & 21 & 32 & 43 \\ -32 & -23 & -14 & -4 & 6 & 17 & 28 & 39 \\ -29 & -20 & -11 & -2 & 7 & 16 & 27 & 38 \\ -28 & -19 & -10 & -1 & 8 & 17 & 26 & 36 \\ -27 & -18 & -9 & 0 & 9 & 18 & 27 & 36 \\ -20 & -13 & -6 & 1 & 8 & 15 & 22 & 29 \\ -13 & -8 & -3 & 2 & 7 & 12 & 17 & 22 \\ -11 & -8 & -5 & -1 & 3 & 7 & 12 & 17 \\ -7 & -4 & -1 & 2 & 5 & 8 & 11 & 15 \\ -6 & -3 & 0 & 3 & 6 & 9 & 12 & 15 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 1 & 1 & 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 & 1 & 1 & 1 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}
\]

Fig. 1. Reduction for \( n = 8 \) and \( S = \{ 8 \cdot 0 + 3, 8 \cdot 2 + 2, 8 \cdot 2 + 5, 8 \cdot 2 + 6, 8 \cdot 5 + 2, 8 \cdot 5 + 6, 8 \cdot 7 + 1, 8 \cdot 7 + 4 \} \).
5 Data structure for partial Monge matrices

Our goal in this section is to extend the solution described in Section 3 to partial Monge matrices. Recall that in a partial Monge matrix $M$, for any $i < j$ and $k < \ell$, the condition $M[i, k] + M[j, \ell] \geq M[i, \ell] + M[j, k]$ holds only if all $M[i, k], M[j, \ell], M[i, \ell], M[j, k]$ are defined. Not all entries in $M$ are defined, but the defined entries in every row and every column are contiguous. Let $s_i$ and $t_i$ denote the first and last columns containing defined entries in the $i$'th row respectively. We assume that we know the coordinates of at least one of the defined entries. This allows us to find all $s_i$'s and $t_i$'s in $O(n \log n)$ time.

We begin with noting that subcolumn (and subrow) maximum queries can be implemented on partial Monge matrices in the same bounds as full Monge matrices (i.e., the bounds of Theorem 3 and Lemma 3 also apply to partial Monge matrices). This is because we can implicitly fill appropriate constants instead of the undefined entries to turn a partial Monge matrix into a full Monge matrix (see [16] for details). Upon subcolumn query (a column $c$ and a range of rows $R$) we first restrict $R$ to the defined entries in the column $c$ and only then query the data structure. For submatrix queries however, this trick only works if the query range is entirely defined. In general, it does not work because the defined entries in the query range do not necessarily form a submatrix. Handling submatrix queries is therefore more complicated. We describe our solution next.

We will need the following preliminary lemma, that follows quite easily from the persistent predecessor structure of Chan [9].

Lemma 5. A collection $S$ of $O(n)$ weighted points on an $n \times n$ grid can be preprocessed in $O(n \log \log n)$ time and $O(n)$ space, so that, given any $(x, y)$, the maximum weight of a point $(x', y') \in S$ such that $x' \geq x$ and $y' \geq y$ can be calculated in $O(\log \log n)$ time.

Proof. We use the standard geometric idea of sweeping the grid with a horizontal line while maintaining a data structure describing the current situation. The data structure is made partially persistent, so that after sweeping given a query $(x, y)$ we can retrieve the version of the structure corresponding to a horizontal line passing through $(x, y)$. Querying that version of the data structure will allow us to answer the request. The data structure will be a predecessor structure made persistent using the result of Chan [9]. See Theorem 5 of [20] for a more detailed description of a similar lemma.

