(Co)-Algebraic and Analytical Aspects of Weighted Automata Minimisation and Equivalence (Part 2)

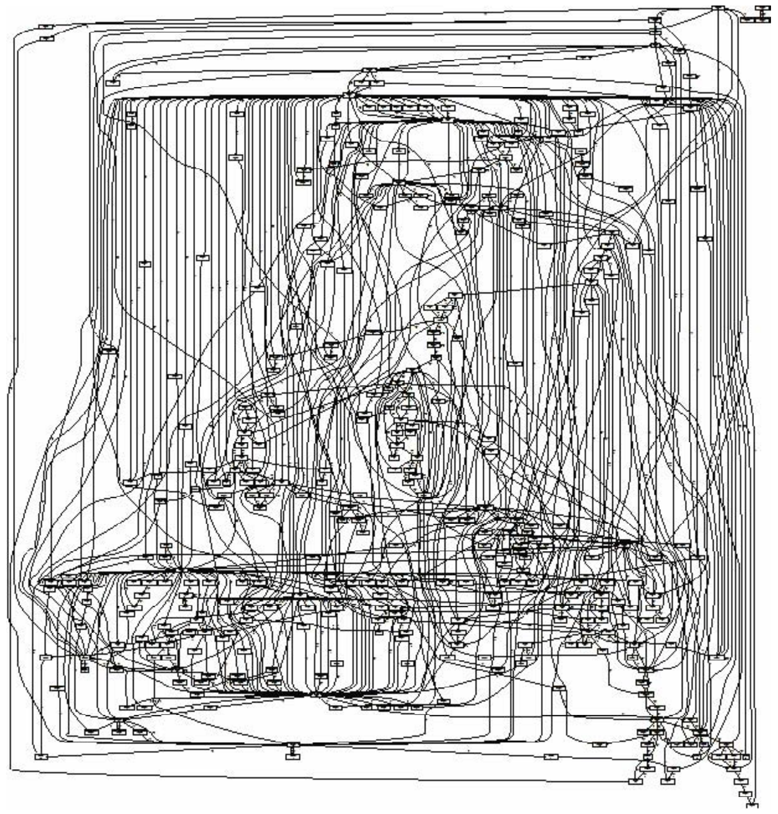
Borja Balle

Data Science | Lancaster University

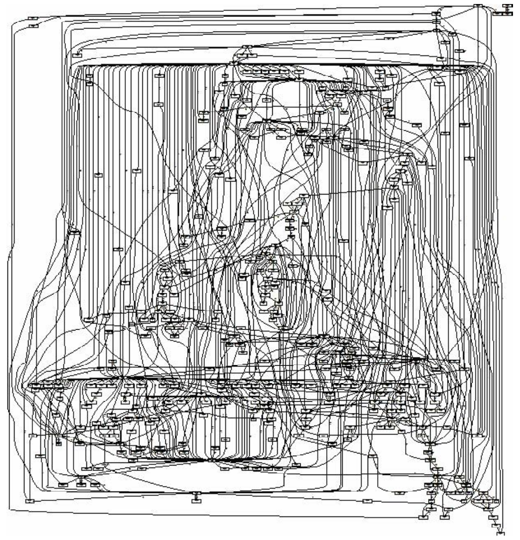
Mathematics & Statistics | Lancaster University

[CMCS Tutorial, Apr 2 2016]

