UCSB, Physics 129L, Computational Physics: Practice Final Exam Solution

Zihang Wang (UCSB), zihangwang@ucsb.edu

March 16, 2025

Final Exam Instructions

Location: Web 1100
Time: Tuesday, March 18, 2025 4:00 PM - 7:00 PM

Before you start, **write your name and perm number clearly on the blue book cover**. This is a closed-book, pen-and-paper exam. Please label which question you are answering and write your answer clearly in the blue book.

There are 20 questions and 1 extra credit question. You have a maximum of **three hours to complete the exam**. You are allowed to use **a single (front and back) handwritten note sheet**, and you must complete the final exam without consulting any other information sources. Academic dishonesty will not be tolerated and carries significant consequences, including disciplinary action and loss of credibility at the university level.

If you finish the exam early, you can turn in your blue book and leave early. You may keep the final exam page.

Good luck.

1

What is the major difference between a scripting language and compile language? Write an example for each type. (5 points)

Scripting language executes instructions line-by-line (Python) while compile language compiles instructions into an executable and executes at once (C++).

2

What is a Fetch operation in Git? Explain the concept and draw a schematic that demonstrates this operation between a remote branch and a local branch. (10 points)

The Fetch operations are able to acquire all remote commits that are absent in the local main branch. See the image from the lecture note.

3

Consider a harmonic potential surface derivative,

$$x^2 + y^2 + \frac{z^2}{3} = V, \quad z \geq 0, \tag{1}$$

where V is a function of x, y, z.

Using the projector method ($\mathbf{N}$ is the surface normal),

$$\mathbf{P} = \mathbf{I} - \mathbf{N}\mathbf{N}^T, \quad \nabla_s \phi = P_{ij}\partial_j \phi, \tag{2}$$

find the surface derivative ∇_s on the following two functions, $\phi = V(x, y, z)$ and $\phi = x^4$. (10 points

---

We first find the gradient,

$$\nabla V = \left(2x,\, 2y,\, \frac{2z}{3}\right).$$

Then, the unit normal is given by the ratio,

$$\mathbf{N} = \frac{\nabla V}{\|\nabla V\|} = \frac{(x, y, z/3)}{\sqrt{V}}.$$

The projection operator is given by,

$$\mathbf{P} = \mathbf{I} - \mathbf{N}\mathbf{N}^T,$$

so that for any function $\phi(x, y, z)$ the surface (tangential) gradient is given by

$$\nabla_s \phi = \nabla \phi - (\nabla \phi \cdot \mathbf{N})\, \mathbf{N}.$$

Notice that,

$$\mathbf{N} = \frac{\nabla V}{\|\nabla V\|},$$

which implies,

$$\nabla V \cdot \mathbf{N} = \|\nabla V\|.$$

Thus, the surface gradient becomes,

$$\nabla_s V = \nabla V - \|\nabla V\| \mathbf{N} = \mathbf{0}.$$

For $\phi = V(x, y, z)$ we have $\nabla_s V(x, y, z) = 0$.

The full gradient is

$$\nabla(x^4) = (4x^3,\, 0,\, 0).$$

Its dot product with the unit normal is

$$\nabla(x^4) \cdot \mathbf{N} = (4x^3, 0, 0) \cdot \frac{(x,\, y,\, z/3)}{\sqrt{V}} = \frac{4x^4}{\sqrt{V}}.$$

Therefore, the surface gradient is,

$$\nabla_s(x^4) = (4x^3,\, 0,\, 0) - \frac{4x^4}{\sqrt{V}} \cdot \frac{(x,\, y,\, z/3)}{\sqrt{V}}.$$

Finally, we have,

$$\nabla_s(x^4) = \left(4x^3 - \frac{4x^5}{V},\, \frac{-4x^4 y}{V},\, \frac{-4x^4 z}{3V}\right).$$

4

What is the difference between surface (facet) norm and vertex norm? (5 points) **Surface norms are defined by the area while vertex norms are weighted sum of surface norms, with respect to a given shared vertex.**

5

a) What is the meaning of "parallel" in parallel transport? b) What are geodesics curves, and what happen if you parallel transport their tangent vectors? (10 points) **Parallel in parallel transport means that the vector direction and magnitude remains inner product remain consistent with the local metric: covariant derivative along the curve is zero. Geodesics curves parallel transport their own tangent vectors.**

6

What is the difference between a deterministic and non-deterministic Turing machine? (5 points)

Turing machine, which follows a single, deterministic path of steps based on its current state and tape symbol, a nondeterministic Turing machine explores all computational paths simultaneously.

