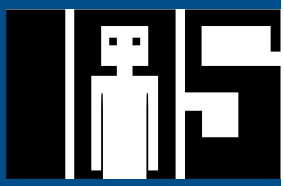
Variational Inference for Switching Dynamics

Variiierende Inferenz für wechselnde Dynamiken
Master thesis by Thomas Lautenschläger
Date of submission: August 31, 2020

1. Review: M.Sc. Hany Abdulsamad
2. Review: Prof. Jan Peters, Ph.D.
Darmstadt

TECHNISCHE
UNIVERSITÄT
DARMSTADT



Erklärung zur Abschlussarbeit gemäß §22 Abs. 7 und §23 Abs. 7 APB der TU Darmstadt

Hiermit versichere ich, Thomas Lautenschläger, die vorliegende Masterarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Mir ist bekannt, dass im Fall eines Plagiats (§38 Abs. 2 APB) ein Täuschungsversuch vorliegt, der dazu führt, dass die Arbeit mit 5,0 bewertet und damit ein Prüfungsversuch verbraucht wird. Abschlussarbeiten dürfen nur einmal wiederholt werden.

Bei der abgegebenen Thesis stimmen die schriftliche und die zur Archivierung eingereichte elektronische Fassung gemäß §23 Abs. 7 APB überein.

Bei einer Thesis des Fachbereichs Architektur entspricht die eingereichte elektronische Fassung dem vorgestellten Modell und den vorgelegten Plänen.

Darmstadt, 31. August 2020

Thomas Lautenschläger

Abstract

Learning and inferring the dynamics of a system ideally requires a model design that expresses the system's behavior, does not overfit to the training data and uses few computational resources. Bayesian probabilistic models have the powerful feature to express models as simple as possible while maximizing the model's information. Moreover, being Bayesian makes the model's uncertainty accessible. This quantity can be advantageous by integrating a learned dynamics model into policy search for planning. In this thesis, we provide a Bayesian extension for recurrent hidden Markov models and analyze their performance capabilities by training them with data sampled from dynamical systems. Since hidden Markov models capture switching regime behavior from data emitted by dynamical systems, we introduce the policy extension Switching actor (SWAC), a policy extension that incorporates the switching regime detection into policy search. This extension supplies additional structure to the policy. We present a performance comparison of SWAC with the non SWAC equipped baselines. In addition, we give an outlook to future work. The complete mathematical derivation is attached as supplementary material.

Zusammenfassung

Bayes'sche probabilistische Modelle haben die mächtige Eigenschaft, die Modelle so simpel wie möglich zu Gestalten und gleichzeitig den Informationsgehalt der Modelle zu maximieren. Außerdem bieten Bayes'sche probabilistische Modelle auf die Ungewissheit des Modells zu quantifizieren und diese auch zugänglich zu machen. In dieser Thesis definieren wir eine Bayes'sche Erweiterung für rekurrente verdeckte Markowmodelle und analysieren deren Leistungsverhalten. Zuvor werden diese Modelle mit Daten trainiert, die aus dynamischen Systemen stammen. Da verdeckte Markowmodelle die Eigenschaft haben, ein wechselndes Schaltverhalten aus Daten, die aus einem dynamischen System emittiert sind, festzustellen, integrieren wir diese Modelle für die Policysuche. Diese Methode führen wir unter der Bezeichnung SWAC ein. Diese Erweiterung stattet eine Policy mit zusätzlicher Struktur aus. Wir präsentieren einen Leistungsvergleich zwischen SWAC und den nicht mit SWAC ausgestatteten Referenzmodellen. Die gesamte mathematische Herleitung ist im Anhang beigefügt.

Contents

Figures

List of Figures

Abbreviations

List of Abbreviations

Notation	Description
VBAR-HMM	Variational Bayes autoregressive hidden Markov model
VBHMM	Variational Bayes hidden Markov model
VBrAR-HMM	Variational Bayes recurrent autoregressive hidden Markov model
A2C	Advantage actor-critic
AR-HMM	Autoregressive hidden Markov model
BNN	Bayesian neural network
ELBO	Evidence lower bound
EM	Expectation maximization
FNN	Feed-forward neural network
GMM	Gaussian mixture model
GP	Gaussian process

HMM	Hidden Markov model
i.i.d.	independently and identically distributed
LSTM	Long short-term memory neural networks
MAP	Maximum a posteriori probability
MBRL	Model-based reinforcement learning
MDP	Markov decision property
rAR-HMM	Recurrent autoregressive hidden Markov model
RNN	Recurrent neural network
SAC	Soft actor-critic
SLDS	switching linear-Gaussian dynamical systems
SOTA	state-of-the-art
SWAC	Switching actor
VAE	Variational autoencoders
w.r.t.	with respect to

1. Introduction

The optimal control for a plane to fly stable depends on specific target scenarios. The plane is starting, being in target height or on landing approach. As the scenarios differ, the control needs to be adapted for each scenario. This example motivates this thesis's goal to find a model that is able to detect the different scenarios and apply the optimal control on each.

Policy optimization for control of complex dynamical systems requires a policy search framework that is capable of capturing the system's dynamics. Since the dynamical systems can be broken into simpler units [1], the policy can be extended with additional structure that makes use of the simpler system units. This is comparable to a car's transmission system where the gear is chosen according to the specific scenario (driving backwards, accelerating, constant speed etc.).

This thesis proposes a Bayesian extension of HMM. These models are able to capture the complex system dynamics into simpler units with switching latent states. Furthermore, we combine these models with policy search to obtain a policy with additional structure.

In Chapter 2, we give an overview of the work related to ours.

In Chapter 3, we explain the basics of HMMs by showing the model structure, how to train and do inference with them. Moreover, we outline the method of variational inference and how this method can be applied by making use of mean-field approximation. Lastly, we describe a recent method that is able to reduce variance for score-function estimator gradient [2].

In Chapter 4, we derive the Bayesian HMMs. Starting the derivation from the classical Variational Bayes hidden Markov model (VBHMM), extending this derivation for the VBAR-HMM and finalizing it with the VBrAR-HMM.

In Chapter 5, we propose a policy search algorithm, called SWAC that requires a switching state space model. SWAC is internally structured with multiple policies. The number of

policies corresponds to the number of the switching state spaces. Besides, we show how to extend A2C and Soft actor-critic (SAC) with SWAC.

Chapter 6 gives a switching state detection and a prediction performance analysis of the introduced Bayesian HMMs from Chapter 4. We compare their prediction performance to SOTA models. In addition, we train Recurrent autoregressive hidden Markov model (rAR-HMM)s from true environment observations. We perform policy search on the rAR-HMM dynamics and evaluate the trained policies on the true dynamics. Lastly, we compare the learning performance of A2C and SAC with the extension of SWAC and without this extension. We use the pendulum and the cart-pole [3] environments for the evaluation.

Chapter 7 gives an outlook of future work and in Chapter 8 we conclude this thesis.

2. Related work

2.1. Bayesian HMMs

[4] derives a solution for Bayesian HMMs with discrete observations. He showed the advantages of being Bayesian compared to non Bayesian HMMs. The HMM is Bayesian in the sense that a prior distribution is put over the parameters and the Evidence lower bound (ELBO) is maximized with respect to (w.r.t.) the assumed distribution parameters. Beal shows the performance of automatically detecting the amount of hidden states that are at least required to describe the data. He assumes a Dirichlet distribution over the state transitions. Makes use of the property of the Dirichlet distribution that cancels out the number of states that are not required to express the data without overfitting. In addition, being Bayesian has the advantage to express the uncertainty of the learned model and is thus is less prone to overfitting. The obtained uncertainty can be used for exploration tasks.

2.2. Model-based reinforcement learning

The goal of reinforcement learning is to compute a policy that returns an optimal control signal for an underlying dynamical model to achieve a defined task such as a stabilization task.

Learning the dynamical model and integrating it into the policy search procedure is called Model-based reinforcement learning (MBRL). Making use of the learned model can result to a more sample efficient policy search.

[5] combines the learned dynamical model and uses a deterministic policy to obtain an analytic gradient for the policy parameter update. They use Gaussian process (GP) to capture the model dynamics. Moreover, they make use of the uncertainty of the GPs and

incorporate this quantity in a planning step. The planning step involves to use the learned the dynamics model to predict the dynamics behavior with a control signal obtained from the policy.

[6] formulate a general formulation for the combination of a deterministic policy and learning the model dynamics. The long-term reward is $J_{\boldsymbol{\theta}} = \sum_{t=1}^{T} \gamma^t r(\mathbf{x}_t)$ where $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t)$. $f(.)$ is the learned dynamical model, $\mathbf{x}$ are the observations and $\mathbf{u}$ is the control signal from the deterministic policy. Applying the chain rule on $J_{\boldsymbol{\theta}}$ w.r.t. to the policy parameters $\boldsymbol{\theta}$ gives:

$$\frac{\partial J_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T} \gamma^t \frac{\partial r(\mathbf{x}_t)}{\partial \mathbf{x}_t} \left(\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{x}_t}{\partial \mathbf{u}_t} \frac{\partial \mathbf{u}_t}{\partial \boldsymbol{\theta}} \right) \tag{2.1}$$

where we can differentiate through the learned dynamics model and the deterministic policy.

Recent research makes use of Variational autoencoders (VAE) [7, 8, 9] as a general model that captures the model dynamics. VAEs enable the differentiation through the learned dynamics model in a simple fashion that does not require specific model specification [10].

2.3. Switching linear dynamical systems

Learning time series data emitted from a dynamical system, requires a model design that is capable of capturing the dynamical behavior over time. [1] introduce the extension of a switching linear-Gaussian dynamical systems (SLDS)[11, 12, 13, 14, 15] with a recurrent connection where the next latent discrete state depends on the current continuous latent states. The recurrent connection is captured using logistic regression. This extension results in a more interpretable model. The data generated with this model better expresses the true dynamics.

2.4. Pathwise gradient estimator

Gradient estimation for parameter optimization of a distribution function can lead to gradients with a large variance when using estimators such as the score function estimator [2]. A less general solution are pathwise gradient estimators. They make use of the

structure of the problem and can thus lead to minimizing the variance of the estimated gradient.
The idea is to push in the drawn sample from the target distribution into the cost function and differentiate the cost function w.r.t. the distribution function parameters. The drawn sample needs to be reparameterized such that the differentiation is possible. The sampling process becomes:

$$\hat{\mathbf{x}} \sim p(\mathbf{x}; \boldsymbol{\theta}) \quad \equiv \quad \hat{\mathbf{x}} = g(\hat{\boldsymbol{\epsilon}}, \boldsymbol{\theta}),\ \hat{\boldsymbol{\epsilon}} \sim p(\boldsymbol{\epsilon})$$

where $g(.)$ is a differentiable function with parameters $\boldsymbol{\theta}$ and $p(\boldsymbol{\epsilon})$ is independent of the parameters $\boldsymbol{\theta}$.

[16] shows the reparameterization procedure for Gaussian distribution for continuous data. [17] and [18] simultaneously introduced a relaxation of the uniform distribution that made gradient passing possible for discrete data.

[19] and [20] make use of the pathwise gradient estimator to profit from the reduced variance of this gradient estimator. In Section 3.3 we explain their approach in detail.

3. Background

3.1. Hidden Markov models

HMMs are probabilistic models that expresses the switching dynamics of time series data under the Markov decision property (MDP) assumption. HMMs find their application such as in a wide range of fields such as capturing robot dynamics, weather forecasting, speech recognition, financial time series prediction.

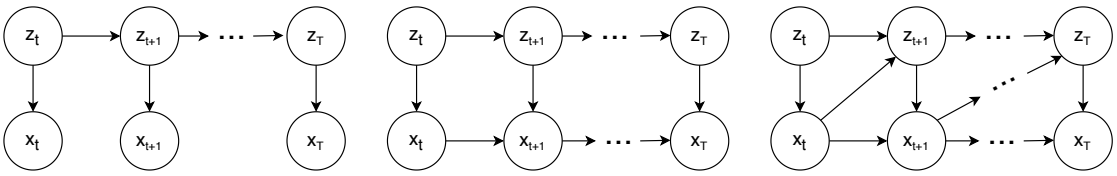
This section provides the basics for HMMs. We present the basic parameters and show how to apply HMMs for learning and inference on Gaussian independently and identically distributed (i.i.d.) data.

3.1.1. Overview of HMM types

In this thesis we cover three types of HMMs. The classical HMM depicted in Figure 3.1a where the observation $\mathbf{x}_t$ are emitted from the discrete latent states z_t. A linear dependence from the current observation $\mathbf{x}_t$ to the next observation $\mathbf{x}_{t+1}$ is the Autoregressive hidden Markov model (AR-HMM) as represented in Figure 3.1b. Figure 3.1c shows the rAR-HMM which extends the AR-HMM with a recurrent link of the current observation $\mathbf{x}_t$ to the next discrete latent state z_{t+1}.

3.1.2. Structure of HMMs

A HMM has captures the transitions between the latent variables $\mathbf{z}$. The observations $\mathbf{x}$ are emissions from the latent variables. Figure 3.1a shows the dependencies of a HMM.



(a) HMM (b) AR-HMM (c) rAR-HMM

Figure 3.1.: Three different types of HMMs. The models expressiveness is ordered from left to right.

In this thesis we focus on HMMs with Gaussian i.i.d. observations. The latent variables $\mathbf{z}$ absorb the Gaussian parameters. This model is similar to a Gaussian mixture model (GMM). The difference is the dependence to discrete time steps between of the data points.

Joint-distribution

The joint-distribution of a HMM is [21]:

$$p(\mathbf{z}_{1:T}, \mathbf{x}_{1:T}) = p(z_1) \prod_{t=2}^{T} p(z_t|z_{t-1}) p(\mathbf{x}_t|z_t)$$

where $p(z_1)$ is the initial state probability vector, $p(z_t|z_{t-1})$ is the transition probability and $p(\mathbf{x}_t|z_t)$ are the observation probabilities.
The initial state density is:

$$p(z_1) = \prod_{k=1}^{K} \pi_k^{\mathbb{I}(z_1=k)}$$

The density for Gaussian observation becomes:

$$p(\mathbf{x}_t|z_t) = \prod_{k=1}^{K} \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)^{\mathbb{I}(z_t=k)}$$

and the transition density becomes:

$$p(z_t|z_{t-1}) = \prod_{k=1}^{K} \prod_{j=1}^{K} a_{kj}^{\mathbb{I}(z_t=j, z_{t-1}=k)}, \quad a_{kj} = p(z_t = j|z_{t-1} = k)$$

where $z_t \in 1, ..., K$. K is the maximum number of discrete latent states in the model and $\mathbf{A}$ is the transition matrix with $a_{kj} \in \mathbf{A}$. We abbreviate the model parameters with $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{A})$

3.1.3. Baum-Welch algorithm - learning a HMM

The Baum-Welch algorithm first introduced in [22] is the extension of the EM algorithm to HMMs. The Baum-Welch algorithm iteratively learns estimates of the model probabilities shown in the joint-distribution equation above.

3.1.3.1. E-step

Using the above defined model densities, we get the estimated log complete data likelihood from:

$$\begin{aligned} \log p(\mathbf{z}_{1:T}, \mathbf{x}_{1:T}|\boldsymbol{\theta}) = &\sum_{k=1}^{K} \mathbb{E}\left[\mathbb{I}(z_1 = k)\right] \log \pi_k \\ &+ \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{j=1}^{K} \mathbb{E}\left[\mathbb{I}(z_{t-1} = k, z_t = j)\right] \log a_{kj} \\ &+ \sum_{t=1}^{T} \sum_{k=1}^{K} \mathbb{E}\left[\mathbb{I}(z_t = k)\right] \log p(\mathbf{x_t}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \end{aligned} \tag{3.1}$$

We make use of dynamic programming to solve the estimates from Equation (3.1).
First, we compute the forward probabilities where we define the log posterior of the hidden states as follows:

$$\alpha_t(k) = p(z_t = k|\mathbf{x}_{1:t}) = \frac{p(z_t = k|\mathbf{x}_{1:t-1})p(\mathbf{x}_t|z_t = k)}{\sum_{j=1}^{K} p(z_t = j|\mathbf{x}_{1:t-1})p(\mathbf{x}_t|z_t = j)}$$

where $p(z_t = k|\mathbf{x}_{1:t-1})$ is the one-step ahead predictive from the start $t = 1$ until density given by:

$$
\begin{aligned}
p(z_t = k|\mathbf{x}_{1:t-1}) &= \sum_{j=1}^{K} p(z_t = k|z_{t-1} = j)p(z_{t-1} = j|\mathbf{x}_{1:t-1}) \\
&= \sum_{j=1}^{K} p(z_t = k|z_{t-1} = j)\alpha_{t-1}(j)
\end{aligned}
$$

We set the prior $\alpha_1(z_1) = 1$ and obtain the final forward pass solution:

$$
\alpha_t(k) \propto p(\mathbf{x}_t|z_t = k) \sum_{j=1}^{K} p(z_t = k|z_{t-1} = j)\alpha_{t-1}(j)
$$

Similar to the forward message passing, we compute the backward messages from $t = T, ..., 1$ which is the conditional likelihood of future evidence:

$$
\begin{aligned}
\beta_t(k) &= p(\mathbf{x}_{t+1:T}|z_t) \\
&= \sum_{j=1}^{K} p(z_t = j, \mathbf{x}_t, \mathbf{x}_{t+1:T}|z_{t-1} = k) \\
&= \sum_{j=1}^{K} p(\mathbf{x}_{t+1:T}|z_t = j)p(z_t = j, \mathbf{x}_t|z_{t-1} = k) \\
&= \sum_{j=1}^{K} p(\mathbf{x}_{t+1:T}|z_t = j)p(\mathbf{x}_t|z_t = j)p(z_t = j|z_{t-1} = k) \\
&= \sum_{j=1}^{K} \beta_t(j)p(\mathbf{x}_t|z_t = j)p(z_t = j|z_{t-1} = k)
\end{aligned}
$$

where the end condition is $\beta_T(k) = 1$.

We combine α_t and β_t and obtain the state marginal:

$$
\begin{aligned}
p(z_t = k|\mathbf{x}_{1:T}) &\propto p(z_t = k|\mathbf{x}_{1:t})p(\mathbf{x}_{t+1:T}|z_t = k) \\
&= \alpha_t(k)\beta_t(k)
\end{aligned}
\tag{3.2}
$$

and the pairwise state marginal:

$$
\begin{aligned}
p(z_{t-1} = k, z_t = k'|\mathbf{x}_{1:T}) &\propto p(z_{t-1} = k|\mathbf{x}_{1:t-1})p(\mathbf{x}_t|z_t = k')p(\mathbf{x}_t|z_t = k')p(\mathbf{x}_{t+1:T}|z_t = k') \\
&= \alpha_{t-1}(k)p(z_t = k'|z_t = k)p(\mathbf{x}_t|z_t = k')\beta_t(k')
\end{aligned}
\tag{3.3}
$$

Using the results from the forward-backward messages, we can solve the expectations from equation (3.1). We rewrite the expectations:

$$\gamma_t(k) = \mathbb{E}\left[\mathbb{I}(z_t = k)\right]$$
$$\xi_{t-1,t}(k,k') = \mathbb{E}\left[\mathbb{I}(z_{t-1} = k, z_t = k')\right]$$

and solve using the results from Equations (3.2) and (3.3):

$$\gamma_t(k) = \frac{\alpha_t(k)\beta_t(k)}{\sum_{j=1}^{K}\alpha_t(j)\beta_t(j)}$$

which is the posterior density of being in state k at time step t and

$$\xi_{t-1,t}(k,k') = \frac{\alpha_{t-1}(k)a_{kk'}p(\mathbf{x}_t|z_t = k', \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})\beta_t(k')}{\sum_{j=1}^{K}\sum_{j'}^{K}\alpha_{t-1}(j)a_{jj'}p(\mathbf{x}_t|z_t = j', \boldsymbol{\mu}_{j'}, \boldsymbol{\Sigma}_{j'})\beta_t(j')}$$

which is the state transition posterior density.

3.1.3.2. M-step

In the M-step we update the model parameters to a maximize the likelihood of the observations. We use the expectation results from the E-step for the model parameter updates. The M-step becomes:

$$\hat{\boldsymbol{\pi}}: \ \hat{\pi}_k = \gamma_1(k)$$
$$\hat{\boldsymbol{\mu}}: \ \hat{\boldsymbol{\mu}}_k = \frac{\sum_{t=1}^{T}\gamma_t(k)\mathbf{x}_t}{\sum_{t=1}^{T}\gamma_t(k)}$$
$$\hat{\boldsymbol{\Sigma}}: \ \hat{\boldsymbol{\Sigma}}_k = \frac{\sum_{t=1}^{T}\gamma_t(k)\mathbf{x}_t\mathbf{x}_t^T - \sum_{t=1}^{T}\gamma_t(k)\hat{\boldsymbol{\mu}}_k\hat{\boldsymbol{\mu}}_k^T}{\sum_{t=1}^{T}\gamma_t(k)}$$
$$\hat{\mathbf{A}}: \ \hat{a}_{kk'} = \frac{\sum_{t=2}^{T}\xi_{t-1,t}(k,k')}{\sum_{t=2}^{T}\sum_{k'=1}^{K}\xi_{t-1,t}(k,k')}$$

This EM routine converges to a local optimum regarding the model likelihood given the observations.

3.1.4. Viterbi algorithm - maximum likely states

The Viterbi algorithm comes in hand if we require the maximum likely latent state sequence given the observations and the trained HMM:

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}_{1:T}} p(\mathbf{z}_{1:T}|\mathbf{x}_{1:T})$$

where $\hat{\mathbf{z}}$ is the most likely path from the trained model.

3.2. Variational inference

Variational inference [23, 24, 25, 26, 27, 28] tackles the problem to find an approximate density posterior model given an assumed prior model. We assume a probabilistic graphical model with observations $\mathbf{x}$ that depend on parameterized distributions. The parameters are absorbed in the latent variables $\mathbf{z}$. If we are interested in the posterior $p(\mathbf{z}|\mathbf{x})$, we need to apply Bayes' rule and solve:

$$\begin{aligned} p(\mathbf{z}|\mathbf{x}) &= \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{\int p(\mathbf{x}|\mathbf{z}_i)p(\mathbf{z}_i)\,\mathrm{d}\mathbf{z}_i} \\ &= \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{x})} \end{aligned} \tag{3.4}$$

Marginalizing out $\mathbf{z}$ in the denominator can be intractable to compute, because the number of the latent variables $\mathbf{z}$ can be infinitely large.

