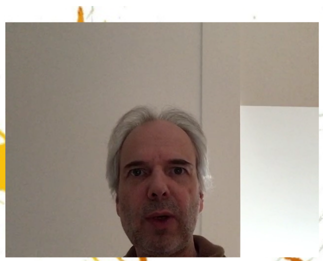


TIME SERIES ANALYSIS & RECURRENT NEURAL NETWORKS #11

- Generative RNN models
- Extended Kalman filter, particle filters
- Laplace approximation

Main lecture: Daniel Durstewitz

Exercises: Leonard Bereska, Manuel Brenner, Daniel Kramer, Georgia Koppe

Heidelberg University

Generative RNN models

$X = \{x_1, \dots x_T\}$ obs. $\quad x_t \uparrow$ generate $\quad z_t$

$z_t = \phi(W z_{t-1} + h)$

$\rightarrow p_\theta(z_t | z_{t-1})$

$Z = \{z_1, \dots z_T\}$ latent

$\rightarrow$ state space models

ini. cond. $z_0 \sim N(\mu_0, \Sigma_0)$

obs. eq. $x_t \sim N(B z_t, \Gamma)$

process eq. $z_t \sim N(A z_{t-1} + h, \Sigma)$

$z_t \sim N(\phi(W z_{t-1} + h), \Sigma)$

Unsupervised learning

Log-Likelihood $\rightarrow$ max. LL

$$\underbrace{p(X|\theta)}_{p_\theta(X)} = \int_z p(X, z)\, dz$$

$$\geq \int_z q(z) \log \frac{p(X,z)}{q(z)} dz \qquad p(X,z) = p(z|X)\,p(X), \quad q(z) = q(z|X)$$

$$= E_q[\log p(X, z)] + H[q(z)]$$

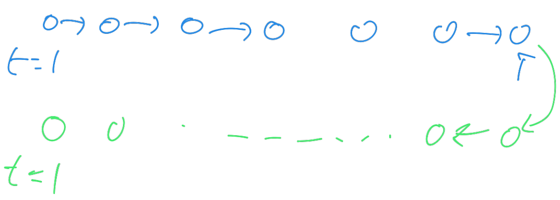
$$= p(X|\theta) - KL[q(z) \,\|\, p(z|X)] \qquad q(z) = q(z|X)$$

Expectation Max. (EM)

E-step: $q^{(n+1)} := \underset{q}{\arg\max}\, ELBO[q, \theta^{(n)}]$ given fixed $\theta^{(n)}$

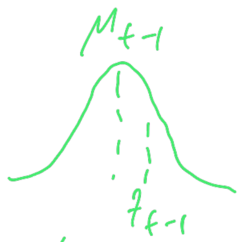
M-step: $\theta^{(n+1)} := \underset{\theta}{\arg\max}\, ELBO[q^{(n)}, \theta]$ given fixed q

Extended Kalman Filter (EKF)



$p(z_t | x_1, \dots x_t) = N(\mu_t, V_t)$ filter

$p(z_t | x_1, \dots x_T) = N(\tilde{\mu}_t, \tilde{V}_t)$ smoother



obs. eq. $x_t = B z_t + \eta_t, \quad \eta_t \sim N(0, \Gamma)$

ini. cond. $z_1 \sim N(\mu_0, \Sigma_0)$

process $z_t = F_\theta(z_{t-1}) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma)$

Kalman filter

$$p(z_t | x_1, \dots x_t) = \frac{\overbrace{p(x_t | z_t)}^{N(B z_t, \Gamma)} \int_{z_{t-1}} \overbrace{p(z_t | z_{t-1})}^{N(F_\theta(z_{t-1}), \Sigma)}\, p(z_{t-1} | x_1, \dots x_{t-1})\, d z_{t-1}}{p(x_t | x_1 \dots x_{t-1})}$$