7

Given the following recurrence expression, use the Master Theorem to calculate the time complexity to the leading order,

$$T(n) = 2T\left(\frac{n}{2}\right) + O(n).$$

(10 points)

$$a = 2, \quad b = 2, \quad c = 1, \quad d_{\max} = \log_b n, \tag{3}$$

Using case 2, we have,

$$T(n) \sim n \sum_{d=0}^{d_{\max}-1} \left(\frac{2}{2}\right)^d = \mathcal{O}(n d_{\max}) = \mathcal{O}(n \log n). \tag{4}$$

8

Consider the following matrix A:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

Use the Gram-Schmidt process to find the QR decomposition of matrix A, where Q is an orthogonal matrix and R is an upper triangular matrix. Construct the upper triangular matrix R using the formula,

$$R = Q^T A.$$

(15 points)

Let the columns of A be denoted by

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{a}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

Let's look at the first vector,

$$\|\boldsymbol{a}_1\| = \sqrt{1^2 + 1^2 + 0^2} = \sqrt{2},$$

and the first orthonormal vector is

$$\boldsymbol{q}_1 = \frac{\boldsymbol{a}_1}{\|\boldsymbol{a}_1\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Using the above result, we can calculate the second orthogonal vector,

$$\boldsymbol{u}_2 = \begin{bmatrix}1\\2\\1\end{bmatrix} - \frac{3}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}-1\\1\\2\end{bmatrix}.$$

with norm $\|\boldsymbol{u}_2\| = \sqrt{3/2}$. Therefore,

$$\boldsymbol{q}_2 = \frac{\boldsymbol{u}_2}{\|\boldsymbol{u}_2\|} = \sqrt{\frac{1}{6}}\begin{bmatrix}-1\\1\\2\end{bmatrix}.$$

We, compute

$$\boldsymbol{u}_3 = \boldsymbol{a}_3 - \mathrm{proj}_{\boldsymbol{q}_1}(\boldsymbol{a}_3) - \mathrm{proj}_{\boldsymbol{q}_2}(\boldsymbol{a}_3).$$

First,

$$\boldsymbol{q}_1^T\boldsymbol{a}_3 = \frac{1}{\sqrt{2}}(0+1+0) = \frac{1}{\sqrt{2}},$$

so

$$\mathrm{proj}_{\boldsymbol{q}_1}(\boldsymbol{a}_3) = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\1\\0\end{bmatrix}.$$

Next, compute

$$\boldsymbol{q}_2^T\boldsymbol{a}_3 = \frac{5}{\sqrt{6}}.$$

Thus,

$$\mathrm{proj}_{\boldsymbol{q}_2}(\boldsymbol{a}_3) = \frac{5}{6}\begin{bmatrix}-1\\1\\2\end{bmatrix}.$$

$$\boldsymbol{u}_3 = \begin{bmatrix}0\\1\\2\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} - \frac{5}{6}\begin{bmatrix}-1\\1\\2\end{bmatrix} = \frac{1}{3}\begin{bmatrix}1\\-1\\1\end{bmatrix}.$$

with norm $\|\boldsymbol{u}_3\| = 1/\sqrt{3}$. Therefore,

$$\boldsymbol{q}_3 = \sqrt{3}\frac{1}{3}\begin{bmatrix}1\\-1\\1\end{bmatrix} = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix}.$$

The matrix Q whose columns are $\boldsymbol{q}_1$, $\boldsymbol{q}_2$, and $\boldsymbol{q}_3$ is given by

$$Q = \begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}}\\ 0 & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{3}}\end{bmatrix}.$$

Since

$$R = Q^TA,$$

its entries are given by $r_{ij} = \boldsymbol{q}_i^T\boldsymbol{a}_j$ for $i \le j$ (and $r_{ij} = 0$ for $i > j$).