One way to obtain a posterior solution is to utilize Monte Carlo sampling methods. This however can be slow to compute, because of the large amount of samples required to approximate or converge to the exact posterior [29].

Variational inference comes at hand if the goal is to approximate a posterior solution in a decent amount of time. To approximate the posterior, we need to introduce a variational distribution q and seek to minimize the Kullback–Leibler divergence (KL divergence) of p and q. This reformulates the integration problem of finding a posterior solution to an optimization problem as follows:

$$q^*(\mathbf{z}) = \underset{q(\mathbf{z})\in\mathcal{D}}{\arg\min}\, D_{\mathrm{KL}}(q(\mathbf{z})||\, p(\mathbf{z}|\mathbf{x}))$$

where $\mathcal{D}$ is a set of approximate distributions.

In Addition, we assume the priors are conjugate to the posteriors which is the case for distributions of the exponential families. Variational inference for non-conjugate models is discussed in [30].

3.2.1. The evidence lower bound

The objective (3.4) is hard to compute since the KL divergence is:

$$D_{\mathrm{KL}}(q(\mathbf{z})||\, p(\mathbf{z}|\mathbf{x})) = \mathbb{E}_q\left[\log q(\mathbf{z})\right] - \mathbb{E}_q\left[\log p(\mathbf{z},\mathbf{x})\right] + \log p(\mathbf{x}) \tag{3.5}$$

where we have the dependence on $\log p(\mathbf{x})$.

An equivalent formulation of the variational optimization problem instead of minimizing (3.5) is to maximize the evidence lower bound (ELBO) defined as:

$$\mathrm{ELBO}(q) = \mathbb{E}_q\left[\log p(\mathbf{z},\mathbf{x})\right] - \mathbb{E}_q\left[\log q(\mathbf{z})\right] \tag{3.6}$$

A property of the ELBO is that it lower-bounds the log evidence $\log p(\mathbf{x}) \geq \mathrm{ELBO}(q)$ which explains the name.

As [27] shows, combining the equations (3.5) and (3.6) using Jensen's inequality gives:

$$\log p(\mathbf{x}) = D_{\mathrm{KL}}(q(\mathbf{z})||\, p(\mathbf{z}|\mathbf{x})) + \mathrm{ELBO}(q)$$

This makes it clear that maximizing the ELBO is equivalent to minimizing the KL divergence of (3.5). We focus on the finding a solution w.r.t. maximizing the ELBO.

3.2.2. Mean-field approximation

We need to restrict the set of approximate distributions $q(\mathbf{z})$ to obtain a tractable solution for this optimization problem. Assuming that $q(\mathbf{z})$ factorizes as follows:

$$q(\mathbf{z}) = \prod_{i=1}^{M} q_i(\mathbf{z}_i) \tag{3.7}$$

where the elements of $\mathbf{z}$ are partitioned into disjoint groups.
Using this as the only assumption of $q(\mathbf{z})$. Bishop [31] derives a general solution for the variational optimization problem. The general solution is:

$$\log q_j(\mathbf{z}_j) = \mathbb{E}_{i \neq j}[\log p(\mathbf{x}, \mathbf{z})] + \text{const.} \tag{3.8}$$

The intuition of the general solution (3.8) is that we fix the the disjoint variational solution for $\log q_j(\mathbf{z}_j)$ and average over all other disjoint groups. The non fixed disjoint groups are called the free energy in this context as the word "free" describes the non fixed property.

3.3. Minimizing the gradient variance

We seek to compute the optimal control policy parameters $\boldsymbol{\theta}$ of a policy $\pi_{\boldsymbol{\theta}}$ using gradient based methods. This section gives an overview of the problems that arise applying gradient-based methods on model-free policy search and how to overcome them.

Score-function gradient estimator (REINFORCE)

The score-function estimator [2] is a general applicable solution to obtain policy parameter gradients:

$$\frac{\partial J(\pi_{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T} \frac{\partial \log \pi_{\boldsymbol{\theta}}(a_t|s_t)}{\partial \boldsymbol{\theta}} \sum_{t'=t}^{T} r(a_{t'}, s_{t'})$$

This is an unbiased estimator but it yields high variance because it does not use any information of $r(s)$ (reward function) to compute the gradient.

Variance reduction using control variates

Since $r(s)$ can point in arbitrary directions and thus causes a high variate gradient, we need to limit the arbitrariness of $r(s)$.
Control variates tackle this problem. As [32] shows, the ideal control variate for this variance problem correlates at the maximum with r(s). To obtain this result, we can use the state value function, that we can train using additive path reward samples of r(s).

LAX surrogate gradient

To remove additional variance from the estimated gradient, [20] introduced a control variate in form of a neural network with the objective to minimize the variance of the estimated reinforce gradient. The extension of reinforce with LAX is as follows:

$$\hat{g}_{\boldsymbol{\theta}}^{LAX} = \sum_{t=1}^{T} \frac{\partial \log \pi_{\boldsymbol{\theta}}(a_t|s_t)}{\partial \boldsymbol{\theta}} \left[\sum_{t'=t}^{T} r(a_{t'}, s_{t'}) - b(s_{t'}) - c_{\boldsymbol{\phi}}(a_{t'}, s_{t'}) \right] + \frac{\partial c_{\boldsymbol{\phi}}(a_{t'}, s_{t'})}{\partial \boldsymbol{\theta}}$$

As a_t is sampled from $\pi_{\boldsymbol{\theta}}$, we need to make reparameterizable to obtain the gradient of of $c_{\boldsymbol{\phi}}$ w.r.t. the parameters $\boldsymbol{\theta}$. This reparameterization yields the following form of $a_t = a(\epsilon_t, s_t, \boldsymbol{\theta})$ where $\epsilon_t \sim p(\epsilon_t)$ and ϵ_t does not depend on $\boldsymbol{\theta}$. The last term is forms a pathwise derivative

We compute the gradient for the neural network control variate as follows:

$$\hat{g}_{\boldsymbol{\phi}} = \frac{\partial \left(\hat{g}_{\boldsymbol{\theta}}^{LAX} \right)^2}{\partial \boldsymbol{\phi}}$$

This formulation of the gradient computation is implicitly defined to minimize the variance of the reinforce gradient.

4. Bayesian hidden Markov models

A HMM from Section 3.1 is able to approximate the dynamics of a model. Since we approximate the model dynamics, we do not know how uncertain the model behaves comparing to the true dynamics. Knowing the model uncertainty would be useful when we integrate the trained dynamics model into policy search e.g. for planning or exploration [33, 34, 5, 6, 35].

Making the model Bayesian by putting prior distributions on the model parameters enables to access uncertainty of the model.

In this chapter we derive Bayesian HMMs for all HMMs from Figure 3.1.

4.1. Variational Bayes HMM

In this section, we define a Variational Bayes HMM and derive the Expectation maximization (EM) update steps. The model we derive in this section extends the HMM model from section 3.1. We extend the model with priors on the model parameters. The complete derivation is in the Appendix B.1.

4.1.1. Model priors

We extend the model from section 3.1 by putting priors over the parameters $\boldsymbol{\pi}$, $\mathbf{A}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

A natural choice for the priors on $\boldsymbol{\pi}$ and on the rows of $\mathbf{A}$ are Dirichlet priors. In addition, the Dirichlet prior has the effect to implicitly adjust the number of the latent states that are required for the HMM to be express the data. The latent states that are not used

are virtually canceled out [4]. Details of the Dirichlet distribution can be found in the appendix C.1. For $p(\boldsymbol{\pi})$ giving:

$$p(\boldsymbol{\pi}) = Dir(\boldsymbol{\pi}|\boldsymbol{\omega}_{\mathbf{0}}^{(\boldsymbol{\pi})}) = C(\boldsymbol{\omega}_{\mathbf{0}}^{(\boldsymbol{\pi})}) \prod_{k=1}^{K} \boldsymbol{\pi}_k^{\omega_0^{(\boldsymbol{\pi})}-1}$$

and for $p(\mathbf{A})$ we get:

$$p(\mathbf{A}) = \prod_{k=1}^{K} Dir(a_{k1}, ..., a_{kK}|\boldsymbol{\omega}_{\mathbf{0}}^{(\mathbf{A})}) = \prod_{k=1}^{K} C(\boldsymbol{\omega}_{\mathbf{0}}^{(\mathbf{A})}) \prod_{j=1}^{K} a_{kj}^{\omega_0^{(\mathbf{A})}-1}$$

where the parameters are $\boldsymbol{\omega}_0^{(\boldsymbol{\pi})}$ and $\boldsymbol{\omega}_0^{(\mathbf{A})}$. The normalization constant is defined by $C(\boldsymbol{\omega}_0)$. A small value of ω_0 affects the posterior rather by the data than by the prior and this applies vice versa for large value of ω_0.

Moreover, we assume a Gaussian-Wishart prior on the mean vector and precision matrix for each Gaussian component, given by:

$$\begin{aligned} p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= p(\boldsymbol{\mu}|\boldsymbol{\Sigma})p(\boldsymbol{\Sigma}) \\ &= \prod_{k=1}^{K} \mathcal{N}\left(\boldsymbol{\mu}_k|\mathbf{m}_0, (\beta_0\boldsymbol{\Sigma}_k)^{-1}\right) \mathcal{W}\left(\boldsymbol{\Sigma}_k|\mathbf{W}_0, \nu_0\right) \end{aligned}$$

Details of the Wishart Distribution are defined in C.4.

Finally, the graphical model is given by:

$$\begin{aligned} \boldsymbol{\pi} &\sim Dir(\boldsymbol{\omega}_0^{(\boldsymbol{\pi})}) \\ \mathbf{A} &\sim Dir(\boldsymbol{\omega}_0^{(\mathbf{A})}) \\ \boldsymbol{\Sigma} &\sim \mathcal{W}(\mathbf{W}_0, \nu_0) \\ \boldsymbol{\mu} &\sim \mathcal{N}(\mathbf{m}_0, (\beta_0, \boldsymbol{\Sigma})^{-1}) \\ \mathbf{x} &\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}) \end{aligned}$$

Regarding the prior parameters of the graphical model, the posterior parameters with

dimensions are:

$$\begin{aligned}
\boldsymbol{\omega}_k^{(\boldsymbol{\pi})} &: 1 \\
\boldsymbol{\omega}_k^{(\mathbf{A})} &: K \\
\mathbf{W}_k &: D \times D \\
\nu_k &: 1 \\
\mathbf{m}_k &: D \times 1 \\
\beta_k &: 1
\end{aligned}$$

where k denotes the latent state of the VBAR-HMM, D the observation data dimension.

4.1.2. Target distribution

Using the above defined priors, the log target distribution of the VBHMM becomes:

$$\begin{aligned}
\log p(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = &\log p(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \mathbf{A}) \\
&+ \log p(\mathbf{A}) + \log p(\boldsymbol{\pi}) + \log p(\boldsymbol{\mu}|\boldsymbol{\Sigma}) \\
&+ \log p(\boldsymbol{\Sigma}) \\
\propto &\sum_{t=1}^{T} \sum_{k=1}^{K} [z_t = k] \log \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right) \\
&+ \sum_{k=1}^{K} [z_1 = k] \log \pi_k + \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{j=1}^{K} [z_{t-1} = k]\,[z_t = j] \log a_{kj} \\
&+ \log Dir(\boldsymbol{\pi}|\boldsymbol{\omega}_0^{(\boldsymbol{\pi})}) + \sum_{k=1}^{K} \log Dir(\mathbf{A}|\boldsymbol{\omega}_{0k}^{(\mathbf{A})}) \\
&+ \log \sum_{k=1}^{K} \mathcal{N}\left(\boldsymbol{\mu}_k|\mathbf{m}_0, (\beta_0 \boldsymbol{\Sigma}_k)^{-1}\right) + \log \sum_{k=1}^{K} \mathcal{W}\left(\boldsymbol{\Sigma}_k|\mathbf{W}_0, \nu_0\right)
\end{aligned}$$

With this assumption of the true model, we can define a variational distribution in the next section.

4.1.3. Mean-field approximation

We consider a variational posterior that factorizes between the latent variables and the model parameters, giving:

$$p(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{z}_{1:T}|\mathbf{x}_{1:T}) \approx q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})q(\mathbf{z}_{1:T})$$

Using this factorization, we apply the general solution from (3.8) to derive a solution for $q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $q(\mathbf{z}_{1:T})$ in the next section.

4.1.4. Variational E-step

To compute the expectations of parameters that are required for the E-step updates, we need to apply the general result from Equation (3.8) to solve $q(\mathbf{z}_{1:T})$ giving:

$$\begin{aligned}
\log q(\mathbf{z}_{1:T}) &\propto \mathbb{E}_{q(\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{A})}[\log p(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})] \\
&= \mathbb{E}_{q(\boldsymbol{\pi})}\left[\sum_{k=1}^{K} \mathbb{I}(z_1 = k) \log \boldsymbol{\pi}_k\right] + \mathbb{E}_{q(\mathbf{A})}\left[\sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{I}(z_{t-1} = k, z_t = j) \log a_{kj}\right] \\
&\quad + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\sum_{t=1}^{T}\sum_{k=1}^{K} \mathbb{I}(z_t = k) \log \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right)\right] \\
&= \sum_{k=1}^{K} \mathbb{I}(z_1 = k)\mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \boldsymbol{\pi}_k\right] + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{I}(z_{t-1} = k, z_t = j)\mathbb{E}_{q(\mathbf{A})}\left[\log a_{kj}\right] \\
&\quad + \sum_{t=1}^{T}\sum_{k=1}^{K} \mathbb{I}(z_t = k) \\
&\quad \left\{\frac{1}{2}\mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\log|\boldsymbol{\Sigma}_k|\right] - \frac{D}{2}\log(2\pi) - \frac{1}{2}\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[(\mathbf{x}_t - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x}_t - \boldsymbol{\mu}_k)\right]\right\}
\end{aligned}$$

It is simple to solve the expectations $\mathbb{E}_{q(\boldsymbol{\pi})}$, $\mathbb{E}_{q(\mathbf{A})}$ and $\mathbb{E}_{q(\boldsymbol{\Sigma})}$ because we can look up the expectations for the corresponding distributions that are written in C. This expectation

formulations yield:

$$\begin{aligned}
\log \widetilde{\boldsymbol{\pi}}_k \equiv \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \boldsymbol{\pi}_k\right] &= \int Dir(\boldsymbol{\pi}_k|\boldsymbol{\omega}^{(\boldsymbol{\pi})}) \log \boldsymbol{\pi}_k \, \mathrm{d}\boldsymbol{\pi}_k \\
&= \psi(\omega_k^{(\boldsymbol{\pi})}) - \psi(\sum_{k=1}^{K} \omega_k^{(\boldsymbol{\pi})}), \;\; \sum_{k=1}^{K} \widetilde{\boldsymbol{\pi}}_k \leq 1 \\
\log \widetilde{a}_{kj} \equiv \mathbb{E}_{q(\mathbf{A})}\left[\log a_{kj}\right] &= \int Dir(a_{kj}|\boldsymbol{\omega}^{(\mathbf{A})}) \log a_{kj} \, \mathrm{d}a_{kj} \\
&= \psi(\omega_{kj}^{(\mathbf{A})}) - \psi(\sum_{j=1}^{K} \omega_{kj}^{(\mathbf{A})}), \;\; \sum_{j=1}^{K} \widetilde{a}_{kj} \leq 1 \\
\log \widetilde{\boldsymbol{\Sigma}}_k \equiv \mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\log |\boldsymbol{\Sigma}_k|\right] &= \int \mathcal{W}\left(\boldsymbol{\Sigma}_k|\mathbf{W}_\mathbf{k}, \nu_k\right) \log |\boldsymbol{\Sigma}_k| \, \mathrm{d}\boldsymbol{\Sigma}_k \\
&= \sum_{i=1}^{D} \psi\left(\frac{\nu_k + 1 - i}{2}\right) + D\log 2 + \log |\mathbf{W}_k|
\end{aligned}$$

We derive a solution $\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}$ as follows:

$$\begin{aligned}
&\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[(\mathbf{x_t} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x_t} - \boldsymbol{\mu}_k)\right] \\
&\quad = \int \mathbb{E}_{q(\boldsymbol{\mu})}\left[(\boldsymbol{\mu}_k - \mathbf{x_t})^T \boldsymbol{\Sigma}_k (\boldsymbol{\mu}_k - \mathbf{x_t})\right] q(\boldsymbol{\Sigma}_k) \, \mathrm{d}\boldsymbol{\Sigma}_k
\end{aligned}$$

Using the equation (380) from [36] to solve the inner expectation with respect to $\boldsymbol{\mu}$ yields:

$$\begin{aligned}
\mathbb{E}_{q(\boldsymbol{\mu})}\left[(\boldsymbol{\mu}_k - \mathbf{x_t})^T \boldsymbol{\Sigma}_k (\boldsymbol{\mu}_k) - \mathbf{x_t}\right] &= (\mathbf{m}_k - \mathbf{x}_t)^T \boldsymbol{\Sigma}_k (\mathbf{m}_k - \mathbf{x}_t) + \mathbf{Tr}\left(\boldsymbol{\Sigma}_k (\beta_k^{-1} \boldsymbol{\Sigma}_k)^{-1}\right) \\
&= (\mathbf{m}_k - \mathbf{x}_t)^T \boldsymbol{\Sigma}_k (\mathbf{m}_k - \mathbf{x}_t) + D\beta_k^{-1}
\end{aligned}$$

Plugging this term back in the equation

$$\begin{aligned}
&\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[(\mathbf{x_t} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x_t} - \boldsymbol{\mu}_k)\right] \\
&\quad = \int \left\{(\mathbf{m}_k - \mathbf{x}_t)^T \boldsymbol{\Sigma}_k (\mathbf{m}_k - \mathbf{x}_t) + D\beta_k^{-1}\right\} q(\boldsymbol{\Sigma}_k) \, \mathrm{d}\boldsymbol{\Sigma}_k \\
&\quad = D\beta_k^{-1} + \nu_k (\mathbf{m}_k - \mathbf{x}_t)^T \mathbf{W}_k (\mathbf{m}_k - \mathbf{x}_t)
\end{aligned}$$

We used the expectation of the Wishart distribution C.4 for the solution.

4.1.5. Variational M-step

For the M-step updates we need to solve $\log q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Again make use of the the general solution from equation (3.8) giving:

$$\begin{aligned}
\log q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &\propto \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] \\
&= \log Dir(\boldsymbol{\pi}|\boldsymbol{\omega}_{\mathbf{0}}^{(\boldsymbol{\pi})}) + \sum_{k=1}^{K} \log Dir(\mathbf{A}|\boldsymbol{\omega}_{0k}^{(\mathbf{A})}) \\
&\quad + \sum_{k=1}^{K} \log \mathcal{N}\left(\boldsymbol{\mu}_k|\boldsymbol{m}_0, (\beta_0\boldsymbol{\Sigma}_k)^{-1}\right) + \sum_{k=1}^{K} \log \mathcal{W}\left(\boldsymbol{\Sigma}_k|\mathbf{W}_{\mathbf{0}}, \nu_0\right) \\
&\quad + \sum_{t=1}^{T}\sum_{k=1}^{K} \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\mathbb{I}(z_t = k)\right] \log \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right) \\
&\quad + \sum_{k=1}^{K} \left\{\mathbb{E}_{q(z_1)}\left[\mathbb{I}(z_1 = k)\right] \log \pi_k\right\} \\
&\quad + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\mathbb{I}(z_{t-1} = k, z_t = j)\right] \log a_{kj}
\end{aligned} \tag{4.1}$$

Since we have a HMM structure, we can solve the expectations from equation (B.1) using message passing. We use solutions from the equations (**??**) and (**??**). We rewrite equation (B.1):

$$\begin{aligned}
\log q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &\propto \log Dir(\boldsymbol{\pi}|\boldsymbol{\omega}_{\mathbf{0}}^{(\boldsymbol{\pi})}) + \sum_{k=1}^{K} \log Dir(\mathbf{A}|\boldsymbol{\omega}_{0k}^{(\mathbf{A})}) \\
&\quad + \sum_{k=1}^{K} \log \mathcal{N}\left(\boldsymbol{\mu}_k|\boldsymbol{m}_0, (\beta_0\boldsymbol{\Sigma}_k)^{-1}\right) + \sum_{k=1}^{K} \log \mathcal{W}\left(\boldsymbol{\Sigma}_k|\mathbf{W}_{\mathbf{0}}, \nu_0\right) \\
&\quad + \sum_{t=1}^{T}\sum_{k=1}^{K} \gamma_t(k) \log \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right) + \sum_{k=1}^{K} \gamma_1(k) \log \pi_k \\
&\quad + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \xi_{t-1,t}(k, j) \log a_{kj}
\end{aligned} \tag{4.2}$$

Onwards we directly plug in the densities from the message passing step when expectations are of the above form.

We observe the right-hand side decomposes into a sum of terms only involving $\boldsymbol{\pi}$ together, $\mathbf{A}$ together and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ together. This observation implies that the variational posterior $q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ factorizes to give $q(\boldsymbol{\pi})q(\mathbf{A})q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = q(\boldsymbol{\pi})q(\mathbf{A}) \prod_{k=1}^{K} q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{4.3}$$

Using this factorization and combine it with the results of the applied mean-field factorization, we obtain the solutions for the VBM step.