Taylor series

$$F_\theta(z_{t-1}) \approx F_\theta(\mu_{t-1}) + \underbrace{F'_\theta(\mu_{t-1})}_{\text{Jacobi matrix}}(z_{t-1} - \mu_{t-1}) \qquad \in \mathbb{R}^{M \times 1}$$

$$\frac{\partial F_\theta(\mu_{t-1})}{\partial \mu_{t-1}}$$

$$p(z_t | z_{t-1}) \approx (2\pi)^{-M/2} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(z_t - F}$$

M - step

$z_1 \sim N(\mu_0, \Sigma_0)$

$x_t = B z_t + \eta_t$

$z_t = \phi(W z_{t-1} + h)$

$\to z_t = W \phi(z_{t-1}) + h$

$$E_q[\log p(x,z)] = E\left[-\frac{1}{2}(z_1-\mu_0)^T \Sigma^{-1} (z_1-\mu_0)\right]$$

$$+ E\left[-\frac{1}{2}\sum_{t=2}^{T}(z_t - W\phi(z_{t-1}) - h)^T \Sigma^{-1}(\dots)\right]$$

$$+ E\left[-\frac{1}{2}\sum_{t=1}^{T}(x_t - B z_t)^T \Gamma^{-1}(\dots)\right]$$

$$-\frac{T}{2}\log|\Sigma| - \frac{T}{2}\log|\Gamma| + \text{const.}$$

lin. case: $E[z_t]$, $E[z_t z_t^T]$, $E[z_t z_{t-1}^T]$

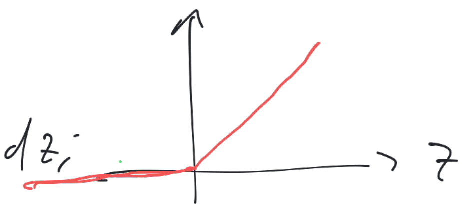
on top: $E[\phi(z_t)]$, $E[z_t \phi(z_{t-1})^T]$,

$E[\phi(z_t)\phi(z_t)^T]$

e.g. $E[\phi(z_t)] = \int_{-\infty}^{+\infty} N(\tilde{\mu}_t, \tilde{V}_t)\, \phi(z_t)\, dz_t$

→ need to do these integrals!!

Example: $\phi(z) = ReLU = \max[z, 0]$



$E[\phi(z_i)] =$

$$\int_{-\infty}^{+\infty} N(z_i; \tilde{\mu}_i, \tilde{\sigma}_i^2) \max[z_i, 0]\, dz_i$$

$$= \int_{0}^{+\infty} N(z_i; \tilde{\mu}_i, \tilde{\sigma}_i^2)\, z_i\, dz_i$$

$$= \int_{0}^{+\infty} (2\pi\tilde{\sigma}_i^2)^{-1/2} e^{-\frac{1}{2}(z_i - \tilde{\mu}_i)^2/\tilde{\sigma}_i^2}\, z_i\, dz_i$$