Denote the points by $(x_i, y_i)$ and their corresponding weights by $w_i$. We assume that the weight are distinct. We sweep the grid with a horizontal line starting at $y = n$. The predecessor structure stores $x$-coordinates of some of the already seen points. $x_i$ is stored in the predecessor structure iff $y_i \geq y$ and there is no $i'$ such that $y_{i'} \geq y_i$, $x_{i'} \geq x_i$ and $w_{i'} > w_i$. This is because otherwise the $i'$-th point is a better answer than the $i$-th point for any query processed using this or any future version of the data structure. Consequently, the points whose $x$-coordinates are stored in the predecessor structure can be arranged so that their $x$-coordinates are increasing and the weights decreasing. Then it follows that locating the maximum weight of a point $(x', y') \in S$ such that $x' \geq x$ and $y' \geq y$ can be done by finding the successor of $x$ in the version of the predecessor structure corresponding to $y$. Maintaining the structure while sweeping the grid is also done with a predecessor search. After having seen a new point $(x_i, y_i)$ we locate the predecessor of $x_i$. If the weight of the corresponding point is smaller than $w_i$, we remove it from the structure and repeat.

A persistent predecessor search structure can be implemented in space $O(n)$ while keeping the query time $O(\log \log n)$ [9]. Consequently, we can build in $O(n \log \log n)$ time a structure of size $O(n)$ answering queries in $O(\log \log n)$ time.

Before we handle arbitrary partial Monge matrix we describe a solution to a restricted type of partial Monge matrix called a staircase matrix. In a staircase matrix, the defined entries in every row either all start in the first column, or all end in the last column. We begin with a weaker result, which is that one can answer submatrix maximum queries on an $n \times n$ staircase matrix in $O(\log \log n)$ time with a structure of size $O(n \log n)$. We will then show how to reduce the space to $O(n)$, and finally how to use the solution for staircase matrices in order to handle arbitrary partial Monge matrices.

Theorem 6. Given an $n \times n$ staircase Monge matrix $M$, a data structure of size $O(n \log n)$ can be constructed in $O(n \log n)$ time to answer submatrix maximum queries in $O(\log \log n)$ time.
Proof. Because of left-right symmetry, we can assume that the defined entries in row \( i \) start in the first column and end in column \( t_i \). Notice that either \( t_1 \leq t_2 \leq \ldots \leq t_n \) or \( t_1 \geq t_2 \geq \ldots \geq t_n \). Without loss of generality we will assume the latter. This is enough because we will not be explicitly using the Monge property in our solution, except for applying Theorem 2 on a copy of \( M \) (called \( \tilde{M} \)) where the undefined entries are appropriately filled.

We partition \( M \) into full Monge matrices using a standard method: First, create a full Monge matrix by taking the upper-left fragment \([1, n/2] \times [1, t_{n/2}]\) of \( M \). Then, recursively decompose the staircase matrices created by taking the upper-right fragment \([1, n/2] \times [t_{n/2}+1, n]\) and the lower-left fragment \([n/2+1, n] \times [1, n]\) of \( M \). See Figure 2. It is easy to verify that the decomposition consists of at most \( 2n \) full Monge matrices (called fragments). The decomposition has other useful properties on which we elaborate further.

Consider a query range \([i_0, i_1] \times [j_0, j_1]\). To find the maximum (defined) \( M[i,j] \) over all \( i \in [i_0, i_1] \) and \( j \in [j_0, j_1] \) we proceed as follows. The simple case is when the query range is fully within the defined part of \( M \). To handle this case, we apply Theorem 2 on a copy of \( M \) (denoted \( \tilde{M} \)) where the undefined entries are appropriately (and implicitly) filled. This allows us to do submatrix queries in \( O(\log \log n) \) time when the query range is fully defined. Otherwise, we decompose the query into three parts. The first part, which we call a dominance maximum query, is to find the maximum \( M[i,j] \) over all \( i \geq i' \) and \( j \geq j' \), for \( i', j' \) to be defined shortly. The other two are submatrix maximum queries fully within the defined part of \( M \) (and hence can be processed by querying the structure built for \( \tilde{M} \) in \( O(\log \log n) \) time). The decomposition is performed in \( O(1) \) time by setting \( j' = t_{i_1} + 1 \) and choosing the smallest \( i' \geq i_0 \) such that \( t_{i'} < j_1 \) (which can be preprocessed for every possible \( j_1 \) in \( O(n) \) space). The two submatrix maximum queries are therefore over the full Monge matrices \([i_0, i'] \times [j_0, j'-1] \) and \([i_0, i'-1] \times [j', j_1] \). Hence, it is enough to focus on answering dominance maximum queries.