$$r_{11} = \|\boldsymbol{a}_1\| = \frac{1}{\sqrt{2}}(1+1+0) = \sqrt{2},$$

$$r_{12} = \boldsymbol{q}_1^T\boldsymbol{a}_2 = \frac{1}{\sqrt{2}}(1+2+0) = \frac{3}{\sqrt{2}},$$

$$r_{13} = \boldsymbol{q}_1^T\boldsymbol{a}_3 = \frac{1}{\sqrt{2}}(0+1+0) = \frac{1}{\sqrt{2}},$$

$$r_{22} = \|\boldsymbol{u}_2\| = \sqrt{\frac{3}{2}},$$

$$r_{23} = \boldsymbol{q}_2^T\boldsymbol{a}_3 = \frac{5\sqrt{6}}{6},$$

$$r_{33} = \|\boldsymbol{u}_3\| = \frac{1}{\sqrt{3}}.$$

Thus, the upper triangular matrix R is

$$R = \begin{bmatrix}\sqrt{2} & \frac{3}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ 0 & \sqrt{\frac{3}{2}} & \frac{5\sqrt{6}}{6}\\ 0 & 0 & \frac{1}{\sqrt{3}}\end{bmatrix}.$$

9

Let's consider a Markov chain on the following state space $\{S_1, S_2, S_3\}$ with the following transition matrix,

$$P = \begin{bmatrix}\frac{1}{2} & \frac{1}{2} & 0\\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4}\\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4}\end{bmatrix}.$$

a) If we start from the state S_1, what is the probability remaining at state S_1 after three Markov steps? b) What is the stationary distribution of this Markov chain? (20 points)

The probability of being in state S_1 after three steps, is the $(1,1)$ entry of P^3. First, we compute $P^2 = P \cdot P$,

$$P^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

Since we know that only the first row of P^2 contributes to the transition for P^3, We have,

$$P^2_{11} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} = \frac{3}{8}$$
$$P^2_{12} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot \frac{1}{4} = \frac{1}{2}$$
$$P^2_{13} = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} = \frac{1}{8}$$

.

Then, we extract $(P^3)_{11}$,

$$(P^3)_{11} = \frac{1}{2} \cdot \frac{3}{8} + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{8} = \frac{3}{8},$$

which is the probability of being in S_1 after three steps given we started in S_1 .

The stationary distribution satisfies the following,

$$\boldsymbol{\pi}(P - I) = 0, \quad \boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3),$$

where

$$P - I = \frac{1}{4} \begin{bmatrix} -2 & 2 & 0 \\ 1 & -2 & 1 \\ 2 & 1 & -3 \end{bmatrix}.$$

The third column gives,

$$\pi_2 = 3\pi_3,$$

and plug in to the first equation,

$$-2\pi_1 + \pi_2 + 2\pi_3 = 0, \tag{5}$$

such that,

$$\pi_1 = \frac{5}{2}\pi_3. \tag{6}$$

We have,

$$\boldsymbol{\pi} = (5, 6, 2)\pi_3, \tag{7}$$

and we absorb the fraction into π_3. Since we know the normalization condition,

$$\|\boldsymbol{\pi}\| = 1, \tag{8}$$

this give,

$$25 + 36 + 4 = \pi_3^2 = 65. \tag{9}$$

Therefore, we finally have the stationary distribution,

$$\boldsymbol{\pi} = \frac{1}{\sqrt{65}}(5, 6, 2). \tag{10}$$

10

Let W_t be a standard Brownian motion and consider the following quadratic function,

$$g(t, W_t) = \mu_1 t + \mu_2 t^2 + \sigma^2 t (W_t)^2. \tag{11}$$

Calculate the differential dg using the Ito's Lemma (10 points)

Ito's Lemma gives:

$$dg = \left(\frac{\partial g}{\partial t}\right) dt + \left(\frac{\partial g}{\partial W_t}\right) dW_t + \frac{1}{2}\left(\frac{\partial^2 g}{\partial W_t^2}\right)(dW_t)^2 \tag{12}$$

Substituting the partial derivatives and using $(dW_t)^2 = dt$:

$$dg = \left[\mu_1 + (2\mu_2 + \sigma^2)t + \sigma^2 W_t^2\right] dt + 2\sigma^2 t W_t \, dW_t. \tag{13}$$

The mean is

11

In the Langevin equation, the stochastic differential of the velocity dv is given by **Note that a typo in** $e^{-\frac{\gamma}{m}t}$,

$$dv = -\frac{\gamma}{m} v\, dt + e^{-\frac{\gamma}{m}t} \frac{\sqrt{D}}{m}\, W_t\, dW_t, \tag{14}$$

where γ is the friction coefficient, and D is the diffusion constant. Calculate the following ensemble average, a) $\langle v_t \rangle$. b) $\langle v_t^2 \rangle - \langle v_t \rangle^2$. (20 points)

