Hankel Matrices for Weighted Visibly Pushdown Automata

Nadia Labai  Vienna University of Technology
Johann A. Makowsky  Technion - Israel Institute of Technology

Prague, LATA 2016
Nested words

- Linear order
- Matching relation

leaving calls and entering returns

\[
b \rightarrow a \rightarrow a \rightarrow b \rightarrow b
\]
Nested words

- Linear order
- Matching relation leaving \textit{calls} and entering \textit{returns}

\[
b \longrightarrow a \longrightarrow a \overset{\sim} \longrightarrow b \longrightarrow b
\]

Nested words model data with a nested hierarchical structure which can be given sequentially.
Classic examples: XML trees, execution paths.
Visibly Pushdown Automata (VPA)

The input alphabet consists of letters that cause pushs, letters that cause pops, and letters that only update the automaton's state.

\[ \Sigma = \Sigma_{\text{push}} \cup \Sigma_{\text{pop}} \cup \Sigma_{\text{int}} \]

VPA were introduced in 2004 by Alur and Madhusudan and have since seen enormous amounts of follow-up research.

N. Labai & J.A. Makowsky
Visibly Pushdown Automata (VPA)

The input alphabet consists of letters that cause pushes,
Visibly Pushdown Automata (VPA)

The input alphabet consists of letters that cause pushes, letters that cause pops,

\[ \Sigma = \Sigma_{\text{push}} \cup \Sigma_{\text{pop}} \cup \Sigma_{\text{int}} \]

VPA were introduced in 2004 by Alur and Madhusudan and have since seen enormous amounts of follow-up research.

N. Labai & J.A. Makowsky
Visibly Pushdown Automata (VPA)

The input alphabet consists of letters that cause pushes, letters that cause pops, and letters that only update the automaton’s state.
Visibly Pushdown Automata (VPA)

The input alphabet consists of letters that cause pushes, letters that cause pops, and letters that only update the automaton’s state.

\[ \Sigma = \Sigma_{push} \cup \Sigma_{pop} \cup \Sigma_{int} \]
Visibly Pushdown Automata (VPA)

The input alphabet consists of letters that cause pushes, letters that cause pops, and letters that only update the automaton’s state.

$$\Sigma = \Sigma_{push} \cup \Sigma_{pop} \cup \Sigma_{int}$$

VPA were introduced in 2004 by Alur and Madhusudan and have since seen enormous amounts of follow-up research.
A weighted VPA is VPA whose transitions are assigned weights from a numeric domain $\mathcal{F}$. 
Answering quantitative questions

A **weighted VPA** is VPA whose transitions are assigned weights from a numeric domain \( \mathcal{F} \).

- value assigned to input \( \equiv \) the sum of the weights of all possible runs
- weight of a run \( \equiv \) the product of the weights of the transitions taken
Answering quantitative questions

A weighted VPA is VPA whose transitions are assigned weights from a numeric domain $\mathcal{F}$.

value assigned to input $\quad = \quad$ the sum of the weights of all possible runs

weight of a run $\quad = \quad$ the product of the weights of the transitions taken

A function is recognizable by a weighted VPA if there is a weighted VPA which computes it.
Which functions can weighted VPA recognize?
Which functions can weighted VPA recognize?

- Mathissen gave a logical answer using weighted MSOL.
Which functions can weighted VPA recognize?

- Mathissen gave a logical answer using weighted MSOL.
- We give an *algebraic* characterization.
Similarities between VPA and word automata

- Share nice closure properties
  union, intersection, Kleene-$\star$ and more.

- Share decidable problems
  emptiness, inclusion, equivalence.

- Their weighted versions are logically characterized using similar formalisms.
  Droste and Gastin’s weighted MSOL.
Similarities between VPA and word automata

- Share nice closure properties
  union, intersection, Kleene-$*$ and more.

- Share decidable problems
  emptiness, inclusion, equivalence.

- Their weighted versions are logically characterized using similar formalisms.
  Droste and Gastin’s weighted MSOL.

