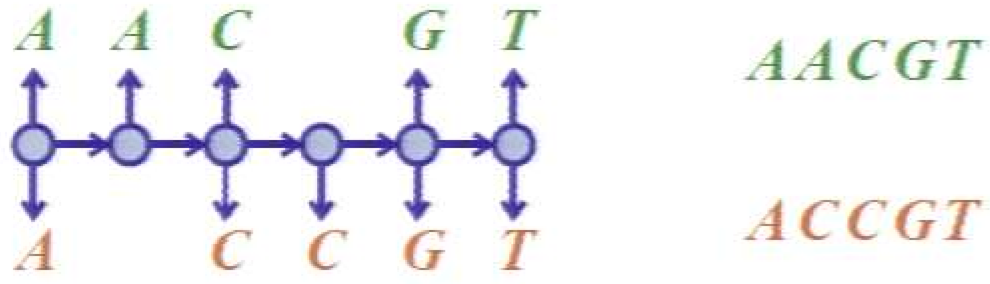
Unsupervised Learning of Finite State Transducers

Raphaël BAILLY, *Xavier CARRERAS, Ariadna QUATTONI*, Universitat Politecnica de Catalunya
Spotlight session 1, December 6th, 2013, Poster Fri67



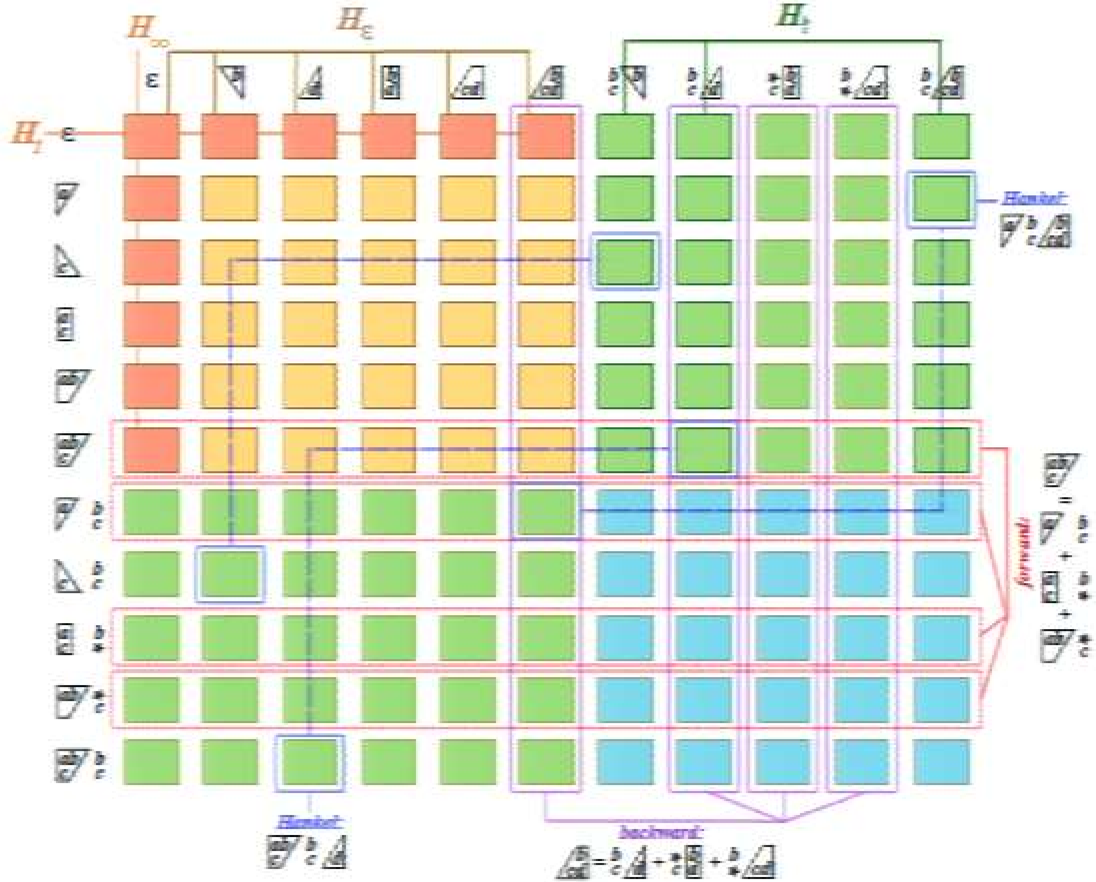
- Sequential graphical model with latent states
- Two types of observations : input / output