subst. $y := z_i - \tilde{\mu}_i$

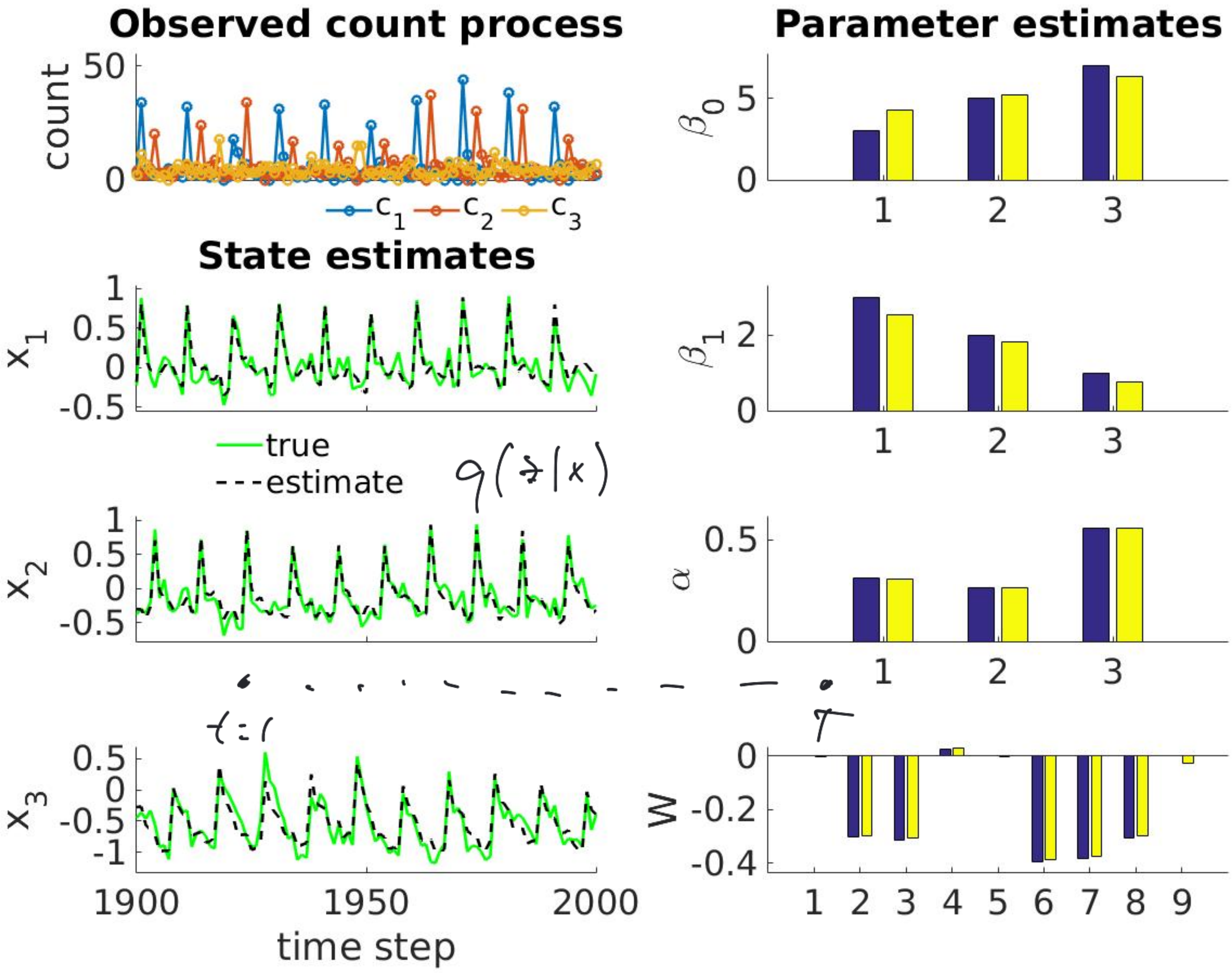
$$= \int_{-\tilde{\mu}_i}^{+\infty} (y + \tilde{\mu}_i)(2\pi\tilde{\sigma}_i^2)^{-\frac{1}{2}} e^{-\frac{1}{2}y^2/\tilde{\sigma}_i^2} \overbrace{\frac{dz_i}{dy}}^{=1} dy$$

$$= \underbrace{\int_{-\tilde{\mu}_i}^{+\infty} y\,(2\pi\tilde{\sigma}_i^2)^{-\frac{1}{2}} e^{-\frac{1}{2}y^2/\tilde{\sigma}_i^2}\, dy}_{-\tilde{\sigma}_i^2 (2\pi\tilde{\sigma}_i^2)^{-\frac{1}{2}} e^{-\frac{1}{2}y^2/\tilde{\sigma}_i^2}\Big|_{-\tilde{\mu}_i}^{+\infty}} + \tilde{\mu}_i \underbrace{\int_0^{+\infty} N(z_i; \tilde{\mu}_i, \tilde{\sigma}_i^2)\, dz_i}_{1 - CDF(0)}$$

$$= \underbrace{\tilde{\sigma}_i^2 (2\pi\tilde{\sigma}_i^2)^{-\frac{1}{2}} e^{-\frac{1}{2}\tilde{\mu}_i^2/\tilde{\sigma}_i^2}}_{N(0, \tilde{\mu}_i, \tilde{\sigma}_i^2)} + \tilde{\mu}_i \int_0^{+\infty} N(z_i; \tilde{\mu}_i, \tilde{\sigma}_i^2)\, dz_i$$