Now we also have:

- Their weighted versions have similar algebraic characterizations
  via finiteness conditions on Hankel matrices.
An algebraic characterization using nested Hankel matrices

Theorem (L and Makowsky, 2016)

Let $f$ be a function of well-nested words over $\mathbb{R}$ or $\mathbb{C}$. $f$ is recognized by some weighted VPA iff the nested Hankel matrix of $f$ has finite rank.
An algebraic characterization using nested Hankel matrices

Theorem (L and Makowsky, 2016)

Let $f$ be a function of well-nested words over $\mathbb{R}$ or $\mathbb{C}$. $f$ is recognized by some weighted VPA iff the nested Hankel matrix of $f$ has finite rank.

We should define

- well-nested words
- nested Hankel matrices
Well-nested words and nested Hankel matrices

\[ \Sigma = \Sigma_{push} \cup \Sigma_{pop} \cup \Sigma_{int} \]

A word over \( \Sigma \) is well-nested if there are no unmatched pushes or unmatched pops.
Well-nested words and nested Hankel matrices

\[ \Sigma = \Sigma_{push} \cup \Sigma_{pop} \cup \Sigma_{int} \]

A word over \( \Sigma \) is well-nested if there are no unmatched pushes or unmatched pops.

The nested Hankel matrix \( nH_f \) of a function \( f \) on well-nested words is an infinite matrix where
Well-nested words and nested Hankel matrices

$$\Sigma = \Sigma_{push} \cup \Sigma_{pop} \cup \Sigma_{int}$$

A word over $\Sigma$ is well-nested if there are no unmatched pushs or unmatched pops.

The nested Hankel matrix $nH_f$ of a function $f$ on well-nested words is an infinite matrix where

- the rows and columns are indexed with well-nested words and
Well-nested words and nested Hankel matrices

\[ \Sigma = \Sigma_{\text{push}} \cup \Sigma_{\text{pop}} \cup \Sigma_{\text{int}} \]

A word over \( \Sigma \) is well-nested if there are no unmatched \texttt{pushs} or unmatched \texttt{pops}.

The nested Hankel matrix \( nH_f \) of a function \( f \) on well-nested words is an infinite matrix where

- the rows and columns are indexed with well-nested words and
- for the entry at the row labeled \( u \) and the column labeled \( v \) we have

\[ nH_f(u, v) = f(uv) \]
Well-nested words and nested Hankel matrices

\[ \Sigma = \Sigma_{\text{push}} \cup \Sigma_{\text{pop}} \cup \Sigma_{\text{int}} \]

A word over \( \Sigma \) is well-nested if there are no unmatched pushes or unmatched pops.

The nested Hankel matrix \( nH_f \) of a function \( f \) on well-nested words is an infinite matrix where

- the rows and columns are indexed with well-nested words and
- for the entry at the row labeled \( u \) and the column labeled \( v \) we have

\[ nH_f(u, v) = f(uv) \]

A nested Hankel matrix has finite rank if there is a finite set of rows which linearly span it.
Example

\[ \Sigma = \{ \langle a, a, a \rangle \} \]

\( f(w) = \) number of pairs of parentheses in \( w \)

<table>
<thead>
<tr>
<th></th>
<th>( \varepsilon )</th>
<th>( a )</th>
<th>( \langle aa \rangle )</th>
<th>( aa )</th>
<th>( \langle aaa \rangle )</th>
<th>( \ldots )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varepsilon )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( a )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \langle aa \rangle )</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( aa )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \langle aaa \rangle )</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \langle a\langle aa \rangle a \rangle )</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
</tbody>
</table>

N. Labai & J.A. Makowsky
Example

\[ \Sigma = \{ \langle a, a, a \rangle \} \]

\[ f(w) = \text{number of pairs of parentheses in } w \]

<table>
<thead>
<tr>
<th></th>
<th>( \varepsilon )</th>
<th>( a )</th>
<th>( \langle aa \rangle )</th>
<th>( aa )</th>
<th>( \langle aaa \rangle )</th>
<th>( \ldots )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varepsilon )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( a )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \langle aa \rangle )</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( aa )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \langle aaa \rangle )</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \langle a\langle aa \rangle a \rangle )</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
</tbody>
</table>

\[ f(\langle aaa \rangle \langle aa \rangle) = 2 \]
Main result, more precisely

Let \( \mathcal{F} = \mathbb{R} \) or \( \mathcal{F} = \mathbb{C} \).