Regarding the right-hand side of equation (4.2) that depend on $\boldsymbol{\pi}$, we get

$$\begin{aligned}
\log q(\boldsymbol{\pi}) &\propto (\omega_0^{(\boldsymbol{\pi})} - 1) \sum_{k=1}^{K} \log \boldsymbol{\pi}_k + \sum_{k=1}^{K} \gamma_1(k) \log \boldsymbol{\pi}_k \\
&= \sum_{k=1}^{K} \log \boldsymbol{\pi}_k \left\{ \left(\omega_0^{(\boldsymbol{\pi})} - 1 \right) + \gamma_1(k) \right\}
\end{aligned}$$

Taking the exponential of both sides, we obtain

$$\begin{aligned}
q(\boldsymbol{\pi}) &= \text{const.} \cdot \prod_{k=1}^{K} \boldsymbol{\pi}_k^{\omega_0^{(\boldsymbol{\pi})} - 1 + \gamma_1(k)} \\
&= \text{const.} \cdot \prod_{k=1}^{K} \boldsymbol{\pi}_k^{\omega_k^{(\boldsymbol{\pi})} - 1} \\
&= Dir(\boldsymbol{\pi} | \boldsymbol{\omega}^{(\boldsymbol{\pi})})
\end{aligned}$$

where const. is the normalization term of the Dirichlet distribution and $\boldsymbol{\omega}^{(\boldsymbol{\pi})}$ has components $\omega_k^{(\boldsymbol{\pi})}$ given by:

$$\omega_k^{(\boldsymbol{\pi})} = \omega_0^{(\boldsymbol{\pi})} + \gamma_1(k)$$

Applying the same principle on $\log q(\mathbf{A})$ we get

$$\begin{aligned}\log q(\mathbf{A}) &\propto \sum_{k=1}^{K}\sum_{j=1}^{K}(\omega_0^{(\mathbf{A})}-1)\log a_{kj} + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K}\xi_{t-1,t}(k,j)\log a_{kj} \\ &= \sum_{k=1}^{K}\sum_{j=1}^{K}\log a_{kj}\left\{(\omega_0^{(\mathbf{A})}-1)+\sum_{t=2}^{T}\xi_{t-1,t}(k,j)\right\}\end{aligned}$$

Taking the exponential on both sides gives

$$\begin{aligned}q(\mathbf{A}) &= \prod_{k=1}^{K}\text{const.}\cdot\prod_{j=1}^{K}a_{kj}^{\omega_0^{(\mathbf{A})}-1+\sum_{t=2}^{T}\xi_{t-1,t}(k,j)} \\ &= \prod_{k=1}^{K}\text{const.}\cdot\prod_{j=1}^{K}a_{kj}^{\omega_{kj}^{(\mathbf{A})}-1} \\ &= \prod_{k=1}^{K}Dir(\mathbf{A}|\boldsymbol{\omega}_k^{(\mathbf{A})})\end{aligned}$$

where $\boldsymbol{\omega}_{\boldsymbol{k}}^{(\mathbf{A})}$ has components $\omega_{kj}^{(\mathbf{A})}$ given by:

$$\omega_{kj}^{(\mathbf{A})} = \omega_0^{(\mathbf{A})} + \sum_{t=2}^{T}\xi_{t-1,t}(k,j)$$

To obtain a solution for $q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, we use the result from (4.2) and take only the terms that depend on $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$.

$$
\begin{aligned}
\log q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \propto & \log \mathcal{N}\left(\boldsymbol{\mu}_k | \boldsymbol{m}_0, (\beta_0 \boldsymbol{\Sigma}_k)^{-1}\right) + \log \mathcal{W}\left(\boldsymbol{\Sigma}_k | \mathbf{W}_0, \nu_0\right) \\
& + \sum_{t=1}^{T} \gamma_t(k) \log \mathcal{N}\left(\mathbf{x}_t | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right) \\
= & -\frac{D}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{\beta_0}{2}(\boldsymbol{\mu}_k - \boldsymbol{m}_0)^T \boldsymbol{\Sigma}_k (\boldsymbol{\mu}_k - \boldsymbol{m}_0) \\
& - \frac{1}{2}\mathbf{Tr}\left\{\boldsymbol{\Sigma}_k \mathbf{W}_0^{-1}\right\} + \frac{(\nu_0 - D - 1)}{2}\log|\boldsymbol{\Sigma}_k| \\
& - \frac{1}{2}\sum_{t=1}^{T} \gamma_t(k)(\mathbf{x}_t - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x}_t - \boldsymbol{\mu}_k) \\
& + \frac{1}{2}\left(\sum_{t=1}^{T} \gamma_t(k)\right)\log|\boldsymbol{\Sigma}_k|
\end{aligned}
$$

We make use of the product rule to obtain $\log q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \log q(\boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k) + \log q(\boldsymbol{\Sigma}_k)$. First we write $\log q(\boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k)$ by only making use of the terms that depend on $\boldsymbol{\mu}_k$. We separate the dependencies in quadratic and linear terms.

$$
\begin{aligned}
\log q(\boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k) = & \underbrace{-\frac{1}{2}\left(\boldsymbol{\mu}_k^T \left(\beta_0 \boldsymbol{\Sigma}_k\right) \boldsymbol{\mu}_k\right) - \frac{1}{2}\sum_{t=1}^{T} \gamma_t(k) \boldsymbol{\mu}_k^T \boldsymbol{\Sigma} \boldsymbol{\mu}_k}_{\text{quadratic term}} \\
& \underbrace{+\boldsymbol{\mu}_k^T \left(\beta_0 \boldsymbol{\Sigma}_k \mathbf{m}_0\right) + \sum_{t=1}^{T} \gamma_t(k) \boldsymbol{\mu}_k^T (\boldsymbol{\Sigma}_k \mathbf{x}_t)}_{\text{linear term}} \\
= & \underbrace{-\frac{1}{2}\left\{\boldsymbol{\mu}_k^T \left(\beta_0 + \sum_{t=1}^{T} \gamma_t(k)\right) \boldsymbol{\Sigma}_k \boldsymbol{\mu}_k\right\}}_{\text{quadratic term}} \\
& \underbrace{+\boldsymbol{\mu}_k^T \boldsymbol{\Sigma}_k \left(\beta_0 \mathbf{m}_0 + \sum_{t=1}^{T} \gamma_t(k) \mathbf{x}_t\right)}_{\text{linear term}}
\end{aligned}
$$

The defined form of $q(\boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k)$ is Gaussian, giving

$$
q(\boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k) = \mathcal{N}\left(\boldsymbol{\mu}_k | \mathbf{m}_k, \beta_k \boldsymbol{\Sigma}_k\right)
$$

We separate this Gaussian in quadratic and linear terms and use this result to obtain equations to compute β_k and $\mathbf{m}_k$. Starting to equate the quadratic terms yields:

$$\beta_k = \beta_0 + \sum_{t=1}^{T} \gamma_t(k)$$

We repeat this procedure to obtain a solution for $\mathbf{m}_k$ by equating the linear terms.

$$\mathbf{m}_k = \frac{1}{\beta_k}\left(\beta_0 \mathbf{m}_0 + \sum_{t=1}^{T} \gamma_t(k)\mathbf{x}_t\right)$$

To obtain $q(\boldsymbol{\Sigma}_k)$ we subtract $q(\boldsymbol{\mu}_k|\boldsymbol{\Sigma}_k)$ from $q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ focusing only on the terms that depend on $\boldsymbol{\Sigma}_k$

$$\begin{aligned}
\log q(\boldsymbol{\Sigma}_k) &= \log \mathcal{W}\left(\boldsymbol{\Sigma}_k|\mathbf{W}_k, \nu_k\right) \\
&= \log q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) - \log q(\boldsymbol{\mu}_k|\boldsymbol{\Sigma}_k) \\
&\propto \frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{\beta_0}{2}(\boldsymbol{\mu}_k - \boldsymbol{m}_0)^T\boldsymbol{\Sigma}_k(\boldsymbol{\mu}_k - \boldsymbol{m}_0) \\
&\quad - \frac{1}{2}\mathbf{Tr}\left\{\boldsymbol{\Sigma}_k\mathbf{W}_0^{-1}\right\} + \frac{(\nu_0 - D - 1)}{2}\log|\boldsymbol{\Sigma}_k| \\
&\quad - \frac{1}{2}\sum_{t=1}^{T}\gamma_t(k)(\mathbf{x}_t - \boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k(\mathbf{x}_t - \boldsymbol{\mu}_k) \\
&\quad + \frac{1}{2}\left(\sum_{t=1}^{T}\gamma_t(k)\right)\log|\boldsymbol{\Sigma}_k| \\
&\quad + \frac{\beta_k}{2}(\boldsymbol{\mu}_k - \mathbf{m}_k)^T\boldsymbol{\Sigma}_k(\boldsymbol{\mu}_k - \mathbf{m}_k) - \frac{1}{2}\log|\boldsymbol{\Sigma}_k| \\
&= \frac{(\nu_k - D - 1)}{2}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}\mathbf{Tr}\left(\boldsymbol{\Sigma}_k\mathbf{W}_k^{-1}\right)
\end{aligned}$$

where $\mathbf{W}_k^{-1}$ is defined by:

$$\mathbf{W}_k^{-1} = \mathbf{W}_0^{-1}N_k\mathbf{S}_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(\overline{\mathbf{x}}_k - \mathbf{m}_0)(\overline{\mathbf{x}}_k - \mathbf{m}_0)^T$$

where we defined:

$$
\begin{aligned}
N_k &= \sum_{t=1}^{T} \gamma_t(k) \\
\overline{\mathbf{x}}_k &= \frac{1}{N_k} \sum_{t=1}^{T} \gamma_t(k) \mathbf{x}_t \\
\mathbf{S}_k &= \frac{1}{N_k} \sum_{t=1}^{T} \gamma_t(k) (\mathbf{x}_t - \overline{\mathbf{x}}_k)(\mathbf{x}_t - \overline{\mathbf{x}}_k)^T
\end{aligned}
$$

We arrange the terms to obtain the form of a Wishart distribution. This rearrangement gives the solution for ν_k:

$$
\nu_k = \nu_0 + \sum_{t=1}^{T} \gamma_t(k)
$$

The EM updates for the VBHMM are summarized in the Appendix A.1.

4.1.6. The lower bound

The ELBO for the VBHMM is:

$$
\begin{aligned}
\text{ELBO}_{\text{VBHMM}} = \int\int\int\int q\left(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \Bigg[\log \frac{p(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})} \\
+ \sum_{\mathbf{z}_{1:T}} q(\mathbf{z}_{1:T}) \log \frac{p(\mathbf{x}_{1:T}, \mathbf{z}_{1:T} | \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{q(\mathbf{z}_{1:T})} \Bigg] \mathrm{d}\boldsymbol{\pi}\, \mathrm{d}\mathbf{A}\, \mathrm{d}\boldsymbol{\mu}\, \mathrm{d}\boldsymbol{\Sigma} \\
= \mathbb{E}\left[\log p(\mathbf{x}_{1:T} | \mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] + \mathbb{E}\left[\log p(\mathbf{z}_{1:T} | \boldsymbol{\pi}, \mathbf{A})\right] + \mathbb{E}\left[\log p(\boldsymbol{\pi})\right] \\
+ \mathbb{E}\left[\log p(\mathbf{A})\right] + \mathbb{E}\left[\log p(\boldsymbol{\mu}, \boldsymbol{\Sigma})\right] - \mathbb{E}\left[\log q(\boldsymbol{\pi})\right] - \mathbb{E}\left[\log q(\mathbf{A})\right] \\
- \mathbb{E}\left[\log q(\boldsymbol{\mu}, \boldsymbol{\Sigma})\right] - \mathbb{E}\left[\log q(\mathbf{z}_{1:T})\right]
\end{aligned}
$$

All expectations are w.r.t. $q(\mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. The solutions for the expectations are derived in the variational E- and M-steps.

4.2. Variational Bayes ARHMM

The VBAR-HMM is an extension of the VBHMM. We add a linear dependency between each observation and its successive observation as illustrated in Figure 3.1b. In addition, the next observation becomes dependent on a control signal $\mathbf{u}$. In this section we define and derive a solution for the model extension. The complete derivation is in the Appendix B.2.

4.2.1. Model priors

Since the VBAR-HMM is an extension of the VBHMM, we can carry over the model priors from Section 4.1.1. In addition, we assume a linear weight matrix $\mathbf{U}$ for every component k s.t. $\mathbf{x}_t = \mathbf{U}_k \hat{\mathbf{x}}_{t-1}$ for $z_t = k$ where:

$$\hat{\mathbf{x}}_{t-1} = \begin{bmatrix} \mathbf{x}_{t-1} & \mathbf{u}_t & 1 \end{bmatrix}^T$$

We stack the previous observation $\mathbf{x}_{t-1}$, the successive control signal $\mathbf{u}_t$ with the constant value 1 to obtain a linear transition matrix $\mathbf{U}$ that contains all dependency information for $\mathbf{x}_{t+1}$

We make use of this reformulation to implicitly learn the linear regression constant in $\mathbf{U}$. Moreover

A Matrix-Normal prior is a typical choice for linear weight matrix as [37] shows. We choose a Matrix Normal prior with unknown row covariance matrix $\mathbf{V}$ giving:

$$\begin{aligned} p(\mathbf{U}, \mathbf{V}) &= p(\mathbf{U}|\mathbf{V})p(\mathbf{V}) \\ &= \prod_{k=1}^{K} \mathcal{MN}\left(\mathbf{U}_k|\mathbf{M}_0, \mathbf{V}_k^{-1}, \mathbf{K}_0\right) \mathcal{W}\left(\mathbf{V}_k|\mathbf{P}_0, \eta_0\right) \end{aligned}$$

where we assume a Wishart prior on $\mathbf{V}$. The prior parameter choices of the $\mathbf{P}_0$ and η_0 are equivalent to the choices for the Wishart prior parameters of Equation (4.1.1). For symmetry, we naturally set $\mathbf{M}_0 = 0$. Small values for the column covariance matrix $\mathbf{K}_0$ affect the density rather by the data than by the prior.

The updated graphical model for the VBAR-HMM is now given by:

$$
\begin{aligned}
\boldsymbol{\pi}_k &\sim Dir(\boldsymbol{\omega}_0^{(\boldsymbol{\pi})}) \\
\mathbf{A}_k &\sim Dir(\boldsymbol{\omega}_0^{(\mathbf{A})}) \\
\boldsymbol{\Sigma}_k &\sim \mathcal{W}(\mathbf{W}_0, \nu_0) \\
\boldsymbol{\mu}_k &\sim \mathcal{N}(\mathbf{m}_0, (\beta_0, \boldsymbol{\Sigma}_k)^{-1}) \\
\mathbf{V}_k &\sim \mathcal{W}(\mathbf{P}_0, \eta_0) \\
\mathbf{U}_k &\sim \mathcal{MN}(\mathbf{M}_0, \mathbf{V}_k^{-1}, \mathbf{K}_0) \\
\mathbf{x_1} &\sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}) \\
\mathbf{x_{2:T}} &\sim \mathcal{N}(\mathbf{U}_k \hat{\mathbf{x}}_{1:T-1}, \mathbf{V}_k^{-1})
\end{aligned}
$$

The resulting posterior parameters with dimensions for this graphical model are:

$$
\begin{aligned}
\omega_k^{(\boldsymbol{\pi})} &: 1 \\
\boldsymbol{\omega}_k^{(\mathbf{A})} &: K \\
\mathbf{W}_k &: D \times D \\
\nu_k &: 1 \\
\mathbf{m}_k &: D \times 1 \\
\beta_k &: 1 \\
\eta_k &: 1 \\
\mathbf{M}_k &: D \times (D + C + 1) \\
\mathbf{K}_k &: (D + C + 1) \times (D + C + 1)
\end{aligned}
$$

where k denotes the latent state of the VBAR-HMM, D the observation data dimension and C the dimension of the control signal $\mathbf{u}$.

4.2.2. Target distribution

With the above the defined priors, the log target distribution of the VBAR-HMM becomes:

$$
\begin{aligned}
&\log p(\mathbf{x}_{1:T}, \hat{\mathbf{x}}_{2:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V}) \\
&\quad = \log p(\mathbf{x}_1 | z_1, \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \log p(\mathbf{x}_{2:T} | \hat{\mathbf{x}}_{2:T}, \mathbf{z}_{2:T}, \mathbf{U}, \mathbf{V}) \\
&\qquad + \log p(\mathbf{z}_{1:T} | \boldsymbol{\pi}, \mathbf{A}) + \log p(\boldsymbol{\pi}) + \log p(\mathbf{A}) \\
&\qquad + \log p(\boldsymbol{\mu} | \boldsymbol{\Sigma}) + \log p(\boldsymbol{\Sigma}) + \log p(\mathbf{U} | \mathbf{V}) + \log p(\mathbf{V}) \\
&\quad \propto \sum_{k=1}^{K} [z_1 = k] \log \mathcal{N}\left(\mathbf{x}_1 | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right) \\
&\qquad + \sum_{t=2}^{T} \sum_{k=1}^{K} [z_t = k] \log \mathcal{N}\left(\mathbf{x}_t | \mathbf{U}_k \hat{\mathbf{x}}_{t-1}, \mathbf{V}_k^{-1}\right) \\
&\qquad + \sum_{k=1}^{K} \{\mathbb{I}(z_1 = k) \log \pi_k\} + \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{j=1}^{K} \mathbb{I}(z_{t-1} = k, z_t = j) \log a_{kj} \\
&\qquad + \log Dir(\boldsymbol{\pi} | \boldsymbol{\omega}_{\mathbf{0}}^{(\boldsymbol{\pi})}) + \sum_{k=1}^{K} \log Dir(\mathbf{A} | \boldsymbol{\omega}_{0k}^{(\mathbf{A})}) \\
&\qquad + \log \sum_{k=1}^{K} \mathcal{N}\left(\boldsymbol{\mu}_k | \mathbf{m}_0, (\beta_0 \boldsymbol{\Sigma}_k)^{-1}\right) + \log \sum_{k=1}^{K} \mathcal{W}\left(\boldsymbol{\Sigma}_k | \mathbf{W}_0, \nu_0\right) \\
&\qquad + \sum_{k=1}^{K} \log \mathcal{MN}\left(\mathbf{U}_k | \mathbf{M}_0, \mathbf{V}_k^{-1}, \mathbf{K}_0\right) + \sum_{k=1}^{K} \log \mathcal{W}\left(\mathbf{V}_k | \mathbf{P}_0, \eta_0\right)
\end{aligned}
\tag{4.4}
$$

4.2.3. Mean-field approximation

We consider the same factorization assumption as in Section 4.1.3 giving:

$$p(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V}\mathbf{z}_{1:T} | \mathbf{x}_{1:T}) \approx q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V}) q(\mathbf{z}_{1:T})$$

Again, we apply the general result from Equation (3.8) to obtain solutions for the variational distributions.

4.2.4. Variational E-step

To compute the expectations of parameters that are required for the M-step updates, we need to apply the general result from Equation (3.8) to solve $q(\mathbf{z}_{1:T})$ giving:

$$
\begin{aligned}
\log q(\mathbf{z}_{1:T}) &\propto \mathbb{E}_{q(\boldsymbol{\pi},\mathbf{A},\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{U},\mathbf{V})}\left[\log p(\mathbf{x}_{1:T}, \hat{\mathbf{x}_{2:T}}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})\right] \\
&= \mathbb{E}_{q(\boldsymbol{\pi})}\left[\sum_{k=1}^{K} \mathbb{I}(z_1 = k) \log\left(\boldsymbol{\pi}_k\right)\right] \\
&\quad + \mathbb{E}_{q(\mathbf{A})}\left[\sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{I}(z_{t-1} = k, z_t = j) \log a_{kj}\right] \\
&\quad + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\sum_{k=1}^{K} [z_1 = k] \log \mathcal{N}\left(\mathbf{x}_1 | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)\right] \\
&\quad + \mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[\sum_{t=2}^{T}\sum_{k=1}^{K} \mathbb{I}(z_t = k) \log \mathcal{N}\left(\mathbf{x}_t | \mathbf{U}_k \hat{\mathbf{x}}_{t-1}, \mathbf{V}_k^{-1}\right)\right] \\
&= \sum_{k=1}^{K} \mathbb{I}(z_1 = k) \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \boldsymbol{\pi}_k\right] + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{I}(z_{t-1} = k, z_t = j) \mathbb{E}_{q(\mathbf{A})}\left[\log a_{kj}\right] \\
&\quad + \sum_{k=1}^{K} \mathbb{I}(z_1 = k) \\
&\quad \left\{\frac{1}{2}\mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\log |\boldsymbol{\Sigma}_k|\right] - \frac{D}{2}\log\left(2\pi\right) - \frac{1}{2}\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\left(\mathbf{x}_1 - \boldsymbol{\mu}_k\right)^T \boldsymbol{\Sigma}_k \left(\mathbf{x}_1 - \boldsymbol{\mu}_k\right)\right]\right\} \\
&\quad + \sum_{t=2}^{T}\sum_{k=1}^{K} \mathbb{I}(z_t = k) \\
&\quad \left\{\frac{1}{2}\mathbb{E}_{q(\mathbf{V})}\left[\log |\mathbf{V}_k|\right] - \frac{D}{2}\log\left(2\pi\right)\right. \\
&\quad \left.-\frac{1}{2}\mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[\left(\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1}\right)^T \mathbf{V}_k \left(\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1}\right)\right]\right\}
\end{aligned}
$$

We can simply take the expectation results from Section 4.1.4 that are:

$$\log \widetilde{\boldsymbol{\pi}}_k \equiv \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \boldsymbol{\pi}_k\right] = \psi(\omega_k^{(\boldsymbol{\pi})}) - \psi(\sum_{k=1}^{K} \omega_k^{(\boldsymbol{\pi})}), \;\; \sum_{k=1}^{K} \widetilde{\boldsymbol{\pi}}_k \leq 1 \tag{4.5}$$

$$\log \widetilde{a}_{kj} \equiv \mathbb{E}_{q(\mathbf{A})}\left[\log a_{kj}\right] = \psi(\omega_{kj}^{(\mathbf{A})}) - \psi(\sum_{j=1}^{K} \omega_{kj}^{(\mathbf{A})}), \;\; \sum_{j=1}^{K} \widetilde{a}_{kj} \leq 1 \tag{4.6}$$

$$\log \widetilde{\boldsymbol{\Sigma}}_k \equiv \mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\log |\boldsymbol{\Sigma}_k|\right] = \sum_{i=1}^{D} \psi\left(\frac{\nu_k + 1 - i}{2}\right) + D\log 2 + \log |\mathbf{W}_k| \tag{4.7}$$

$$\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[(\mathbf{x_1} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x_1} - \boldsymbol{\mu}_k)\right] = D\beta_k^{-1} + \nu_k (\mathbf{m}_k - \mathbf{x}_1)^T \mathbf{W}_k (\mathbf{m}_k - \mathbf{x}_1) \tag{4.8}$$

Since we put a Wishart prior on $\mathbf{V}_k$ the solution is equivalent to Equation (4.7) giving:

$$\log \widetilde{\mathbf{V}}_k \equiv \mathbb{E}_{q(\mathbf{V})}\left[\log |\mathbf{V}_k|\right] = \sum_{i=1}^{D} \psi\left(\frac{\eta_k + 1 - i}{2}\right) + D\log 2 + \log |\mathbf{P}_k|$$

We are left to solve the expectation for:

$$\mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[(\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1})^T \mathbf{V}_k (\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1})\right]$$

We rewrite:

$$\begin{aligned} &\mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[(\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1})^T \mathbf{V}_k (\mathbf{U}_k \hat{\mathbf{x}}_{t-1})\right] \\ &\quad = \int \mathbb{E}_{q(\mathbf{U})}\left[(\mathbf{U}_k \hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T \mathbf{V}_k (\mathbf{U}_k \hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)\right] q(\mathbf{V}_k)\, \mathrm{d}\mathbf{V}_k \end{aligned} \tag{4.9}$$