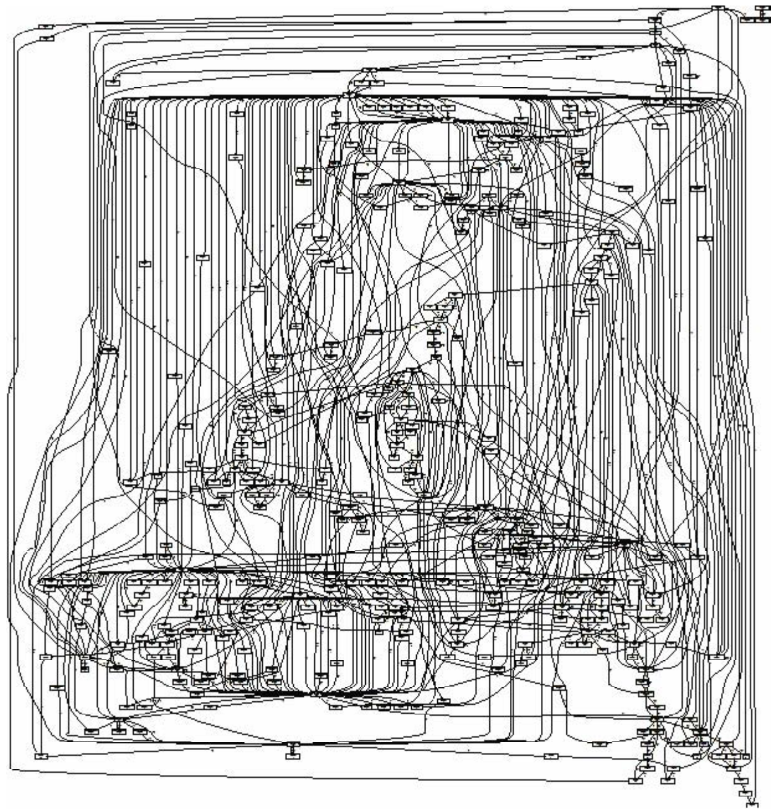
Minimisation vs Approximation



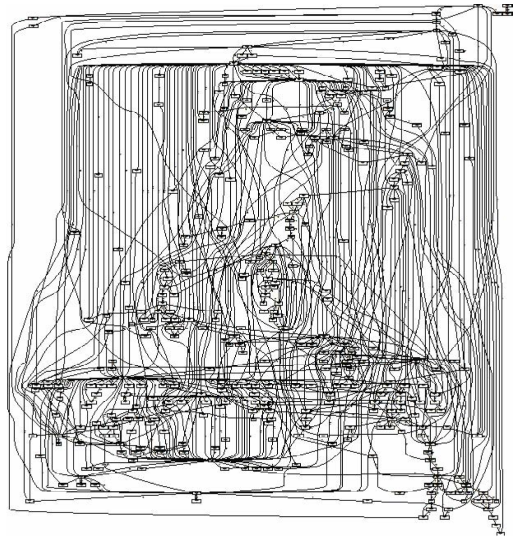
minimisation



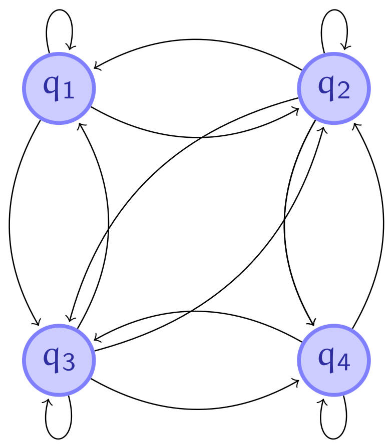
Minimisation vs Approximation



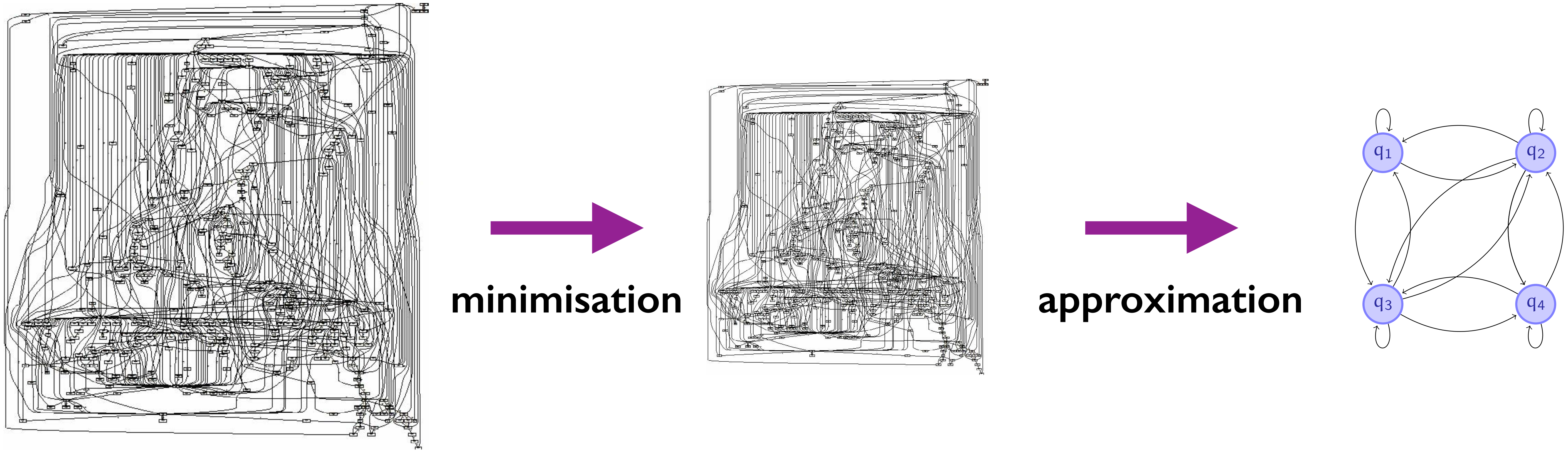
minimisation



approximation



Minimisation vs Approximation



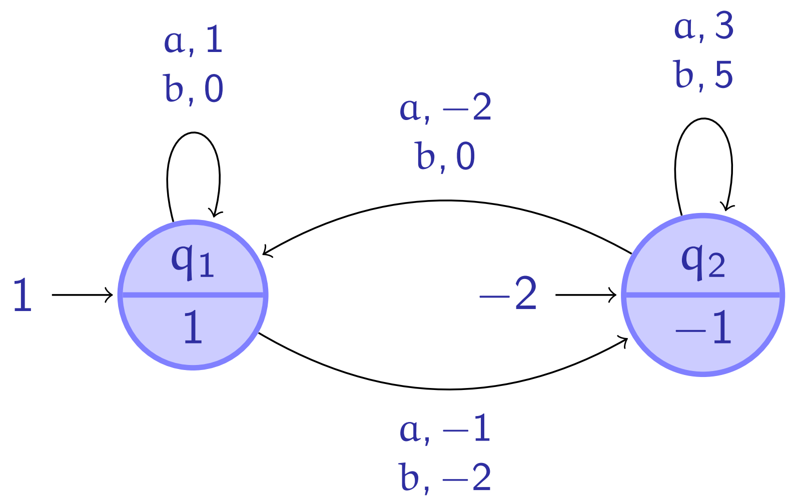
- Minimisation is naturally related to *(co-)algebraic* properties
- Approximation questions need *analytical* tools (eg. metrics)

Motivations for Approximation

1. Efficient (approximate) computation with WA
   - Compute faster and pay a controlled price in terms of accuracy
2. Machine learning for WA
   - Understand learning bias, design better algorithms
3. Interesting mathematical questions

WA and Rational Functions

$\mathcal{A}$

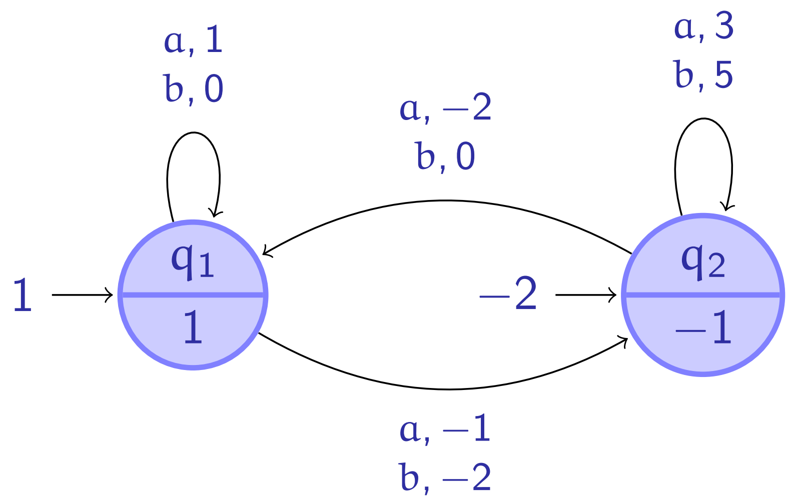


$$\mathcal{A} = \langle \alpha, \beta, \{T_a\}_{a \in \Sigma} \rangle$$
$$\alpha, \beta \in \mathbb{R}^n, T_a \in \mathbb{R}^{n \times n}$$

$$\alpha = \begin{bmatrix} 1 & -2 \end{bmatrix} \qquad \beta = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad T_a = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \qquad T_b = \begin{bmatrix} 0 & -2 \\ 0 & 5 \end{bmatrix}$$

WA and Rational Functions

$\mathcal{A}$



$$\mathcal{A} = \langle \alpha, \beta, \{T_a\}_{a\in\Sigma} \rangle$$
$$\alpha, \beta \in \mathbb{R}^n, T_a \in \mathbb{R}^{n\times n}$$

$f_{\mathcal{A}} : \Sigma^\star \to \mathbb{R}$

sum of weights of all paths labelled by the input string

$$f_{\mathcal{A}}(x) = \alpha T_{x_1} \cdots T_{x_t} \beta$$
$$= \alpha T_x \beta$$

(Pseudo-)Metrics for Approximation

How do we measure the quality of WA approximations?

