Lecture Notes 3/12

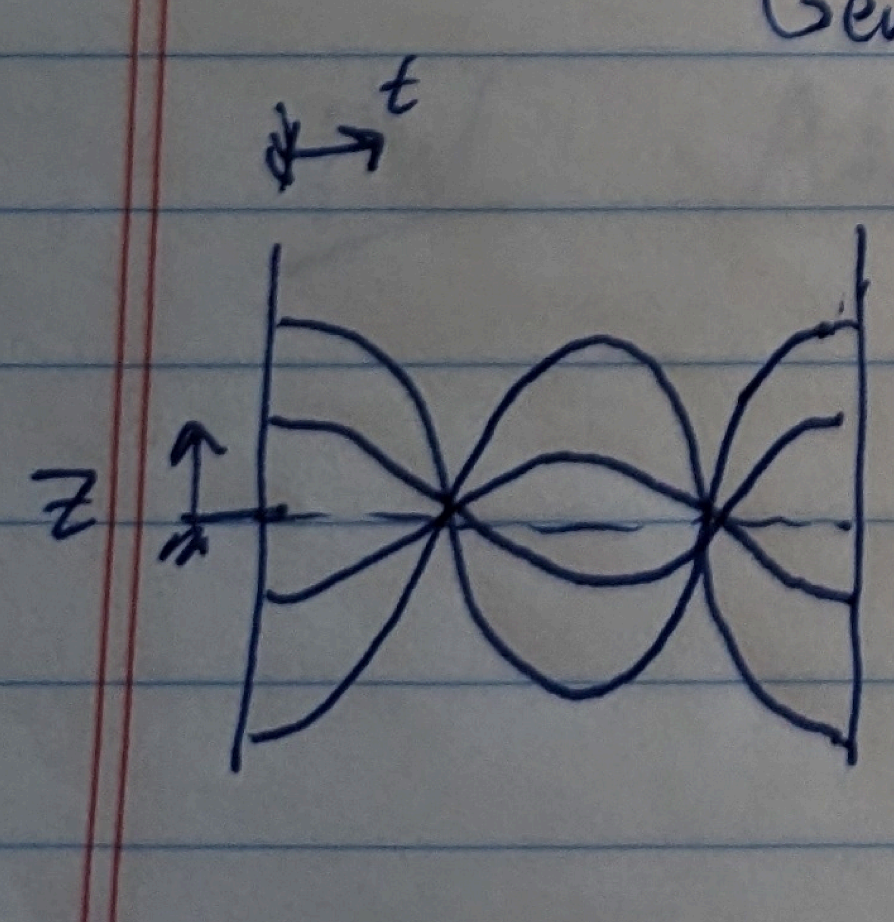
Next HW due 3/28 (same day as exam)
- out thurs
- two primarily coding problems
  - some light exercises on manipulating the ELBO
  - implementation of EM to fit a Gaussian Mixture
  - modification of the code that we're covering in class today