To answer a dominance maximum query (i.e., to find the maximum \( M[i,j] \) over all \( i \geq i' \) and \( j \geq j' \)) we use the partition of \( M \) into full Monge matrices (fragments). Every such fragment is either fully inside the query range, fully outside of the query range, or intersected by the query range boundary.

**Fragments inside the query range.** A fragment \([r_0, r_1] \times [c_0, c_1]\) is fully inside the query range iff \( r_0 \geq i' \) and \( c_0 \geq j' \). This observation allows us to reduce computing the maximum over all matrices fully inside the query to the problem defined in Lemma 5. The reduction is simply that for every fragment \([r_0, r_1] \times [c_0, c_1]\) we create a point \((r_0, c_0)\) and set its weight to be the maximum inside the fragment. As a result, we create

![Figure 2](image-url)

(a) A staircase \( n \times n \) Monge matrix partitioned into \( 2n \) smaller full Monge matrices (fragments). (b) A query range \([i_0, i_1] \times [j_0, j_1]\) decomposed into two full Monge matrices \( A \) and \( B \) and one dominance query \( C \). (c) The dominance query as vertical and horizontal lines (the green fragment is fully inside the range and the blue and red fragment intersect the horizontal line).
at most $O(n)$ points on the $n \times n$ grid. Using Theorem 2 on $\tilde{M}$ to create every point separately takes total $O(n \log \log n)$ in the preprocessing time, so in $O(n \log \log n)$ time we can construct a structure of size $O(n)$ answering queries in $O(\log \log n)$ time.

**Fragments intersected by the query range.** We are left only with finding the maximum over all fragments intersected by the boundary of our dominance maximum query. We partition these fragments into three groups. The first consists of the single fragment containing $M[i', j']$. The maximum there can be found with a submatrix maximum query on $M$ in $O(\log \log n)$ time. All other fragments intersected by the boundary are either intersected by the horizontal line $y = i'$ or the vertical line $x = j'$, but not both. We show how to find the maximum over all matrices intersected by the horizontal line $y = i'$ and fully to the right of the vertical line $x = j'$ (the other case is symmetric).

By the properties of our decomposition scheme, there are at most $\log n$ fragments intersected by any horizontal line, and they can be arranged in the natural left-to-right order. For every possible horizontal line, we store these at most $\log n$ fragments in an array. For every fragment we store the coordinates of its corresponding submatrix of $M$ and the maximum in all of its entries below the horizontal line. The array is additionally equipped with the maximum over all maxima in each one of its suffixes. Such preprocessed data allows us to find the maximum over all fragments intersected by a horizontal line $y = i'$ and fully on the right of a vertical line $x = j'$ in $O(\log \log n)$ time: First, we binary search over the array stored for $y = i'$ to locate the leftmost fragment completely on the right of $x = j'$. Then we return the stored corresponding maximum. Notice that the binary search also allows us to locate the fragment containing $M[i', j']$. Consequently, the whole query time is $O(\log \log n)$ using $O(n \log n)$ space for this part of the implementation. To guarantee $O(n \log n)$ preprocessing time, we run the SMAWK algorithm on every fragment in the decomposition in total $O(n \log n)$ time. This gives us the maximum in every row of every fragment. This is then enough to construct all arrays in $O(n \log n)$ time.

We now proceed to improving Theorem 6 so that the structure needs just linear space. The main idea is to partition the $n \times n$ staircase matrix $M$ into cells of size $\log n \times \log n$ and then define a new smaller $(n/\log n) \times (n/\log n)$ staircase matrix $M'$ (whose entries correspond to cell-maxima in $M$) on which we apply Theorem 6.