(Pseudo-)Metrics for Approximation

How do we measure the quality of WA approximations?

$$\mathrm{dist}_p(\mathcal{A}, \mathcal{B}) = \|f_{\mathcal{A}} - f_{\mathcal{B}}\|_p \qquad \|f\|_p = \left(\sum_{x \in \Sigma^\star} |f(x)|^p \right)^{1/p}$$

(Pseudo-)Metrics for Approximation

How do we measure the quality of WA approximations?

$$\mathrm{dist}_p(\mathcal{A}, \mathcal{B}) = \|f_{\mathcal{A}} - f_{\mathcal{B}}\|_p$$

pseudo-metric

$$\|f\|_p = \left(\sum_{x \in \Sigma^\star} |f(x)|^p \right)^{1/p}$$

(Pseudo-)Metrics for Approximation

How do we measure the quality of WA approximations?

$$\mathrm{dist}_p(\mathcal{A}, \mathcal{B}) = \|f_{\mathcal{A}} - f_{\mathcal{B}}\|_p \qquad \|f\|_p = \left(\sum_{x \in \Sigma^\star} |f(x)|^p \right)^{1/p}$$

Important linear subspaces of $\mathbb{R}^{\Sigma^\star}$

$$\mathsf{Rat} = \{f : \Sigma^\star \to \mathbb{R} \mid \textbf{rational}\}$$

$$\ell^p = \{f : \Sigma^\star \to \mathbb{R} \mid \|f\|_p < \infty\}$$

$$\ell^p_{\mathsf{Rat}} = \mathsf{Rat} \cap \ell^p = \{f : \Sigma^\star \to \mathbb{R} \mid \textbf{rational}, \|f\|_p < \infty\}$$

(Pseudo-)Metrics for Approximation

How do we measure the quality of WA approximations?

$$\mathrm{dist}_p(\mathcal{A}, \mathcal{B}) = \|f_{\mathcal{A}} - f_{\mathcal{B}}\|_p \qquad \|f\|_p = \left(\sum_{x \in \Sigma^\star} |f(x)|^p \right)^{1/p}$$

Important linear subspaces of $\mathbb{R}^{\Sigma^\star}$

$$\mathsf{Rat} = \{f : \Sigma^\star \to \mathbb{R} \mid \textbf{rational}\}$$

$$\ell^p = \{f : \Sigma^\star \to \mathbb{R} \mid \|f\|_p < \infty\}$$

Banach/Hilbert

$$\ell^p_{\mathsf{Rat}} = \mathsf{Rat} \cap \ell^p = \{f : \Sigma^\star \to \mathbb{R} \mid \textbf{rational}, \|f\|_p < \infty\}$$

not complete!

An Approximation Result

There is an algorithm that given a minimal WA $\mathcal{A}$ with n states and $k < n$, runs in time $O(|\Sigma| n^6)$ and returns a WA $\hat{\mathcal{A}}$ with k states such that

$$\mathrm{dist}_2(\mathcal{A}, \hat{\mathcal{A}}) \leqslant \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_n^2}$$

[Balle, Panangaden, Precup 2015, Kulesza, Jiang, Singh 2015]

An Approximation Result

There is an algorithm that given a minimal WA $\mathcal{A}$ with n states and $k < n$, runs in time $O(|\Sigma| n^6)$ and returns a WA $\hat{\mathcal{A}}$ with k states such that

$$\mathrm{dist}_2(\mathcal{A}, \hat{\mathcal{A}}) \leqslant \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_n^2}$$

singular values of the *Hankel matrix* of $f_{\mathcal{A}}$

[Balle, Panangaden, Precup 2015, Kulesza, Jiang, Singh 2015]

The Plan

1. Hankel Matrices and Singular Value Automata (SVA)
2. Algorithms for Computing SVA
3. WA Approximation via SVA Truncation

The Hankel Matrix

$$f : \Sigma^\star \to \mathbb{R}$$

$$H_f \in \mathbb{R}^{\Sigma^\star \times \Sigma^\star}$$

$$H_f(p, s) = f(ps)$$

The Hankel Matrix

$$f : \Sigma^\star \to \mathbb{R}$$

$$H_f \in \mathbb{R}^{\Sigma^\star \times \Sigma^\star}$$

$$H_f(p, s) = f(ps)$$

$$\begin{array}{c@{\quad}c}
 & \begin{array}{cccccc} \varepsilon & a & b & \cdots & s & \cdots \end{array} \\
\begin{array}{c} \varepsilon \\ a \\ b \\ \vdots \\ p \\ \vdots \end{array} &
\left[\begin{array}{cccccc}
f(\varepsilon) & f(a) & f(b) & & \vdots & \\
f(a) & f(aa) & f(ab) & & \vdots & \\
f(b) & f(ba) & f(bb) & & \vdots & \\
 & & & & & \\
\cdots & \cdots & \cdots & & f(ps) & \\
 & & & & &
\end{array}\right]
\end{array}$$

The Hankel Matrix

$$f : \Sigma^\star \to \mathbb{R}$$

$$H_f \in \mathbb{R}^{\Sigma^\star \times \Sigma^\star}$$

$$H_f(p, s) = f(ps)$$

$$
\begin{array}{c|c}
 & \begin{array}{cccccc} \varepsilon & a & b & \cdots & s & \cdots \end{array} \\
\begin{array}{c} \varepsilon \\ a \\ b \\ \vdots \\ p \\ \vdots \end{array} &
\begin{bmatrix}
f(\varepsilon) & f(a) & f(b) & & \vdots & \\
f(a) & f(aa) & f(ab) & & \vdots & \\
f(b) & f(ba) & f(bb) & & \vdots & \\
 & & & & & \\
\cdots & \cdots & \cdots & & f(ps) & \\
 & & & & &
\end{bmatrix}
\end{array}
$$

Theorem: $\mathrm{rank}(H_f) = n$ iff f computed by minimal WA with n states

[Kronecker 1881, Carlyle & Paz 1971, Fliess 1974]

Proof (upper bound $\mathrm{rank}(H_{f_{\mathcal{A}}}) \leqslant |\mathcal{A}|$)

$$H_f(p, s) = f(ps) = \alpha T_{p_1} \cdots T_{p_t} T_{s_1} \cdots T_{s_{t'}} \beta = \alpha_p \cdot \beta_s$$

Proof (upper bound $\mathsf{rank}(H_{f_{\mathcal{A}}}) \leqslant |\mathcal{A}|$)

$$H_f(p,s) = f(ps) = \alpha T_{p_1} \cdots T_{p_t} T_{s_1} \cdots T_{s_{t'}} \beta = \alpha_p \cdot \beta_s$$

$$\begin{array}{cc} & \begin{array}{ccccc} & & s & & \end{array} \\ p & \begin{bmatrix} & & \vdots & & \\ & & \vdots & & \\ & & \vdots & & \\ \cdots & \cdots & \bullet & \cdots & \cdots \\ & & \vdots & & \end{bmatrix} \end{array} = \begin{array}{cc} p & \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \bullet & \bullet & \bullet \\ \cdot & \cdot & \cdot \end{bmatrix} \end{array} \begin{array}{c} \begin{array}{ccccc} & & s & & \end{array} \\ \begin{bmatrix} \cdot & \cdot & \bullet & \cdot & \cdot \\ \cdot & \cdot & \bullet & \cdot & \cdot \\ \cdot & \cdot & \bullet & \cdot & \cdot \end{bmatrix} \end{array}$$

Proof (upper bound $\mathrm{rank}(H_{f_{\mathcal{A}}}) \leqslant |\mathcal{A}|$)

$$H_f(p, s) = f(ps) = \alpha T_{p_1} \cdots T_{p_t} T_{s_1} \cdots T_{s_{t'}} \beta = \alpha_p \cdot \beta_s$$

$$\begin{array}{cc} & s \\ p & \begin{bmatrix} & & \vdots & & \\ & & \vdots & & \\ & & \vdots & & \\ \cdots & \cdots & \bullet & \cdots & \cdots \\ & & \vdots & & \end{bmatrix} \end{array} = \begin{array}{cc} & \\ & \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \bullet & \bullet & \bullet \\ \cdot & \cdot & \cdot \end{bmatrix} \\ p & \end{array} \begin{array}{c} s \\ \begin{bmatrix} \cdot & \cdot & \bullet & \cdot & \cdot \\ \cdot & \cdot & \bullet & \cdot & \cdot \\ \cdot & \cdot & \bullet & \cdot & \cdot \end{bmatrix} \end{array}$$

$$H_f = P \cdot S$$

Proof (equality $\mathrm{rank}(H_f) = \min_{f_{\mathcal{A}}=f} |\mathcal{A}|$)