---

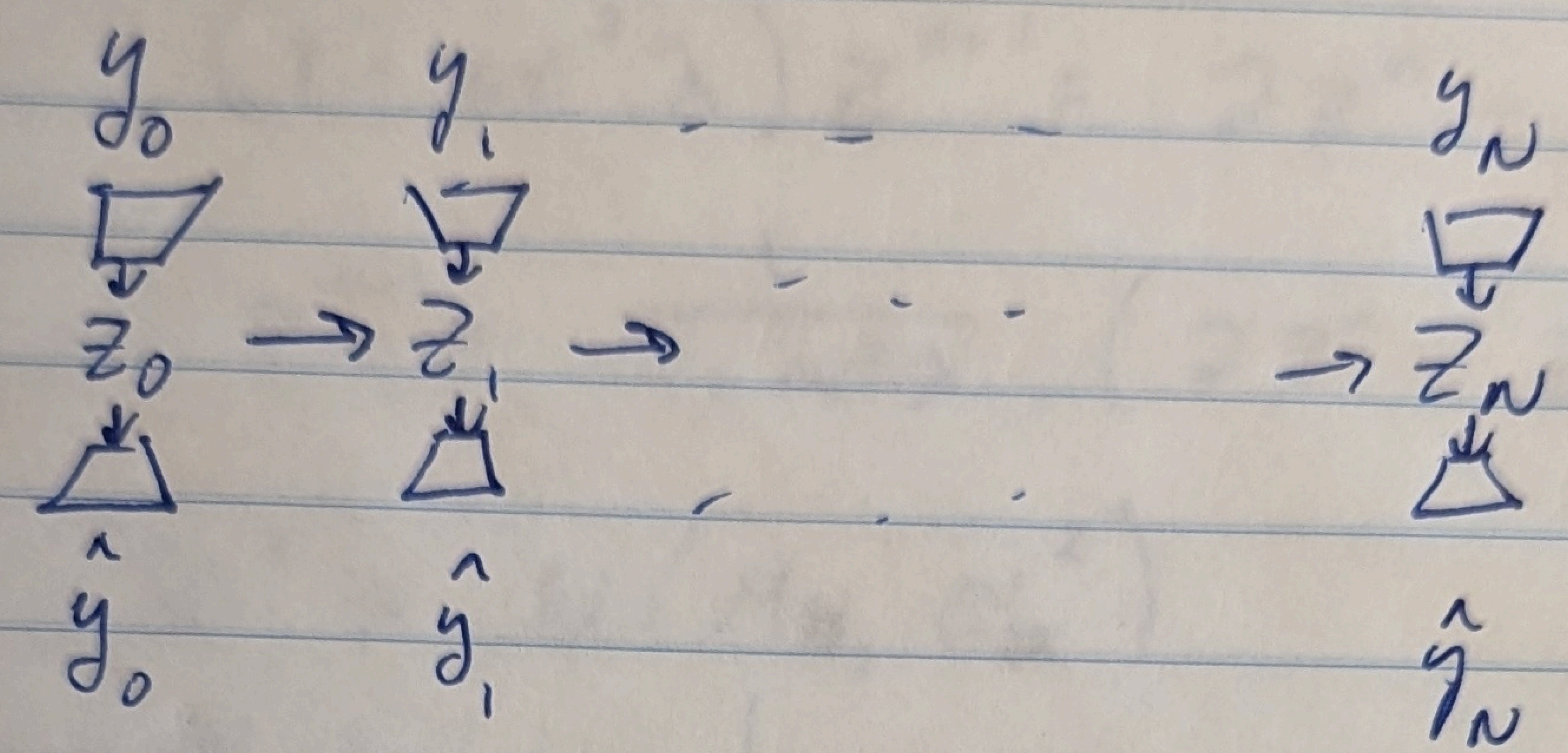
Recall from before break:

Hidden ODE
$$\ddot{z} = -\lambda z$$
With noisy observations
$$y = \begin{pmatrix} \sin z(t) \\ -\cos z(t) \end{pmatrix} + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I)$$

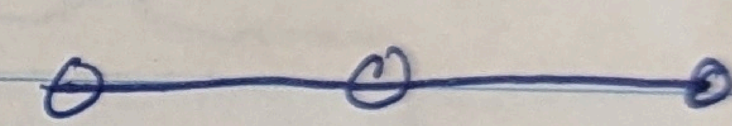
Generate data by solving analytically
$$z_i(t) = z_i(t_0) \cos\left(\tfrac{1}{\theta} t\right)$$
$$z_i(t_0) \sim U[-1, 1]$$