$= E[\phi(z_i)]$

EKF posterior state & parameter estimates in RNN

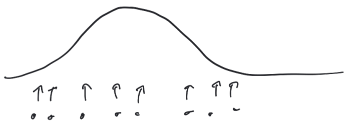


Source: Durstewitz (2017) *Advanced Data Analysis in Neuroscience.* Springer.

Sampling-based methods: particle filters

→ consistent estimates

→ pot. high comp. costs!



- can represent arbitrary distrib.

recall: $$p(z_t|x_1\dots x_t) = \frac{p(x_t|z_t) \overbrace{\int_{z_{t-1}} p(z_t|z_{t-1}) \, p(z_{t-1}|x_1\dots x_{t-1}) \, dz_{t-1}}^{p(z_t|x_1\dots x_{t-1})}}{p(x_t|x_1\dots x_{t-1})}$$

$\{\tilde{z}_t^{(k)}\}$

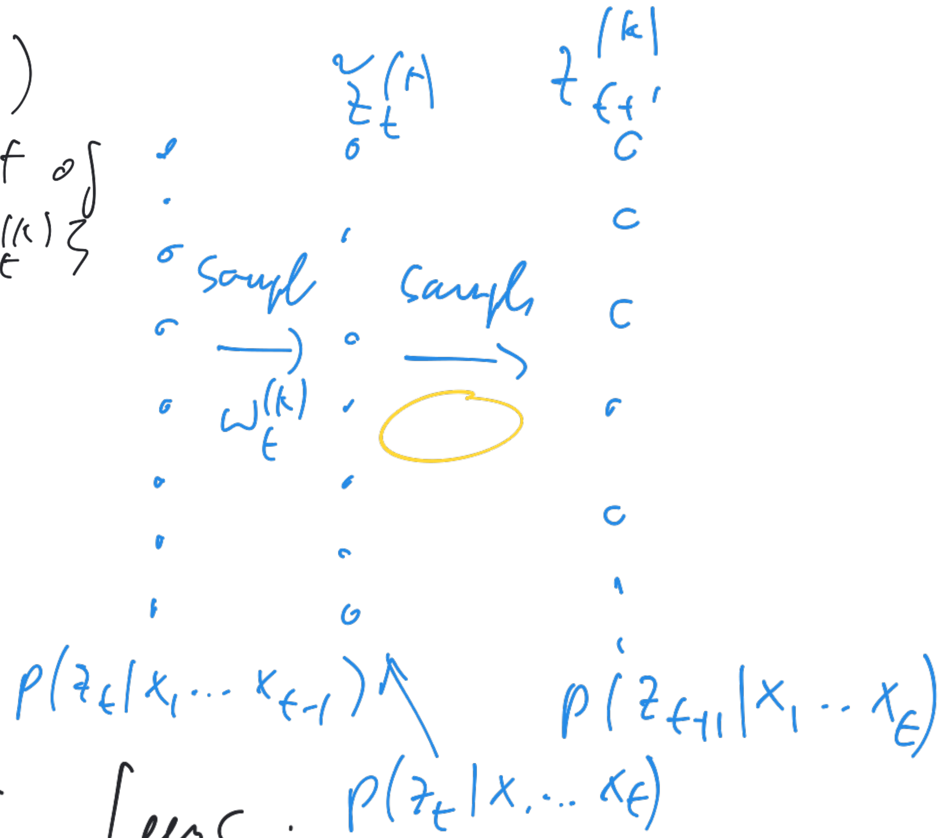
bootstrap PF:

- represent 1-step forward dens. by a set of K particles (samples) $\{z_t^{(1)}, z_t^{(2)}, \dots, z_t^{(K)}\} \sim p(z_t|x_1\dots x_{t-1})$
- to represent full post. $p(z_t|x_1\dots x_t)$ def. sampling weights $\{w_t^{(1)}, \dots, w_t^{(K)}\}$ as

$$w_t^{(k)} := \frac{p(x_t|z_t^{(k)})}{\underbrace{\sum_{l=1}^{K} p(x_t|z_t^{(l)})}_{\text{"consistency" } z_t^{(k)} \text{ with obs. } x_t}}, \quad \sum w_t^{(k)} = 1$$

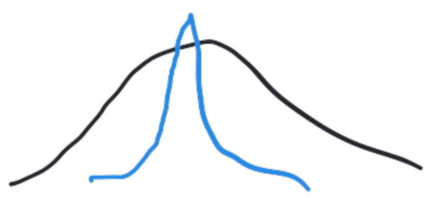
PF algo.

- at $t=1$ draw K samples $\{z_1^{(k)}\} \sim p(z_1)$
- for $t>1$ draw K samples from prev. set of particles $\{z_t^{(k)}\}$ acc. to weights $\{w_t^{(k)}\}$ with replacement → $\{\tilde{z}_t^{(k)}\}$
- combine with trans. model $\{z_{t+1}^{(k)}\} \sim p(z_{t+1}|\tilde{z}_t^{(k)})$



→ eval. expec. of nonlin. func.:

$$E[\phi(z_t)] \approx \sum_{k=1}^{K} w_t^{(k)} \phi(z_t^{(k)})$$



caveats: - comp. expensive!!!

→ scales expon. with dim.

→ filter collapse or impoverishment

consistency: for $K \to \infty$: $q(z|x) \to p(z|x)$

Laplace approximation

LDS

$X = \{x_1, \dots x_T\}, \quad x_t \in \mathbb{R}^{N\times 1}$

$Z = \{z_1, \dots\dots z_T\}, \quad z_t \in \mathbb{R}^{M\times 1}$

ini. $z_1 \sim N(\mu_0, \Sigma_0)$

process $z_t = W z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma)$

obs. $x_t = B z_t + \eta_t, \quad \eta_t \sim N(0, \Gamma)$

E-step: $p(z_t | x_1, \dots x_T) = N(\tilde{\mu}_t, \tilde{V}_t)$

$\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_T \end{pmatrix}$ $\quad p(Z|X) = N(\underset{M\cdot T\times 1}{\mu}, \underset{(M\cdot T)\cdot(M\cdot T)}{V})$

$\mu := E[Z|X]$

$V := \mathrm{cov}[Z|X]$

μ = mode

$$E[Z|X] = \arg\max_z p(Z|X) \quad = \frac{p(Z,X)}{p(X)}$$

$$= \arg\max_z \left[\log \frac{p(Z,X)}{p(X)}\right]$$

$$= \arg\max_z \left[\log p(Z,X) - \cancel{\log p(X)}\right] \quad = p(X|Z)\,p(Z)$$

$$= \arg\max_z \left[\log p(z_1) + \sum_{t=2}^{T} \log p(z_t|z_{t-1}) + \sum_{t=1}^{T} \log p(x_t|z_t)\right]$$

$$= \arg\max_z \Big[-\tfrac{1}{2}(z_1-\mu_0)^T\Sigma^{-1}(z_1-\mu_0)$$
$$-\tfrac{1}{2}\sum_{t=2}^{T}(z_t - W z_{t-1})^T\Sigma^{-1}(z_t - W z_{t-1})$$
$$-\tfrac{1}{2}\sum_{t=1}^{T}(x_t - B z_t)^T\Gamma^{-1}(x_t - B z_t) + \text{const.}\Big]$$