Unsupervised Learning :

- FST : $\alpha_1 \in \mathbb{R}^d$, $\beta_\infty \in \mathbb{R}^d$, $M_{\substack{x\\y}} \in \mathbb{R}^{d^2}$, $p\binom{x_1 x_2 x_3}{y_1 y_2 y_3} = \alpha_1^\top M_{\substack{x_1\\y_1}} M_{\substack{x_2\\y_2}} M_{\substack{x_3\\y_3}} \beta_\infty$
- Expectation-Maximization (*forward-backward*) : prone to local minimum

Spectral Method

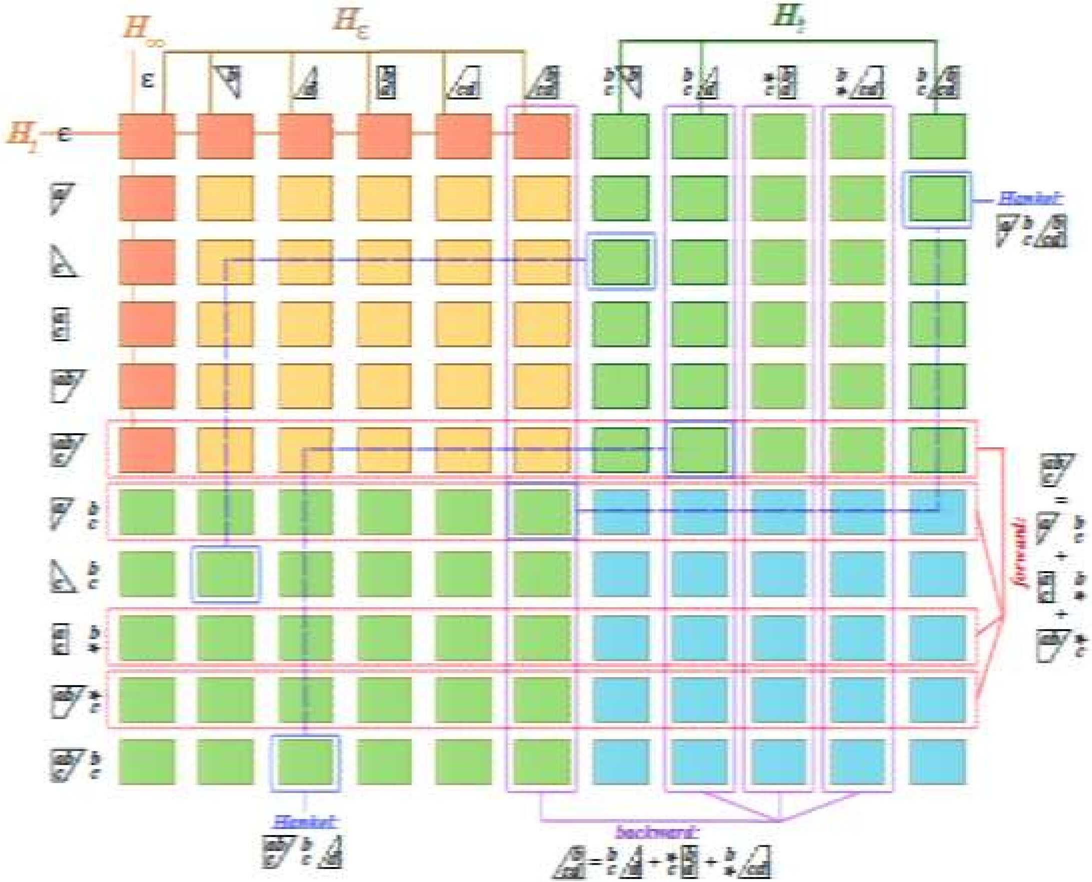
- Low-rank approximation of H_ϵ (SVD, ...)
- Computed FST : $\alpha_1^\top = H_1 H_\epsilon^+, \beta_\infty = H_\infty, M_{\frac{x}{y}} = H_{\frac{x}{y}} H_\epsilon^+$
- Statistical consistency



