Compositional Functions and Bilinear Operators

- Compositional functions defined in terms of recurrence relations
- Consider a sequence $\mathtt{abaccb}$

$$
\begin{aligned}
f(\mathtt{abaccb}) &= \boldsymbol{\alpha}_f(\mathtt{ab}) \cdot \boldsymbol{\beta}_f(\mathtt{accb}) \\
&= \boldsymbol{\alpha}_f(\mathtt{ab}) \cdot \mathbf{A}_{\mathtt{a}} \cdot \boldsymbol{\beta}_f(\mathtt{ccb}) \\
&= \boldsymbol{\alpha}_f(\mathtt{aba}) \cdot \mathbf{A}_{\mathtt{c}} \cdot \boldsymbol{\beta}_f(\mathtt{cb})
\end{aligned}
$$

where

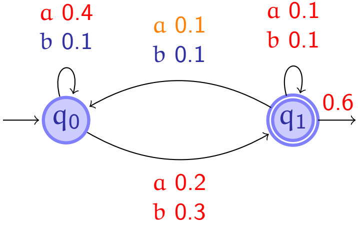
- n is the dimension of the model
- $\boldsymbol{\alpha}_f$ maps prefixes to $\mathbb{R}^n$
- $\boldsymbol{\beta}_f$ maps suffixes to $\mathbb{R}^n$
- $\mathbf{A}_{\mathtt{a}}$ is a bilinear operator in $\mathbb{R}^{n \times n}$

Problem

How to estimate $\boldsymbol{\alpha}_f$, $\boldsymbol{\beta}_f$ and $\mathbf{A}_{\mathtt{a}}, \mathbf{A}_{\mathtt{b}}, \ldots$ from "samples" of f?

Weighted Finite Automata (WFA)

Example with 2 states and alphabet $\Sigma = \{a, b\}$



Operator Representation

$$\boldsymbol{\alpha}_0 = \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix}$$

$$\boldsymbol{\alpha}_\infty = \begin{bmatrix} 0.0 \\ 0.6 \end{bmatrix}$$

$$\mathbf{A}_a = \begin{bmatrix} 0.4 & 0.2 \\ 0.1 & 0.1 \end{bmatrix}$$

$$\mathbf{A}_b = \begin{bmatrix} 0.1 & 0.3 \\ 0.1 & 0.1 \end{bmatrix}$$

$$f(ab) = \boldsymbol{\alpha}_0^\top \mathbf{A}_a \mathbf{A}_b \boldsymbol{\alpha}_\infty$$

Weighted Finite Automata (WFA)

Notation:

- Σ: alphabet – finite set
- n: number of states – positive integer
- $\boldsymbol{\alpha}_0$: initial weights – vector in $\mathbb{R}^n$ (features of empty prefix)
- $\boldsymbol{\alpha}_\infty$: final weights – vector in $\mathbb{R}^n$ (features of empty suffix)
- $\mathbf{A}_\sigma$: transition weights – matrix in $\mathbb{R}^{n \times n}$ ($\forall \sigma \in \Sigma$)

Definition: WFA with n states over Σ

$$A = \langle \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_\infty, \{\mathbf{A}_\sigma\} \rangle$$

Compositional Function: Every WFA A defines a function $f_A : \Sigma^\star \to \mathbb{R}$

$$f_A(x) = f_A(x_1 \dots x_T) = \boldsymbol{\alpha}_0^\top \mathbf{A}_{x_1} \cdots \mathbf{A}_{x_T} \boldsymbol{\alpha}_\infty = \boldsymbol{\alpha}_0^\top \mathbf{A}_x \boldsymbol{\alpha}_\infty$$

Example – Hidden Markov Model

- Assigns probabilities to strings $f(x) = \mathbb{P}[x]$
- Emission and transition are conditionally independent given state

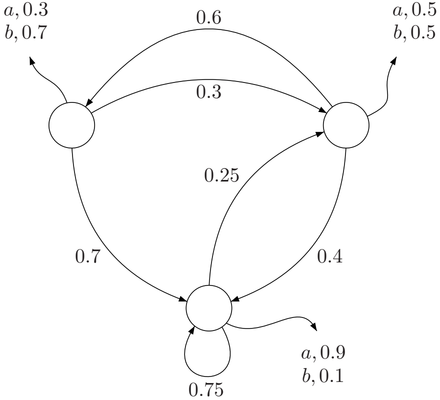
$$\boldsymbol{\alpha}_0^\top = [0.3 \;\; 0.3 \;\; 0.4]$$

$$\boldsymbol{\alpha}_\infty^\top = [1 \;\; 1 \;\; 1]$$

$$\mathbf{A}_a = \mathbf{O}_a \cdot \mathbf{T}$$

$$\mathbf{T} = \begin{bmatrix} 0 & 0.7 & 0.3 \\ 0 & 0.75 & 0.25 \\ 0 & 0.4 & 0.6 \end{bmatrix}$$

$$\mathbf{O}_a = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.9 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}$$