The inner expectation corresponds to a Matrix-Normal distribution. Thus, we make use of transformations derived in [38] and get two expectations:

$$\begin{aligned} \mathbb{E}\left[\mathbf{U}_k\right] &= \mathbf{M}_k \\ \mathbb{E}\left[\mathbf{U}_k \mathbf{Q} \mathbf{U}_k^T\right] &= \mathbf{M}_k \mathbf{U} \mathbf{M}_k^T + \mathbf{Tr}\left\{\mathbf{K}_k^{-1}\mathbf{Q}\right\} \mathbf{V}_k^{-1} \end{aligned}$$

Applying this transformations on the inner expectation from Equation (B.2) gives:

$$\begin{aligned} &\mathbb{E}_{q(\mathbf{U})}\left[(\mathbf{U}_k \hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T \mathbf{V}_k (\mathbf{U}_k \hat{\mathbf{x}}_{t-1} - \mathbf{x}_t\right] \\ &\quad = (\mathbf{M}_k \hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T \mathbf{V}_k (\mathbf{M}_k \hat{\mathbf{x}}_{t-1} - \mathbf{x}_t) + \mathbf{Tr}\left\{\mathbf{K}_k^{-1} \hat{\mathbf{x}}_{t-1} \hat{\mathbf{x}}_{t-1}^T\right\} \end{aligned}$$

Substituting back yields the final solution:

$$\begin{aligned}\mathbb{E}_{q(\mathbf{U},\mathbf{V})}&\left[(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})^T\,\mathbf{V}_k\,(\mathbf{U}_k\hat{\mathbf{x}}_{t-1})\right]\\ &= \eta_k(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T\mathbf{P}_k(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t) + \mathbf{Tr}\left\{\mathbf{K}_k^{-1}\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\right\}\end{aligned}$$

4.2.5. Variational M-step

For the M-step updates, we need to solve $\log q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})$. Applying the general result from Equation (3.8) gives:

$$\begin{aligned}\log q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V}) &\propto \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{x}_{1:T}, \mathbf{u}_{2:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})\right]\\ &= \log Dir(\boldsymbol{\pi}|\boldsymbol{\omega}_{\mathbf{0}}^{(\boldsymbol{\pi})}) + \sum_{k=1}^{K}\log Dir(\mathbf{A}|\boldsymbol{\omega}_{0k}^{(\mathbf{A})})\\ &\quad + \sum_{k=1}^{K}\gamma_1(k)\log\mathcal{N}\left(\mathbf{x}_1|\boldsymbol{\mu}_k, (\beta_0\boldsymbol{\Sigma}_k)^{-1}\right)\\ &\quad + \sum_{t=2}^{T}\sum_{k=1}^{K}\gamma_t(k)\log\mathcal{N}\left(\mathbf{x}_t|\mathbf{U}_k\hat{\mathbf{x}}_{t-1}, \mathbf{V}_k^{-1}\right)\\ &\quad + \sum_{k=1}^{K}\left\{\gamma_1(k)\log\pi_k\right\}\\ &\quad + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K}\xi_{t-1,t}(k,j)\log a_{ij}\\ &\quad + \sum_{k=1}^{K}\log\mathcal{MN}\left(\mathbf{U}_k|\mathbf{M}_0, \mathbf{V}_k^{-1}, \mathbf{K}_0\right) + \sum_{k=1}^{K}\log\mathcal{W}\left(\mathbf{V}_k|\mathbf{P}_0, \eta_0\right)\end{aligned} \tag{4.10}$$

We observe the right-hands side decomposes into a sum of terms only involving $\boldsymbol{\pi}$ together, $\mathbf{A}$ together, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ together and $\mathbf{U}$ and $\mathbf{V}$ together. This observation implies that the variational posterior $q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})$ factorizes to:

$$q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V}) = q(\boldsymbol{\pi})q(\mathbf{A})\prod_{k=1}^{K}q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\prod_{k=1}^{K}q(\mathbf{U}_k, \mathbf{V}_k) \tag{4.11}$$

As we already derived solutions for $q(\boldsymbol{\pi})$, $q(\mathbf{A})$ and $q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ in Section 4.1.4, we are left to solve $q(\mathbf{U}_k, \mathbf{V}_k)$.

Regarding the target distribution from Equation (4.3.2), we assume the model for $q(\mathbf{U}, \mathbf{V})$ has the form:

$$
\begin{aligned}
q(\mathbf{U}, \mathbf{V}) &= \prod_{k=1}^{K} q(\mathbf{U}_k, \mathbf{V}_k) = \prod_{k=1}^{K} q(\mathbf{U}_k|\mathbf{V}_k) q(\mathbf{V}_k) \\
&= \prod_{k=1}^{K} \mathcal{MN}\left(\mathbf{U}_k|\mathbf{M}_k, \mathbf{V}_k^{-1}, \mathbf{K}_k\right) \mathcal{W}\left(\mathbf{V}_k|\mathbf{P}_k, \eta_k\right) \\
&= \prod_{k=1}^{K} \frac{|\mathbf{K}_k|^{d/2}|\mathbf{V}_k|^{m/2}}{(2\pi)^{m/2}} \exp\left\{-\frac{1}{2}\mathbf{Tr}\left\{(\mathbf{U}_k - \mathbf{M}_k)^T \mathbf{V}_k (\mathbf{U}_k - \mathbf{M}_k)\mathbf{K}_k\right\}\right\} \\
&\qquad B(\mathbf{P}_k, \eta_k)|\mathbf{V}_k|^{(\eta_k - D - 1)/2} \exp\left\{-\frac{1}{2}\mathbf{Tr}\left\{\mathbf{P}_k^{-1}\mathbf{V}_k\right\}\right\}
\end{aligned}
$$

where $B(\mathbf{P}_k, \eta_k)$ is defined in C.4.

Solving $q(\mathbf{U}, \mathbf{V})$ requires to make use of result from Equation (4.11) and take only the terms from Equation (4.10) that depend on $\mathbf{U}_k$ and $\mathbf{V}_k$ giving:

$$
\begin{aligned}
\log q(\mathbf{U}_k, \mathbf{V}_k) \propto{}& \log \mathcal{MN}\left(\mathbf{U}_k|\mathbf{M}_0, \mathbf{V}_k^{-1}, \mathbf{K}_0\right) + \log \mathcal{W}\left(\mathbf{V}_k|\mathbf{P}_0, \eta_0\right) \\
&+ \sum_{t=2}^{T} \gamma_t(k) \log \mathcal{N}\left(\mathbf{x}_t|\mathbf{U}_k \hat{\mathbf{x}}_{t-1}, \mathbf{V}_k^{-1}\right) \\
={}& -\frac{1}{2}\mathbf{Tr}\left\{(\mathbf{U}_k - \mathbf{M}_0)^T \mathbf{V}_k (\mathbf{U}_k - \mathbf{M}_0)\mathbf{K}_0\right\} - \frac{m}{2}\log(|\mathbf{V}_k|) \\
&- \frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k \mathbf{W}_0^{-1}\right\} + \frac{(\eta_0 - D - 1)}{2}\log|\mathbf{V}_k| \\
&- \frac{1}{2}\sum_{t=2}^{T} \gamma_t(k)(\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1})^T \mathbf{V}_k (\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1}) \\
&+ \frac{1}{2}\left(\sum_{t=2}^{T} \gamma_t(k)\right)\log|\mathbf{V}_k|
\end{aligned}
$$

We make use of the product rule to obtain $\log q(\mathbf{U}_k, \mathbf{V}_k) = \log q(\mathbf{U}_k|\mathbf{V}_k) + \log q(\mathbf{V}_k)$. First we write $\log q(\mathbf{U}_k|\mathbf{V}_k)$ by only making use of the terms that depend on $\mathbf{U}_k$. We gather all terms in (4.2.5) that contain $\mathbf{U}_k$ giving:

$$
\begin{aligned}
\log q(\mathbf{U}_k|\mathbf{V}_k) \propto & -\frac{1}{2}\mathbf{Tr}\left\{(\mathbf{U}_k-\mathbf{M}_0)^T\mathbf{V}_k(\mathbf{U}_k-\mathbf{M}_0)\mathbf{K}_0\right\} \\
& -\frac{1}{2}\sum_{t=2}^{T}\gamma_t(k)(\mathbf{x}_t-\mathbf{U}_k\hat{\mathbf{x}}_t)^T\mathbf{V}_k(\mathbf{x}_t-\mathbf{U}_k\hat{\mathbf{x}}_t) \\
= & -\frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k(\mathbf{U}_k\mathbf{K}_0\mathbf{U}_k^T-\mathbf{U}_k\mathbf{K}_0\mathbf{M}_0^T-\mathbf{M}_0\mathbf{K}_0\mathbf{U}_k^T+\mathbf{M}_0\mathbf{K}_0\mathbf{M}_0^T)\right\} \\
& -\frac{1}{2}\sum_{t=2}^{T}\gamma_t(k)\mathbf{Tr}\left\{\mathbf{V}_k(\mathbf{x}_t\mathbf{x}_t^T-\mathbf{x}_t\hat{\mathbf{x}}_t^T\mathbf{U}_k^T-\mathbf{U}_k\hat{\mathbf{x}}_t\mathbf{x}_t^T+\mathbf{U}_k\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t^T\mathbf{U}_k^T)\right\}
\end{aligned}
$$

We separate the dependencies in quadratic and linear terms which yields:

$$
\begin{aligned}
\log q(\mathbf{U}_k|\mathbf{V}_k) = & \underbrace{-\frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k\mathbf{U}_k\left[\sum_{t=2}^{T}\gamma_t(k)\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t^T+\mathbf{K}_0\right]\mathbf{U}_k^T\right\}}_{\text{quadratic term}} \\
& \underbrace{-\frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k\mathbf{U}_k\left[-\sum_{t=2}^{T}\gamma_t(k)\mathbf{x}_t\hat{\mathbf{x}}_{t-1}^T-\mathbf{M}_0\mathbf{K}_0\right]\right\}}_{\text{linear term}}
\end{aligned}
$$

The defined form of $q(\mathbf{U}_k|\mathbf{V}_k)$ is Matrix Normal giving:

$$
\mathcal{MN}\left(\mathbf{U}_k|\mathbf{M}_k,\mathbf{V}_k^{-1},\mathbf{K}_k\right)
$$

We obtain a solution for the posterior parameter updates using the defined form and bringing Using this form, we obtain a solution for the posterior parameter updates.

$$
\mathbf{K}_k=\sum_{t=2}^{T}\gamma_t(k)\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T+\mathbf{K}_0
$$

We apply this procedure on the linear terms to obtain a solution for $\mathbf{M}_k$

$$\mathbf{M}_k = \left[\sum_{t=2}^{T} \gamma_t(k)\mathbf{x}_t\hat{\mathbf{x}}_{t-1}^T + \mathbf{M}_0\mathbf{K}_0\right]\mathbf{K}_k^{-1}$$

To obtain the parameters $\mathbf{P}_k$ and $\boldsymbol{\eta}_k$ of the Wishart distribution $q(\mathbf{V}_k)$, we subtract the conditional distribution from the joint distribution giving:

$$\begin{aligned}\log q(\mathbf{V}_k) &= \log \mathcal{W}\left(\mathbf{U}_k|\mathbf{P}_k, \eta_k\right) \\ &= \log q(\mathbf{U}_k, \mathbf{V}_k) - \log q(\mathbf{U}_k|\mathbf{V}_k) \\ &\propto \frac{(\eta_k - D - 1)}{2}\log|\mathbf{V}_k| - \frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k\mathbf{P}_k^{-1}\right\}\end{aligned}$$

where $\mathbf{P}_k^{-1}$ is defined as:

$$\mathbf{P}_k^{-1} = \mathbf{P}_0^{-1} + \mathbf{M}_0\mathbf{K}_0\mathbf{M}_0^T + \sum_{t=2}^{T}\gamma_t(k)\mathbf{x}_t\mathbf{x}_t^T - \mathbf{M}_k\mathbf{K}_k\mathbf{M}_k^T$$

We arange the terms to obtain a solution for η_k and $\mathbf{V}_k$. For η_k we get:

$$\eta_k = \eta_0 + \sum_{t=2}^{T}\gamma_t(k)$$

The EM updates for the VBAR-HMM are summarized in the Appendix A.2.

4.2.6. The evidence lower bound

The ELBO for the VBAR-HMM is:

$$\begin{aligned}
\text{ELBO}_{\text{VBAR-HMM}} = \int\int\int\int\int\int q\left(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V}\right) \Bigg[\log \frac{p(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})}{q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}), \mathbf{U}, \mathbf{V}} \\
+ \sum_{\mathbf{z}_{1:T}} q(\mathbf{z}_{1:T}) \log \frac{p(\mathbf{x}_{1:T}, \mathbf{z}_{1:T} | \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})}{q(\mathbf{z}_{1:T})} \Bigg] \\
\mathrm{d}\boldsymbol{\pi}\, \mathrm{d}\mathbf{A}\, \mathrm{d}\boldsymbol{\mu}\, \mathrm{d}\boldsymbol{\Sigma}\, \mathrm{d}\mathbf{U}\, \mathrm{d}\mathbf{V} \\
= \mathbb{E}\left[\log p(\mathbf{x}_1 | z_1, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] + \mathbb{E}\left[\log p(\mathbf{x}_{2:T} | \hat{\mathbf{x}}_{1:T-1} \mathbf{z}_{2:T}, \mathbf{U}, \mathbf{V})\right] \\
+ \mathbb{E}\left[\log p(\mathbf{z}_{1:T} | \boldsymbol{\pi}, \mathbf{A})\right] + \mathbb{E}\left[\log p(\boldsymbol{\pi})\right] + \mathbb{E}\left[\log p(\mathbf{A})\right] \\
+ \mathbb{E}\left[\log p(\boldsymbol{\mu}, \boldsymbol{\Sigma})\right] + \mathbb{E}\left[\log p(\mathbf{U}, \mathbf{V})\right] - \mathbb{E}\left[\log q(\boldsymbol{\pi})\right] - \mathbb{E}\left[\log q(\mathbf{A})\right] \\
- \mathbb{E}\left[\log q(\boldsymbol{\mu}, \boldsymbol{\Sigma})\right] - \mathbb{E}\left[\log q(\mathbf{U}, \mathbf{V})\right] - \mathbb{E}\left[\log q(\mathbf{z}_{1:T})\right]
\end{aligned}$$

All expectations are w.r.t. $q(\mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})$. The solutions for the expectations are derived in the variational E- and M-steps.

4.3. Variational Bayes rAR-HMM

The VBrAR-HMM similar to the VBAR-HMM. The difference is the added dependency link the observation $\mathbf{x}_t$ to the next state z_{t+1} as Figure 3.1c illustrates. Since z_{t+1} now depends on $\mathbf{x}_t$ and z_t, the latent state transition function has the property of being nonlinear.

4.3.1. Model priors

We introduce the function $f_{\boldsymbol{\phi}}(.)$ for the recurrent state transition which defines a Bayesian neural network with parameters $\boldsymbol{\phi}$. We can take the model priors from Section 4.2.1 and need to change the Dirichlet transition prior with a Gaussian prior we put on the neural

network weights ϕ giving the graphical model:

$$
\begin{aligned}
\boldsymbol{\pi}_k &\sim Dir(\boldsymbol{\omega}_0^{(\boldsymbol{\pi})}) \\
\boldsymbol{\phi} &\sim \mathcal{N}(\boldsymbol{\zeta}_0, \mathbf{Y}_0) \\
\boldsymbol{\Sigma}_k &\sim \mathcal{W}(\mathbf{W}_0, \nu_0) \\
\boldsymbol{\mu}_k &\sim \mathcal{N}(\mathbf{m}_0, (\beta_0, \boldsymbol{\Sigma}_k)^{-1}) \\
\mathbf{V}_k &\sim \mathcal{W}(\mathbf{P}_0, \eta_0) \\
\mathbf{U}_k &\sim \mathcal{MN}(\mathbf{M}_0, \mathbf{V}_k^{-1}, \mathbf{K}_0) \\
\mathbf{x_1} &\sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}) \\
\mathbf{x_{2:T}} &\sim \mathcal{N}(\mathbf{U}_k \hat{\mathbf{x}}_{1:T-1}, \mathbf{V}_k^{-1})
\end{aligned}
$$

The posterior parameters are equal to the posterior parameters from Subsection 4.2.1. Except the Dirichlet posterior $\omega^{(\mathbf{A})}$ is exchanged with the Bayesian neural network (BNN) posteriors $\boldsymbol{\zeta}$ and $\mathbf{Y}$. The dimensions of the BNN parameter posteriors corresponds to the number of hidden layers and parameters per layer.

4.3.2. Target distribution

With the above the defined priors, the log target distribution of the VBAR-HMM becomes:

$$
\begin{aligned}
&\log p(\mathbf{x}_{1:T},\mathbf{z}_{1:T}, \boldsymbol{\pi}, \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V}) \\
&\quad = \log p(\mathbf{x}_1|z_1, \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \log p(\mathbf{x}_{2:T}|\hat{\mathbf{x}}_{2:T}, \mathbf{z}_{2:T}, \mathbf{U}, \mathbf{V}) \\
&\qquad + \log p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \hat{\mathbf{x}}_{2:T}, \boldsymbol{\phi}) + \log p(\boldsymbol{\pi}) + \log p(\boldsymbol{\phi}) \\
&\qquad + \log p(\boldsymbol{\mu}|\boldsymbol{\Sigma}) + \log p(\boldsymbol{\Sigma}) + \log p(\mathbf{U}|\mathbf{V}) + \log p(\mathbf{V}) \\
&\quad \propto \sum_{k=1}^{K} [z_1 = k] \log \mathcal{N}\left(\mathbf{x}_1|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right) \\
&\qquad + \sum_{t=2}^{T}\sum_{k=1}^{K} [z_t = k] \log \mathcal{N}\left(\mathbf{x}_t|\mathbf{U}_k\hat{\mathbf{x}}_{t-1}, \mathbf{V}_k^{-1}\right) \\
&\qquad + \sum_{k=1}^{K} \{\mathbb{I}(z_1 = k) \log \pi_k\} + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{I}(z_{t-1} = k, z_t = j) \log \psi_{kj} \\
&\qquad + \log Dir(\boldsymbol{\pi}|\boldsymbol{\omega}_{\mathbf{0}}^{(\boldsymbol{\pi})}) + \sum_{k=1}^{K} \log \mathcal{N}\left(\boldsymbol{\phi}|\boldsymbol{\zeta}_0, \mathbf{Y}_0\right) \\
&\qquad + \log \sum_{k=1}^{K} \mathcal{N}\left(\boldsymbol{\mu}_k|\mathbf{m}_0, (\beta_0\boldsymbol{\Sigma}_k)^{-1}\right) + \log \sum_{k=1}^{K} \mathcal{W}\left(\boldsymbol{\Sigma}_k|\mathbf{W}_0, \nu_0\right) \\
&\qquad + \sum_{k=1}^{K} \log \mathcal{MN}\left(\mathbf{U}_k|\mathbf{M}_0, \mathbf{V}_k^{-1}, \mathbf{K}_0\right) + \sum_{k=1}^{K} \log \mathcal{W}\left(\mathbf{V}_k|\mathbf{P}_0, \eta_0\right)
\end{aligned}
$$

where $\psi_{k,j}$ is defined as follows:

$$
\psi_{k,j} = \frac{\exp(f_{\boldsymbol{\phi}}(\mathbf{x}_{t-1}, \mathbf{u}_t)_{k,j})}{\sum_{k'=1}^{K} \exp(f_{\boldsymbol{\phi}}(\mathbf{x}_{t-1}, \mathbf{u}_t)_{k,k'})} \tag{4.12}
$$

We sum over the columns for each row of the output matrix to obtain normalized transition probabilities.

4.3.3. Mean-field approximation

The model of the rARHMM only changes w.r.t. the transition function. The rARHMM model is equipped with a non-linear function. Thus, the factorization model remains

identical in all parts but the transition function. Applying the mean-field approximation to the rARHMM we get the following result:

4.3.4. Variational E-step

$$
\begin{aligned}
\log q(\mathbf{z}_{1:T}) \propto\ & \mathbb{E}_{q(\boldsymbol{\pi},\boldsymbol{\phi},\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{U},\mathbf{V})}\left[\log p(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})\right] \\
=\ & \sum_{k=1}^{K} \mathbb{I}(z_1 = k) \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \boldsymbol{\pi}_k\right] + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{I}(z_{t-1} = k, z_t = j) \mathbb{E}_{q(\boldsymbol{\phi})}\left[\log \psi_{kj}\right] \\
& + \sum_{k=1}^{K} \mathbb{I}(z_1 = k) \\
& \left\{ \frac{1}{2}\mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\log |\boldsymbol{\Sigma}_k|\right] - \frac{D}{2}\log(2\pi) - \frac{1}{2}\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[(\mathbf{x}_1 - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x}_1 - \boldsymbol{\mu}_k)\right] \right\} \\
& + \sum_{t=2}^{T}\sum_{k=1}^{K} \mathbb{I}(z_t = k) \\
& \left\{ \frac{1}{2}\mathbb{E}_{q(\mathbf{V})}\left[\log |\mathbf{V}_k|\right] - \frac{D}{2}\log(2\pi) \right. \\
& \left. -\frac{1}{2}\mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})^T \mathbf{V}_k (\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})\right] \right\}
\end{aligned}
$$

Since the rAR-HMM changes w.r.t. the transition function, only the estimation of the transition function differs from the AR-HMM model. We can not solve the estimation term in closed form due to its non linearity property as equation (4.12) shows. We make use of a non-linear function approximation framework to approximate the optimal solution. To remain in the Bayesian fashion, we choose a BNN [39, 40, 41] for this task.

We solve the estimation for $\mathbb{E}_{q(\boldsymbol{\phi})}\left[\log \psi_{ij}\right]$ by plugging in the observations $\mathbf{x}_{t-1}$ and $\mathbf{u}_t$ into the bayesian neural network $f_{\boldsymbol{\phi}}(.)$ and normalize the result to get proper probabilities.

4.3.5. Variational M-step

To optimize the parameters ϕ of the BNN, we make use of stochastic gradient function approximations. For this task, we choose to maximize the part of the lower bound that

depends on the network output. We write the objective as follows:

$$\max_{\boldsymbol{\phi}} \log p(\mathbf{z}_{2:T}|\boldsymbol{\phi}) = \max_{\boldsymbol{\phi}} \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{j=1}^{K} \xi_{t-1,t}(k,j) \log \psi_{kj} \tag{4.13}$$

This objective can be plugged in any gradient solver. Thus, a BNN is not mandatory for this task. The EM updates for the VBrAR-HMM are summarized in the Appendix A.3.