To build prior distribution on form of dynamics recall time integrators



$$\ddot{z} = -\lambda z$$



centered difference

$$\frac{z^{n+1} - 2z^{n} + z^{n-1}}{\Delta t^2} = -\lambda z^{?}$$

z^{n+1}
Implicit Euler
Unconditionally stable, dissipative

z^{n}
Explicit Euler
conditionally stable
$\frac{\lambda}{\Delta t^2} < C$

HW Adopt code for explicit euler to implicit Euler case w/ implicit Euler

$$z^{n+1} - 2z^{n} + z^{n-1} = -\Delta t^{2} \lambda z^{n+1}$$

rearrange

$$(1 + \Delta t^{2} \lambda) z^{n+1} = 2z^{n} - z^{n-1}$$

$$z^{n+1} = \frac{1}{1 + \Delta t^{2} \lambda} \left(2z^{n} - z^{n-1} \right)$$

Taking $z^{n} \sim N(\mu_{n}, \sigma_{n}^{2})$

$$\alpha = \frac{1}{1 + \Delta t^{2} \lambda}$$

$$z^{n+1} \mid \lambda \sim N\left(\overbrace{\alpha(2\mu_{n} - \mu_{n-1})}^{\mu_n}, \underbrace{\alpha^{2}(4\sigma_{n}^{2} + \sigma_{n-1}^{2})}_{S_n} \right)$$

<u>HW</u> repeat w/ explicit euler

ELBO

Maximize $\mathcal{E} = \underbrace{\overset{(1)}{\mathbb{E}_{q(z|y)}\left[\log p(y|z)\right]}}_{} - \overset{(2)}{KL\left(q(z|y)\,||\,p(z)\right)}$

$$p(\vec{y}|\vec{z}) = N(\hat{y}, I)$$

$$-\sum_d \frac{1}{2}(y_d - \hat{y}_d)^2$$

$$(2) = \sum_{n=2}^{N} \mathbb{E}_q\left[\log q(z_n|y_n)\right] - \mathbb{E}_q\left[\log p(\cancel{z}_n)\right]$$

HW2

Recall $\mathbb{E}_f[\log g]$, $f = N(\mu_1, \sigma_1^2)$
$g = N(\mu_2, \sigma_2^2)$

$$= -\frac{1}{2}\sum_j \left[\log 2\pi\sigma_{2,j}^2 + \frac{\sigma_{1,j}^2}{\sigma_{2,j}^2} + \frac{(\mu_{1j} - \mu_{2j})^2}{\sigma_{2j}^2}\right]$$

Evaluate w/ $q(z_n|y_n) \approx N(\mu_n, \sigma_n^2)$

encoder prob.

$p(z_n) = N(m_n, \overset{S_n}{\cancel{\Lambda_n}})$

ODE prior

Feedback for exam:

Any topics that you'd like reviewed, submit to survey to be linked w/ HW4.

Revisiting multivariate Gaussians

Now that we have some experience working w/ Gaussians in regression & physics discovery contexts, we'll dig into some deeper identities we need for Gaussian process regression and active learning

def $\mathcal{N}(y|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)\right)$

Marginals + Conditionals of MVN (Murphy §3.2.3)

Let $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

precision matrix $\Lambda = \Sigma^{-1} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}$

Then marginals $p(y_1) = \mathcal{N}(\mu_1, \Sigma_{11})$

$p(y_2) = \mathcal{N}(\mu_2, \Sigma_{22})$

pf Let $S = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$ $Sy \sim \mathcal{N}(S\mu, S^T \Sigma S)$

Schur Complements

Upper Block Triangular

$$M = \begin{pmatrix} I & A \\ 0 & I \end{pmatrix}$$

$$M^{-1} = \begin{pmatrix} I & -A \\ 0 & I \end{pmatrix}$$

For block structured diagonal matrix

$$M = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \quad M^{-1} = \begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix}$$

But for $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$

$$M^{-1} = \begin{pmatrix} (M \backslash D)^{-1} & -(M \backslash D)^{-1} B D^{-1} \\ -D^{-1} C (M \backslash D)^{-1} & D^{-1} + D^{-1} C (M \backslash D)^{-1} B D^{-1} \end{pmatrix}$$

where $M \backslash D := A - B D^{-1} C$

→ so only need inverses of D, $M \backslash D$

Similarly can write

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1} B (M \backslash A)^{-1} C A^{-1} & -A^{-1} B (M \backslash A)^{-1} \\ -(M \backslash A)^{-1} C A^{-1} & (M \backslash A)^{-1} \end{pmatrix}$$

for $(M \backslash A) = D - C A^{-1} B$

$M \backslash A$ → Schur complement of M wrt A

$M \backslash D$ " wrt D

Think of Gauss elimination

Proof Objective is to rewrite M into block diagonal form

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

$$\begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A - BD^{-1}C & 0 \\ C & D \end{pmatrix}$$