To implement this idea we need a number of additional auxiliary data structures, which take $O(n)$ space in total. We start with an auxiliary lemma, which will be used to provide constant-time access to entries of $M'$.

**Lemma 6.** Given an $n \times n$ Monge matrix $M$ partitioned into $\log n \times \log n$ cells, a data structure of size $O(n)$ can be constructed in $O(n \log n)$ time to find the maximum in a given cell in $O(1)$ time.

**Proof.** We partition $M$ into $n/\log n$ horizontal slices, each consisting of $\log n$ rows (and all columns). Consider a single slice, which is a $\log n \times n$ Monge matrix. We store its breakpoints $c_1 < c_2 < \ldots < c_k$ (where $k \leq \log n$) in an atomic heap, consequently allowing predecessor queries in $O(1)$ time (this is exactly how the structure from Lemma 3 works). Additionally, similarly to Lemma 2, for every $i \geq 2$ we precompute the value of

$$m_i = \max_{j \in [c_{i-1}, c_i]} M[r(c_{i-1}), j]$$

and augment these values with a (one dimensional) range maximum data structure. Here, $r(c)$ denotes the row containing the maximum element in the $c$-th column of the slice in question. Using two predecessor queries and one range maximum query, the problem of finding the maximum in a given cell (which is fully contained in a single horizontal slice) reduces in $O(1)$ time to finding the maximum in at most two rows. The total space is $O(n/\log n \cdot \log n) = O(n)$ and the bottleneck in the preprocessing is computing the breakpoints for all slices. The breakpoints of a single slice can be computed in $O(\log^2 n)$ by adding one row at a time, as done in the proof of Lemma 1. In total, this takes $O(n/\log n \cdot \log^2 n) = O(n \log n)$ total time.

We repeat the above reasoning on the transpose of $M$. As a result, we either already know the maximum element, or we have isolated at most two rows and at most two columns, such that the maximum lies in one of these rows and one of these columns. This gives us at most four candidates for the maximum, which can be retrieved and compared naively.

□
We are now ready to present our linear-space improvement to Theorem 6.

**Theorem 7.** Given an \( n \times n \) staircase Monge matrix \( M \), a data structure of size \( O(n) \) can be constructed in \( O(n \log n) \) time to answer submatrix maximum queries in \( O(\log \log n) \) time.

**Proof.** As in the proof of Theorem 6, we can assume that the defined entries in row \( i \) start in the first column and end in column \( t_i \), and that \( t_1 \geq t_2 \geq \ldots \geq t_n \).

We partition \( M \) into cells of size \( \log n \times \log n \) and then define a smaller \( (n/\log n) \times (n/\log n) \) staircase matrix \( M' \). Notice that, unlike Lemma 6, \( M \) is a staircase Monge matrix (not a full Monge matrix). This means that there are three types of cells in \( M \): fully defined, partially defined, and fully undefined. An entry of \( M' \) is defined iff its corresponding cell in \( M \) is fully defined. In this case the entry is equal to the maximum in the corresponding cell. The undefined entries of \( M' \) are the ones corresponding to either partially defined or fully undefined cells of \( M \). We appropriately (and implicitly) fill these entries to turn \( M' \) into a full Monge matrix, on which we apply Lemma 6. This gives us constant-time access to the entries of \( M' \), so finally we can apply Theorem 6 to preprocess it in \( O(n) \) space and \( O(n \log n) \) time to answer submatrix maximum queries in \( O(\log \log n) \) time.

