Probabilistic models for neural data: From single neurons to population dynamics

NEUROBIO 316QC

Jan Drugowitsch

jan_drugowitsch@hms.harvard.edu

Session 6: The Laplace approximation & state space models

Today

Q&A about previous session

Paper discussion (~1h)

Laplace approximation & state space models (~30min)

Overview

Laplace approximation

State space models

General structure, and dependence/independence

Filtering (forward pass)

Smoothing (forward & backward pass)

Overview

Laplace approximation

State space models

- General structure, and dependence/independence
- Filtering (forward pass)
- Smoothing (forward & backward pass)

The Laplace approximation

Likelihood $p(\boldsymbol{z})$ non-Gaussian in $\mathbf{z}$ & potentially no conjugate prior (e.g., logistic regression)

Aim: approximate likelihood by Gaussian $q(\boldsymbol{z}) = N(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$; which mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$?

Approach: observe that Gaussian mode at $\boldsymbol{\mu}$ & log-probability is quadratic function

1. Match modes (optimization; e.g., gradient ascent)

$$\boldsymbol{\mu} = \mathrm{argmax}_{\boldsymbol{z}} p(\boldsymbol{z}) \longrightarrow \nabla_{\boldsymbol{z}} \log p(\boldsymbol{z})\Big|_{\boldsymbol{z}=\boldsymbol{\mu}} = 0$$

(as $\boldsymbol{\mu}$ at maximum $p(\boldsymbol{z})$)

2. Second-order Taylor expand $\log p(\boldsymbol{z})$ around $\boldsymbol{z} = \boldsymbol{\mu}$ to find quadratic log-probability

$$\log p(\boldsymbol{z}) \approx \log p(\boldsymbol{z})\Big|_{\boldsymbol{z}=\boldsymbol{\mu}} + \cancelto{0}{\nabla_{z} \log p(\boldsymbol{z})\Big|_{\boldsymbol{z}=\boldsymbol{\mu}}} (\boldsymbol{z}-\boldsymbol{\mu}) - \frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})^T \boldsymbol{A} (\boldsymbol{z}-\boldsymbol{\mu})$$

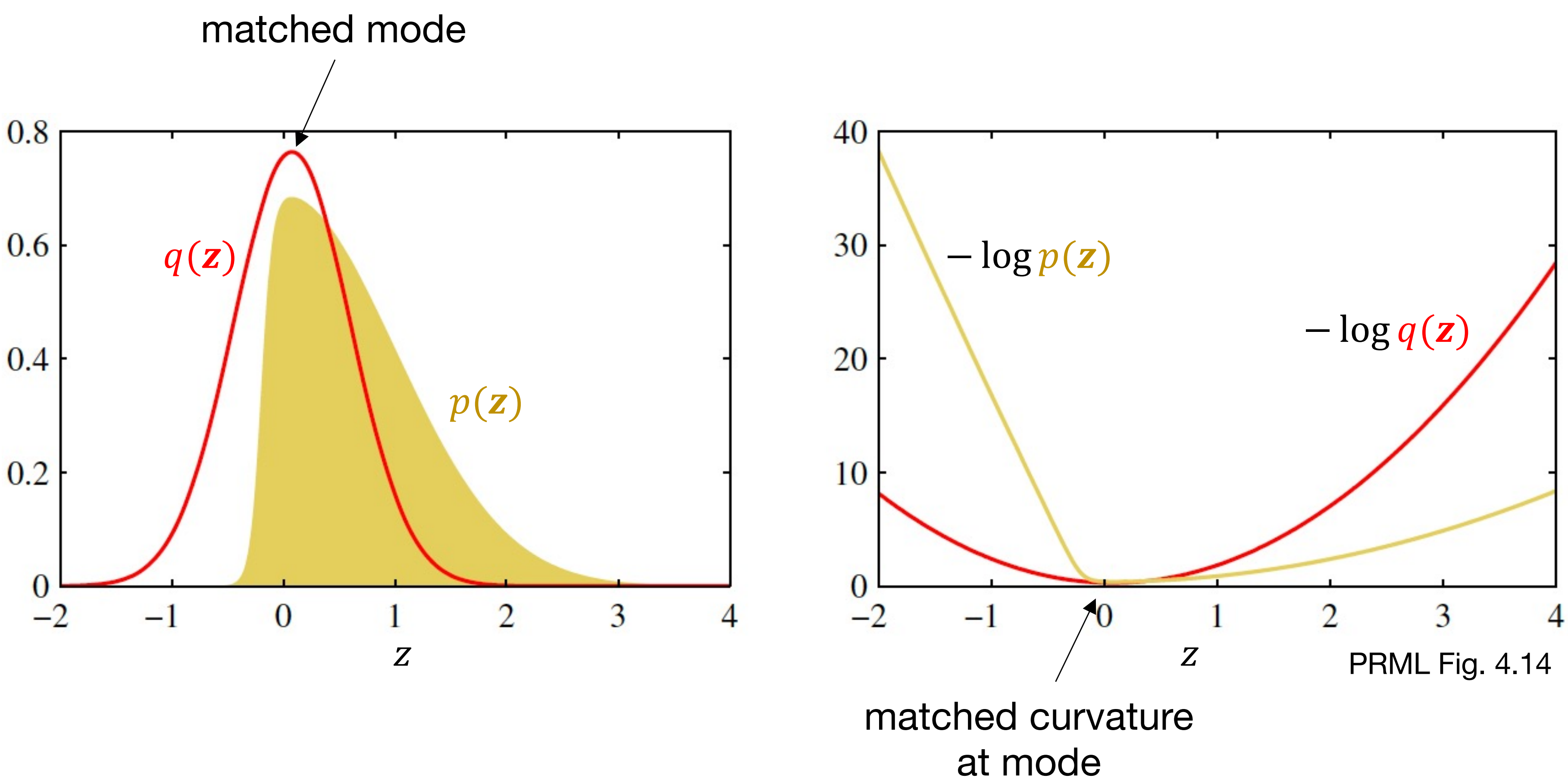
$$\boldsymbol{A} = -\nabla\nabla_{\boldsymbol{z}} \log p(\boldsymbol{z})\Big|_{\boldsymbol{z}=\boldsymbol{\mu}}$$ (neg. Hessian)