1. Suppose $H_f = P \cdot S$ is rank factorisation
2. For $a \in \Sigma$ define $H_f^a(p,s) = f(pas)$ and note $H_f^a = R_a H_f$ where $(R_a f)(x) = f(xa)$
3. Observe that because S has full row rank $\mathrm{colspan}(R_a P) = \mathrm{colspan}(H_f^a) \subseteq \mathrm{colspan}(H_f) = \mathrm{colspan}(P)$
4. Hence there exists T_a such that $R_a P = P T_a$
5. Let $\mathcal{A} = \langle \alpha = P(\varepsilon, -), \beta = S(-, \varepsilon), \{T_a\} \rangle$
6. Note $f_{\mathcal{A}} = f$ since $P(\varepsilon,-)T_x S(-,\varepsilon) = R_x P(\varepsilon,-)S(-,\varepsilon) = P(x,-)S(-,\varepsilon) = H_f(x,\varepsilon)$

A Useful Correspondence

$\mathcal{A} = \langle \alpha, \beta, \{T_a\}_{a \in \Sigma} \rangle$ $\quad$ $Q \in \mathbb{R}^{n \times n}$ invertible

$$\mathcal{A}^Q = \langle \alpha Q, Q^{-1}\beta, \{Q^{-1} T_a Q\}_{a \in \Sigma} \rangle$$

A Useful Correspondence

$\mathcal{A} = \langle \alpha, \beta, \{T_a\}_{a\in\Sigma} \rangle$ $\quad$ $Q \in \mathbb{R}^{n\times n}$ invertible

$$\mathcal{A}^Q = \langle \alpha Q, Q^{-1}\beta, \{Q^{-1} T_a Q\}_{a\in\Sigma} \rangle$$

$$(\alpha Q)(Q^{-1} T_{x_1} Q) \cdots (Q^{-1} T_{x_t} Q)(Q^{-1}\beta) = \alpha T_{x_1} \cdots T_{x_t} \beta$$

A Useful Correspondence

$\mathcal{A} = \langle \alpha, \beta, \{T_a\}_{a \in \Sigma} \rangle$ $\quad Q \in \mathbb{R}^{n \times n}$ invertible

$$\mathcal{A}^Q = \langle \alpha Q, Q^{-1}\beta, \{Q^{-1} T_a Q\}_{a \in \Sigma} \rangle$$

$$(\alpha Q)(Q^{-1} T_{x_1} Q) \cdots (Q^{-1} T_{x_t} Q)(Q^{-1}\beta) = \alpha T_{x_1} \cdots T_{x_t} \beta$$

Theorem: For any rational $f : \Sigma^\star \to \mathbb{R}$

Minimal WA $\mathcal{A}$ computing f $\xleftrightarrow{\text{bijection}}$ Rank factorizations of the form $H_f = P \cdot S$

[Balle, Carreras, Luque, Quattoni 2014]

SVD of Hankel Matrices

$$H_f = UDV^\top = \begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_n \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_n^\top & \cdots \end{bmatrix}$$

SVD of Hankel Matrices

$$H_f = UDV^\top = \begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_n \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_n^\top & \cdots \end{bmatrix}$$

Theorem:

H_f with f rational admits an SVD iff $\|f\|_2 < \infty$

[Balle, Panangaden, Precup 2015, 201?]

SVD of Hankel Matrices

$$H_f = UDV^\top = \begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_n \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_n^\top & \cdots \end{bmatrix}$$

Theorem:

H_f with f rational admits an SVD iff $\|f\|_2 < \infty$

Key Idea: see Hankel matrix as an operator in a Hilbert space

$$H_f : \ell^2 \to \ell^2$$

$$g \mapsto H_f g$$

$$(H_f g)(x) = \sum_{y \in \Sigma^\star} f(xy) g(y)$$

[Balle, Panangaden, Precup 2015, 201?]

The Singular Value Automaton

From Correspondence Theorem:

If $\mathrm{H_f}$ admits an SVD $\mathrm{H_f} = \mathrm{UDV}^\top$, then there exists a WA $\mathcal{A}$ for f inducing $\mathrm{P} = \mathrm{UD}^{1/2}$ and $\mathrm{S} = \mathrm{D}^{1/2}\mathrm{V}^\top$

[Balle, Panangaden, Precup 2015]

The Singular Value Automaton

<u>From Correspondence Theorem:</u> ***singular value automaton***

If $\mathrm{H_f}$ admits an SVD $\mathrm{H_f = UDV^\top}$, then there exists a WA $\mathcal{A}$ for f inducing $\mathrm{P = UD^{1/2}}$ and $\mathrm{S = D^{1/2}V^\top}$

[Balle, Panangaden, Precup 2015]

The Singular Value Automaton

From Correspondence Theorem: *singular value automaton*

If H_f admits an SVD $H_f = UDV^\top$, then there exists a WA $\mathcal{A}$ for f inducing $P = UD^{1/2}$ and $S = D^{1/2}V^\top$

$$\mathcal{A} = \langle \alpha, \beta, \{T_a\}_{a \in \Sigma} \rangle$$

WA computing sing. vect. can be obtained from SVA:

$$\mathcal{U}_i = \langle \alpha, e_i / \sqrt{\sigma_i}, \{T_a\} \rangle \qquad f_{\mathcal{U}_i} = u_i$$

$$\mathcal{V}_i = \langle e_i / \sqrt{\sigma_i}, \beta, \{T_a\} \rangle \qquad f_{\mathcal{V}_i} = v_i$$

[Balle, Panangaden, Precup 2015]

Consequences of SVA

1. For every $f \in \ell^2_{\mathsf{Rat}}$ the SVA provides a “unique” canonical WA representation

Consequences of SVA

1. For every $f \in \ell^2_{\mathsf{Rat}}$ the SVA provides a “unique” canonical WA representation

2. The vector subspace $\ell^2_{\mathsf{Rat}} \subset \mathbb{R}^{\Sigma^\star}$ is closed under taking left/right singular vectors of Hankel matrices

The Plan

1. Hankel Matrices and Singular Value Automata (SVA)
2. Algorithms for Computing SVA
3. WA Approximation via SVA Truncation

Two Important Matrices

Reachability Gramian: $G_P = P^\top \cdot P$

Observability Gramian: $G_S = S \cdot S^\top$

Two Important Matrices

Reachability Gramian: $\mathbf{G}_\mathbf{P} = \mathbf{P}^\top \cdot \mathbf{P}$

Observability Gramian: $\mathbf{G}_\mathbf{S} = \mathbf{S} \cdot \mathbf{S}^\top$

<u>**Theorem:**</u> [Balle, Panangaden, Precup 2015, 201?]