**Theorem (L and Makowsky, 2016)**

Let \( f \) be an \( \mathcal{F} \)-valued function on well-nested words. Then \( f \) is recognized by a weighted VPA with \( n \) states iff the nested Hankel matrix \( nH_f \) has rank \( \leq n^2 \).

**Theorem (Carlyle and Paz, 1971)**

Let \( f \) be an \( \mathcal{F} \)-valued function on words. Then \( f \) is recognized by a weighted automaton with \( n \) states iff the Hankel matrix \( H_f \) has rank \( n \).
Main result, more precisely

Let $\mathcal{F} = \mathbb{R}$ or $\mathcal{F} = \mathbb{C}$.

**Theorem (L and Makowsky, 2016)**

Let $f$ be an $\mathcal{F}$-valued function on well-nested words. Then $f$ is recognized by a weighted VPA with $n$ states iff the nested Hankel matrix $nH_f$ has rank $\leq n^2$.

**Theorem (Carlyle and Paz, 1971)**

Let $f$ be an $\mathcal{F}$-valued function on words. Then $f$ is recognized by a weighted automaton with $n$ states iff the Hankel matrix $H_f$ has rank $n$. 
Main result, more precisely

Let $\mathcal{F} = \mathbb{R}$ or $\mathcal{F} = \mathbb{C}$.

**Theorem (L and Makowsky, 2016)**

Let $f$ be an $\mathcal{F}$-valued function on well-nested words. Then $f$ is recognized by a weighted VPA with $n$ states iff the nested Hankel matrix $nH_f$ has rank $\leq n^2$.

**Theorem (Carlyle and Paz, 1971)**

Let $f$ be an $\mathcal{F}$-valued function on words. Then $f$ is recognized by a weighted automaton with $n$ states iff the Hankel matrix $H_f$ has rank $n$.

Why only for $\mathbb{R}$ and $\mathbb{C}$?
Let us look at the computations:

\[
f(w_{i,j}) = \alpha^T \cdot M_{w_{i,j}} \cdot \eta
\]

\[
f(a) = \alpha^T \cdot M_o \cdot \eta
\]

\[
f(cw_{i,j}) = \alpha^T \left( \sum_{\gamma \in \Gamma} M_{\text{call}}(\gamma) \cdot M_{w_{i,j}} \cdot M^{(r,\gamma)} \right) \eta
\]

**roof.** Consider the matrix \(N \in \mathcal{F}^{n \times n}\) defined as \(N(i,j) = f(w_{i,j}).\) By Theorem 4, since \(N\) has rank 1, there exist vectors \(x, y \in \mathcal{F}^n\) such that \(N = xy^T.\)

\[\text{vec}(\eta) = y \quad \text{and} \quad \alpha = x, \quad \text{and} \quad M_{w_{i,j}} = \beta_{i,j} \cdot A^{(i,j)}, \quad \text{where} \quad \beta_{i,j} = f(w_{i,j})f(w_{i,j})^{-1}\]

Note that for \(w_{i,j}\), we have \(\beta_{i,j} = 1\) and \(M_{w_{i,j}} = A^{(i,j)}.\)

We need to show that \(\alpha^T \cdot M_{w_{i,j}} \cdot \eta = f(w_{i,j})\) for \(w_{i,j} \in \mathcal{B}\). Since the entries \(M_{w_{i,j}}\) are zero except for entry \((i,j)\), we have

\[
\alpha^T \cdot M_{w_{i,j}} \cdot \eta = \alpha(i) \cdot \beta_{i,j} \cdot \eta(j)
\]

Hence \(\alpha(i) \eta(j) = f(w_{i,j}),\) we have

\[
\alpha^T \cdot M_{w_{i,j}} \cdot \eta = \beta_{i,j} \cdot f(w_{i,j}) = f(w_{i,j}) \cdot f(w_{i,j})^{-1} \cdot f(w_{i,j}) = f(w_{i,j})
\]

Equation 1 holds.