$$\log q(\boldsymbol{z}) = \log \mathrm{N}(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}-\boldsymbol{\mu}) + \mathrm{const.}$$

Other use of (neg.) Hessian: **quantify posterior uncertainty** (implicit Laplace approximation)

approximate posterior covariance by $$\left(-\nabla\nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|\boldsymbol{X})\Big|_{\boldsymbol{\theta}=\mathrm{argmax}_{\theta} p(\boldsymbol{\theta}|\boldsymbol{X})}\right)^{-1}$$

Illustrating the Laplace approximation



PRML Fig. 4.14

Good approximation	“Gaussian-like” distribution (light tail, ~symmetric)
Bad approximation	Heavily skewed, heavy tails

Overview

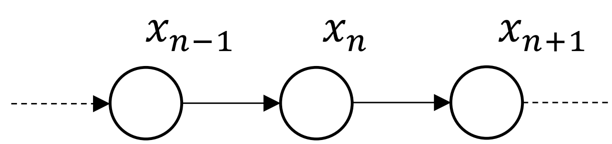
Laplace approximation

State space models

- General structure, and dependence/independence
- Filtering (forward pass)
- Smoothing (forward & backward pass)

Markov chain

Assumes $p(x_{n+1}|x_1, \ldots, x_n) = p(x_{n+1}|x_n)$



Interpretation of x_n full characterization of a system's state, everything that needs to be known to predict future states

Examples Full-information board games (e.g., chess, backgammon, …) Brownian motion

Violations x = location for movement with momentum (velocity missing) activity of individual neuron in network