Assuming $\mathcal{A}$ is minimal, $\mathbf{G}_\mathbf{P}$ and $\mathbf{G}_\mathbf{S}$ are well-defined iff $\|f_\mathcal{A}\|_2 < \infty$

Two Important Matrices

Reachability Gramian: $G_P = P^\top \cdot P$

Observability Gramian: $G_S = S \cdot S^\top$

For SVA

$$P = UD^{1/2} \quad \Rightarrow \quad G_P = D$$

$$S = D^{1/2}V^\top \quad \Rightarrow \quad G_S = D$$

Two Important Matrices

Reachability Gramian: $$G_P = P^\top \cdot P$$

Observability Gramian: $$G_S = S \cdot S^\top$$

For SVA

$$P = UD^{1/2} \quad \Rightarrow \quad G_P = D$$
$$S = D^{1/2}V^\top \quad \Rightarrow \quad G_S = D$$

For $\mathcal{A}^Q$

$$G_P^Q = Q^\top G_P Q$$
$$G_S^Q = Q^{-1} G_S Q^{-\top}$$

Computing the SVA

Given minimal WA $\mathcal{A}$ computing $f \in \ell^2_{\mathsf{Rat}}$ do:

1. Compute Gramians G_P and G_S
2. Diagonalize $M = G_S G_P$
   (i.e. find Q s.t. $Q^{-1} M Q = D^2$)
3. Obtain SVA as $\mathcal{A}^Q$

Computing the SVA

Given minimal WA $\mathcal{A}$ computing $f \in \ell^2_{\mathsf{Rat}}$ do:

1. Compute Gramians G_P and G_S
2. Diagonalize $M = G_S G_P$
   (i.e. find Q s.t. $Q^{-1} M Q = D^2$)
3. Obtain SVA as $\mathcal{A}^Q$

note:

Step 2 is a finite eigenvalue problem :-)

Step 1 involves products of infinite matrices :-o

Computing the Gramians

Fixed-point Equations for Gramians

$$G_S = \beta \cdot \beta^\top + \sum_{a \in \Sigma} T_a \cdot G_S \cdot T_a^\top$$

$$G_P = \alpha^\top \cdot \alpha + \sum_{a \in \Sigma} T_a^\top \cdot G_P \cdot T_a$$

Computing the Gramians

Fixed-point Equations for Gramians

$$G_S = \beta \cdot \beta^\top + \sum_{a \in \Sigma} T_a \cdot G_S \cdot T_a^\top$$

$$G_P = \alpha^\top \cdot \alpha + \sum_{a \in \Sigma} T_a^\top \cdot G_P \cdot T_a$$

Closed-form Algorithm

Total time $O(n^6 + |\Sigma| n^4)$

$$vec(G_S) = \left(I - \sum_{a \in \Sigma} T_a \otimes T_a \right)^{-1} (\beta \otimes \beta)$$

Computing the Gramians

Fixed-point Equations for Gramians

$$G_S = \beta \cdot \beta^\top + \sum_{a \in \Sigma} T_a \cdot G_S \cdot T_a^\top$$

$$G_P = \alpha^\top \cdot \alpha + \sum_{a \in \Sigma} T_a^\top \cdot G_P \cdot T_a$$

Closed-form Algorithm

Total time $O(n^6 + |\Sigma| n^4)$

$$vec(G_S) = \left(I - \sum_{a \in \Sigma} T_a \otimes T_a \right)^{-1} (\beta \otimes \beta)$$

note: there is also an iterative algorithm with $O(|\Sigma| n^{2.4})$ flops per iteration

The Plan

1. Hankel Matrices and Singular Value Automata (SVA)
2. Algorithms for Computing SVA
3. WA Approximation via SVA Truncation

Approximation (Attempt 1)

$$\underbrace{\begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_n \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_n^\top & \cdots \end{bmatrix}}_{H_f} \approx \underbrace{\begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_{\color{red}k} \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_{\color{red}k} \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_{\color{red}k}^\top & \cdots \end{bmatrix}}_{\hat{H}}$$

Approximation (Attempt 1)

$$\underbrace{\begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_n \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_n^\top & \cdots \end{bmatrix}}_{H_f} \approx \underbrace{\begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_{\color{red}k} \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_{\color{red}k} \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_{\color{red}k}^\top & \cdots \end{bmatrix}}_{\hat{H}}$$

$\hat{H} \rightarrow$ **best rank k approximation**

Approximation (Attempt 1)

$$\underbrace{\begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_n \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_n^\top & \cdots \end{bmatrix}}_{H_f} \approx \underbrace{\begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_k \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_k^\top & \cdots \end{bmatrix}}_{\hat{H}}$$

$\hat{H} \rightarrow$ **best rank k approximation**

Claim:

If $\hat{f} : \Sigma^\star \to \mathbb{R}$ is such that $H_{\hat{f}} = \hat{H}$, then $\mathrm{rank}(\hat{f}) = k$

and $\|f - \hat{f}\|_2 \leqslant \|H_f - \hat{H}\|_{op} = \sigma_{k+1}$

Approximation (Attempt 1)

$$\underbrace{\begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_n \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_n^\top & \cdots \end{bmatrix}}_{H_f} \approx \underbrace{\begin{bmatrix} \vdots & & \vdots \\ u_1 & \cdots & u_k \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} \cdots & v_1^\top & \cdots \\ & \vdots & \\ \cdots & v_k^\top & \cdots \end{bmatrix}}_{\hat{H}}$$

best rank k approximation

Claim:

not Hankel!

If $\hat{f} : \Sigma^\star \to \mathbb{R}$ is such that $H_{\hat{f}} = \hat{H}$, then $\mathrm{rank}(\hat{f}) = k$ and $\|f - \hat{f}\|_2 \leqslant \|H_f - \hat{H}\|_{op} = \sigma_{k+1}$

Approximation (Attempt 2)

$$H_f = \begin{bmatrix} \vdots & & \\ f & & \\ \vdots & & \end{bmatrix} = \underbrace{\begin{bmatrix} \vdots & & \vdots \\ \sqrt{\sigma_1}u_1 & \cdots & \sqrt{\sigma_n}u_n \\ \vdots & & \vdots \end{bmatrix}}_{UD^{1/2}} \underbrace{\begin{bmatrix} \beta_1 & & \\ \vdots & & \\ \beta_n & & \end{bmatrix}}_{D^{1/2}V^\top}$$

Approximation (Attempt 2)

$$H_f = \begin{bmatrix} \vdots & & \\ f & & \\ \vdots & & \end{bmatrix} = \underbrace{\begin{bmatrix} \vdots & & \vdots \\ \sqrt{\sigma_1}u_1 & \cdots & \sqrt{\sigma_n}u_n \\ \vdots & & \vdots \end{bmatrix}}_{UD^{1/2}} \underbrace{\begin{bmatrix} \beta_1 & & \\ \vdots & & \\ \beta_n & & \end{bmatrix}}_{D^{1/2}V^\top} \qquad f = \sum_{i=1}^{n} \beta_i \sqrt{\sigma_i} u_i$$

