FACTOR ANALYSED HIDDEN MARKOV MODELS

Antti-Veikko Rosti & Mark Gales

Cambridge University Engineering Department

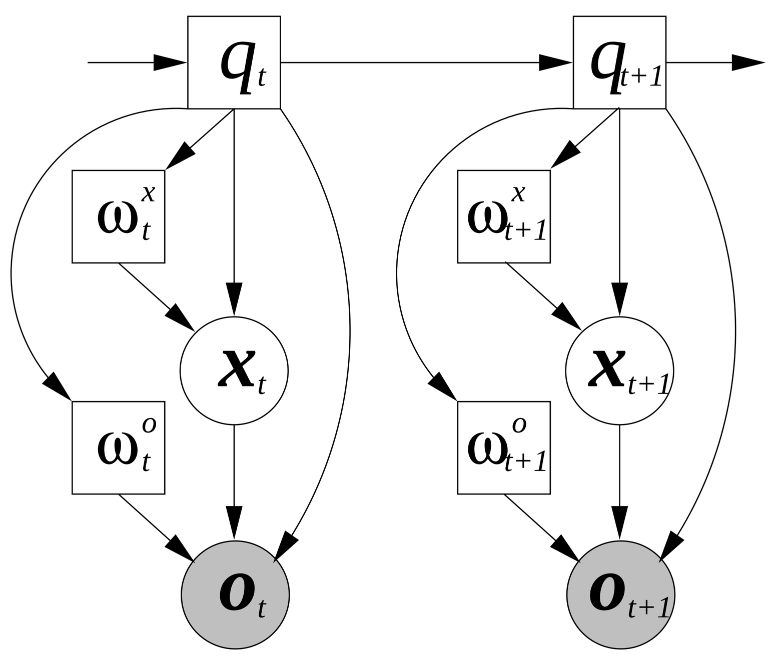
{*avir2, mjfg*}@eng.cam.ac.uk

INTRODUCTION

- A general form of acoustic model, FAHMM, is presented:
  - extends HMMs to provide better intra-frame correlation modelling;
  - fewer parameters than full covariance matrix HMMs;
  - relates to many standard schemes.

WHAT IS AN FAHMM?

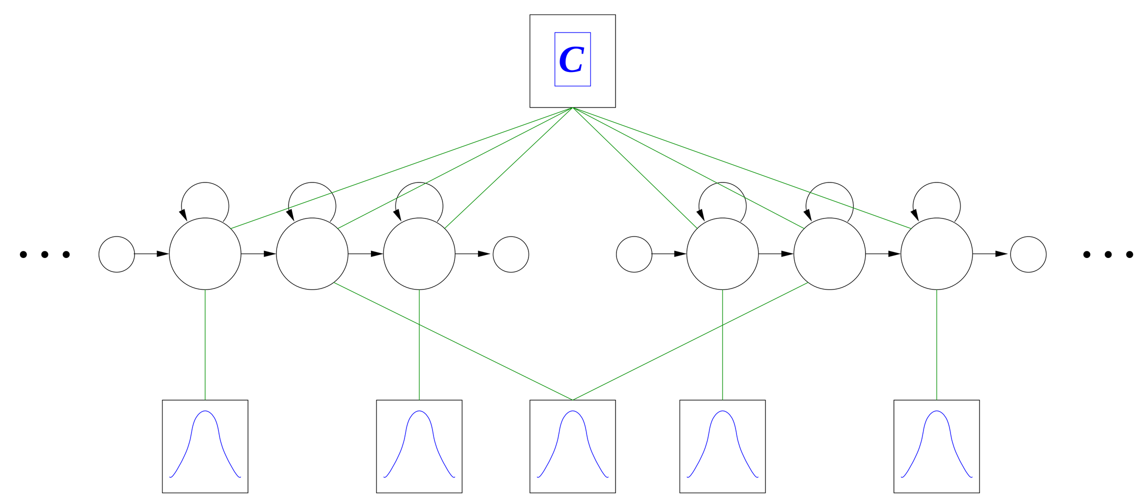
- A dynamic Bayesian network representing an FAHMM:



Generative model

$\boldsymbol{x}_t \sim \mathcal{M}^{hmm}$	$\mathcal{M}^{hmm} = \{a_{ij}, c^{(x)}_{jn}, \boldsymbol{\mu}^{(x)}_{jn}, \boldsymbol{\Sigma}^{(x)}_{jn}\}$
$\boldsymbol{o}_t = \boldsymbol{C}\boldsymbol{x}_t + \boldsymbol{v}_t$	$\boldsymbol{v}_t \sim \sum_m c^{(o)}_{jm}\mathcal{N}(\boldsymbol{\mu}^{(o)}_{jm}, \boldsymbol{\Sigma}^{(o)}_{jm})$

- The generative model is presented above with:
  - k dimensional state vectors $\boldsymbol{x}_t$;
  - p dimensional observation vectors $\boldsymbol{o}_t$;
  - $M^{(x)}$ state distribution components;
  - $M^{(o)}$ observation noise components;
  - diagonal covariance matrices $\boldsymbol{\Sigma}^{(o)}_{jm}$ and $\boldsymbol{\Sigma}^{(x)}_{jn}$;
  - observation matrix $\boldsymbol{C}$ which can be arbitrarily shared.

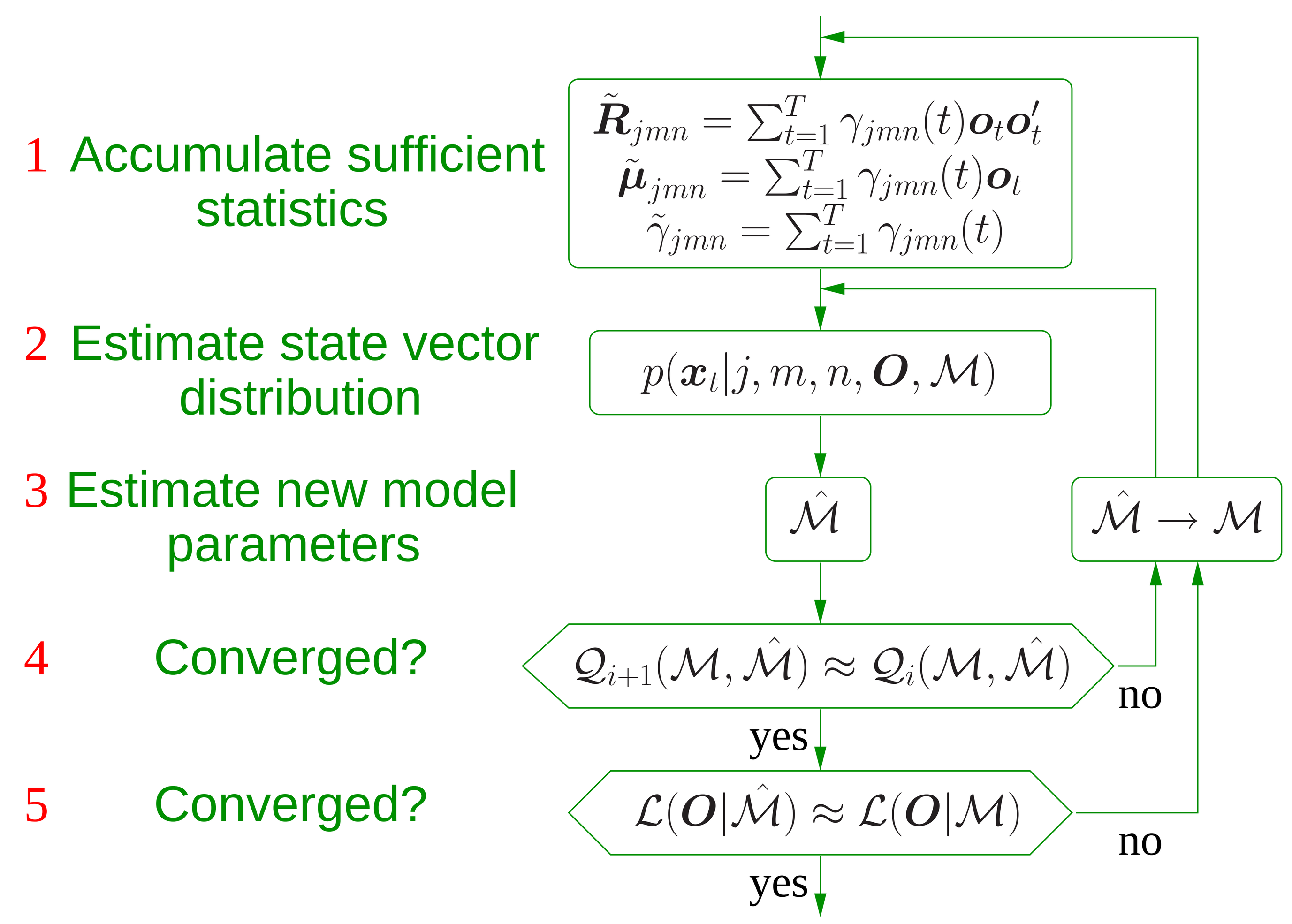


- Effectively $M^{(o)}M^{(x)}$ component full covariance matrix system:
  - mean vectors $\boldsymbol{\mu}_{jmn} = \boldsymbol{C}\boldsymbol{\mu}^{(x)}_{jn} + \boldsymbol{\mu}^{(o)}_{jm}$;
  - covariance matrices $\boldsymbol{\Sigma}_{jmn} = \boldsymbol{C}\boldsymbol{\Sigma}^{(x)}_{jn}\boldsymbol{C}' + \boldsymbol{\Sigma}^{(o)}_{jm}$.

- Related standard systems include:
  - $M^{(x)} = 0 \quad \mapsto$ hidden Markov model;
  - $M^{(x)} = 1 \quad \mapsto$ shared factor analysis;
  - $M^{(o)} = 1 \quad \mapsto$ dynamic version of independent factor analysis;
  - $k = p$ and $\boldsymbol{v}_t = \boldsymbol{0} \quad \mapsto$ semi-tied covariance matrix HMM;
  - $k > p$ and $\boldsymbol{v}_t = \boldsymbol{0} \quad \mapsto$ extended MLLT.

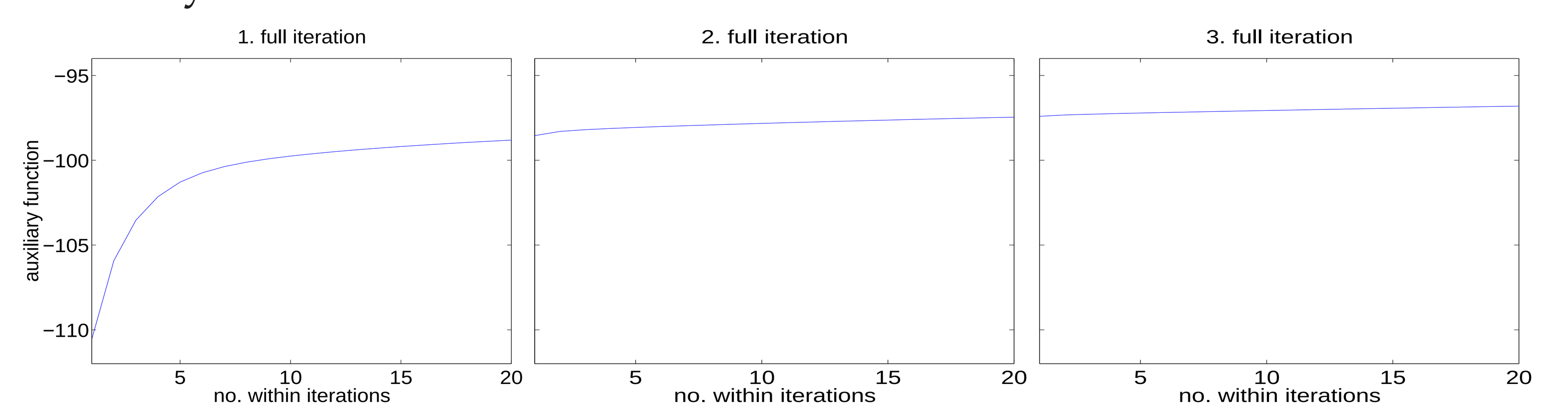
TRAINING FAHMMS

- Expectation maximisation algorithm:
  - auxiliary function $\mathcal{Q}(\mathcal{M}, \hat{\mathcal{M}}) = E\Big\{\log p(\boldsymbol{O}, \Omega^o, \boldsymbol{X}, \Omega^x, Q|\hat{\mathcal{M}})\Big|\boldsymbol{O}, \mathcal{M}\Big\}$;