4.3.6. The evidence lower bound

The ELBO for the VBAR-HMM is:

$$\begin{aligned}
\text{ELBO}_{\text{VBAR-HMM}} = \int\int\int\int\int\int q\left(\boldsymbol{\pi}, \boldsymbol{phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V}\right) \Bigg[\log \frac{p(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})}{q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}), \mathbf{U}, \mathbf{V}} \\
+ \sum_{\mathbf{z}_{1:T}} q(\mathbf{z}_{1:T}) \log \frac{p(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}|\boldsymbol{\pi}, \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})}{q(\mathbf{z}_{1:T})} \Bigg] \\
\mathrm{d}\boldsymbol{\pi}\, \mathrm{d}\boldsymbol{\phi}\, \mathrm{d}\boldsymbol{\mu}\, \mathrm{d}\boldsymbol{\Sigma}\, \mathrm{d}\mathbf{U}\, \mathrm{d}\mathbf{V} \\
= \mathbb{E}\left[\log p(\mathbf{x}_1|z_1, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] + \mathbb{E}\left[\log p(\mathbf{x}_{2:T}|\hat{\mathbf{x}}_{2:T}\mathbf{z}_{2:T}, \mathbf{U}, \mathbf{V})\right] \\
+ \mathbb{E}\left[\log p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \boldsymbol{\phi})\right] + \mathbb{E}\left[\log p(\boldsymbol{\pi})\right] + \mathbb{E}\left[\log p(\boldsymbol{\phi})\right] \\
+ \mathbb{E}\left[\log p(\boldsymbol{\mu}, \boldsymbol{\Sigma})\right] + \mathbb{E}\left[\log p(\mathbf{U}, \mathbf{V})\right] - \mathbb{E}\left[\log q(\boldsymbol{\pi})\right] - \mathbb{E}\left[\log q(\boldsymbol{\phi})\right] \\
- \mathbb{E}\left[\log q(\boldsymbol{\mu}, \boldsymbol{\Sigma})\right] - \mathbb{E}\left[\log q(\mathbf{U}, \mathbf{V})\right] - \mathbb{E}\left[\log q(\mathbf{z}_{1:T})\right]
\end{aligned}$$

The expectations have the same solutions as VBAR-HMM in Subsection 4.2.6. Only the expectation of the BNN parameters changed. The solution is derived in the EM-steps.

5. Switching actor

The idea of SWAC is to train multiple policies with few parameters rather than training one global policy with a large number of parameters. This procedure is inspired by the option-critic architecture from [42]. Ideally, we can choose linear policies that capture local linearities. Having such policies can be seen as a single global policy with extended internal structure. To achieve this policy transformation, we make use of a trained HMM that mimics the system dynamics. This adaption makes SWAC to a MBRL because of the dynamics model we plug in this algorithm.

In this section, we show the implementation of SWAC and extend one of the recent actor-critic algorithms with SWAC.

5.1. Implementation

The design of SWAC makes SWAC is generally applicable to all actor-critic frameworks where we collect environment sample trajectories with a policy and maximize the reward signal.

Algorithm 1 shows the implementation of SWAC. The training procedure of SWAC in 1 is similar to REINFORCE or any actor-critic framework. The main difference to typical model free policy search frameworks is that we make use of a dynamics model and we initialize multiple policies that are disjointly attached to the latent states of the dynamics model.

To generate samples with the policies, we first compute a belief vector (normalized probabilities for each state). Using the belief vector $\mathbf{b}$ we sample the policy that samples the action for the current observation.

Input: step size: α, number of switching states: Z, max horizon: H

$\tau_{slds} \leftarrow$ collect randomly sampled trajectories from environment;
$slds \leftarrow$ train dynamics model with Z switching states from τ_{slds};
$\pi_{\boldsymbol{\theta}}^{(z)} \leftarrow$ initialize Z policies where $z \in \{1, ..., Z\}$;
while *not converged* **do**
  for *max horizon not reached* **do**
    $\mathbf{b} \leftarrow$ compute state belief vector with $slds$;
    $z \leftarrow$ randomly choose state from $\mathbf{b}$;
    $\tau \leftarrow$ collect samples with policy $\pi_{\boldsymbol{\theta}}^{(z)}(\mathbf{a}_t|\mathbf{s}_t)$;
  end
  foreach *latent state* z **do**
    $\hat{g}_{\boldsymbol{\theta}}^{(z)} \leftarrow \sum_{t=1}^{T} \mathbb{I}(z_t = z) \nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}^{(z)}(\mathbf{a}_t|\mathbf{s_t}) \left[\sum_{t'=t}^{T} \mathbb{I}(z_{t'} = z) r(\mathbf{s}_{t'}, \mathbf{a}_{t'})\right]$;
    $\boldsymbol{\theta}^{(z)} \leftarrow \boldsymbol{\theta}^{(z)} + \alpha \hat{g}_{\boldsymbol{\theta}}^{(z)}$;
  end
end

Algorithm 1: SWAC: Switching actor

For the policy parameter update step, we only take the actions into account that are sampled from the policies. The indicator functions ensure the mapping to the matching policies.

In addition, we can exchange the reward with a discounted reward, add a baseline, exchange the reward with an advantage function [43] etc. For the evaluation task, we test SWAC on A2C [44] and on SAC [45]. Both extension use a single global value function.

5.2. Extending A2C with SWAC

The implementation of A2C with SWAC is similar to the proposed algorithm 1. We refer to this extension as SWAC-A2C. A2C requires the computation of the advantage function. Thus, we add a global value function network and initialize multiple policies. The gradient computation becomes:

$$\hat{g}_{\boldsymbol{\theta}}^{(z)} \leftarrow \sum_{t=1}^{T} \mathbb{I}(z_t = z) \nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}^{(z)}(\mathbf{a}_t|\mathbf{s_t}) \left[Q(\mathbf{s}_t, \mathbf{a}_t) - V(s_t)\right]$$

where $Q(.)$ is the state-action value function obtained by computing the discounted reward.

If the HMM detects a latent state less frequently, the policy linked to that state samples only few samples which can lead to gradient distortion. To prevent from this behavior, we attach each policy with a own learn rate. During each policy update step, we measure the policy change by computing the KL divergence of the policy parameters before and after the policy update. The KL divergence needs to remain within a fixed value which is set as a hyper parameter. A too small KL divergence increases the learn rate and a too large KL divergence increases the learn rate for the next iterations.

5.3. Extending SAC with SWAC

The extension of SAC with SWAC is similar to the SWAC extension of A2C. For this implementation we choose a fixed learn rate since the policy gradient has naturally less variance than A2C due to its architecture.

In addition, we need to adapt the computation of the policy log probabilities from the soft value function stated in equation (6) in [45]. Our solution for this problem is to average the log probability over all policies and thus relinquish to use exclusive policies for this computation.

Using the exclusive policy log probabilities affected the accuracy of the soft value function and lead to bad performance.

6. Evaluation

For the experiments, we show the switching state behavior of the VBrAR-HMMs and compare their prediction performance to state of the art neural network architectures.

Moreover, we train policies to control previously trained SLDS and evaluate their performance on the true dynamics.

Finally, we equip actor-critic architectures with SWAC, measure the learning performance and compare it to the non SWAC equipped variants.

We evaluate the experiments on the pendulum and the cart-pole [46] environments. These are classical benchmark systems from the control literature. The swing up task is set as the objective for both environments. For the observation model we represent the angles in the trigonometric space.

6.1. Bayesian HMMs

The experiments of this section exemplarily show the dynamic switching behavior of the VBrAR-HMMs. Additionally, we compare the prediction performance of the VBAR-HMM and VBrAR-HMM to state-of-the-art models.

6.1.1. Viterbi state detection

In this experiment, we evaluate the Viterbi algorithm on randomly sampled trajectories using the likelihoods of previously trained VBrAR-HMMs. The VBrAR-HMMs are initialized with seven latent states to capture the switching dynamics for each environment. From the pendulum environment, the VBrAR-HMM is given 25 training rollouts with a horizon

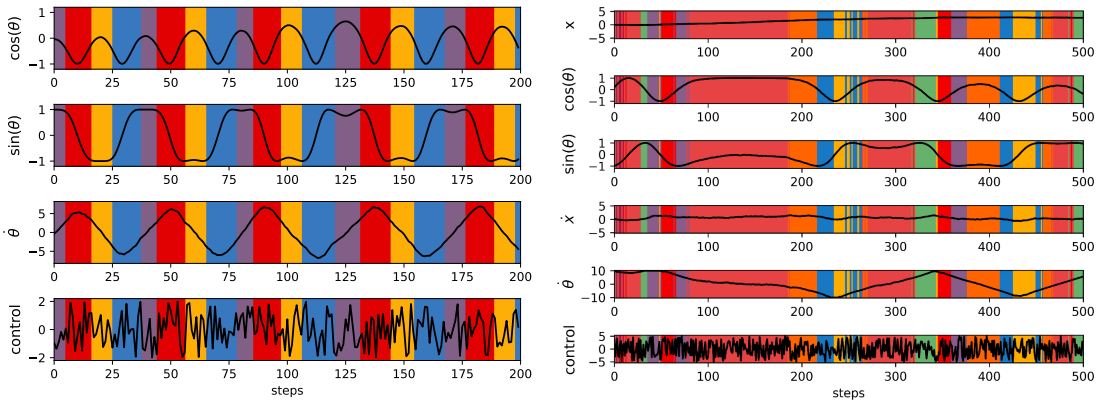


Figure 6.1.: Randomly sampled trajectories on the pendulum (left) and cart-pole (right) environment. Each color represents a detected latent state from the VBrAR-HMM using the Viterbi algorithm. The state transitions match the repeating dynamic patterns.

of 200 each. The VBrAR-HMM imitating the cart-pole dynamics, is given 45 training rollouts with a horizon of 250 each.

Figure 3.1.4 shows the Viterbi state detection of the VBrAR-HMMs. The horizon for the pendulum trajectory is 200 and 500 for the cart-pole trajectory. Each color is disjointly mapped to a latent state of the model. The VBrAR-HMM only uses 4 of its 7 states to capture the switching dynamics of this trajectory. Similarly, the VBrAR-HMM requires 6 of the total 7 states to express the cart-pole dynamics of the given trajectory.

We observe that the VBrAR-HMM learns a sensible assignment of the states, especially regarding the repeating patterns of the pendulum trajectory.

6.1.2. Model learning performance

In this experiment, we compare the out-of-sample predictive performance of the VBAR-HMM, VBrAR-HMM. Evaluating two types of the VBrAR-HMMs. One uses a BNN and the other uses a classic Feed-forward neural network (FNN) for the latent state transition. We compare the performance of our models with FNN, Recurrent neural network (RNN) and Long short-term memory neural networks (LSTM)[47]. A prediction performance overview of the non Bayesian HMMs can be found in [46].

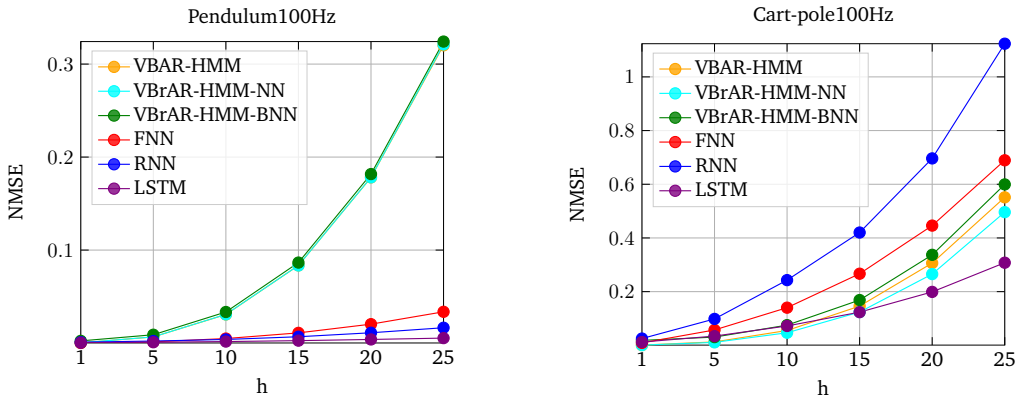


Figure 6.2.: Out-of-sample predictive performance of the VBAR-HMMs and VBrAR-HMMs to SOTA models. Evaluation on the pendulum (left) and cart-pole (right) environment.

Figure 6.2 shows the evaluation results for the out-of-sample prediction task on the pendulum and cart-pole environment. The s, RNNs and LSTMs use a single hidden layer with 64 neurons each. The VBrAR-HMMs use a single layer with 24 neurons for the transition network. The predictive performance is evaluated on the horizons $h \in \{1, 5, 10, 15, 20, 25\}$. Each model is trained with a 10 trajectories with 250 steps each. The performance is averaged over 20 differently trained models. For the pendulum task, the other models perform better than ours. All three of our presented models perform equally on this task. The logistic link of the VBrAR-HMM does not improve the models predictive accuracy. However, for the cart-pole task, the logistic link improves the prediction accuracy but only for the variant without using the BNN. Our models outperform the other models on this task except the performs better.

6.2. Learning control on rAR-HMM

In this experiment, we perform model-free policy search on trained rAR-HMMs. We use the non Bayesian rAR-HMMs from [46] because they are well tested on the pendulum and cart-pole environment. We train the models with these environments. The goal is to showcase a potential performance gain of SWAC by using a reliable SLDS model.

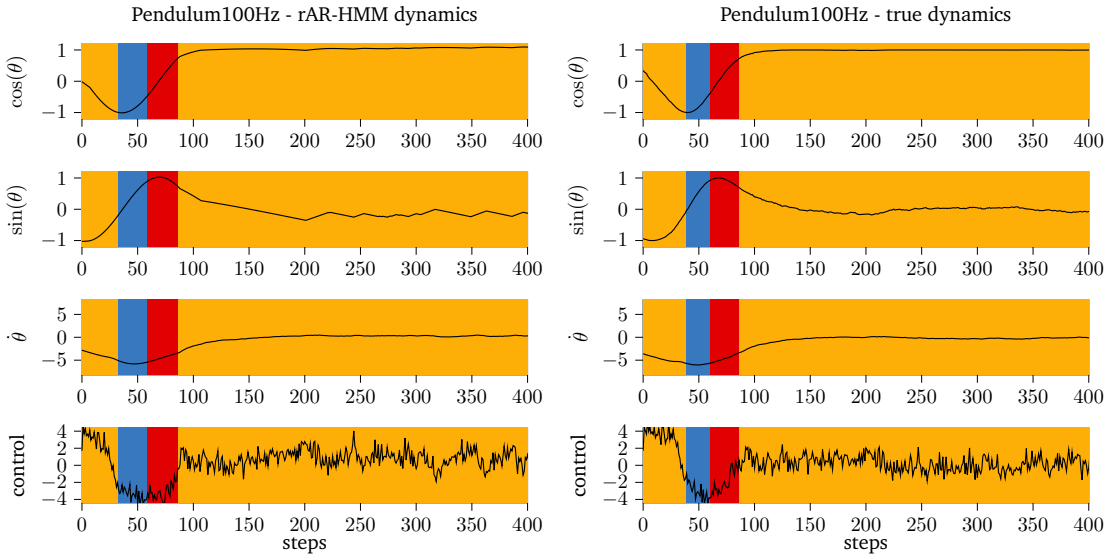


Figure 6.3.: Policy evaluation on the rAR-HMM dynamics (left) and on the true dynamics (right) of the pendulum environment. The policy training observations are sampled from the rAR-HMM dynamics.

We evaluate the performance of the trained policy on hybrid environments with rAR-HMM dynamics and on the true dynamics models.

The approximation gaps of the trained models can lead to variance increasement of the policy gradients. Thus, we use LAX [20] as the policy search framework for this task.

6.2.1. Policy evaluation on the pendulum task

For the pendulum environment, the rAR-HMM is trained with 25 trajectories each with a horizon of 200. The policy is trained with 800 epochs, each with 5000 samples.

Figure 6.3 illustrates the pendulum policy evaluation on the rAR-HMM dynamics and on the true dynamics. The latent state transitions are detected with the trained rAR-HMM that is used as the hybrid dynamics model on the left. The model dynamics are nearly identical for both systems. Thus, the policy manages to succeed to swing up and stabilize the pendulum on both systems. The policy achieves an expected reward with **-2310.73** $\pm$ **1983.48** on the rAR-HMM dynamics and on the true dynamics the expected reward

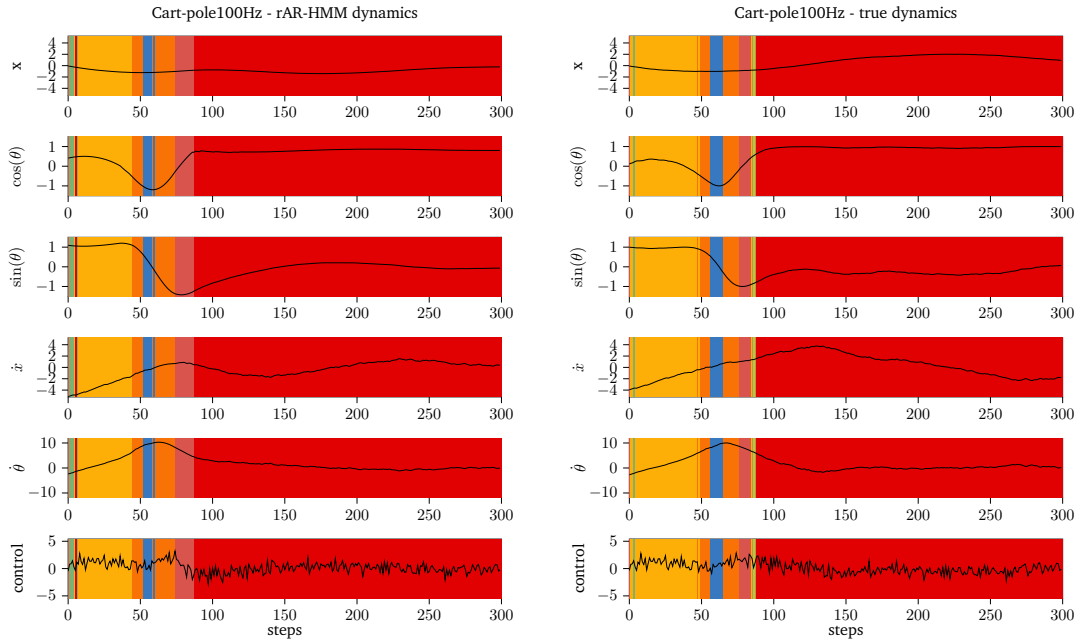


Figure 6.4.: Policy evaluation on the rAR-HMM dynamics (left) and on true dynamics (right)of the pendulum environment. The policy training observations are sampled from the rAR-HMM dynamics.

using the same policy is **-2300.90** $\pm$ **2513.69**. We averaged over 20 rollouts, each with a horizon of 1000.

6.2.2. Policy evaluation on the cart-pole task

For the cart-pole environment, the rAR-HMM is trained with 45 trajectories each with a horizon of 250. The policy is trained with 500 epochs, each with 5000 samples.

Figure 6.4 illustrates the cart-pole policy evaluation on the rAR-HMM dynamics and on the true dynamics. The latent state transitions are detected with the trained rAR-HMM that is used as the hybrid dynamics model on the left. Other than in the pendulum task, a difference in the dynamics between the two models is observable. Owing to a more difficult prediction of the cart-pole dynamics. However, the policy still manages to achieve the swing up and stabilization task on both models. The policy achieves an expected

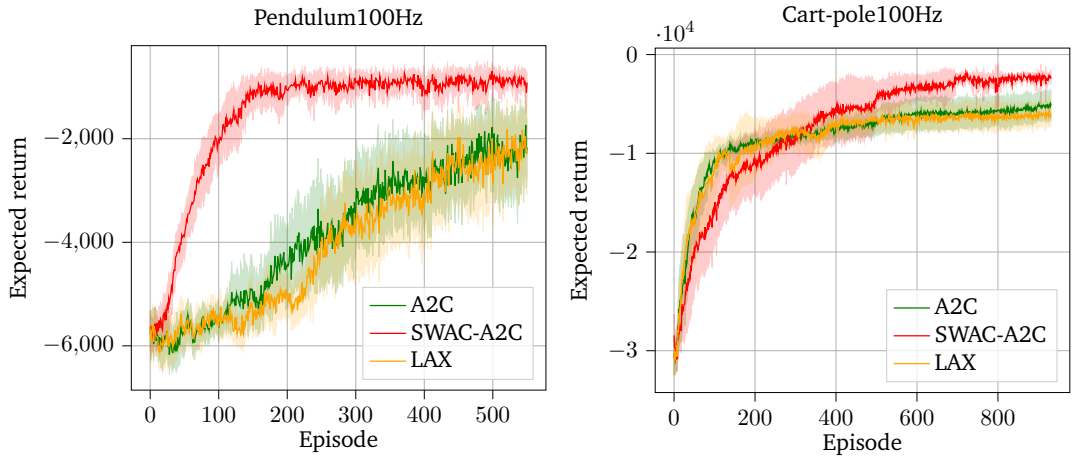


Figure 6.5.: Policy evaluation performance comparison of A2C, SWAC-A2C and LAX on the pendulum (left) and on the cart-pole (right) environment.

reward with **-2993.20** $\pm$ **1918.37** on the rAR-HMM dynamics and on the true dynamics the expected reward using the same policy is **-3389.59** $\pm$ **1785.15**. We averaged over 20 rollouts, each with a horizon of 5000.

6.3. SWAC-A2C

In this experiment, we compare the learning performance of SWAC-A2C, A2C and LAX. The performance is evaluated on 20 different seeds for each algorithm on each environment. The underlying rAR-HMMs of each environment for SWAC-A2C are trained with the same settings as in Section 6.2. For a fair comparison, we have added the learn rate regularization of the KL divergence between the old policy and the updated policy parameters to A2C and (LAX). to A2C.