Approximation (Attempt 2)

$$H_f = \begin{bmatrix} \vdots & \\ f & \\ \vdots & \end{bmatrix} = \underbrace{\begin{bmatrix} \vdots & & \vdots \\ \sqrt{\sigma_1}u_1 & \cdots & \sqrt{\sigma_n}u_n \\ \vdots & & \vdots \end{bmatrix}}_{UD^{1/2}} \underbrace{\begin{bmatrix} \beta_1 & \\ \vdots & \\ \beta_n & \end{bmatrix}}_{D^{1/2}V^\top} \qquad f = \sum_{i=1}^{n} \beta_i \sqrt{\sigma_i} u_i$$

Claim:

Taking $\hat{f} = \sum_{i=1}^{k} \beta_i \sqrt{\sigma_i} u_i$ we have a rational function sum of k orthogonal rational functions and $\|f - \hat{f}\|_2^2 = \sum_{i=k+1}^{n} \beta_i^2 \sigma_i \leqslant \sum_{i=k+1}^{n} \sigma_i^2$

Approximation (Attempt 2)

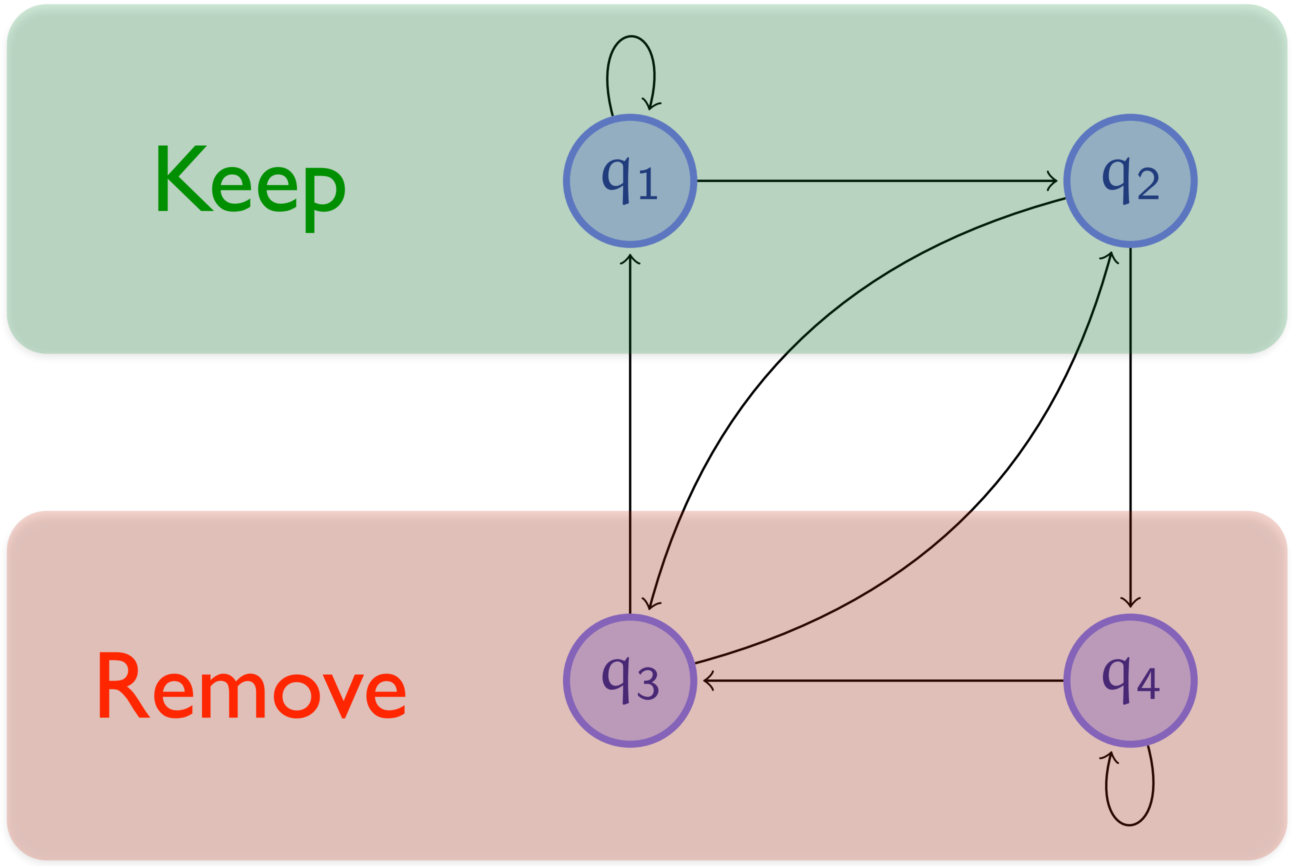
$$H_f = \begin{bmatrix} \vdots & \\ f & \\ \vdots & \end{bmatrix} = \underbrace{\begin{bmatrix} \vdots & & \vdots \\ \sqrt{\sigma_1}u_1 & \cdots & \sqrt{\sigma_n}u_n \\ \vdots & & \vdots \end{bmatrix}}_{UD^{1/2}} \underbrace{\begin{bmatrix} \beta_1 & \\ \vdots & \\ \beta_n & \end{bmatrix}}_{D^{1/2}V^\top} \qquad f = \sum_{i=1}^{n} \beta_i \sqrt{\sigma_i} u_i$$

has rank n, but want k!

Claim:

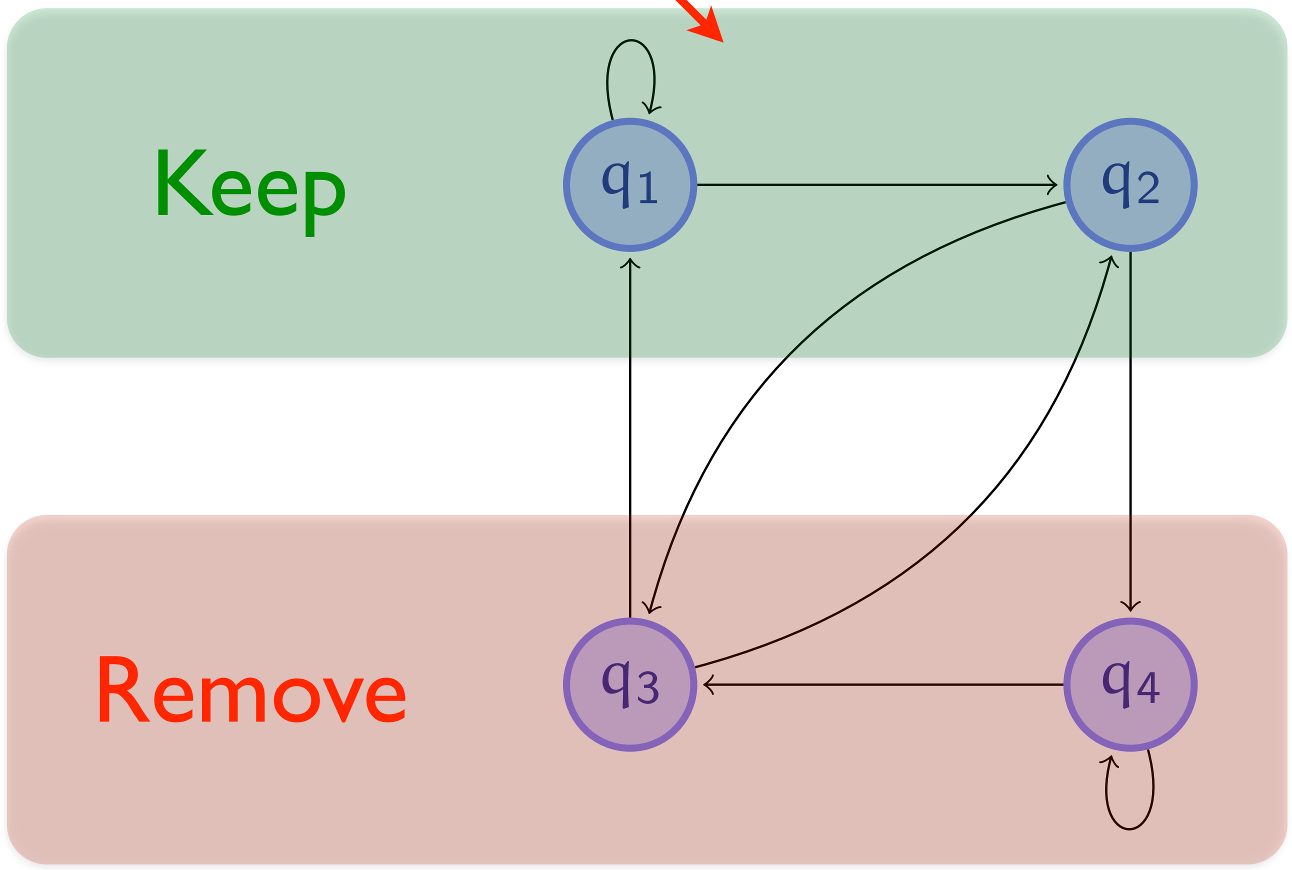
Taking $\hat{f} = \sum_{i=1}^{k} \beta_i \sqrt{\sigma_i} u_i$ we have a rational function sum of k orthogonal rational functions and $\|f - \hat{f}\|_2^2 = \sum_{i=k+1}^{n} \beta_i^2 \sigma_i \leqslant \sum_{i=k+1}^{n} \sigma_i^2$

