Generative AI and Foundation Models, M3309.001800 001
E. Ryu
Spring 2024

Homework 3
Due 5pm, Friday, June 21, 2024

Problem 1: *Derivation of zero-order hold.* Let $A \in \mathbb{C}^{n\times n}$ and $B \in \mathbb{C}^{n\times k}$. Consider the ODE

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x(0) = x_0 \in \mathbb{C}^n.$$

Let $t \geq 0$ and $\Delta t \geq 0$.

(a) Show

$$x(t+\Delta t) = e^{\Delta t A}x(t) + \int_0^{\Delta t} e^{(\Delta t - s)A}Bu(t+s)\, ds.$$

(b) Show that if $u(t+s) = u(t)$ for $s \in [0, \Delta t]$, then

$$x(t+\Delta t) = e^{\Delta t A}x(t) + \Big(\int_0^{\Delta t} e^{sA}\, ds\Big)Bu(t).$$

(c) Show that if A is invertible, then

$$e^{\Delta t A}x(t) + \Big(\int_0^{\Delta t} e^{sA}\, ds\Big)Bu(t) = e^{\Delta t A}x(t) + \left(e^{\Delta t A} - I\right)A^{-1}Bu(t).$$

Hint. For (c), integrate the power series definition of e^{sA}.

Problem 2: *Laguerre polynomials with Gram–Schmidt.* Consider the Laguerre inner product and norm

$$\langle f, g\rangle = \int_0^\infty f(x)g(x)e^{-x}\, dx, \qquad \|f\|^2 = \langle f, f\rangle.$$

Let

$$m_0(x) = 1, \qquad m_1(x) = x, \qquad m_2(x) = x^2, \qquad m_3(x) = x^3$$

be the first three monomials. Let

$$\begin{aligned}
\tilde{L}_0 &= m_0 \\
\tilde{L}_1 &= m_1 - \frac{\langle m_1, \tilde{L}_0\rangle}{\|\tilde{L}_0\|^2}\tilde{L}_0 \\
\tilde{L}_2 &= m_2 - \frac{\langle m_2, \tilde{L}_0\rangle}{\|\tilde{L}_0\|^2}\tilde{L}_0 - \frac{\langle m_2, \tilde{L}_1\rangle}{\|\tilde{L}_1\|^2}\tilde{L}_1 \\
\tilde{L}_3 &= m_3 - \frac{\langle m_3, \tilde{L}_0\rangle}{\|\tilde{L}_0\|^2}\tilde{L}_0 - \frac{\langle m_3, \tilde{L}_1\rangle}{\|\tilde{L}_1\|^2}\tilde{L}_1 - \frac{\langle m_3, \tilde{L}_2\rangle}{\|\tilde{L}_2\|^2}\tilde{L}_2.
\end{aligned}$$

Compute

$$L_0 = \frac{\tilde{L}_0}{\|\tilde{L}_0\|}, \qquad L_1 = \frac{\tilde{L}_1}{\|\tilde{L}_1\|}, \qquad L_2 = \frac{\tilde{L}_2}{\|\tilde{L}_2\|}, \qquad L_3 = \frac{\tilde{L}_3}{\|\tilde{L}_3\|}.$$

Problem 3: *DALLE 2 without prior.* In the training of DALLE 2, consider training the decoder h_ψ to take in as input CLIP text embeddings, rather than CLIP image embeddings.

(a) How should the training of h_ψ need to be modified?

(b) How should the text-to-image generation be modified and why would the prior p_ω no longer be necessary?

(c) The original DALLE 2 model has the capability to produce image variations, image interpolations, and text diffs. Under the proposed modifications, which of these capabilities would be lost and which would be retained? Justify your answers.

Hint. For (c), consider using a "bipartite representation" (X_T, Z^{text}).