Regarding partially defined cells, we observe that there are at most \( 2n/\log n \) of them. Furthermore, they can be arranged in a linear order, so that if the part of \( M \) corresponding to the \( i \)-th partially defined cell is \( [r_i, r'_i] \times [c_i, c'_i] \), then for all \( i \) either \( r_i, r'_i = [r_{i+1}, r'_{i+1}] \) and \( c'_i + 1 = c_{i+1} \) or \( r_i = r'_{i+1} + 1 \) and \( [c_i, c'_i] = [c_{i+1}, c'_{i+1}] \) (to be more precise, we might need to declare some fully defined cells partially defined to guarantee this property). We create a predecessor structure storing all \( r_i \)'s and a separate predecessor structure storing all \( c_i \)'s. We also compute the maximum in every partially defined cell and store them in an array (arranged in the aforementioned linear order) augmented with a (one dimensional) range maximum structure. Computing the maximum in all partially defined cells is done in \( O(\log n \cdot \log n \cdot \alpha(\log n)) = O(n \cdot \alpha(\log n)) \) time using [18].

By the same reasoning given in the proof of Theorem 6, it is enough to implement dominance maximum queries on \( M \). A dominance maximum query can be decomposed into (i) a dominance maximum query on \( M' \), which can be answered in \( O(\log \log n) \) time, (ii) finding the maximum inside all partially defined cells fully within the query range, and (iii) finding the maximum inside partially defined cells intersected by the boundaries of the query range. All partially defined cells fully within the query range create a contiguous interval in the linear order. The range can be determined in \( O(\log \log n) \) using the predecessor structures storing all \( r \)'s and \( c \)'s, and then the maximum can be found in \( O(1) \) time with a (one dimensional) range maximum query. It remains to calculate the maximum inside partially defined cells intersected by the boundaries of the query range. We will describe how to process all partially defined cells intersected by the horizontal boundary. Handling the vertical boundary is symmetric.

Let the dominance maximum query be specified by \((i', j')\). We want to compute the maximum inside the query range and belonging to a partially defined cell intersected by the horizontal line \( y = i' \). All such cells create a contiguous interval in the linear order, which can be determined with two predecessor queries in \( O(\log \log n) \) time. In the same complexity, we can find the leftmost such cell \( u \) which is not fully on the left of the vertical line \( x = j' \). We decompose the original query into a dominance maximum query inside \( u \), and the remaining part. The remaining part starts at a left boundary of a partially defined cell and consists of the entries at or below \( y = i' \) in all partially defined cells to the right of \( u \). Consequently, the answer can be preprocessed for every point on a left boundary of a partially defined cell using \( O(n/\log n \cdot \log n) = O(n) \) space and \( O(n/\log n \cdot \log n \cdot \alpha(\log n)) = O(n \cdot \alpha(\log n)) \) time using [18]. The bottleneck in the preprocessing is computing the maximum in every row of every partially defined cell.

It remains to describe how to handle the dominance query in \( u \). In other words, after constructing in \( O(n \log n) \) time an \( O(n) \) size structure, we have, in \( O(\log \log n) \) time, reduced an arbitrary dominance maximum query into a dominance maximum query inside a single partially defined cell. This cell is a smaller \( \log n \times \log n \) staircase matrix, and furthermore there are at most \( 2n/\log n \) such cells. By recursing on each of these smaller staircase matrices separately, we construct in additional \( O(n/\log n \cdot \log n \cdot \log \log n) = O(n \log \log n) \) time an \( O(n/\log n \cdot \log n) \) size structure, which reduces the original dominance query, in additional \( O(\log \log \log n) \) time, into a dominance maximum query inside one of \( O(n/\log n \cdot \log n / \log \log n) = O(n/\log \log n) \) tiny \( \log n \times \log \log n \) staircase matrix (each of them
being a submatrix of the original $M$). By recursing again on every tiny staircase matrix separately, we construct in additional $O(n \log \log \log n)$ time an $O(n)$ size structure, which reduces the original arbitrary dominance query in additional $O(\log \log \log \log n)$ time into a dominance maximum query inside an $(\log \log n) \times (\log \log n)$ submatrix of $M$. Such dominance maximum query can be answered naively resulting in $O(\log \log n + (\log \log \log n)^2) = O(\log \log n)$ total query time. □

We are now ready to prove the main theorem of this section, which is that using Theorem 7 we can actually implement submatrix maximum queries on arbitrary (and not just staircase) partial Monge matrices. The idea is to partition the partial Monge matrix into staircase matrices, so that each row and each column belong to $O(1)$ staircase matrices. Such partitioning was used in [1,16]. We build the data structure of Theorem 7 on each staircase matrix in the decomposition, and build an additional data structure for queries spanning more than one staircase matrix.