$\mu_0^T\Sigma^{-1}\mu_0$ $\quad z_t^T B^T \Gamma^{-1} x_t$ $\quad z_{t-1}^T W^T \Sigma^{-1} W z_{t-1}$

$L = -W^T\Sigma^{-1}$

$K = W^T\Sigma^{-1}W + B^T\Gamma^{-1}B + \Sigma^{-1}$

$$= \arg\max_z \underbrace{\left[-\tfrac{1}{2} z^T H z + \tfrac{1}{2}(d^T z + z^T d) + \text{const.}\right]}$$

$$\frac{\partial}{\partial z}[\circ] = -\tfrac{1}{2}(H + H^T) z + d = 0$$

$$\Rightarrow \boxed{\hat{z} = H^{-1} d} = E[Z|X]$$

Kalman filter $O(T)$

$O(T^2) - O(T^3)$

$$\frac{\partial^2 \log N(z; \mu, V)}{\partial z\, \partial z^T} = -V^{-1} = H$$

$$\Rightarrow V = H^{-1}$$

$$z = (\underbrace{z_{11} \dots z_{M1}}_{z_1}\ \underbrace{z_{12} \dots z_{M2}}_{z_2} \dots \underbrace{z_{1T} \dots z_{MT}}_{z_T})^T$$

Matlab

$H \backslash d = H^{-1} d$

$$-\frac{1}{2}(z_1, \dots z_T)\ \underbrace{\begin{array}{c|c} & \begin{array}{cccc} z_1 & z_2 & z_3 & \end{array} \\ \begin{array}{c} z_1 \\ z_2 \\ \\ \\ \end{array} & \begin{pmatrix} K & L & 0 & \\ L^T & K & L & 0 \\ & L^T & K & L \\ 0 & & L^T & K \\ & & & \underbrace{K - W^T\Sigma^{-1}W}_{T} \end{pmatrix} \end{array}}_{H} \begin{pmatrix} z_1 \\ \vdots \\ z_T \end{pmatrix}$$

→ block-tridiag. struc. of H

→ L. Paninski et al. (2010) J. Comp. Neurosci.

→ $O(T)$

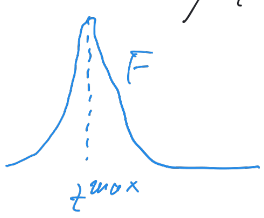
- robust algo.
- exact! for LDS

nonlinear models

$$p(z) = p(F_\theta(z))$$

$$\text{e.g.} = N(\mu_z, V_z)$$

$$\mu_z = F_\theta(z) \quad \text{e.g. RNN}$$



$$F(z) \approx F(z^{max}) + \underbrace{(z - z^{max})^T \left[\frac{\partial F}{\partial z}\right]}_{=0} + \frac{1}{2}(z - z^{max})^T \underbrace{H}_{\frac{\partial^2 F}{\partial z \partial z^T}} (z - z^{max})$$

def.: $h(z) := \log\left[p(x|z)\,p(z)\right]$

$$z = (z_1, \dots z_T)$$

$$\Rightarrow p_\theta(x) = \int_z e^{h(z)} dz \approx e^{h(z^{max})} \int_z e^{-\frac{1}{2}(z - z^{max})^T \overbrace{(-H)}^{V_{max}^{-1}} (z - z^{max})} dz$$

$$= e^{h(z^{max})} (2\pi)^{MT/2} |-H^{max}|^{-\frac{1}{2}} \underbrace{\int_z (2\pi)^{-MT/2} |-H^{max}|^{1/2} e^{-\frac{1}{2}(z - z^{max})^T(-H)(\dots)} dz}_{=1}$$

$$= p(x|z^{max})\, p(z^{max})\, (2\pi)^{MT/2} |-H^{max}|^{-\frac{1}{2}}$$

Laplace approx.