Let \(r_a\) denote the row in \(nH_1\) corresponding to some letter \(a \in \Sigma_m.\) Assume the matrix, so there is a linear combination

\[
r_a = \sum_{1 \leq i,j \leq n} z(a)_{i,j} \cdot r_{w_{i,j}}
\]

In particular, \(f(a) = r_a(\varepsilon) = \sum_{1 \leq i,j \leq n} z(a)_{i,j} f(w_{i,j}).\)

\[
M_a = \sum_{1 \leq i,j \leq n} z(a)_{i,j} M_{w_{i,j}}
\]

\[\text{we need to show that } f(a) = \alpha^T \cdot M_a \cdot \eta. \text{ We have}
\]

\[
\begin{align*}
\alpha^T \cdot M_a \cdot \eta &= \alpha^T \left( \sum_{1 \leq i,j \leq n} z(a)_{i,j} M_{w_{i,j}} \right) \cdot \eta = \sum_{1 \leq i,j \leq n} z(a)_{i,j} \left( \alpha^T M_{w_{i,j}} \eta \right) \\
&= \sum_{1 \leq i,j \leq n} z(a)_{i,j} f(w_{i,j}) = f(a)
\end{align*}
\]

Or, equivalently, that for every \(w \in WNW(\Sigma),\)

\[
r_{w_a}(w) = \alpha^T \cdot M^{(A)}(u) \cdot M^{(A)}(i) \cdot M^{(A)}(w) \cdot \eta
\]

Consider the linear combination:

\[
\nu_{w_a} = \sum_{1 \leq i,j \leq n} (M^{(A)}_{u_a}(i) \cdot M^{(A)}_t(i,j) \cdot \nu^{(i,j)}) = \sum_{1 \leq i,k \leq n} M^{(A)}_{u_a}(i,k) \cdot M^{(A)}_t(k,j) \cdot \nu^{(i,j)}
\]

Then, for \(w \in WNW(\Sigma)\) we have

\[
\nu_{w_a}(w) = \sum_{1 \leq i,k \leq n} M^{(A)}_{u_a}(i,k) \cdot M^{(A)}_t(i,k) \cdot \nu^{(i,j)}(w)
\]

\[= \sum_{1 \leq i,k \leq n} M^{(A)}_{u_a}(i,k) \cdot M^{(A)}_t(i,k) \cdot \left( \alpha^T \cdot A^{(i,k)}M^{(A)}_{u_a} \cdot \eta \right)
\]

Note that for \(N = A^{(i,j)}M^{(A)}_{u_a},\) the row \(i\) of \(N\) is row \(j\) of \(M^{(A)}_{u_a}\) and all other rows are zero. Then

\[
\nu_{w_a}(w) = \sum_{1 \leq i,k \leq n} M^{(A)}_{u_a}(i,k) \cdot M^{(A)}_t(i,k) \cdot \left( \sum_{l=1}^n \alpha(i) M^{(A)}_t(j,l) \cdot \nu^{(i,j)}(l) \right)
\]

\[= \sum_{1 \leq i,k \leq n} \alpha(i) \cdot M^{(A)}_{u_a}(i,k) \cdot M^{(A)}_t(k,j) \cdot M^{(A)}_t(j,l) \cdot \eta(l)
\]

\[= \alpha^T \cdot M^{(A)}_{u_a} \cdot M^{(A)}_t \cdot M^{(A)}_t \cdot \eta = r_{w_a}(w)
\]
Let us look at the computations:

\[ f(w_{i,j}) = \alpha^T \cdot M_{w_{i,j}} \cdot \eta \]

\[ f(a) = \alpha^T \cdot M_a \cdot \eta \]

\[ f(cw_{i,j}r) = \alpha^T \left( \sum_{z \in I} M_{zd}^{(c,z)} \cdot M_{w_{i,j}} \cdot M_{r_{j+1}}^{(r,z)} \right) \eta \]

\[ \alpha^T \cdot M_{w_{i,j}} \cdot \eta = \alpha^T \left( \sum_{1 \leq i,j \leq n} z(a)_{i,j} M_{w_{i,j}} \right) \cdot \eta = \sum_{1 \leq i,j \leq n} z(a)_{i,j} (\alpha^T M_{w_{i,j}} \cdot \eta) \]

\[ = \sum_{1 \leq i,j \leq n} z(a)_{i,j} f(w_{i,j}) = f(a) \]