Figure 6.5 shows the learning performance on both environments with a confidence interval at 95%. Since rAR-HMMs are trained with 7 states, SWAC-A2C is equipped with 7 policies each with a single hidden layer and 14 neurons. A2C and LAX use policies with 2 hidden layers and 64 neurons for each layer. The value function neural network is the same for all three algorithms. 2 hidden layers each with 64 neurons. For each training episode 5000 training samples are generated. The target KL divergence for the policy parameter

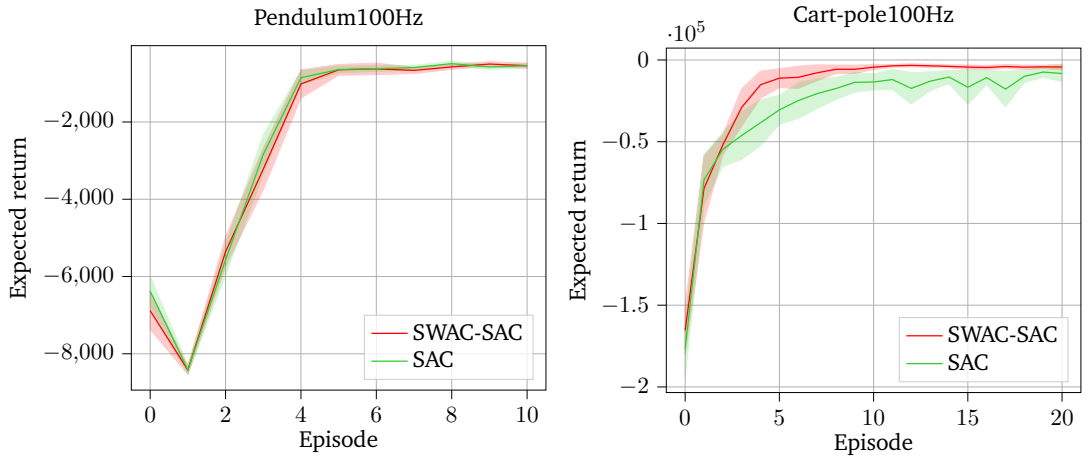


Figure 6.6.: Policy evaluation performance comparison of SAC and SWAC-SAC on the pendulum (left) and on the cart-pole (right) environment.

updates is set to 0.02 for all algorithms. SWAC-A2C clearly outperforms the baseline algorithms on the pendulum swing up task. Whereas, on the cart-pole environment SWAC-A2C shows a worse performance at the beginning of the training but performs better in the long run compared to the baseline. LAX shows the same performance as A2C on both environments.

6.4. SWAC-SAC

In this experiment, we compare the learning performance of SWAC-SAC and SAC. The performance is evaluated on 20 different seeds for each algorithm on each environment. The underlying rAR-HMMs of each environment for SWAC-SAC are trained with the same settings as in Section 6.2.

Figure 6.6 shows the learning performance on both environments with a confidence interval at 95%. As in experiment of subsection 5.2 the rAR-HMMs are trained with 7 states, SWAC-SAC is equipped with 7 policies each with a single hidden layer and 14 neurons. SAC uses a policy network with 2 hidden layers and 64 neurons for each layer. The value function neural network is the same for all three algorithms. 2 hidden layers each with 64 neurons. 5,000 training samples are generated for the pendulum and 10,000

training samples are generated for the cart-pole task for each training episode. SWAC-SAC and SAC perform identically on the pendulum task. For the cart-pole task SWAC-SAC outperforms SAC. SWAC-SAC succeeds in approximately 4 training episodes to achieve the task and remains stable for the further training episodes. SAC on the other hand approximately requires 10 training episodes to achieve the cart-pole task.

6.5. Discussion

The results show that the variational Bayes variants of the HMMs are capable of capturing the system dynamics of a given model. The weaker performance of the VBrAR-HMM-BNN compared to the VBAR-HMM and VBrAR-HMM-FNN can be due to the difficulties of hyper-parameter tuning for the BNNs [48]. However, using the BNN we can access the uncertainty of the network which could be interesting for further work. Moreover, we have shown that SLDS can be successfully incorporated into policy search problems which is evidenced by our results. The actor-critic frameworks extended with SWAC show a similar or better performance but no worse performance compared to the non SWAC extended actor-critic frameworks. A weakness of the SWAC extension is the computation overhead that occurs during the gradient propagation for each policy. This computation can be parallelized but still requires additional computation resources. Interestingly, LAX shows the same performance as A2C on our selected control environments. In the published paper of LAX [20] the authors used more complex environments to evaluate LAX. Observing this behavior, we conclude that LAX plays its strength for gradient variance reduction on more complex environments.

7. Outlook

The Bayesian HMMs have the property make the access to the models uncertainty possible. For future work it would be interesting to incorporate this quantity into a policy search framework such that exploration can be enhanced. Moreover, further investigation of these models for the planning step in a MBRL setup can improve the learning performance as [5] have done with GPs. A requirement to achieve this task, is to make differentiating through the whole model possible.

For future work of SWAC, it would be interesting to find a way to get SWAC work by only using multiple linear policies. A problem of SWAC is that when a latent state is rarely detected by the switching model, the policy mapped to that latent state produces not enough samples for training. This behavior can be avoided by using as few latent states as possible.

8. Conclusion

In this thesis, we have tackled the problem of equipping a policy with additional internal structure. Initially, we have formulated a Bayesian HMM that makes the model's uncertainty accessible. This uncertainty can be used as an additional quantity for policy search problems. To succeed this thesis's goal, we have successfully integrated a non Bayesian rAR-HMMs to a policy framework that we have defined in this work. This policy framework profits from the switching behavior of the rAR-HMMs and shows superior performance compared to the variants that are not equipped with switching structure of the rAR-HMMs.
In Chapter 4, we have proposed a Bayesian formulation of HMMs, AR-HMMs and rAR-HMMs by making use of variational inference and the mean-field approximation. We have defined the priors of the model parameters and have derived all required steps to make an implementation of these models straightforward using the Baum-Welch algorithm.
In Chapter 5, we have introduced SWAC, an extension for policy search algorithms that brings additional structure to the policy by integrating a switching state space model, in our case rAR-HMMs, into the policy search procedure.
In Chapter 6, we have illustrated the switching behavior of the VBrAR-HMMs.
Moreover, we have compared the predictive performance of the VBAR-HMM and the VBrAR-HMM. On the simpler pendulum environment, the reference models performed better than our proposed models. Whereas, on the more complex cart-pole environment, our models outperformed the reference models except the LSTM.
To show the advantage of minimizing sample complexity by using a learned dynamics model, we have performed policy search on a trained rAR-HMM. The obtained policy is evaluated on the true and on the rAR-HMM dynamics. The policy performed minimally better on the rAR-HMM dynamics.
As a proof of concept for SWAC, we have measured the learning performance of the SWAC to the non SWAC equipped counterparts. For a fair comparison, we have used the same hyper-parameters for both variants. The results of SWAC have shown a better or an equal performance compared to the non SWAC variant.

Finally, we have reviewed a policy search approach that benefits from the Bayesian HMMs by using the model for planning and accessing the model's uncertainty for exploration. Regarding SWAC, we have pointed on the possibility to make it run by using linear policies only.

Bibliography

[1] S. Linderman, M. Johnson, A. Miller, R. Adams, D. Blei, and L. Paninski, "Bayesian Learning and Inference in Recurrent Switching Linear Dynamical Systems," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (A. Singh and J. Zhu, eds.), vol. 54 of *Proceedings of Machine Learning Research*, (Fort Lauderdale, FL, USA), pp. 914–922, PMLR, 20–22 Apr 2017.

[2] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, vol. 8, pp. 229–256, 1992.

[3] A. G. Barto, R. S. Sutton, and C. W. Anderson, "Neuronlike adaptive elements that can solve difficult learning control problems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 5, pp. 834–846, 1983.

[4] M. J. Beal, "Variational algorithms for approximate bayesian inference," tech. rep., 2003.

[5] M. Deisenroth and C. Rasmussen, "Pilco: A model-based and data-efficient approach to policy search," in *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, pp. 465–472, Omnipress, 2011.

[6] M. P. Deisenroth*, G. Neumann*, and J. Peters, "A survey on policy search for robotics," *Foundations and Trends in Robotics*, pp. 388–403, 2013.

[7] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *CoRR*, vol. abs/1906.02691, 2019.

[8] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2014.

[9] C. Doersch, "Tutorial on variational autoencoders," 2016. cite arxiv:1606.05908.

[10] P. Becker-Ehmck, M. Karl, J. Peters, and P. van der Smagt, "Learning to fly via deep model-based reinforcement learning," 2020.

[11] G. Ackerson and K. Fu, "On state estimation in switching environments," *IEEE Transactions on Automatic Control*, vol. 15, no. 1, pp. 10–17, 1970.

[12] C. B. Chang and M. Athans, "State estimation for discrete systems with switching parameters," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-14, no. 3, pp. 418–425, 1978.

[13] J. D. Hamilton, "Analysis of time series subject to changes in regime," 1990.

[14] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "Nonparametric bayesian learning of switching linear dynamical systems," in *Advances in Neural Information Processing Systems 21* (D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, eds.), pp. 457–464, Curran Associates, Inc.

[15] Z. Ghahramani and G. E. Hinton, "Switching state-space models," tech. rep., King's College Road, Toronto M5S 3H5, 1996.

[16] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013. cite arxiv:1312.6114.

[17] E. Jang, S. Gu, and B. Poole, "Categorical reparametrization with gumbel-softmax," in *Proceedings International Conference on Learning Representations 2017*, OpenReviews.net, Apr. 2017.

[18] C. J. Maddison, A. Mnih, and Y. W. Teh, "The concrete distribution: A continuous relaxation of discrete random variables," 2016. cite arxiv:1611.00712.

[19] G. Tucker, A. Mnih, C. J. Maddison, J. Lawson, and J. Sohl-Dickstein, "Rebar: Low-variance, unbiased gradient estimates for discrete latent variable models," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 2627–2636, Curran Associates, Inc., 2017.

[20] W. Grathwohl, D. Choi, Y. Wu, G. Roeder, and D. Duvenaud, "Backpropagation through the void: Optimizing control variates for black-box gradient estimation," 2017. cite arxiv:1711.00123Comment: Published at ICLR 2018.

[21] K. P. Murphy, *Machine learning : a probabilistic perspective*. Cambridge, Mass. [u.a.]: MIT Press, 2013.

[22] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state markov chains," *Ann. Math. Statist.*, vol. 37, pp. 1554–1563, 12 1966.

[23] V. Smídl and A. Quinn, *The Variational Bayes Method in Signal Processing (Signals and Communication Technology)*. Berlin, Heidelberg: Springer-Verlag, 2005.

[24] T. S. Jaakkola and M. I. Jordan, "Bayesian parameter estimation via variational methods," *Statistics and Computing*, vol. 10, no. 1, pp. 25–37, 2000.

[25] T. S. Jaakkola and M. I. Jordan, "Computing upper and lower bounds on likelihoods in intractable networks," pp. 340–348, Morgan Kaufmann, 1996.

[26] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, p. 1–305, Jan. 2008.

[27] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," 2018.

[28] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Mach. Learn.*, vol. 37, p. 183–233, Nov. 1999.

[29] S. Mohamed, M. Rosca, M. Figurnov, and A. Mnih, "Monte carlo gradient estimation in machine learning," 2019. cite arxiv:1906.10652Comment: 59 pages, under review.

[30] C. Wang and D. M. Blei, "Variational inference in nonconjugate models," *J. Mach. Learn. Res.*, vol. 14, p. 1005–1031, Apr. 2013.

[31] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[32] S. M. Ross, *Simulation, Fourth Edition*. USA: Academic Press, Inc., 2006.

[33] R. Dearden, N. Friedman, and D. Andre, "Model-based bayesian exploration," *CoRR*, vol. abs/1301.6690, 2013.

[34] P. Abbeel, M. Quigley, and A. Y. Ng, "Using inaccurate models in reinforcement learning," in *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, (New York, NY, USA), p. 1–8, Association for Computing Machinery, 2006.

[35] J. C. Spall and J. A. Cristion, "Model-free control of nonlinear stochastic systems with discrete-time measurements," *IEEE Transactions on Automatic Control*, vol. 43, no. 9, pp. 1198–1210, 1998.

[36] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Oct. 2008. Version 20081110.

[37] T. P. Minka, "Bayesian linear regression," tech. rep., 3594 Security Ticket Control, 1999.

[38] D. V. rosen, "Moments for matrix normal variables," *Statistics*, vol. 19, no. 4, pp. 575–583, 1988.

[39] D. J. MacKay, "Bayesian neural networks and density networks," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 354, no. 1, pp. 73 – 80, 1995. Proceedings of the Third Workshop on Neutron Scattering Data Analysis.

[40] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural network," in *Proceedings of the 32nd International Conference on Machine Learning* (F. Bach and D. Blei, eds.), vol. 37 of *Proceedings of Machine Learning Research*, (Lille, France), pp. 1613–1622, PMLR, 07–09 Jul 2015.

[41] D. P. Kingma, T. Salimans, and M. Welling, "Variational dropout and the local reparameterization trick," in *Advances in Neural Information Processing Systems 28* (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds.), pp. 2575–2583, Curran Associates, Inc.

[42] P.-L. Bacon, J. Harb, and D. Precup, "The option-critic architecture," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, p. 1726–1734, AAAI Press, 2017.

[43] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-dimensional continuous control using generalized advantage estimation," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2016.

[44] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," *CoRR*, vol. abs/1602.01783, 2016.

[45] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.), vol. 80 of *Proceedings of Machine Learning Research*, (Stockholmsmässan, Stockholm Sweden), pp. 1861–1870, PMLR, 10–15 Jul 2018.

[46] H. Abdulsamad and J. Peters, "Hierarchical decomposition of nonlinear dynamics and control for system identification and policy distillation," in *Proceedings of the 2nd Conference on Learning for Dynamics and Control* (A. M. Bayen, A. Jadbabaie, G. Pappas, P. A. Parrilo, B. Recht, C. Tomlin, and M. Zeilinger, eds.), vol. 120 of *Proceedings of Machine Learning Research*, (The Cloud), pp. 904–914, PMLR, 10–11 Jun 2020.

[47] F. A. Gers, J. A. Schmidhuber, and F. A. Cummins, "Learning to forget: Continual prediction with lstm," *Neural Comput.*, vol. 12, p. 2451–2471, Oct. 2000.

[48] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 6402–6413, Curran Associates, Inc., 2017.

A. Summary VBEM steps

A.1. VBHMM EM steps

E-step updates

$$\log \widetilde{\boldsymbol{\pi}}_k \equiv \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \boldsymbol{\pi}_k\right] = \psi(\omega_k^{(\boldsymbol{\pi})}) - \psi(\sum_{k=1}^{K} \omega_k^{(\boldsymbol{\pi})}), \;\; \sum_{k=1}^{K} \widetilde{\boldsymbol{\pi}}_k \leq 1$$

$$\log \widetilde{a}_{kj} \equiv \mathbb{E}_{q(\mathbf{A})}\left[\log a_{kj}\right] = \psi(\omega_{kj}^{(\mathbf{A})}) - \psi(\sum_{j=1}^{K} \omega_{kj}^{(\mathbf{A})}), \;\; \sum_{j=1}^{K} \widetilde{a}_{kj} \leq 1$$

$$\log \widetilde{\boldsymbol{\Sigma}}_k \equiv \mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\log |\boldsymbol{\Sigma}_k|\right] = \sum_{i=1}^{D} \psi\left(\frac{\nu_k + 1 - i}{2}\right) + D\log 2 + \log |\mathbf{W}_k|$$

$$\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[(\mathbf{x_t} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x_t} - \boldsymbol{\mu}_k)\right] = D\beta_k^{-1} + \nu_k (\mathbf{m}_k - \mathbf{x}_t)^T \mathbf{W}_k (\mathbf{m}_k - \mathbf{x}_t)$$

M-step updates

$$\begin{aligned}
\omega_k^{(\boldsymbol{\pi})} &= \omega_0^{(\boldsymbol{\pi})} + \gamma_1(k) \\
\omega_{kj}^{(\mathbf{A})} &= \omega_0^{(\mathbf{A})} + \sum_{t=2}^{T} \xi_{t-1,t}(k,j) \\
\beta_k &= \beta_0 + N_k \\
\mathbf{m}_k &= \frac{1}{\beta_k}\left(\beta_0 \mathbf{m}_0 + N_k \overline{\mathbf{x}}_t\right) \\
\mathbf{W}_k^{-1} &= \mathbf{W}_0^{-1} N_k \mathbf{S}_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(\overline{\mathbf{x}}_k - \mathbf{m}_0)(\overline{\mathbf{x}}_k - \mathbf{m}_0)^T \\
\nu_k &= \nu_0 + N_k
\end{aligned}$$

A.2. VBAR-HMM EM steps

E-step updates

$$\log \widetilde{\boldsymbol{\pi}}_k \equiv \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \boldsymbol{\pi}_k\right] = \psi(\omega_k^{(\boldsymbol{\pi})}) - \psi(\sum_{k=1}^{K} \omega_k^{(\boldsymbol{\pi})}), \;\; \sum_{k=1}^{K} \widetilde{\boldsymbol{\pi}}_k \leq 1$$

$$\log \widetilde{a}_{kj} \equiv \mathbb{E}_{q(\mathbf{A})}\left[\log a_{kj}\right] = \psi(\omega_{kj}^{(\mathbf{A})}) - \psi(\sum_{j=1}^{K} \omega_{kj}^{(\mathbf{A})}), \;\; \sum_{j=1}^{K} \widetilde{a}_{kj} \leq 1$$

$$\log \widetilde{\boldsymbol{\Sigma}}_k \equiv \mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\log |\boldsymbol{\Sigma}_k|\right] = \sum_{i=1}^{D} \psi\left(\frac{\nu_k + 1 - i}{2}\right) + D\log 2 + \log |\mathbf{W}_k|$$

$$\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[(\mathbf{x_1} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x_1} - \boldsymbol{\mu}_k)\right] = D\beta_k^{-1} + \nu_k (\mathbf{m}_k - \mathbf{x}_1)^T \mathbf{W}_k (\mathbf{m}_k - \mathbf{x}_1)$$

$$\log \widetilde{\mathbf{V}}_k \equiv \mathbb{E}_{q(\mathbf{V})}\left[\log |\mathbf{V}_k|\right] = \sum_{i=1}^{D} \psi\left(\frac{\eta_k + 1 - i}{2}\right) + D\log 2 + \log |\mathbf{P}_k|$$

$$\mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[(\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1})^T \mathbf{V}_k (\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1})\right] = \eta_k (\mathbf{M}_k \hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T \mathbf{P}_k (\mathbf{M}_k \hat{\mathbf{x}}_{t-1} - \mathbf{x}_t) + \mathbf{Tr}\left\{\mathbf{K}_k^{-1} \hat{\mathbf{x}}_{t-1} \hat{\mathbf{x}}_{t-1}^T\right\}$$

M-step updates

$$
\begin{aligned}
\omega_k^{(\boldsymbol{\pi})} &= \omega_0^{(\boldsymbol{\pi})} + \gamma_1(k) \\
\omega_{kj}^{(\mathbf{A})} &= \omega_0^{(\mathbf{A})} + \sum_{t=2}^{T} \xi_{t-1,t}(k,j) \\
\beta_k &= \beta_0 + N_k \\
\mathbf{m}_k &= \frac{1}{\beta_k}\left(\beta_0 \mathbf{m}_0 + N_k \overline{\mathbf{x}}_t\right) \\
\mathbf{W}_k^{-1} &= \mathbf{W}_0^{-1} N_k \mathbf{S}_k + \frac{\beta_0 N_k}{\beta_0 + N_k}(\overline{\mathbf{x}}_k - \mathbf{m}_0)(\overline{\mathbf{x}}_k - \mathbf{m}_0)^T \\
\nu_k &= \nu_0 + N_k \\
\mathbf{K}_k &= \sum_{t=2}^{T} \gamma_t(k) \hat{\mathbf{x}}_{t-1} \hat{\mathbf{x}}_{t-1}^T + \mathbf{K}_0 \\
\mathbf{M}_k &= \left[\sum_{t=2}^{T} \gamma_t(k) \mathbf{x}_t \hat{\mathbf{x}}_{t-1}^T + \mathbf{M}_0 \mathbf{K}_0\right] \mathbf{K}_k^{-1} \\
\mathbf{P}_k^{-1} &= \mathbf{P}_0^{-1} + \mathbf{M}_0 \mathbf{K}_0 \mathbf{M}_0^T + \sum_{t=2}^{T} \gamma_t(k) \mathbf{x}_t \mathbf{x}_t^T - \mathbf{M}_k \mathbf{K}_k \mathbf{M}_k^T \\
\eta_k &= \eta_0 + \sum_{t=2}^{T} \gamma_t(k)
\end{aligned}
$$

A.3. VBrAR-HMM EM steps

E-step updates

$$\log \widetilde{\boldsymbol{\pi}}_k \equiv \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \boldsymbol{\pi}_k\right] = \psi(\omega_k^{(\boldsymbol{\pi})}) - \psi(\sum_{k=1}^{K} \omega_k^{(\boldsymbol{\pi})}), \;\; \sum_{k=1}^{K} \widetilde{\boldsymbol{\pi}}_k \leq 1$$

$$\mathbb{E}_{q(\boldsymbol{\phi})}\left[\log \psi_{ij}\right] = f_{\boldsymbol{\phi}}(\mathbf{x}, \mathbf{u})$$

$$\log \widetilde{\boldsymbol{\Sigma}}_k \equiv \mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\log |\boldsymbol{\Sigma}_k|\right] = \sum_{i=1}^{D} \psi\left(\frac{\nu_k + 1 - i}{2}\right) + D\log 2 + \log |\mathbf{W}_k|$$

$$\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[(\mathbf{x_1} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x_1} - \boldsymbol{\mu}_k)\right] = D\beta_k^{-1} + \nu_k (\mathbf{m}_k - \mathbf{x}_1)^T \mathbf{W}_k (\mathbf{m}_k - \mathbf{x}_1)$$

$$\log \widetilde{\mathbf{V}}_k \equiv \mathbb{E}_{q(\mathbf{V})}\left[\log |\mathbf{V}_k|\right] = \sum_{i=1}^{D} \psi\left(\frac{\eta_k + 1 - i}{2}\right) + D\log 2 + \log |\mathbf{P}_k|$$

$$\mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[(\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1})^T \mathbf{V}_k (\mathbf{x}_t - \mathbf{U}_k \hat{\mathbf{x}}_{t-1})\right] = \eta_k (\mathbf{M}_k \hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T \mathbf{P}_k (\mathbf{M}_k \hat{\mathbf{x}}_{t-1} - \mathbf{x}_t) + \mathbf{Tr}\left\{\mathbf{K}_k^{-1} \hat{\mathbf{x}}_{t-1} \hat{\mathbf{x}}_{t-1}^T\right\}$$