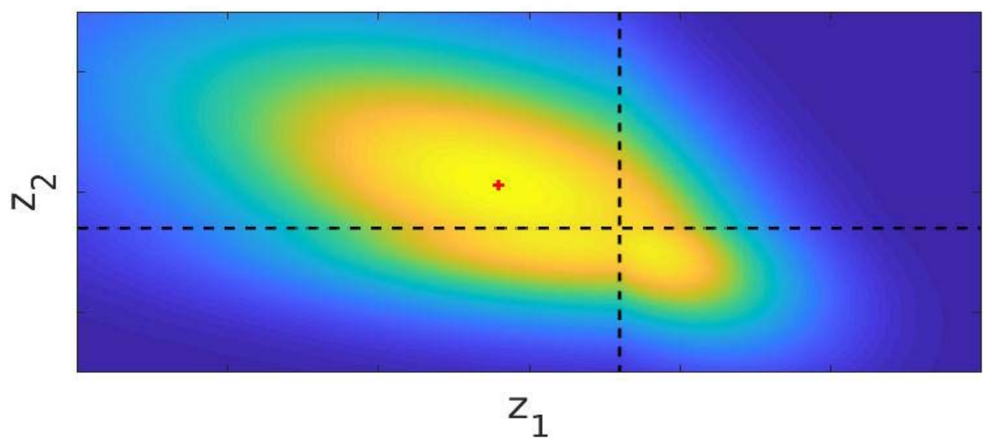
$$\Rightarrow \boxed{\log p_\theta(x) \approx \underbrace{\log p_\theta(x|z^{max})}_{\text{obs.}} + \underbrace{\log p_\theta(z)}_{\text{latent}} - \frac{1}{2}\log|-H^{max}| + \text{const.}}$$

→ numerical techn.

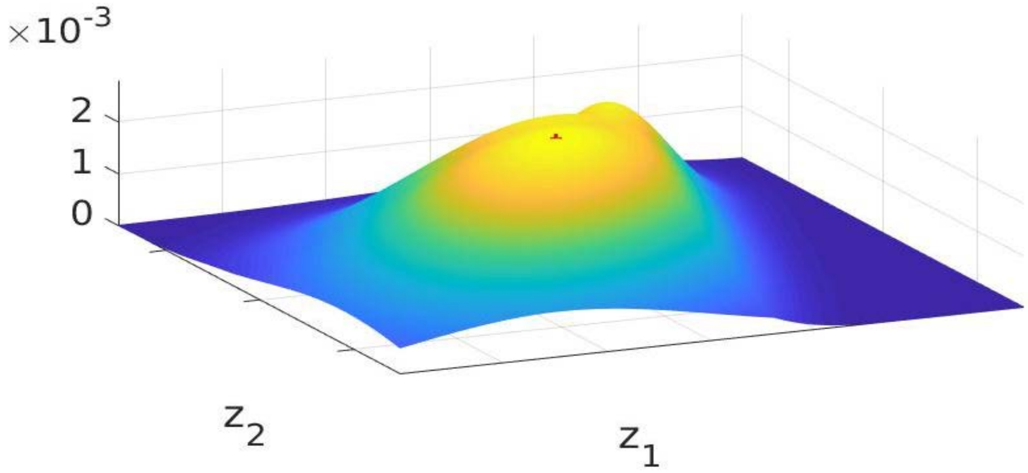
→ conv. E-step: max. problem

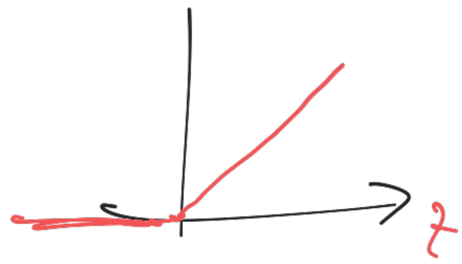
$$E[z|x] \approx z^{max}, \quad \text{cov}[z|x] \approx -H_{max}^{-1}$$

Likelihood landscape



Koppe et al. (2019)
PLoS Comp Biol





$$\phi(z) = \mathrm{ReLU} = \max[0, z]$$