Higher-order Markov chains

2nd order $p(x_{n+1}|x_1, \ldots, x_n) = p(x_{n+1}|x_n, x_{n-1})$

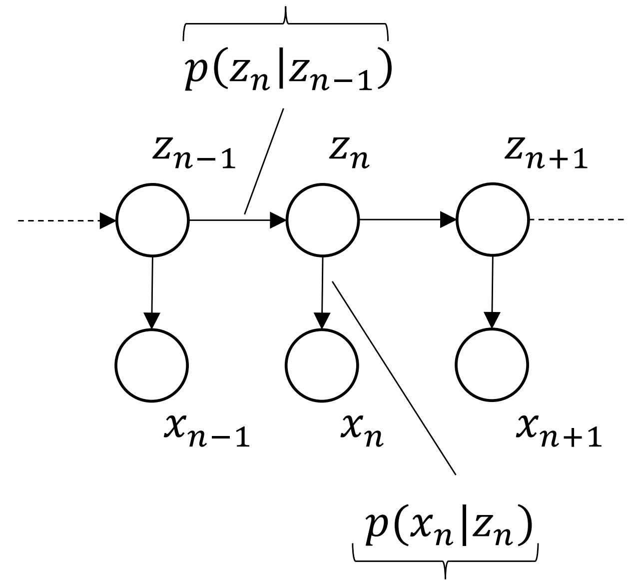
3rd order $p(x_{n+1}|x_1, \ldots, x_n) = p(x_{n+1}|x_n, x_{n-1}, x_{n-2})$

General structure of state space models

(Markovian) *transition model*:
how latent states are assumed to evolve

Sequence of latent states, $z_1, z_2, \ldots$

Sequence of observations, $x_1, x_2, \ldots$



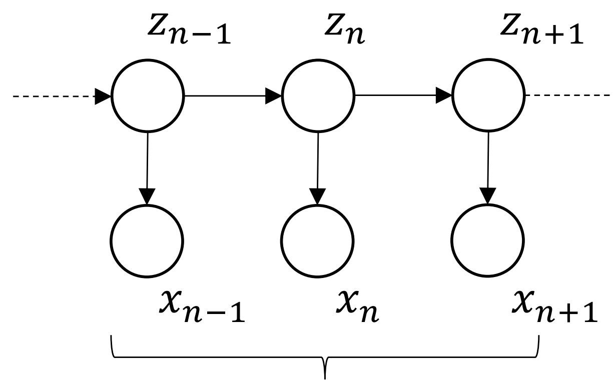
emission model:
how observations are generated from latent states

Special cases

Hidden Markov Model (HMM): Discrete latent states z_n,
such that $p(z_n|z_{n-1})$ specified by transition matrix

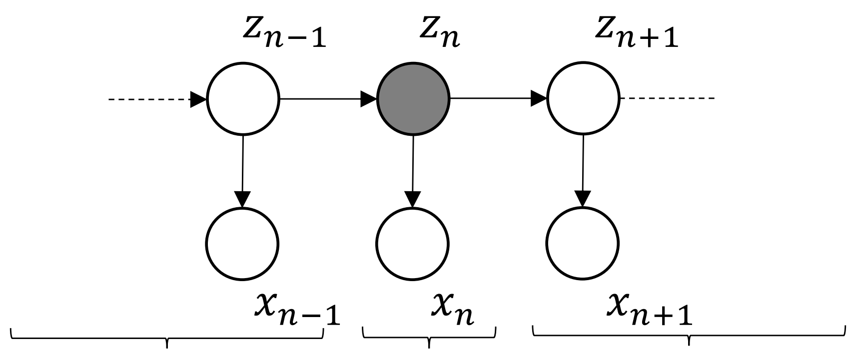
Kalman filter: Continuous observations x_n and latent states z_n,
with linear-Gaussian $p(z_n|z_{n-1})$ and $p(x_n|z_n)$

Dependence/independence in state space models



without latent state conditioning,
observations are all dependent on each other

Conditional on z_n (i.e., hypothesizing or observing its value)

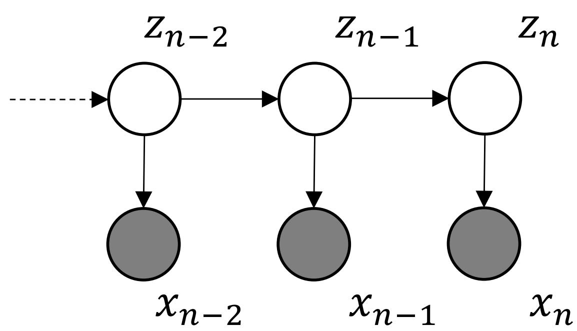


"past" "present" "future" ← conditionally independent

$$p(x_{1:N}|z_n) = p(x_{1:n-1}|z_n)p(x_n|z_n)p(x_{n+1:N}|z_n)$$

Filtering (forward pass)

“What is my estimate of the current latent state z_n, given all observations so far $x_{1:n}$”



$p(z_n|x_{1:n})$?