Or, equivalently, that for every \( w \in WNW(\Sigma) \),

\[ r_{ul}(w) = \alpha^T \cdot M^{(A)}(u) \cdot M^{(A)}(i) \cdot M^{(A)}(w) \cdot \eta \]

Consider the linear combination:

\[ v_{ul} = \sum_{1 \leq i,j \leq n} (M^{(A)}_u \cdot M^{(A)}_i) (i,j) \cdot \nu^{(i,j)} = \sum_{1 \leq i,j \leq n} M^{(A)}_u (i,k) \cdot M^{(A)}_i (k,j) \cdot \nu^{(i,j)} \]

Then, for \( w \in WNW(\Sigma) \) we have

\[ v_{ul}(w) = \sum_{1 \leq i,j,k \leq n} (M^{(A)}_u (i,k) \cdot M^{(A)}_i (k,j) \cdot \nu^{(i,j)} (w)) \]

\[ = \sum_{1 \leq i,j,k \leq n} M^{(A)}_u (i,k) \cdot M^{(A)}_i (k,j) \cdot (\alpha^T \cdot A^{(i,j)} M^{(A)}_i \cdot \eta) \]

Note that for \( N = A^{(i,j)} M^{(A)}_u \), the row \( i \) of \( M^{(A)}_u \) and all other rows are zero. Then

\[ v_{ul}(w) = \sum_{1 \leq i,k,l \leq n} M^{(A)}_u (i,k) \cdot M^{(A)}_i (k,j) \cdot (\sum_{l=1}^n \alpha(i) M^{(A)}_j (j,l) \cdot \eta(l)) \]

\[ = \sum_{1 \leq i,k,l \leq n} \alpha(i) \cdot M^{(A)}_u (i,k) \cdot M^{(A)}_i (k,j) \cdot M^{(A)}_w (j,l) \cdot \eta(l) \]

\[ = \alpha^T \cdot M^{(A)}_u \cdot M^{(A)}_i \cdot M^{(A)}_w \cdot \eta = r_{ul}(w) \]

These computations only work...

if and only if the matrix \( N \) has a decomposition \( N = \sum_{k=1}^n q^{z,k} \cdot q^{r,k} \).
Let us look at the computations:

\[ f(w_{ij}) = \alpha^T \cdot M_{w_{ij}} \cdot \eta \]

\[ f(a) = \alpha^T \cdot M_a \cdot \eta \]

\[ f(cw_{ij}r) = \alpha^T \left( \sum_{g \in F} M_{c(i,j)}^{(g)} \cdot M_{w_{ij}}^{(r,g)} \right) \eta \]

We need to show that \( f(a) = \alpha^T \cdot M_a \cdot \eta \). We have

\[ \alpha^T \cdot M_a \cdot \eta = \alpha^T \left( \sum_{1 \leq i,j \leq n} z(a)_{i,j} M_{w_{ij}} \right) \cdot \eta = \sum_{1 \leq i,j \leq n} z(a)_{i,j} (\alpha^T M_{w_{ij}} \eta) \]

\[ = \sum_{1 \leq i,j \leq n} z(a)_{i,j} f(w_{ij}) = f(a) \]

Or, equivalently, that for every \( w \in WNW(\Sigma) \),

\[ r_{aw}(w) = \alpha^T \cdot M^{(i)}(w) \cdot M^{(j)}(t) \cdot M^{(A)}(w) \cdot \eta \]

Consider the linear combination:

\[ \mathbf{v}_{aw} = \sum_{1 \leq i,j \leq n} (M^{(A)}_a \cdot M^{(A)}_t(i,j) \cdot \mathbf{v}(i,j) = \sum_{1 \leq i,j \leq n} M^{(A)}_a(i,k) \cdot M^{(A)}_t(k,j) \cdot \mathbf{v}(i,j) \]