Forward–Backward Equations for $\mathbf{A}_\sigma$

Any WFA A defines *forward* and *backward* maps

$$\boldsymbol{\alpha}_A, \boldsymbol{\beta}_A : \Sigma^\star \to \mathbb{R}^n$$

such that for any splitting $x = p \cdot s$ one has

$$\begin{aligned}
\boldsymbol{\alpha}_A(p) &= \boldsymbol{\alpha}_0^\top \mathbf{A}_{p_1} \cdots \mathbf{A}_{p_T} \\
\boldsymbol{\beta}_A(s) &= \mathbf{A}_{s_1} \cdots \mathbf{A}_{s_{T'}} \boldsymbol{\alpha}_\infty \\
f_A(x) &= \boldsymbol{\alpha}_A(p) \cdot \boldsymbol{\beta}_A(s)
\end{aligned}$$

Example

- In HMM and PFA one has for every $i \in [n]$

$$\begin{aligned}
[\boldsymbol{\alpha}_A(p)]_i &= \mathbb{P}[p\,,\, h_{+1} = i] \\
[\boldsymbol{\beta}_A(s)]_i &= \mathbb{P}[s \mid h = i]
\end{aligned}$$

Forward–Backward Equations for $\mathbf{A}_\sigma$

Any WFA A defines *forward* and *backward* maps

$$\boldsymbol{\alpha}_A, \boldsymbol{\beta}_A : \Sigma^\star \to \mathbb{R}^n$$

such that for any splitting $x = p \cdot s$ one has

$$\boldsymbol{\alpha}_A(p) = \boldsymbol{\alpha}_0^\top \mathbf{A}_{p_1} \cdots \mathbf{A}_{p_T}$$
$$\boldsymbol{\beta}_A(s) = \mathbf{A}_{s_1} \cdots \mathbf{A}_{s_{T'}} \boldsymbol{\alpha}_\infty$$
$$f_A(x) = \boldsymbol{\alpha}_A(p) \cdot \boldsymbol{\beta}_A(s)$$

Key Observation

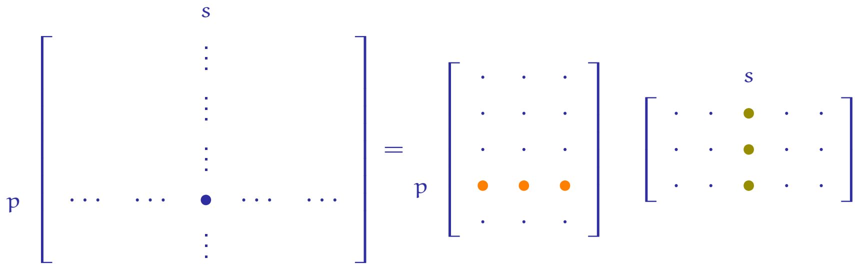
If $f_A(p\sigma s)$, $\boldsymbol{\alpha}_A(p)$, and $\boldsymbol{\beta}_A(s)$ were known for many p, s, then $\mathbf{A}_\sigma$ could be recovered from equations of the form

$$f_A(p\sigma s) = \boldsymbol{\alpha}_A(p) \cdot \mathbf{A}_\sigma \cdot \boldsymbol{\beta}_A(s)$$

Hankel matrices help organize these equations!

Structure of Low-rank Hankel Matrices

$\mathbf{H}_f \in \mathbb{R}^{\Sigma^\star \times \Sigma^\star}$ $\qquad$ $\mathbf{P} \in \mathbb{R}^{\Sigma^\star \times n}$ $\qquad$ $\mathbf{S} \in \mathbb{R}^{n \times \Sigma^\star}$



$$f(p_1 \cdots p_T \cdot s_1 \cdots s_{T'}) = \underbrace{\boldsymbol{\alpha}_0^\top \mathbf{A}_{p_1} \cdots \mathbf{A}_{p_T}}_{\boldsymbol{\alpha}_A(p)} \ \underbrace{\mathbf{A}_{s_1} \cdots \mathbf{A}_{s_{T'}} \boldsymbol{\alpha}_\infty}_{\boldsymbol{\beta}_A(s)}$$

$$\boldsymbol{\alpha}_A(p) = \mathbf{P}(p, \cdot) \qquad \boldsymbol{\beta}_A(s) = \mathbf{S}(\cdot, s)$$

Hankel Factorizations and Operators

$$\mathbf{H}_{\sigma} \in \mathbb{R}^{\Sigma^\star \times \Sigma^\star} \qquad \mathbf{P} \in \mathbb{R}^{\Sigma^\star \times n} \qquad \mathbf{A}_{\sigma} \in \mathbb{R}^{n \times n} \qquad \mathbf{S} \in \mathbb{R}^{n \times \Sigma^\star}$$