M-step updates

$$
\begin{aligned}
\omega_k^{(\boldsymbol{\pi})} &= \omega_0^{(\boldsymbol{\pi})} + \gamma_1(k) \\
\max_{\boldsymbol{\phi}} \log p(\mathbf{z}_{2:T}|\boldsymbol{\phi}) &= \max_{\boldsymbol{\phi}} \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{j=1}^{K} \xi_{t-1,t}(k,j) \log \psi_{kj} \\
\beta_k &= \beta_0 + N_k \\
\mathbf{m}_k &= \frac{1}{\beta_k} \left(\beta_0 \mathbf{m}_0 + N_k \overline{\mathbf{x}}_t\right) \\
\mathbf{W}_k^{-1} &= \mathbf{W}_0^{-1} N_k \mathbf{S}_k + \frac{\beta_0 N_k}{\beta_0 + N_k} (\overline{\mathbf{x}}_k - \mathbf{m}_0)(\overline{\mathbf{x}}_k - \mathbf{m}_0)^T \\
\nu_k &= \nu_0 + N_k \\
\mathbf{K}_k &= \sum_{t=2}^{T} \gamma_t(k) \hat{\mathbf{x}}_{t-1} \hat{\mathbf{x}}_{t-1}^T + \mathbf{K}_0 \\
\mathbf{M}_k &= \left[\sum_{t=2}^{T} \gamma_t(k) \mathbf{x}_t \hat{\mathbf{x}}_{t-1}^T + \mathbf{M}_0 \mathbf{K}_0\right] \mathbf{K}_k^{-1} \\
\mathbf{P}_k^{-1} &= \mathbf{P}_0^{-1} + \mathbf{M}_0 \mathbf{K}_0 \mathbf{M}_0^T + \sum_{t=2}^{T} \gamma_t(k) \mathbf{x}_t \mathbf{x}_t^T - \mathbf{M}_k \mathbf{K}_k \mathbf{M}_k^T \\
\eta_k &= \eta_0 + \sum_{t=2}^{T} \gamma_t(k)
\end{aligned}
$$

B. Derivation VBHMMs

B.1. Derivation VBHMM

Variational E-step full derivation Solving $q(\mathbf{z}_{1:T})$ with mean-field approximation:

$$
\begin{aligned}
\log q(\mathbf{z}_{1:T}) &\propto \mathbb{E}_{q(\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{A})}[\log p(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})] \\
&= \mathbb{E}_{q(\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{A})}[p(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma})p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \mathbf{A})p(\mathbf{A})p(\boldsymbol{\pi})p(\boldsymbol{\mu}|\boldsymbol{\Sigma})p(\boldsymbol{\Sigma})] \\
&= \mathbb{E}_{q(\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{A})}[\log p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \mathbf{A})p(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma})] \\
&= \mathbb{E}_{q(\boldsymbol{\pi},\mathbf{A})}[\log p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \mathbf{A})] + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}[\log p(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma})] \\
&= \mathbb{E}_{q(\boldsymbol{\pi},\mathbf{A})}\left[\log p(z_1|\boldsymbol{\pi})p(\mathbf{z}_{2:T}|z_1, \mathbf{A})\right] + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}[\log p(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma})] \\
&= \mathbb{E}_{q(\boldsymbol{\pi},\mathbf{A})}\left[\log \prod_{k=1}^{K} \left(\boldsymbol{\pi}_k^{\mathbb{I}(z_1=k)}\right) \prod_{t=2}^{T}\prod_{k=1}^{K}\prod_{j=1}^{K} a_{kj}^{\mathbb{I}(z_{t-1}=k,z_t=j)}\right] \\
&\quad + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\log \prod_{t=1}^{T}\prod_{k=1}^{K} \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right)^{\mathbb{I}(z_t=k)}\right] \\
&= \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \prod_{k=1}^{K} \left(\boldsymbol{\pi}_k^{\mathbb{I}(z_1=k)}\right)\right] + \mathbb{E}_{q(\mathbf{A})}\left[\log \prod_{t=2}^{T}\prod_{k=1}^{K}\prod_{j=1}^{K} a_{kj}^{\mathbb{I}(z_{t-1}=k,z_t=j)}\right] \\
&\quad + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\log \prod_{t=1}^{T}\prod_{k=1}^{K} \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right)^{\mathbb{I}(z_t=k)}\right] \\
&= \mathbb{E}_{q(\boldsymbol{\pi})}\left[\sum_{k=1}^{K} \mathbb{I}(z_1 = k) \log \boldsymbol{\pi}_k\right] \\
&\quad + \mathbb{E}_{q(\mathbf{A})}\left[\sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{I}(z_{t-1} = k, z_t = j) \log a_{kj}\right] \\
&\quad + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\sum_{t=1}^{T}\sum_{k=1}^{K} \mathbb{I}(z_t = k) \log \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right)\right]
\end{aligned}
$$

Continuation of the derivation for $\log q(\mathbf{z}_{1:T})$:

$$\begin{aligned}
\log q(\mathbf{z}_{1:T}) \propto & \sum_{k=1}^{K} \mathbb{I}(z_1 = k) \\
& \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \boldsymbol{\pi}_k\right] + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{I}(z_{t-1} = k, z_t = j)\mathbb{E}_{q(\mathbf{A})}\left[\log a_{kj}\right] \\
& + \sum_{t=1}^{T}\sum_{k=1}^{K} \mathbb{I}(z_t = k) \\
& \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{D}{2}\log 2\pi - \frac{1}{2}\left(\mathbf{x}_t - \boldsymbol{\mu}_k\right)^T \boldsymbol{\Sigma}_k \left(\mathbf{x}_t - \boldsymbol{\mu}_k\right)\right] \\
= & \sum_{k=1}^{K} \mathbb{I}(z_1 = k) \\
& \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \boldsymbol{\pi}_k\right] + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K} \mathbb{I}(z_{t-1} = k, z_t = j)\mathbb{E}_{q(\mathbf{A})}\left[\log a_{kj}\right] \\
& + \sum_{t=1}^{T}\sum_{k=1}^{K} \mathbb{I}(z_t = k) \\
& \left\{\frac{1}{2}\mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\log|\boldsymbol{\Sigma}_k|\right] - \frac{D}{2}\log\left(2\pi\right) - \frac{1}{2}\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\left(\mathbf{x}_t - \boldsymbol{\mu}_k\right)^T \boldsymbol{\Sigma}_k \left(\mathbf{x}_t - \boldsymbol{\mu}_k\right)\right]\right\}
\end{aligned}$$

Derving $\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}$:

$$\begin{aligned}
& \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[(\mathbf{x_t} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x_t} - \boldsymbol{\mu}_k)\right] \\
& \quad = \int\int \left((\mathbf{x_t} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x_t} - \boldsymbol{\mu}_k)\right) q(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\, \mathrm{d}\boldsymbol{\mu}_k\, \mathrm{d}\boldsymbol{\Sigma}_k \\
& \quad = \int \left\{\int (\mathbf{x_t} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k (\mathbf{x_t} - \boldsymbol{\mu}_k) q(\boldsymbol{\mu}_k|\boldsymbol{\Sigma}_k)\, \mathrm{d}\boldsymbol{\mu}_k\right\} q(\boldsymbol{\Sigma}_k)\, \mathrm{d}\boldsymbol{\Sigma}_k \\
& \quad = \int \mathbb{E}_{q(\boldsymbol{\mu})}\left[(\boldsymbol{\mu}_k - \mathbf{x_t})^T \boldsymbol{\Sigma}_k (\boldsymbol{\mu}_k - \mathbf{x_t})\right] q(\boldsymbol{\Sigma}_k)\, \mathrm{d}\boldsymbol{\Sigma}_k
\end{aligned}$$

Using the equation (380) from [36] to solve the inner expectation with respect to $\boldsymbol{\mu}$ yields:

$$\begin{aligned}
\mathbb{E}_{q(\boldsymbol{\mu})}\left[(\boldsymbol{\mu}_k - \mathbf{x_t})^T \boldsymbol{\Sigma}_k (\boldsymbol{\mu}_k) - \mathbf{x_t}\right] &= (\mathbf{m}_k - \mathbf{x}_t)^T \boldsymbol{\Sigma}_k (\mathbf{m}_k - \mathbf{x}_t) + \mathbf{Tr}\left(\boldsymbol{\Sigma}_k (\beta_k^{-1}\boldsymbol{\Sigma}_k)^{-1}\right) \\
&= (\mathbf{m}_k - \mathbf{x}_t)^T \boldsymbol{\Sigma}_k (\mathbf{m}_k - \mathbf{x}_t) + D\beta_k^{-1}
\end{aligned}$$

Plugging this term back in the equation

$$
\begin{aligned}
&\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[(\mathbf{x_t}-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k(\mathbf{x_t}-\boldsymbol{\mu}_k)\right] \\
&\qquad = \int \left\{(\mathbf{m}_k-\mathbf{x}_t)^T\boldsymbol{\Sigma}_k(\mathbf{m}_k-\mathbf{x}_t)+D\beta_k^{-1}\right\} q(\boldsymbol{\Sigma}_k)\,\mathrm{d}\boldsymbol{\Sigma}_k \\
&\qquad = D\beta_k^{-1}+\mathbb{E}_{q(\boldsymbol{\Sigma})}\left[(\mathbf{m}_k-\mathbf{x}_t)^T\boldsymbol{\Sigma}_k(\mathbf{m}_k-\mathbf{x}_t)\right] \\
&\qquad = D\beta_k^{-1}+\mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\mathbf{Tr}\left\{\boldsymbol{\Sigma}_k(\mathbf{m}_k-\mathbf{x}_t)(\mathbf{m}_k-\mathbf{x}_t)^T\right\}\right] \\
&\qquad = D\beta_k^{-1}+\mathbf{Tr}\left\{\mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\boldsymbol{\Sigma}_k\right](\mathbf{m}_k-\mathbf{x}_t)(\mathbf{m}_k-\mathbf{x}_t)^T\right\} \\
&\qquad = D\beta_k^{-1}+\mathbf{Tr}\left\{\nu_k\mathbf{W}_k(\mathbf{m}_k-\mathbf{x}_t)(\mathbf{m}_k-\mathbf{x}_t)^T\right\} \\
&\qquad = D\beta_k^{-1}+\nu_k(\mathbf{m}_k-\mathbf{x}_t)^T\mathbf{W}_k(\mathbf{m}_k-\mathbf{x}_t)
\end{aligned}
$$

Variational M-step full derivation Solving $\log q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean-field approximation:

$$
\begin{aligned}
\log q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &\propto \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] \\
&= \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \mathbf{A}) p(\boldsymbol{\pi}) p(\mathbf{A}) p(\boldsymbol{\mu}|\boldsymbol{\Sigma}) p(\boldsymbol{\Sigma})\right] \\
&= \log p(\boldsymbol{\pi}) + \log p(\mathbf{A}) + \sum_{k=1}^{K} \log p(\boldsymbol{\mu}_k|\boldsymbol{\Sigma}_k) p(\boldsymbol{\Sigma}_k) \\
&\quad + \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] + \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \mathbf{A})\right] \\
&= \log p(\boldsymbol{\pi}) + \log p(\mathbf{A}) + \sum_{k=1}^{K} \log p(\boldsymbol{\mu}_k|\boldsymbol{\Sigma}_k) p(\boldsymbol{\Sigma}_k) \\
&\quad + \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] \\
&\quad + \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(z_1|\boldsymbol{\pi}) p(\mathbf{z}_{2:T}|z_1, \mathbf{A})\right] \\
&= \log Dir(\boldsymbol{\pi}|\boldsymbol{\omega}_{\mathbf{0}}^{(\boldsymbol{\pi})}) + \sum_{k=1}^{K} \log Dir(\mathbf{A}|\boldsymbol{\omega}_{0k}^{(\mathbf{A})}) \\
&\quad + \sum_{k=1}^{K} \log \mathcal{N}\left(\boldsymbol{\mu}_k|\boldsymbol{m}_0, (\beta_0 \boldsymbol{\Sigma}_k)^{-1}\right) + \sum_{k=1}^{K} \log \mathcal{W}\left(\boldsymbol{\Sigma}_k|\mathbf{W}_{\mathbf{0}}, \nu_0\right) \\
&\quad + \sum_{t=1}^{T} \sum_{k=1}^{K} \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\mathbb{I}(z_t = k)\right] \log \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right) \\
&\quad + \sum_{k=1}^{K} \left\{\mathbb{E}_{q(z_1)}\left[\mathbb{I}(z_1 = k)\right] \log \pi_k\right\} \\
&\quad + \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{j=1}^{K} \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\mathbb{I}(z_{t-1} = k, z_t = j)\right] \log a_{kj} \\
&= \log Dir(\boldsymbol{\pi}|\boldsymbol{\omega}_{\mathbf{0}}^{(\boldsymbol{\pi})}) + \sum_{k=1}^{K} \log Dir(\mathbf{A}|\boldsymbol{\omega}_{0k}^{(\mathbf{A})}) \\
&\quad + \sum_{k=1}^{K} \log \mathcal{N}\left(\boldsymbol{\mu}_k|\boldsymbol{m}_0, (\beta_0 \boldsymbol{\Sigma}_k)^{-1}\right) + \sum_{k=1}^{K} \log \mathcal{W}\left(\boldsymbol{\Sigma}_k|\mathbf{W}_{\mathbf{0}}, \nu_0\right) \\
&\quad + \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_t(k) \log \mathcal{N}\left(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right) + \sum_{k=1}^{K} \gamma_1(k) \log \pi_k \\
&\quad + \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{j=1}^{K} \xi_{t-1,t}(k, j) \log a_{kj}
\end{aligned}
$$

$$\log q(\mathbf{A}) \propto \sum_{k=1}^{K}\sum_{j=1}^{K}(\omega_0^{(\mathbf{A})}-1)\log a_{kj} + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K}\mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\mathbb{I}(z_{t-1}=k,z_t=j)\right]\log a_{kj}$$

$$= \sum_{k=1}^{K}\sum_{j=1}^{K}\log a_{kj}\left\{(\omega_0^{(\mathbf{A})}-1)+\sum_{t=2}^{T}\mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\mathbb{I}(z_{t-1}=k,z_t=j)\right]\right\}$$

$$-\frac{1}{2}\left(\boldsymbol{\mu}_k^T(\beta_k\boldsymbol{\Sigma}_k)\boldsymbol{\mu}_k\right) = -\frac{1}{2}\left\{\boldsymbol{\mu}_k^T\left(\beta_0+\sum_{t=1}^{T}\gamma_t(k)\right)\boldsymbol{\Sigma}_k\boldsymbol{\mu}_k\right\}$$

$$\beta_k = \beta_0+\sum_{t=1}^{T}\gamma_t(k)$$

$$\boldsymbol{\mu}_k^T\boldsymbol{\Sigma}_k\beta_k\mathbf{m}_k = \boldsymbol{\mu}_k^T\boldsymbol{\Sigma}_k\left(\beta_0\mathbf{m}_0+\sum_{t=1}^{T}\gamma_t(k)\mathbf{x}_t\right)$$

$$\mathbf{m}_k = \frac{1}{\beta_k}\left(\beta_0\mathbf{m}_0+\sum_{t=1}^{T}\gamma_t(k)\mathbf{x}_t\right)$$

Deriving ν_k by rearanging the term to obtain the form of the Wishart distribution:

$$\frac{(\nu_k-D-1)}{2}\log|\boldsymbol{\Sigma}_k| = \frac{(\nu_0-D-1)}{2}\log|\boldsymbol{\Sigma}_k| + \frac{1}{2}(\sum_{t=1}^{T}\gamma_t(k))\log|\boldsymbol{\Sigma}_k|$$

$$\nu_k = \nu_0+\sum_{t=1}^{T}\gamma_t(k)$$

B.2. Derivation VBAR-HMM

Variational E-step full derivation

$$
\begin{aligned}
\log q(\mathbf{z}_{1:T}) &\propto \mathbb{E}_{q(\boldsymbol{\pi},\mathbf{A},\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{U},\mathbf{V})}\left[\log p(\mathbf{x}_{1:T}, \hat{\mathbf{x}}_{2:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})\right] \\
&= \mathbb{E}_{q(\boldsymbol{\pi},\mathbf{A},\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{U},\mathbf{V})}\left[\log(p(\mathbf{x}_{1:T}|\hat{\mathbf{x}}_{2:T}, \mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\mathbf{U}, \mathbf{V})p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \mathbf{A})p(\mathbf{A})p(\boldsymbol{\pi})\right. \\
&\qquad \left. p(\boldsymbol{\mu}|\boldsymbol{\Sigma})p(\boldsymbol{\Sigma}))p(\mathbf{U}|\mathbf{V})p(\mathbf{V}))\right] \\
&= \mathbb{E}_{q(\boldsymbol{\pi},\mathbf{A})}\left[\log p(z_1|\boldsymbol{\pi})p(z_{2:T}|z_1, \mathbf{A})\right] + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}[\log p(\mathbf{x}_1|z_1, \boldsymbol{\mu}, \boldsymbol{\Sigma})] \\
&\qquad + \mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[\log p(\mathbf{x}_{1:T}|\hat{\mathbf{x}}_{2:T}, \mathbf{z}_{1:T}, \mathbf{U}, \mathbf{V})\right] \\
&= \mathbb{E}_{q(\boldsymbol{\pi},\mathbf{A})}\left[\log \prod_{k=1}^{K}\left(\boldsymbol{\pi}_k^{\mathbb{I}(z_1=k)}\right)\prod_{t=2}^{T}\prod_{i=1}^{K}\prod_{j=1}^{K} a_{ij}^{\mathbb{I}(z_t=j,z_{t-1}=i)}\right] \\
&\qquad + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\log \prod_{k=1}^{K}\mathcal{N}\left(\mathbf{x}_1|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right)^{\mathbb{I}(z_1=k)}\right] \\
&\qquad + \mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[\log \prod_{t=2}^{T}\prod_{k=1}^{K}\mathcal{N}\left(\mathbf{x}_t|\mathbf{U}_k\hat{\mathbf{x}}_{t-1}, \mathbf{V}_k^{-1}\right)^{\mathbb{I}(z_t=k)}\right] \\
&= \mathbb{E}_{q(\boldsymbol{\pi})}\left[\log \prod_{k=1}^{K}\left(\boldsymbol{\pi}_k^{\mathbb{I}(z_1=k)}\right)\right] + \mathbb{E}_{q(\mathbf{A})}\left[\log \prod_{t=2}^{T}\prod_{i=1}^{K}\prod_{j=1}^{K} a_{ij}^{\mathbb{I}(z_t=j,z_{t-1}=i)}\right] \\
&\qquad + \qquad + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\log \prod_{k=1}^{K}\mathcal{N}\left(\mathbf{x}_1|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right)^{\mathbb{I}(z_1=k)}\right] \\
&\qquad + \mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[\log \prod_{t=1}^{T}\prod_{k=1}^{K}\mathcal{N}\left(\mathbf{x}_t|\mathbf{U}_k\hat{\mathbf{x}}_{t-1}, \mathbf{V}_k^{-1}\right)^{\mathbb{I}(z_t=k)}\right] \\
&= \mathbb{E}_{q(\boldsymbol{\pi})}\left[\sum_{k=1}^{K}\mathbb{I}(z_1 = k)\log\left(\boldsymbol{\pi}_k\right)\right] \\
&\qquad + \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\sum_{k=1}^{K}\mathbb{I}(z_1 = k)\log\mathcal{N}\left(\mathbf{x}_1|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right)\right] \\
&\qquad + \mathbb{E}_{q(\mathbf{A})}\left[\sum_{t=2}^{T}\sum_{i=1}^{K}\sum_{j=1}^{K}\mathbb{I}(z_t = j, z_{t-1} = i)\log a_{ij}\right] \\
&\qquad + \mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[\sum_{t=2}^{T}\sum_{k=1}^{K}\mathbb{I}(z_t = k)\log\mathcal{N}\left(\mathbf{x}_t|\mathbf{U}_k\hat{\mathbf{x}}_{t-1}, \mathbf{V}_k^{-1}\right)\right]
\end{aligned}
$$

Continuation of the derivation for $\log q(\mathbf{z}_{1:T})$

$$
\begin{aligned}
\log q(\mathbf{z}_{1:T}) \propto & \mathbb{E}_{q(\boldsymbol{\pi})}\left[\sum_{k=1}^{K}\mathbb{I}(z_1=k)\log\left(\boldsymbol{\pi}_k\right)\right] \\
& + \mathbb{E}_{q(\mathbf{A})}\left[\sum_{t=2}^{T}\sum_{i=1}^{K}\sum_{j=1}^{K}\mathbb{I}(z_t=j,z_{t-1}=i)\log a_{ij}\right] \\
& + \sum_{k=1}^{K}\mathbb{I}(z_1=k) \\
& \mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{D}{2}\log 2\pi - \frac{1}{2}\left(\mathbf{x}_1-\boldsymbol{\mu}_k\right)^T\boldsymbol{\Sigma}_k\left(\mathbf{x}_1-\boldsymbol{\mu}_k\right)\right] \\
& + \sum_{t=2}^{T}\sum_{k=1}^{K}\mathbb{I}(z_t=k) \\
& \mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[\frac{1}{2}\log|\mathbf{V}_k| - \frac{D}{2}\log 2\pi - \frac{1}{2}\left(\mathbf{x}_t-\mathbf{U}_k\hat{\mathbf{x}}_{t-1}\right)^T\mathbf{V}_k\left(\mathbf{U}_k\hat{\mathbf{x}}_{t-1}\right)\right] \\
= & \sum_{k=1}^{K}\mathbb{I}(z_1=k)\mathbb{E}_{q(\boldsymbol{\pi})}\left[\log\boldsymbol{\pi}_k\right] + \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K}\mathbb{I}(z_{t-1}=k,z_t=j)\mathbb{E}_{q(\mathbf{A})}\left[\log a_{kj}\right] \\
& + \sum_{k=1}^{K}\mathbb{I}(z_1=k) \\
& \left\{\frac{1}{2}\mathbb{E}_{q(\boldsymbol{\Sigma})}\left[\log|\boldsymbol{\Sigma}_k|\right] - \frac{D}{2}\log\left(2\pi\right) - \frac{1}{2}\mathbb{E}_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}\left[\left(\mathbf{x}_1-\boldsymbol{\mu}_k\right)^T\boldsymbol{\Sigma}_k\left(\mathbf{x}_1-\boldsymbol{\mu}_k\right)\right]\right\} \\
& + \sum_{t=2}^{T}\sum_{k=1}^{K}\mathbb{I}(z_t=k) \\
& \left\{\frac{1}{2}\mathbb{E}_{q(\mathbf{V})}\left[\log|\mathbf{V}_k|\right] - \frac{D}{2}\log\left(2\pi\right)\right. \\
& \left.-\frac{1}{2}\mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[\left(\mathbf{x}_t-\mathbf{U}_k\hat{\mathbf{x}}_{t-1}\right)^T\mathbf{V}_k\left(\mathbf{x}_t-\mathbf{U}_k\hat{\mathbf{x}}_{t-1}\right)\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})^T\,\mathbf{V}_k\,(\mathbf{U}_k\hat{\mathbf{x}}_{t-1})\right] \\
&\quad= \int\int \left((\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})^T\mathbf{V}_k(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})\right) q(\mathbf{U}_k,\mathbf{V}_k)\,\mathrm{d}\mathbf{U}_k\,\mathrm{d}\mathbf{V}_k \\
&\quad= \int\left\{\int (\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})^T\mathbf{V}_k(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1}) q(\mathbf{U}_k|\mathbf{V}_k)\,\mathrm{d}\mathbf{U}_k\right\} q(\mathbf{V}_k)\,\mathrm{d}\mathbf{V}_k \\
&\quad= \int \mathbb{E}_{q(\mathbf{U})}\left[(\mathbf{U}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T\mathbf{V}_k(\mathbf{U}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)\right] q(\mathbf{V}_k)\,\mathrm{d}\mathbf{V}_k
\end{aligned}
$$