Then, for \( w \in WNW(\Sigma) \) we have

\[ \mathbf{v}_{aw}(w) = \sum_{1 \leq i,j,k \leq n} M^{(A)}_a(i,k) \cdot M^{(A)}_t(k,j) \cdot \mathbf{v}(i,j)(w) \]

\[ = \sum_{1 \leq i,j,k \leq n} M^{(A)}_a(i,k) \cdot M^{(A)}_t(k,j) \cdot (\alpha^T \cdot A^{(i,j)} M^{(A)}_a \cdot \eta) \]

Note that for \( N = A^{(i,j)} M^{(A)}_a \), the row \( i \) of \( M^{(A)}_a \) and all other rows are zero. Then

\[ \mathbf{v}_{aw}(w) = \sum_{1 \leq i,j,k \leq n} M^{(A)}_a(i,k) \cdot M^{(A)}_t(k,j) \cdot \left( \sum_{l=1}^n \alpha(i) M^{(A)}_w(j,l) \cdot \eta(l) \right) \]

\[ = \sum_{1 \leq i,j,k \leq n} \alpha(i) \cdot M^{(A)}_w(i,k) \cdot M^{(A)}_t(k,j) \cdot M^{(A)}_w(j,l) \cdot \eta(l) \]

\[ = \alpha^T \cdot M^{(A)}_w \cdot M^{(A)}_t \cdot M^{(A)}_w \cdot \eta = r_{aw}(w) \]

These computations only work...

if and only if the matrix \( N \) has a decomposition \( N = \sum_{k=1}^n q_{z,k} \cdot q_{r,k} \).

Or in other words, iff \( N \) has a singular value decomposition.
We need to make use of the following theorem:

**The Singular Value Decomposition (SVD) Theorem**

Let $N \in \mathcal{F}^{m \times n}$ be a non-zero matrix, where $\mathcal{F} = \mathbb{R}$ or $\mathcal{F} = \mathbb{C}$. Then there exist orthogonal matrices

$$X = [x_1 \ldots x_m] \in \mathcal{F}^{m \times m}, \quad Y = [y_1 \ldots y_n] \in \mathcal{F}^{n \times n}$$

such that

$$Y^T N X = \text{diag}(\sigma_1, \ldots, \sigma_p) \in \mathcal{F}^{m \times n}$$

where $p = \min\{m, n\}$, $\text{diag}(\sigma_1, \ldots, \sigma_p)$ is a diagonal matrix with the values $\sigma_1, \ldots, \sigma_p$, and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p$.

The bottom line*:

the SVD theorem only holds over $\mathbb{R}$ and $\mathbb{C}$.

* Maybe extendable to max-plus algebras, using an analogue theorem by De Schutter and De Moor.
We need to make use of the following theorem:

**The Singular Value Decomposition (SVD) Theorem**

Let $N \in \mathcal{F}^{m \times n}$ be a non-zero matrix, where $\mathcal{F} = \mathbb{R}$ or $\mathcal{F} = \mathbb{C}$. Then there exist orthogonal matrices

$$X = [x_1 \ldots x_m] \in \mathcal{F}^{m \times m}, \quad Y = [y_1 \ldots y_n] \in \mathcal{F}^{n \times n}$$

such that

$$Y^T N X = \text{diag}(\sigma_1, \ldots, \sigma_p) \in \mathcal{F}^{m \times n}$$

where $p = \min\{m, n\}$, $\text{diag}(\sigma_1, \ldots, \sigma_p)$ is a diagonal matrix with the values $\sigma_1, \ldots, \sigma_p$, and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p$.