Now to eliminate bottom ~~...~~ left block

$$\underbrace{\begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} A & B \\ C & D \end{pmatrix}}_{M} \underbrace{\begin{pmatrix} I & 0 \\ -D^{-1}C & I \end{pmatrix}}_{Z} = \underbrace{\begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix}}_{W}$$

$$M = X^{-1} W Z^{-1}$$

or

$$M^{-1} = Z W^{-1} X$$

Remark: useful for either

- solving block diagonal systems
- partitioning matrix in U/L-tri form

Conditional of MVN

$$p(x_1, x_2) = N(\mu, \Sigma), \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$\cancel{p(x_1) = N(\mu_1, \Sigma_{11})} \qquad p(x_2) = N(\mu_2, \Sigma_{22})$$

seek to define via factorization

$$p(x_1, x_2) = \cancel{p(x_2 | x_1)\, p(x_1)} \; p(x_1 | x_2)\, p(x_2)$$

$$p(x_1, x_2) \propto \exp\left(-\frac{1}{2} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right)$$

Applying Schur Complement factorization

$$\propto \exp\left[-\frac{1}{2} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix} \begin{pmatrix} (\Sigma/\Sigma_{22})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right]$$

$$= \exp\left[-\frac{1}{2} \left(x_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right)^T (\Sigma/\Sigma_{22})^{-1} \left(x_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right) \right]$$

$$\times \exp\left[-\frac{1}{2} (x_2 - \mu_2)^T \Sigma_{22}^{-1} (x_2 - \mu_2) \right]$$

$$= N(x_1 | \mu_{1|2}, \Sigma_{1|2})\, N(x_2 | \mu_2, \Sigma_2)$$

where
$$\begin{cases} \mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{cases}$$

To complete proof, check normalization constants

Inferring hidden data from correlation

If we know joint distribution, we can use the conditional to predict

$$\mathbb{E}[y_2 \mid y_1] = \mu_{2|1} = \cancel{\mu_2 + \Sigma_{21}\Sigma_{22}^{-1}}$$

y_2: hidden, y_1: observed

$$\mu_2 + \underbrace{\Sigma_{21}\Sigma_{11}^{-1}(y_1 - \mu_1)}_{\text{correction term}}$$

Next lecture

engineering a model of covariance to give uncertainty estimates under regression. For now → simple linear model

Bayesian Regression w/ Gaussians

Consider noisy measurements y of latent var z under the model

$$p(z) = \mathcal{N}(z \mid \mu_z, \Sigma_z)$$

$$p(y \mid z) = \mathcal{N}(y \mid Wz + b, \Sigma_y)$$

- Then $p(y, z) = \mathcal{N}(\mu, \Sigma)$, $\mu = \begin{pmatrix} \mu_z \\ W\mu_z + b \end{pmatrix}$

$$\Sigma = \begin{pmatrix} \Sigma_z & \Sigma_z W^T \\ W\Sigma_z & \Sigma_y + W^T \Sigma_z W \end{pmatrix}$$

- Condition to get

Note
Conjugate Prior

$$p(z \mid y) = \mathcal{N}(z \mid \mu_{z|y}, \Sigma_{z|y}), \quad \Sigma_{z|y}^{-1} = \Sigma_z^{-1} + W \Sigma_y^{-1} W$$

$$\mu_{z|y} = \Sigma_{z|y}\left[W^T \Sigma_y^{-1}(y - b) + \Sigma_z^{-1}\mu_z\right]$$