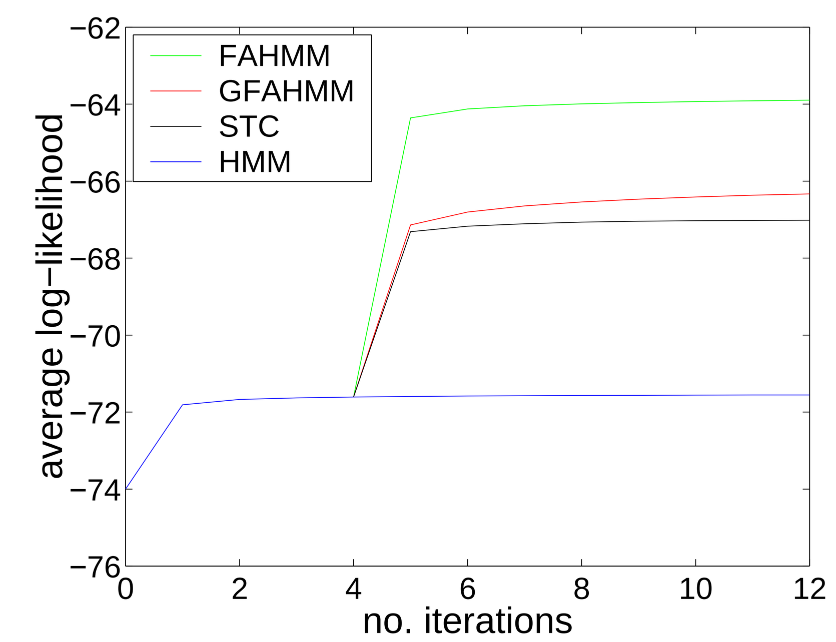


RESOURCE MANAGEMENT EXPERIMENTS

- A standard ARPA Resource Management setup:
  - 3990 sentence training data set (`train+dev_aug`);
  - 39 dimensional feature vectors (MFCC+E+D+A);
  - decision tree clustered cross-word triphone models;
  - global full transform STC;
  - global observation matrix FAHMM with $k = 39$.
- Auxiliary function values for 3 full and 20 within iterations:



- The average log-likelihoods of the training data for every full iteration:



- Evaluation setup:
  - 1200 sentence evaluation data set (`feb89+oct89+feb91+sep92`);
  - word-pair grammar.

WERs in Resource Management experiments

no. comps	1	6
HMM	7.79%	3.99%
STC	7.06%	3.85%
GFAHMM	6.52%	3.77%

SWITCHBOARD MINITRAIN EXPERIMENTS

- Standard Switchboard Minitrain setup:
  - 18 hour training data set;
  - 39 dimensional feature vectors (MF-PLP+A+D);
  - VTLN + mean and variance side based normalisation;
  - decision tree clustered cross-word triphone models;
  - state specific observation matrix FAHMM with $k = 13$.
- Evaluation setup:
  - 10 conversation sides of SWBD-2 and 10 of CHE;
  - best baseline performance with 12 components;
  - baseline performance exceeded with $M^{(o)} = 2$ and $M^{(x)} = 4$ (nfp 741);
  - best performance with $M^{(o)} = 2$ and $M^{(x)} = 6$.

WERs and the order of number of free parameters per state in SWBD task

	WER	nfp
HMM	51.0%	936
FAHMM	50.7%	793

CONCLUSIONS

- A compact and flexible acoustic model with improved intra-frame correlation modelling;
- Investigate different initialising schemes: STC, HLDA;
- Optimise the selection of state-space dimensionality.