SVA Truncation

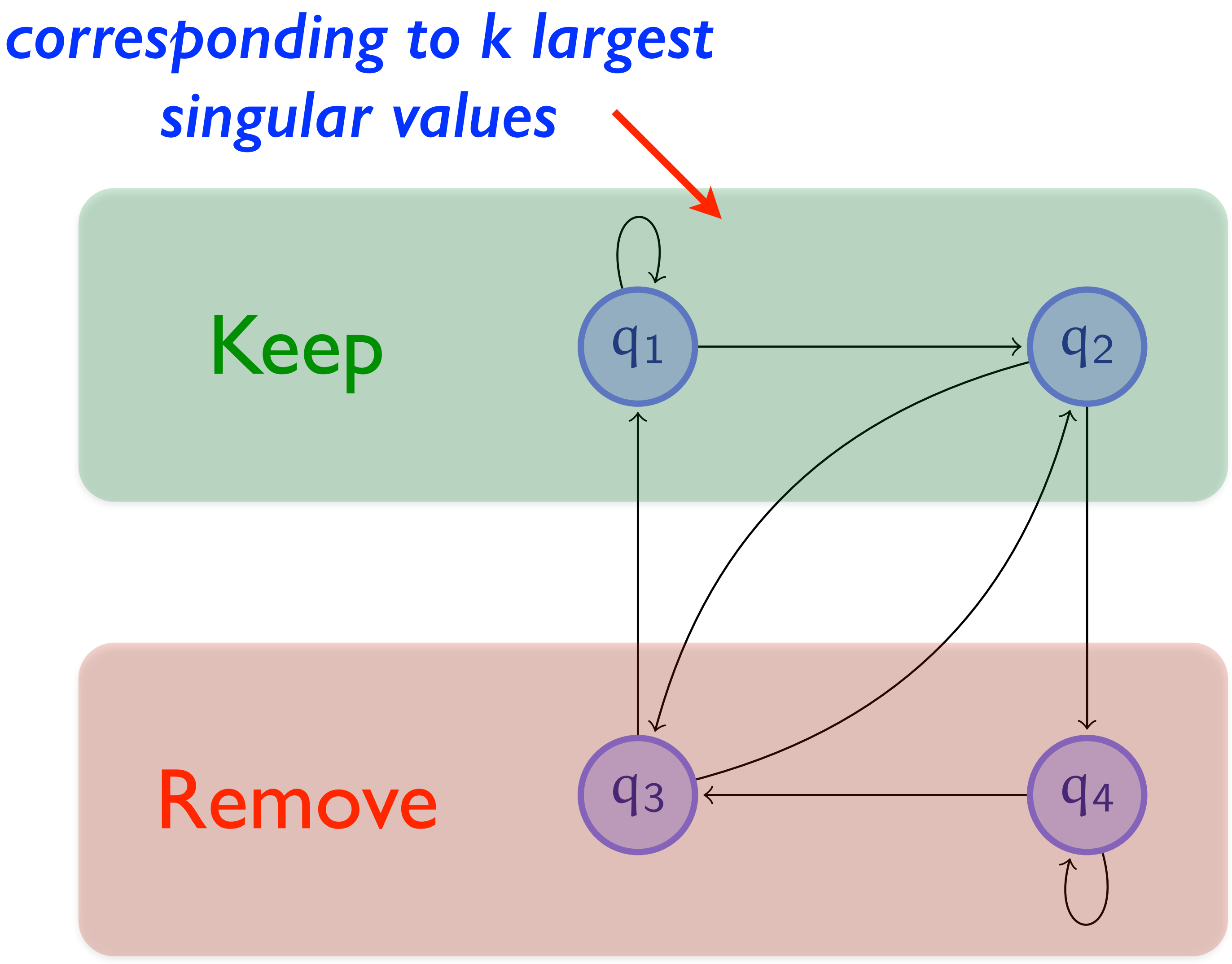


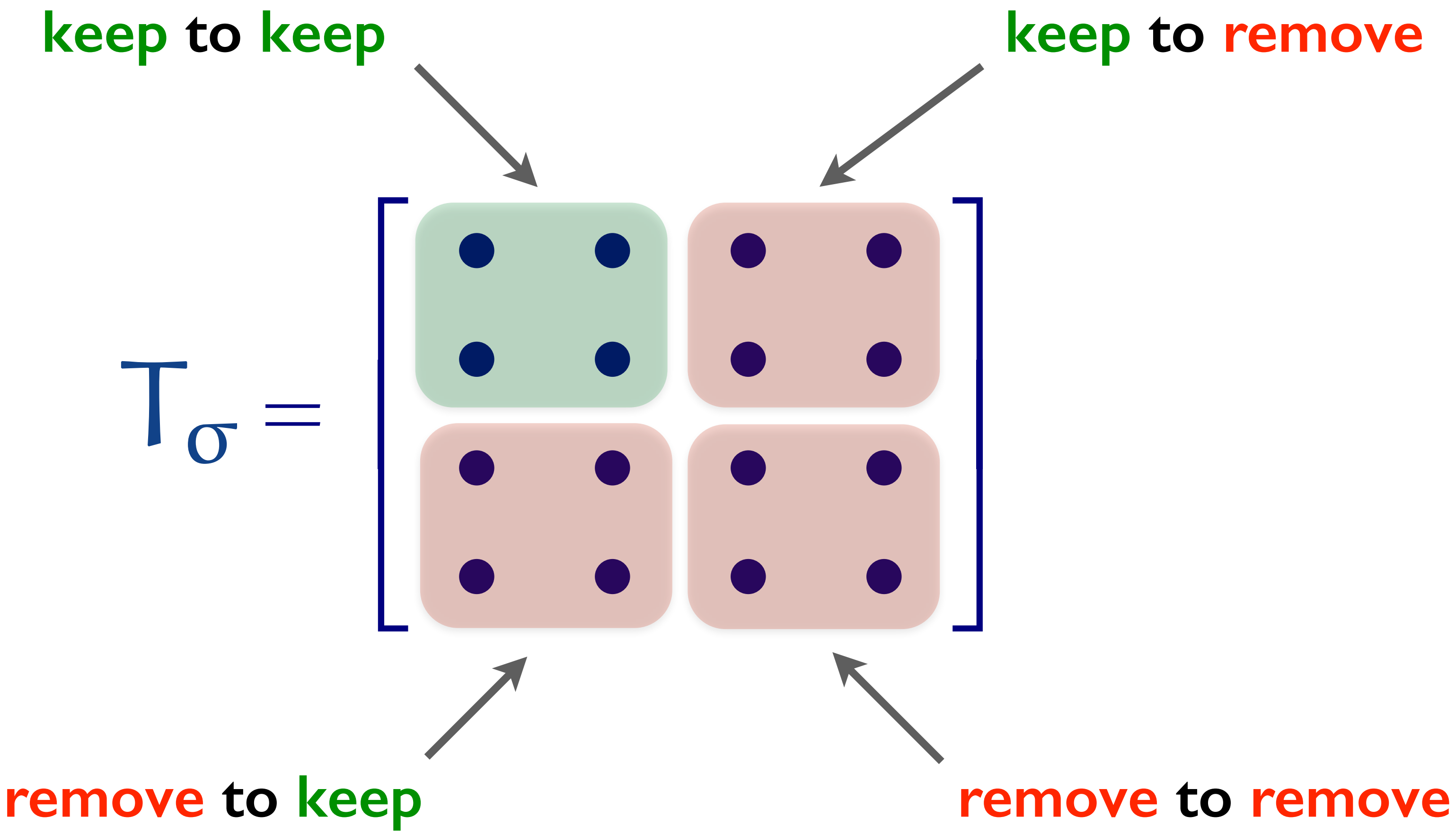
SVA Truncation

corresponding to k largest singular values



SVA Truncation

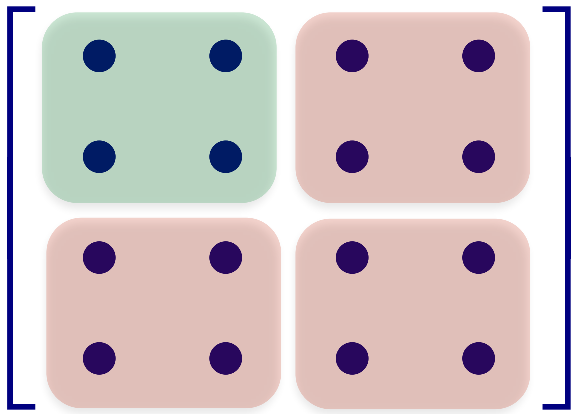




Error in SVA Truncation

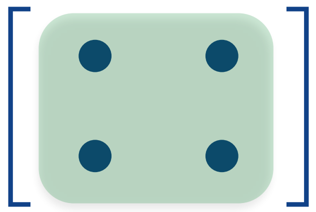
<u>*Original*</u>

$\mathcal{A}$, f, n states

$T_\sigma =$ 

<u>*Truncated*</u>

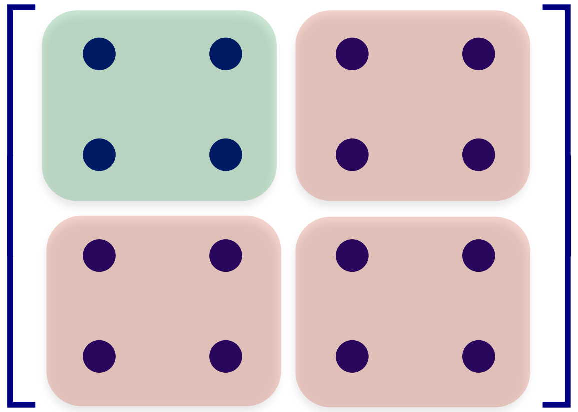
$\hat{\mathcal{A}}$, $\hat{f}$, k states

$\hat{T}_\sigma =$ 

Error in SVA Truncation

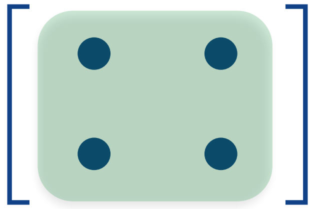
Original

$\mathcal{A}$, f, n states

$T_\sigma =$ 

Truncated

$\hat{\mathcal{A}}$, $\hat{f}$, k states

$\hat{T}_\sigma =$ 

Error Bounds:

$$\|f - \hat{f}\|_2 = \text{big, ugly formula} \leqslant \begin{cases} C_f(\sigma_{k+1} + \cdots + \sigma_n)^{1/4} \\ (\sigma_{k+1}^2 + \cdots + \sigma_n^2)^{1/2} \end{cases}$$

[Balle, Panangaden, Precup 2015, 201?, Kulesza, Jiang, Singh 2015]

Conclusion

Conclusion

- Tools from functional analysis must play a role in a theory of automata approximation

Conclusion

- Tools from functional analysis must play a role in a theory of automata approximation
- Approximation theories can lead to useful algorithms for large-scale problems

Conclusion

- Tools from functional analysis must play a role in a theory of automata approximation
- Approximation theories can lead to useful algorithms for large-scale problems
- Change of basis in WA doesn't change the function but yields automata with different properties (eg. for truncation)

Open Problems

Open Problems

- How can this be extended to other types of automata, state machines, co-algebras, etc?

Open Problems

- How can this be extended to other types of automata, state machines, co-algebras, etc?

- How important is it that WA are closed under reversal for the error analysis? [see Rabusseau, Balle, Cohen 2016]

Open Problems

- How can this be extended to other types of automata, state machines, co-algebras, etc?
- How important is it that WA are closed under reversal for the error analysis? [see Rabusseau, Balle, Cohen 2016]
- Can the AAK theorems in control theory be generalised to WA?

Open Problems

- How can this be extended to other types of automata, state machines, co-algebras, etc?
- How important is it that WA are closed under reversal for the error analysis? [see Rabusseau, Balle, Cohen 2016]
- Can the AAK theorems in control theory be generalised to WA?
- Can we obtain efficient approximation algorithms with extra constraints (eg. sparsity, positivity)?

(Co)-Algebraic and Analytical Aspects of Weighted Automata Minimisation and Equivalence (Part 2)

Borja Balle

Data Science | Lancaster University

Mathematics & Statistics | Lancaster University

[CMCS Tutorial, Apr 2 2016]