**Theorem 8.** Given an $n \times n$ partial Monge matrix $M$, a data structure of size $O(n)$ can be constructed in $O(n \log n)$ time to answer submatrix maximum queries in $O(\log \log n)$ time.

**Fig. 3.** A partial Monge matrix where the defined entries are gray and the undefined white. In black is a partitioning of the matrix into smaller staircase matrices. Each row is covered by at most two staircase matrices, and each column by at most four.

**Proof.** We partition $M$ into staircase matrices as depicted in Figure 3. This partition was used in [1,16]. We define it here formally for completeness: Let the defined entries in the $i$-th row start in the $s_i$-th column and end in the $t_i$-th column (without loss of generality, every row contains at least one defined entry). The
sequence $s_i$ is first non-increasing and then non-decreasing. Similarly, the sequence $t_i$ is first non-increasing and then non-decreasing. Initially, we partition $M$ into three slices. In the first slice, $s_i$ is non-increasing and $t_i$ is non-decreasing. In the second slice, either both $s_i$ and $t_i$ are non-increasing, or both $s_i$ and $t_i$ are non-decreasing. Finally, in the third slice $s_i$ is non-decreasing and $t_i$ is non-increasing. The first and the third slices are then further partitioned into two staircase matrices each. The second slice can be broken into staircase matrices by dividing along alternating rows and columns as shown in Figure 3. It is easy to verify that, in the resulting decomposition, each row is covered by at most two staircase matrices, and each column is covered by at most four staircase matrices. Additionally, the staircase matrices contributed by the second slice can be partitioned into two collections, such that any two matrices in the same collection are row-disjoint and column-disjoint.

The data structure consist of the following components. We apply Theorem 7 on every staircase matrix in our partition. We also store additional data for both collections. By left-right symmetry, we can assume that the ranges of rows and columns of the matrices in the collection are $[r_1, r_2), [r_2, r_3), \ldots$ and $[c_1, c_2), [c_2, c_3), \ldots$, respectively. We create a predecessor structure storing all $r_i$'s and a separate predecessor structure storing all $c_i$'s. We also compute and store the maximum inside every staircase matrix in the collection (this is done in total $O(n \cdot \alpha(n))$ time using the algorithm of Klawe and Kleitman [18]), and augment these maxima with a (one dimensional) range maximum structure.

Now consider a submatrix maximum query $[i_0, i_1] \times [j_0, j_1]$. We first query the $O(1)$ structures built for the staircase matrices corresponding to the first and the third slice of $M$. Next, we consider each of the two collections separately. To find the maximum $M[i, j]$ over all $i \in [i_0, i_1]$ and $j \in [j_0, j_1]$, we use the predecessor structures to determine in $O(\log \log n)$ the following values (without loss of generality, they all exist):

1. $i_0'$ such that $i_0 \in [r_{i_0}', r_{i_0'}+1)$,
2. $i_1'$ such that $i_1 \in [r_{i_1}', r_{i_1'}+1)$,
3. $j_0'$ such that $j_0 \in [c_{j_0}', c_{j_0'}+1)$,
4. $j_1'$ such that $j_1 \in [c_{j_1}', c_{j_1'}+1)$.

We then query the structures built for the $(i_0')$-th, $(i_1')$-th, $(j_0')$-th, and $(j_1')$-th staircase matrix in the collection. Now either we have already found the maximum, or it belongs to one of the staircase matrices fully contained in the query range. Consequently, the maximum can be found in $O(1)$ time with a single (one dimensional) range maximum query.