The inner expectation corresponds to a Matrix-Normal distribution. Thus, we make use of transformations derived in [38] and get two expectations:

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{U}_k\right] &= \mathbf{M}_k \\
\mathbb{E}\left[\mathbf{U}_k\mathbf{Q}\mathbf{U}_k^T\right] &= \mathbf{M}_k\mathbf{U}\mathbf{M}_k^T + \mathbf{Tr}\left\{\mathbf{K}_k^{-1}\mathbf{Q}\right\}\mathbf{V}_k^{-1}
\end{aligned}
$$

Applying this transformations on the inner expectation from Equation (B.2) gives:

$$
\begin{aligned}
&\mathbb{E}_{q(\mathbf{U})}\left[(\mathbf{U}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T\mathbf{V}_k(\mathbf{U}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t\right] \\
&\quad= \mathbb{E}_{q(\mathbf{U})}\left[\mathbf{Tr}\left\{\mathbf{V}_k(\mathbf{U}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)(\mathbf{U}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T\right\}\right] \\
&\quad= \mathbf{Tr}\Big\{\mathbf{V}_k(\mathbb{E}_{q(\mathbf{U})}\left[\mathbf{U}_k\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\mathbf{U}_k^T\right] \\
&\qquad -\mathbb{E}_{q(\mathbf{U})}\left[\mathbf{U}_k\right]\hat{\mathbf{x}}_{t-1}\mathbf{x}_t^T - \mathbf{x}_t\hat{\mathbf{x}}_{t-1}^T\mathbb{E}_{q(\mathbf{U})}\left[\mathbf{U}_k^T\right] + \mathbf{x}_t\mathbf{x}_t^T)\Big\} \\
&\quad= \mathbf{Tr}\Big\{\mathbf{V}_k(\mathbf{M}_k\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\mathbf{M}_k^T + (\mathbf{Tr}\left\{\mathbf{K}_k^{-1}\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\right\}\mathbf{V}_k^{-1}) \\
&\qquad -\mathbf{M}_k\hat{\mathbf{x}}_{t-1}\mathbf{x}_t^T - \mathbf{x}_t\hat{\mathbf{x}}_{t-1}^T\mathbf{M}_k^T + \mathbf{x}_t\mathbf{x}_t^T)\Big\} \\
&\quad= \mathbf{Tr}\left\{\mathbf{V}_k\left((\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T + (\mathbf{Tr}\left\{\mathbf{K}_k^{-1}\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\right\}\mathbf{V}_k^{-1})\right)\right\} \\
&\quad= (\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T\mathbf{V}_k(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t) + \mathbf{Tr}\left\{\mathbf{K}_k^{-1}\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}_{q(\mathbf{U},\mathbf{V})}\left[(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})^T\,\mathbf{V}_k\,(\mathbf{U}_k\hat{\mathbf{x}}_{t-1})\right] \\
&\quad= \int \left[(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T\mathbf{V}_k(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t) + \mathbf{Tr}\left\{\mathbf{K}_k^{-1}\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\right\}\right] q(\mathbf{V}_k)\,\mathrm{d}\mathbf{V}_k \\
&\quad= \mathbb{E}_{q(\mathbf{V})}\left[(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T\mathbf{V}_k(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t) + \mathbf{Tr}\left\{\mathbf{K}_k^{-1}\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\right\}\right] \\
&\quad= \mathbf{Tr}\left\{\mathbb{E}_{q(\mathbf{V})}\left[\mathbf{V}_k\right](\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T\right\} + \mathbf{Tr}\left\{\mathbf{K}_k^{-1}\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\right\} \\
&\quad= \mathbf{Tr}\left\{\eta_k\mathbf{P}_k(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T\right\} + \mathbf{Tr}\left\{\mathbf{K}_k^{-1}\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\right\} \\
&\quad= \eta_k(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t)^T\mathbf{P}_k(\mathbf{M}_k\hat{\mathbf{x}}_{t-1} - \mathbf{x}_t) + \mathbf{Tr}\left\{\mathbf{K}_k^{-1}\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\right\}
\end{aligned}
$$

Variational M-step full derivation

$$
\begin{aligned}
\log q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V}) &\propto \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{x}_{1:T}, \hat{\mathbf{x}}_{2:T}, \mathbf{z}_{1:T}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})\right] \\
&= \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log(p(\mathbf{x}_{1:T}|\hat{\mathbf{x}}_{2:T}, \mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})\right. \\
&\qquad \left. p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \mathbf{A})p(\boldsymbol{\pi})p(\mathbf{A})p(\mathbf{U}|\boldsymbol{\mu})p(\boldsymbol{\Sigma})p(\mathbf{U}|\mathbf{V})p(\mathbf{V}))\right] \\
&= \log p(\boldsymbol{\pi}) + \log p(\mathbf{A}) \\
&\qquad + \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{x}_{1:T}|\hat{\mathbf{x}}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})\right] \\
&\qquad + \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{z}_{1:T}|\boldsymbol{\pi}, \mathbf{A})\right] + \sum_{k=1}^{K}\log p(\boldsymbol{\mu}_k|\boldsymbol{\Sigma}_k)p(\boldsymbol{\Sigma}_k) \\
&\qquad + \sum_{k=1}^{K}\log p(\mathbf{U}_k|\mathbf{V}_k)p(\mathbf{V}_k) \\
&= \log p(\boldsymbol{\pi}) + \log p(\mathbf{A}) \\
&\qquad + \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{x}_1|\mathbf{z}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\right] \\
&\qquad + \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(\mathbf{x}_{2:T}|\hat{\mathbf{x}}_{2:T}, \mathbf{z}_{2:T}, \mathbf{U}, \mathbf{V})\right] \\
&\qquad + \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\log p(z_1|\boldsymbol{\pi})p(\mathbf{z}_{2:T}|z_1, \mathbf{A})\right] \\
&\qquad + \sum_{k=1}^{K}\log p(\boldsymbol{\mu}_k|\boldsymbol{\Sigma}_k)p(\boldsymbol{\Sigma}_k) + \sum_{k=1}^{K}\log p(\mathbf{U}_k|\mathbf{V}_k)p(\mathbf{V}_k)
\end{aligned}
$$

Continuation of deriving $\log q(\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U}, \mathbf{V})$

$$
\begin{aligned}
\log q(\boldsymbol{\pi}, \mathbf{A}, \mathbf{U}, \mathbf{V}) \propto \log Dir(\boldsymbol{\pi}|\boldsymbol{\omega}_{\mathbf{0}}^{(\boldsymbol{\pi})}) + \sum_{k=1}^{K} \log Dir(\mathbf{A}|\boldsymbol{\omega}_{0k}^{(\mathbf{A})}) \\
+ \sum_{k=1}^{K} \mathbb{E}_{q(z_1)}\left[\mathbb{I}(z_1 = k)\right] \log \mathcal{N}\left(\mathbf{x}_1|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^{-1}\right) \\
+ \sum_{t=2}^{T} \sum_{k=1}^{K} \mathbb{E}_{q(\mathbf{z}_{2:T})}\left[\mathbb{I}(z_t = k)\right] \log \mathcal{N}\left(\mathbf{x}_t|\mathbf{U}_k\hat{\mathbf{x}}_{t-1}, \mathbf{V}_k^{-1}\right) \\
+ \sum_{k=1}^{K} \left\{\mathbb{E}_{q(z_1)}\left[\mathbb{I}(z_1 = k)\right] \log \pi_k\right\} \\
+ \sum_{t=2}^{T} \sum_{i=1}^{K} \sum_{j=1}^{K} \mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\mathbb{I}(z_t = j, z_{t-1} = i)\right] \log a_{ij} \\
+ \sum_{k=1}^{K} \log \mathcal{N}\left(\boldsymbol{\mu}_k|\boldsymbol{m}_0, (\beta_0\boldsymbol{\Sigma}_k)^{-1}\right) + \sum_{k=1}^{K} \log \mathcal{W}\left(\boldsymbol{\Sigma}_k|\mathbf{W}_{\mathbf{0}}, \nu_0\right) \\
+ \sum_{k=1}^{K} \log \mathcal{MN}\left(\mathbf{U}_k|\mathbf{M}_0, \mathbf{V}_k^{-1}, \mathbf{K}_0\right) + \sum_{k=1}^{K} \log \mathcal{W}\left(\mathbf{V}_k|\mathbf{P}_0, \eta_0\right)
\end{aligned}
$$

$$
\begin{aligned}
\log q(\mathbf{U}_k|\mathbf{V}_k) &\propto -\frac{1}{2}\mathbf{Tr}\left\{(\mathbf{U}_k - \mathbf{M}_0)^T\mathbf{V}_k(\mathbf{U}_k - \mathbf{M}_0)\mathbf{K}_0\right\} \\
&\quad - \frac{1}{2}\sum_{t=2}^{T} \gamma_t(k)(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})^T\mathbf{V}_k(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1}) \\
&= -\frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k(\mathbf{U}_k - \mathbf{M}_0)\mathbf{K}_0(\mathbf{U}_k - \mathbf{M}_0)^T\right\} \\
&\quad - \frac{1}{2}\sum_{t=2}^{T} \gamma_t(k)\mathbf{Tr}\left\{\mathbf{V}_k(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})^T\right\} \\
&= -\frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k(\mathbf{U}_k\mathbf{K}_0\mathbf{U}_k^T - \mathbf{U}_k\mathbf{K}_0\mathbf{M}_0^T - \mathbf{M}_0\mathbf{K}_0\mathbf{U}_k^T + \mathbf{M}_0\mathbf{K}_0\mathbf{M}_0^T)\right\} \\
&\quad - \frac{1}{2}\sum_{t=2}^{T} \gamma_t(k)\mathbf{Tr}\left\{\mathbf{V}_k(\mathbf{x}_t\mathbf{x}_t^T - \mathbf{x}_t\hat{\mathbf{x}}_{t-1}^T\mathbf{U}_k^T - \mathbf{U}_k\hat{\mathbf{x}}_{t-1}\mathbf{x}_t^T + \mathbf{U}_k\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T\mathbf{U}_k^T)\right\}
\end{aligned}
$$

$$-\frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k\mathbf{U}_k\mathbf{K}_k\mathbf{U}_k^T\right\} = -\frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k\mathbf{U}_k\left[\sum_{t=2}^{T}\mathbb{E}_q\left[\mathbb{I}(z_t=k)\right]\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T + \mathbf{K}_0\right]\mathbf{U}_k^T\right\}$$

$$\mathbf{V}_k\mathbf{U}_k\mathbf{K}_k\mathbf{U}_k^T = \mathbf{V}_k\mathbf{U}_k\left[\sum_{t=2}^{T}\mathbb{E}_q\left[\mathbb{I}(z_t=k)\right]\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T + \mathbf{K}_0\right]\mathbf{U}_k^T$$

$$\mathbf{K}_k = \sum_{t=2}^{T}\mathbb{E}_q\left[\mathbb{I}(z_t=k)\right]\hat{\mathbf{x}}_{t-1}\hat{\mathbf{x}}_{t-1}^T + \mathbf{K}_0$$

$$-\frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k\mathbf{U}_k\mathbf{M}_k\mathbf{K}_k\right\} = -\frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k\mathbf{U}_k\left[-\sum_{t=2}^{T}\mathbb{E}_q\left[\mathbb{I}(z_t=k)\right]\mathbf{x}_t\hat{\mathbf{x}}_{t-1}^T - \mathbf{M}_0\mathbf{K}_0\right]\right\}$$

$$\mathbf{V}_k\mathbf{U}_k\mathbf{M}_k\mathbf{K}_k = \mathbf{V}_k\mathbf{U}_k\left[\sum_{t=2}^{T}\mathbb{E}_q\left[\mathbb{I}(z_t=k)\right]\mathbf{x}_t\hat{\mathbf{x}}_{t-1}^T + \mathbf{M}_0\mathbf{K}_0\right]$$

$$\mathbf{M}_k\mathbf{K}_k = \left[\sum_{t=2}^{T}\mathbb{E}_q\left[\mathbb{I}(z_t=k)\right]\mathbf{x}_t\hat{\mathbf{x}}_{t-1}^T + \mathbf{M}_0\mathbf{K}_0\right]$$

$$\mathbf{M}_k = \left[\sum_{t=2}^{T}\mathbb{E}_q\left[\mathbb{I}(z_t=k)\right]\mathbf{x}_t\hat{\mathbf{x}}_{t-1}^T + \mathbf{M}_0\mathbf{K}_0\right]\mathbf{K}_k^{-1}$$

$$
\begin{aligned}
\log q(\mathbf{V}_k) &= \log \mathcal{W}\left(\mathbf{U}_k|\mathbf{P}_k, \eta_k\right) \\
&= \log q(\mathbf{U}_k, \mathbf{V}_k) - \log q(\mathbf{U}_k|\mathbf{V}_k) \\
&= -\frac{1}{2}\mathbf{Tr}\left\{(\mathbf{U}_k - \mathbf{M}_0)^T\mathbf{V}_k(\mathbf{U}_k - \mathbf{M}_0)\mathbf{K}_0\right\} - \frac{m}{2}\log(|\mathbf{V}_k|) \\
&\quad - \frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k\mathbf{W}_0^{-1}\right\} + \frac{(\eta_0 - D - 1)}{2}\log|\mathbf{V}_k| \\
&\quad - \frac{1}{2}\sum_{t=2}^{T}\mathbb{E}_{q(\mathbf{z}_{2:T})}\left[\mathbb{I}(z_t = k)\right](\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})^T\mathbf{V}_k(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1}) \\
&\quad + \frac{1}{2}\left(\sum_{t=2}^{T}\mathbb{E}_{q(\mathbf{z}_{2:T})}\left[\mathbb{I}(z_t = k)\right]\right)\log|\mathbf{V}_k| \\
&\quad + \frac{m}{2}\log|\mathbf{V}_k| + \frac{1}{2}\mathbf{Tr}\left\{(\mathbf{U}_k - \mathbf{M}_k)^T\mathbf{V}_k(\mathbf{U}_k - \mathbf{M}_k)\mathbf{K}_k\right\} \\
&\propto \frac{(\eta_k - D - 1)}{2}\log|\mathbf{V}_k| - \frac{1}{2}\mathbf{Tr}\left\{\mathbf{V}_k\mathbf{P}_k^{-1}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{P}_k^{-1} &= \sum_{t=2}^{T}\mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\mathbb{I}(z_t = k)\right](\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})(\mathbf{x}_t - \mathbf{U}_k\hat{\mathbf{x}}_{t-1})^T \\
&\quad + (\mathbf{U}_k - \mathbf{M}_0)\mathbf{K}_0(\mathbf{U}_k - \mathbf{M}_0)^T + \mathbf{P}_0^{-1} - (\mathbf{U}_k - \mathbf{M}_k)\mathbf{K}_k(\mathbf{U}_k - \mathbf{M}_k)^T \\
&= \mathbf{P}_0^{-1} + \mathbf{M}_0\mathbf{K}_0\mathbf{M}_0^T + \sum_{t=2}^{T}\gamma_t(k)\mathbf{x}_t\mathbf{x}_t^T - \mathbf{M}_k\mathbf{K}_k\mathbf{M}_k^T
\end{aligned}
$$

$$
\frac{(\eta_k - D - 1)}{2}\log|\mathbf{V}_k| = \frac{(\eta_0 - D - 1)}{2}\log|\mathbf{V}_k| + \frac{1}{2}\left(\sum_{t=1}^{T}\mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\mathbb{I}(z_t = k)\right]\right)\log|\mathbf{V}_k|
$$

$$
\eta_k = \eta_0 + \sum_{t=1}^{T}\mathbb{E}_{q(\mathbf{z}_{1:T})}\left[\mathbb{I}(z_t = k)\right]
$$

C. Distributions

C.1. Dirichlet distribution

Notation and parameters

$$\begin{aligned}\boldsymbol{\pi} &\sim Dir(\boldsymbol{\alpha}) \\ \boldsymbol{\alpha} &= \{\alpha_1, ..., \alpha_k\} \\ \alpha_j > 0;\;\; \alpha_0 &= \sum_{j=1}^{K} \alpha_j\end{aligned}$$

Density function

$$\begin{aligned}p(\boldsymbol{\pi}|\boldsymbol{\alpha}) &= C(\boldsymbol{\alpha})\pi_1^{\alpha_1-1}...\pi_k^{\alpha_k-1} \\ \pi_1, ..., \pi_k \geq 0;\;\; \sum_{j=1}^{K} \pi_j &= 1 \\ C(\boldsymbol{\alpha}) &= \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)...\Gamma(\alpha_k)}\end{aligned}$$

Expectations

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\pi}\right] &= \boldsymbol{\alpha}/\alpha_0 \\ \log \mathbb{E}\left[\pi_j\right] &= \psi(\alpha_j) - \psi(\alpha_0)\end{aligned}$$

KL-divergence

$$D_{\mathrm{KL}}\left(\tilde{\boldsymbol{\alpha}}||\boldsymbol{\alpha}\right) = \ln\frac{\Gamma(\tilde{\alpha}_0)}{\Gamma(\alpha_0)} - \sum_{j=1}^{K}\left[\ln\frac{\Gamma(\tilde{\alpha}_j)}{\Gamma(\alpha_j)} - (\tilde{\alpha}_j - \alpha_j)(\psi(\tilde{\alpha}_j) - \psi(\tilde{\alpha}_0))\right]$$

Entropy

$$H\left[\boldsymbol{\pi}\right] = -\sum_{k=1}^{K}(\alpha_k - 1)(\psi(\alpha_k) - \psi(\alpha_0)) - \ln C(\boldsymbol{\alpha})$$

C.2. Uniform distibution

Notation and parameters

$$\begin{aligned} x &\sim U(a,b) \\ \text{boundaries } & a,b \\ \text{with } & b > a \end{aligned}$$

Density function

$$p(x|a,b) = \frac{1}{b-a}, \quad x \in [a,b]$$

Expectations

$$\mathbb{E}\left[x\right] = \frac{a+b}{2}$$

Entropy

$$H\left[x\right] = \ln(b-a)$$

C.3. Multivariate normal distribution

Notation and parameters

$$\begin{aligned}
\mathbf{x} &\sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\boldsymbol{\mu} &\text{ mean vector} \\
\boldsymbol{\Sigma} &\text{ covariance matrix} \\
\boldsymbol{\Sigma}^{-1} &\text{ precision matrix}
\end{aligned}$$

Density function

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-D/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))\right\}$$

Expectations

$$\begin{aligned}
\mathbb{E}\left[\mathbf{x}\right] &= \boldsymbol{\mu} \\
\mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right] &= \boldsymbol{\Sigma}
\end{aligned}$$

KL-divergence

$$D_{\mathrm{KL}}\left(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}||\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = -\frac{1}{2}\left(\ln|\tilde{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| + \mathbf{Tr}\left\{I - \left[\tilde{\boldsymbol{\Sigma}} + (\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu})(\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu})^T\right]\boldsymbol{\Sigma}^{-1}\right\}\ln e\right)$$

Entropy

$$H[\mathbf{x}] = \frac{1}{2}\ln|\boldsymbol{\Sigma}| + \frac{D}{2}(1+\ln(2\pi))$$

C.4. Wishart distribution

Notation and parameters

$$\begin{aligned}\boldsymbol{\Sigma} &\sim \mathcal{W}(\mathbf{W},\nu)\\ \mathbf{W} &\text{ precision matrix}\\ \nu &\text{ degree of freedom}\end{aligned}$$

Density function

$$p(\boldsymbol{\Sigma}|\mathbf{W},\nu) = B(\mathbf{W},\nu)|\boldsymbol{\Sigma}|^{(\nu-D-1)/2}\exp\left(-\frac{1}{2}\mathbf{Tr}\left\{\mathbf{W}^{-1}\boldsymbol{\Sigma}\right\}\right)$$
$$B(\mathbf{W},\nu) = |\mathbf{W}|^{-\nu/2}\left(2^{\nu D/2}\pi^{D(D-1)/4}\prod_{i=1}^{D}\Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$$

Expectations

$$\mathbb{E}[\boldsymbol{\Sigma}] = \mathbf{W}\nu$$
$$\mathbb{E}[\ln|\boldsymbol{\Sigma}|] = \sum_{i=1}^{D}\psi\left(\frac{\nu+1-i}{2}\right) + D\ln 2 + \ln|\mathbf{W}|$$

KL-divergence

$$D_{\mathrm{KL}}\left(\tilde{\mathbf{W}}, \tilde{\nu} || \mathbf{W}, \nu\right) = \ln \frac{B(\mathbf{W}, \nu)}{\tilde{\mathbf{W}}, \tilde{\nu}} + \frac{\tilde{\nu} - \nu}{2} \mathbb{E}\left[\ln |\mathbf{\Sigma}|\right] + \frac{1}{2} \tilde{\nu} \mathbf{Tr}\left\{\mathbf{W}^{-1} \tilde{\mathbf{W}} - I\right\}$$

Entropy

$$H\left[\mathbf{\Sigma}\right] = -\ln B(\mathbf{W}, \nu) - \frac{\nu - D - 1}{2} \mathbb{E}\left[|\mathbf{\Sigma}|\right] + \frac{\nu D}{2}$$