Assume we know $p(z_{n-1}|x_{1:n-1})$ (i.e., previous latent state given all observations until then)

Prediction step: predict next latent state z_n given observations $x_{1:n-1}$ (i.e., not including x_n)

conditional independence

$$p(z_n|x_{1:n-1}) = \sum_{z_{n-1}} p(z_n, z_{n-1}|x_{1:n-1}) = \sum_{z_{n-1}} \underbrace{p(z_n|z_{n-1}, \cancel{x_{1:n-1}})}_{\text{transition model}} \underbrace{p(z_{n-1}|x_{1_n-1})}_{\text{known from previous step}}$$

Observation step: include knowledge of observed x_n

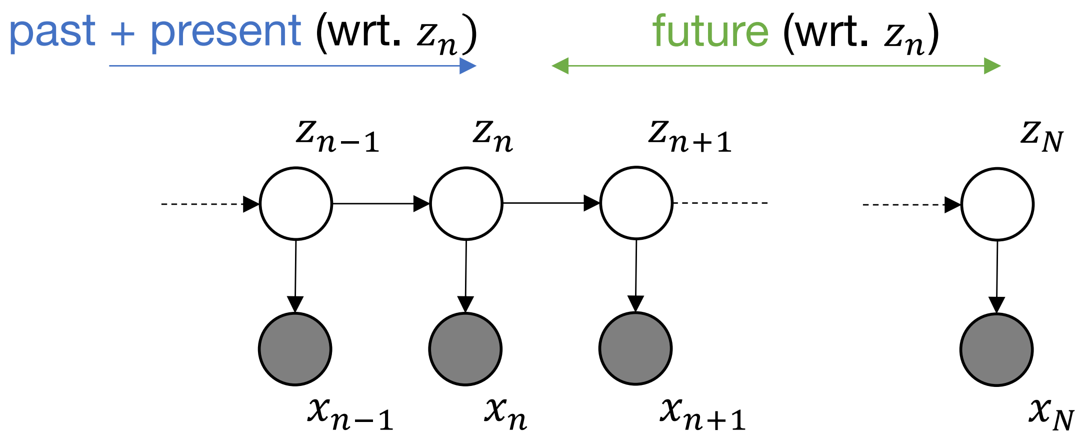
Bayes’ rule

conditional independence

$$p(z_n|x_{1:n}) = p(z_n|x_n, x_{1:n-1}) \propto \underbrace{p(x_n|z_n, \cancel{x_{1:n-1}})}_{\substack{\text{emission model:} \\ \text{likelihood for observation } x_n}} \underbrace{p(z_n|x_{1:n-1})}_{\text{“prior” computed in prediction step}}$$

Smoothing (forward & backward pass)

"What is my estimate of some past latent state z_n, given all observations $x_{1:N}$"



$p(z_n|x_{1:N})$?

splitting into past/present & future

conditional independence

$$p(z_n|x_{1:N}) = \frac{p(z_n, x_{1:N})}{p(x_{1:N})} = \frac{p(z_n, x_{1:n}, x_{n+1:N})}{p(x_{1:N})} = \frac{p(x_{n+1:N}|z_n, \cancel{x_{1:n}})p(z_n|x_{1:n})p(x_{1:n})}{p(x_{1:N})}$$

$$\propto \underbrace{p(x_{n+1:N}|z_n)}_{\text{including information from "future" observations (the "backward" step in HMMs)}}\underbrace{p(z_n|x_{1:n})}_{\text{information from "past, current" observations (from filtering, "forward" step in HMMs)}}$$

Overview

Laplace approximation

State space models

- General structure, and dependence/independence
- Filtering (forward pass)
- Smoothing (forward & backward pass)

Summary

Laplace approximation (Gaussian): quadratic approximation of target log-probability

Markov chains assume that x completely characterizes the system's state

State space models assume latent Markov chain observed through noisy emissions

Special cases: HMMs and Kalman filter

Filtering (forward pass): best estimate given all past

Smoothing (forward & backward pass): best estimate given past & future

Until next week

Read paper and prepare presentation (see notes for Session 7)

Read statistical methods section (see notes for Session 7)

Next session

Q&A for previous session

Paper discussions (~1h)

Kernel methods & Gaussian Processes (~30min)