References

1. A. Aggarwal and M. Klawe. Applications of generalized matrix searching to geometric algorithms. *Discrete Appl. Math.*, 27:3–23, 1990.
2. A. Aggarwal, M. M. Klawe, S. Moran, P. Shor, and R. Wilber. Geometric applications of a matrix-searching algorithm. *Algorithmica*, 2(1):195–208, 1987.
3. S. Alstrup, G. S. Brodal, and T. Rauhe. New data structures for orthogonal range searching. In *41st FOCS*, pages 198–207, 2000.
4. A. Amir, J. Fischer, and M. Lewenstein. Two-dimensional range minimum queries. In *18th CPM*, pages 286–294, 2007.
5. G. S. Brodal, P. Davoodi, M. Lewenstein, R. Raman, and S. S. Rao. Two dimensional range minimum queries and Fibonacci lattices. In *20th ESA*, pages 217–228, 2012.
6. G. S. Brodal, P. Davoodi, and S. S. Rao. On space efficient two dimensional range minimum data structures. In *18th ESA*, pages 171–182, 2010.
7. R. E. Burkard, B. Klinz, and R. Rudolf. Perspectives of Monge properties in optimization. *Discrete Appl. Math.*, 70:95–161, 1996.
8. T. M. Chan, K. G. Larsen, and M. Pătraşcu. Orthogonal range searching on the RAM, revisited. In *27th SOCG*, pages 354–363, 2011.
9. Timothy M. Chan. Persistent predecessor search and orthogonal point location on the word ram. *ACM Trans. Algorithms*, 9(3):22:1–22:22, 2013.
10. B. Chazelle. A functional approach to data structures and its use in multidimensional searching. *SIAM Journal on Computing*, 17:427–462, 1988.
11. B. Chazelle and B. Rosenberg. Computing partial sums in multidimensional arrays. In 5th SOCG, pages 131–139, 1989.

12. E. D. Demaine, G. M. Landau, and O. Weimann. On Cartesian trees and range minimum queries. Algorithmica, 68(3):610–625, 2014.

13. A. Farzan, J. I. Munro, and R. Raman. Succinct indices for range queries with applications to orthogonal range maxima. In 39th ICALP, pages 327–338, 2012.

14. M.L. Fredman and D.E. Willard. Trans-dichotomous algorithms for minimum spanning trees and shortest paths. J. Comput. Syst. Sci., 48(3):533–551, 1994.

15. H. Gabow, J. L. Bentley, and R.E Tarjan. Scaling and related techniques for geometry problems. In 16th STOC, pages 135–143, 1984.

16. P. Gawrychowski, S. Mozes, and O. Weimann. Improved submatrix maximum queries in Monge matrices. In 41st ICALP, pages 525–537, 2014.

17. H. Kaplan, S. Mozes, Y. Nussbaum, and M. Sharir. Submatrix maximum queries in Monge matrices and Monge partial matrices, and their applications. In 23rd SODA, pages 338–355, 2012.

18. M. M. Klawe and D J. Kleitman. An almost linear time algorithm for generalized matrix searching. SIAM Journal Discret. Math., 3(1):81–97, 1990.

19. Tsvi Kopelowitz and Moshe Lewenstein. Dynamic weighted ancestors. In 18th SODA, pages 565–574, 2007.

20. Moshe Lewenstein. Orthogonal range searching for text indexing. CoRR, abs/1306.0615, 2013.

21. Y. Nekrich. Orthogonal range searching in linear and almost-linear space. Comput. Geom., 42(4):342–351, 2009.

22. Mihai Pătraşcu and Mikkel Thorup. Time-space trade-offs for predecessor search. In 38th STOC, pages 232–240, 2006.

23. H. Yuan and M. J. Atallah. Data structures for range minimum queries in multidimensional arrays. In 21st SODA, pages 150–160, 2010.