**The bottom line**: the SVD theorem only holds over $\mathbb{R}$ and $\mathbb{C}$.

N. Labai & J.A. Makowsky
We need to make use of the following theorem:

**The Singular Value Decomposition (SVD) Theorem**

Let \( N \in \mathcal{F}^{m \times n} \) be a non-zero matrix, where \( \mathcal{F} = \mathbb{R} \) or \( \mathcal{F} = \mathbb{C} \). Then there exist orthogonal matrices

\[
X = [x_1 \ldots x_m] \in \mathcal{F}^{m \times m}, \quad Y = [y_1 \ldots y_n] \in \mathcal{F}^{n \times n}
\]

such that

\[
Y^T N X = \text{diag}(\sigma_1, \ldots, \sigma_p) \in \mathcal{F}^{m \times n}
\]

where \( p = \min\{m, n\} \), \( \text{diag}(\sigma_1, \ldots, \sigma_p) \) is a diagonal matrix with the values \( \sigma_1, \ldots, \sigma_p \), and \( \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \).

**The bottom line**: the SVD theorem only holds over \( \mathbb{R} \) and \( \mathbb{C} \).

* Maybe extendable to max-plus algebras, using an analogue theorem by De Schutter and De Moor.

N. Labai & J.A. Makowsky
Learning weighted VPA

(In Angluin’s active model of learning)
Learning model

- Target function - in our case, a weighted VPA
- Learner - keeps a conjectured weighted VPA
- Teacher - knows the target weighted VPA

learner

\[ A' \]

conjecture

teacher

\[ A \]

target

N. Labai & J.A. Makowsky
Learning model - membership queries

learner

$A'$

conjecture

$f_A(w) = ?$

teacher

$A$

target

N. Labai & J.A. Makowsky
Learning model - membership queries

learner

\( \mathcal{A}' \)

\textit{conjecture}

\[ f_\mathcal{A}(w) = ? \]

\[ f_\mathcal{A}(w) = \alpha \]

teacher

\( \mathcal{A} \)

\textit{target}

N. Labai & J.A. Makowsky
Learning model - membership queries

learner

\[ A'' \]

conjecture

teacher

\[ A \]

target
Learning model - equivalence queries

learner

$A'$

conjecture

$A' \equiv A$

teacher

$A$

target

N. Labai & J.A. Makowsky
Learning model - equivalence queries

learner

\[ A' \]

conjecture

\[ A' \cong A \]

YES

teacher

\[ A \]

target

N. Labai & J.A. Makowsky
Learning model - equivalence queries

learner

A'

conjecture

\[ A' \ ? \ A \]

\[ w \]

teacher

A

target

N. Labai & J.A. Makowsky
Learning model - equivalence queries

learner

\[ A'' \]

\textit{conjecture}

teacher

\[ A \]

\textit{target}
Learning algorithms have been developed for word and tree automata. Not only for the classical versions, but also for weighted versions. Not only in the active model of learning, but also in PAC and statistical learning.
Learning weighted VPA

Learning algorithms have been developed for word and tree automata. Not only for the classical versions, but also for weighted versions. Not only in the active model of learning, but also in PAC and statistical learning.

In contrast, we have no learning algorithms for weighted VPA.
Learning weighted VPA

Learning algorithms have been developed for word and tree automata. Not only for the classical versions, but also for weighted versions. Not only in the active model of learning, but also in PAC and statistical learning.

In contrast, we have no learning algorithms for weighted VPA.

▶ Plenty of work still to be done with traditional models?
▶ Not enough applications that provide motivation?
Possible starting point

- Many of the algorithms exploit the connection between the Hankel matrix and the weighted automaton.

We define the weighted VPA from the entries of a full-rank submatrix of $\mathcal{H}_f$. 
Many of the algorithms exploit the connection between the Hankel matrix and the weighted automaton.

We define the weighted VPA from the entries of a full-rank submatrix of $nH_f$.

In recent years, spectral methods which use SVDs of Hankel matrices have been developed.

The construction of the weighted VPA relies on the SVD of matrices defined from $nH_f$. 
Many of the algorithms exploit the connection between the Hankel matrix and the weighted automaton.

We define the weighted VPA from the entries of a full-rank submatrix of $nH_f$.

In recent years, spectral methods which use SVDs of Hankel matrices have been developed.

The construction of the weighted VPA relies on the SVD of matrices defined from $nH_f$.

Encouraging signs that an algorithm can be extracted from the proof.
Thanks!