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

$$f(p_1 \cdots p_T \cdot \sigma \cdot s_1 \cdots s_{T'}) = \underbrace{\boldsymbol{\alpha}_0^\top \mathbf{A}_{p_1} \cdots \mathbf{A}_{p_T}}_{\boldsymbol{\alpha}_A(p)} \cdot \mathbf{A}_{\sigma} \cdot \underbrace{\mathbf{A}_{s_1} \cdots \mathbf{A}_{s_{T'}} \boldsymbol{\alpha}_\infty}_{\boldsymbol{\beta}_A(s)}$$

$$\mathbf{H}_{\sigma} = \mathbf{P}\ \mathbf{A}_{\sigma}\ \mathbf{S} \implies \mathbf{A}_{\sigma} = \mathbf{P}^{+}\ \mathbf{H}_{\sigma}\ \mathbf{S}^{+}$$

Note: Works with finite sub-blocks as well (assuming $\mathrm{rank}(\mathbf{P}) = \mathrm{rank}(\mathbf{S}) = n$)

Finite Sub-blocks of Hankel Matrices

Parameters:

- Set of rows (prefixes) $\mathcal{P} \subset \Sigma^\star$
- Set of columns (suffixes) $\mathcal{S} \subset \Sigma^\star$

$$\begin{array}{cc}
\Sigma^\star & \begin{array}{cccccc} \lambda & a & b & aa & ab & \dots \end{array} \\
\begin{array}{c} \lambda \\ a \\ b \\ aa \\ ab \\ \vdots \end{array} &
\begin{bmatrix}
1 & 0.3 & 0.7 & 0.05 & 0.25 & \dots \\
0.3 & 0.05 & 0.25 & 0.02 & 0.03 & \dots \\
0.7 & 0.6 & 0.1 & 0.03 & 0.2 & \dots \\
0.05 & 0.02 & 0.03 & 0.017 & 0.003 & \dots \\
0.25 & 0.23 & 0.02 & 0.11 & 0.12 & \dots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\end{array}$$

H

H_a

- $\mathbf{H}$ for finding $\mathbf{P}$ and $\mathbf{S}$
- $\mathbf{H}_\sigma$ for finding $\mathbf{A}_\sigma$
- $\mathbf{h}_{\lambda,\mathcal{S}}$ for finding $\boldsymbol{\alpha}_0$
- $\mathbf{h}_{\mathcal{P},\lambda}$ for finding $\boldsymbol{\alpha}_\infty$

Low-rank Approximation and Factorization

Parameters:

- Desired number of states n
- Block $\mathbf{H} \in \mathbb{R}^{\mathcal{P} \times \mathcal{S}}$ of the empirical Hankel matrix

Low-rank Approximation: compute truncated SVD of rank n

$$\underbrace{\mathbf{H}}_{\mathcal{P} \times \mathcal{S}} \approx \underbrace{\mathbf{U}_n}_{\mathcal{P} \times n} \underbrace{\mathbf{\Lambda}_n}_{n \times n} \underbrace{\mathbf{V}_n^\top}_{n \times \mathcal{S}}$$

Factorization: $\mathbf{H} \approx \mathbf{P}\mathbf{S}$ already given by SVD

$$\mathbf{P} = \mathbf{U}_n \mathbf{\Lambda}_n \quad \Rightarrow \quad \mathbf{P}^+ = \mathbf{\Lambda}_n^{-1} \mathbf{U}_n^\top \left(= (\mathbf{H}\mathbf{V}_n)^+ \right)$$

$$\mathbf{S} = \mathbf{V}_n^\top \quad \Rightarrow \quad \mathbf{S}^+ = \mathbf{V}_n$$

Computing the WFA

Parameters:

- Factorization $\mathbf{H} \approx (\mathbf{U}\boldsymbol{\Lambda})\mathbf{V}^\top$
- Hankel blocks $\mathbf{H}_\sigma$, $\mathbf{h}_{\lambda,\mathcal{S}}$, $\mathbf{h}_{\mathcal{P},\lambda}$

$$
\begin{aligned}
\mathbf{A}_\sigma &= \boldsymbol{\Lambda}^{-1}\mathbf{U}^\top\mathbf{H}_\sigma\mathbf{V} \left(= (\mathbf{H}\mathbf{V})^+\mathbf{H}_\sigma\mathbf{V}\right) \\
\boldsymbol{\alpha}_0 &= \mathbf{V}^\top\mathbf{h}_{\lambda,\mathcal{S}} \\
\boldsymbol{\alpha}_\infty &= \boldsymbol{\Lambda}^{-1}\mathbf{U}^\top\mathbf{h}_{\mathcal{P},\lambda} \left(= (\mathbf{H}\mathbf{V})^+\mathbf{h}_{\mathcal{P},\lambda}\right)
\end{aligned}
$$