$$H\left({}_{c},{}^{bb}_{c}\right) = H\left({}^{b}_{cc},{}^{b}\right)$$
$$= p_S\left({}^{bb}_{cc}\right)$$

Unsupervised Spectral Method

- Observable statistics = top/left border coefficients of H_ϵ
- Completion : minimize $rk(H)$ such that H satisfies the constraints O, K, F, B
- Statistical consistency



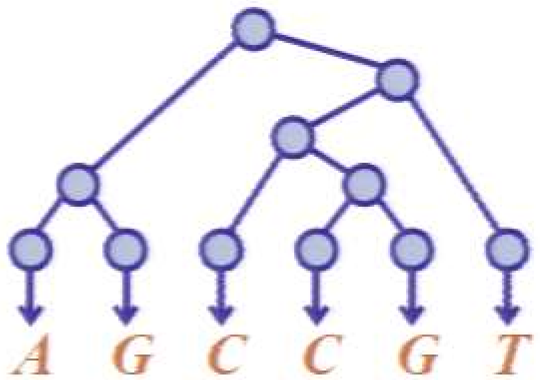
$$H\left({}_{c},{}^{bb}_{c}\right) = H\left({}^{b}_{cc},{}^{b}\right)$$

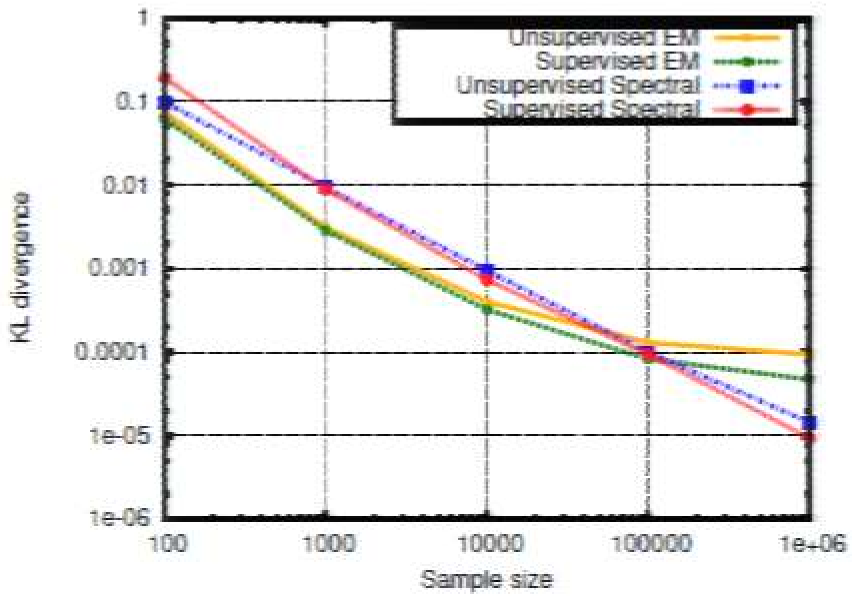
$$=?$$

Solving the rank minimization problem

- Convex relaxation $\|H\|_*$ of the objective $rk(H)$ (e.g. FISTA)
- Greedy algorithms

Other applications





- Nested sequences
- PCFGs (cf. **Spectral Learning Workshop** December 10, 2013)