---

a) The mean is zero because the stochastic integration will have zero mean $\langle dW_t \rangle = 0$. Therefore,

$$\langle dv \rangle = -\frac{\gamma}{m} \langle v \rangle\, dt$$

This gives,

$$\langle v \rangle = v_0 \exp\left(-\frac{\gamma}{m} t\right).$$

b) Using Ito's lemma, we have,

$$d(v^2) = 2v dv + (dv)^2$$

The lowest order contribution in $(dv)^2$ would be,

$$(dv)^2 = e^{-2\frac{\gamma}{m}t} \frac{D}{m^2}\, W_t^2\, dt$$

. since we know, $(dW_t)^2 = dt$. The term $2vdv$,

$$2vdv = -2\frac{\gamma}{m} v^2\, dt + 2e^{-\frac{\gamma}{m}t} \frac{\sqrt{D}}{m} v\, W_t\, dW_t,$$

We then take the ensemble average,

$$\frac{d\langle v^2 \rangle}{dt} = \left(-2\frac{\gamma}{m} \langle v^2 \rangle + e^{-2\frac{\gamma}{m}t} \frac{D}{m^2}\, t\right),$$

where we use that $\langle W_t^2 \rangle = t$. Then, we use the integrating factor method,

$$\frac{d}{dt}\left(\langle v^2 \rangle\, e^{2\frac{\gamma}{m}t}\right) = \frac{D}{m^2}\, t.$$

Then, we integrate both sides, and the variance become,

$$\langle v^2 \rangle = v_0^2 e^{-2\frac{\gamma}{m}t} + \frac{D}{2m^2} t^2 e^{-2\frac{\gamma}{m}t}.$$

Finally, we have note that the first term is the velocity square,

$$\langle v^2 \rangle - \langle v \rangle^2 = \frac{D}{2m^2} t^2 e^{-2\frac{\gamma}{m}t}.$$

12

Let's consider a system with a fixed energy, and the multiplicity is given by $\Omega(E) = \frac{(E-E_0)^2}{2}$. Why can you still define the system temperature? Write the temperature as a function of energy $T(E)$. Discuss the implications for the following case for $E < E_0$, $E > E_0$, $E = E_0$. (10 points)

We define temperature using the thermodynamic relation:

$$\frac{1}{T} = \frac{\partial S}{\partial E}, \quad \text{where} \quad S = k_B \ln \Omega(E).$$

The temperature as a function of energy is:

$$T(E) = \frac{E - E_0}{2k_B}.$$

Let's look at three distinct cases based on the value of E relative to E_0,

- For $E > E_0$:

  $$T(E) = \frac{E - E_0}{2k_B} > 0.$$

  The temperature is positive. This corresponds to typical thermodynamic behavior, where increasing energy leads to increasing temperature.

- For $E < E_0$:

$$T(E) = \frac{E - E_0}{2k_B} < 0.$$

  The temperature is negative. Negative temperatures occur in certain quantum systems characterized by population inversion, where higher-energy states have greater occupation probabilities than lower-energy states.

- For $E = E_0$: The multiplicity becomes zero, and there are no accessible microstates. Therefore, negative and positive temperature cannot be smoothly connected as it crosses $E = E_0$.

13

What is the Central Limit Theorem? Give an intuitive explanation on why Central Limit Theorem is valid? (5 points)

For i.i.d. random variables, in the large N limit, Central Limit Theorem states that the sample mean is normally distributed around the true expectation value. One way to understand this is to consider the asymptotic behavior of cumulant expansion: as $N \to \infty$, the first N-dependent non-trivial term in the expansion is second order, and the remaining higher order cumulants can be consider vanishing. Gaussian is the only distribution that only has non-zero cumulants upto the second order.

14

What are the differences between Bayesian inference and Frequentist interpretation? (10 points)

Bayesian assumes a prior distribution of hyperpameters while Frequentist does not.

15

The Gamma distribution is a continuous probability distribution with two parameters: the shape parameter k (also denoted as α) and the rate parameter λ (also denoted as $\beta = \frac{1}{\lambda}$, the scale parameter). The probability density function (PDF) of the Gamma distribution is:

$$\gamma(x, k, \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)} \quad \text{for } x > 0,$$

where $\Gamma(k)$ is the Gamma function, and $k > 0$ and $\lambda > 0$. Let's consider a given a set of n independent observations $\{x_1, x_2, \ldots, x_n\}$ drawn from a Gamma distribution, and you want to estimate what are the maximum likelihood of the hyperparameter k, λ. a) Write down the log-likelihood function. b) find the maximum likelihood estimator for $\hat{k}, \hat{\lambda}$. You do not need to solve the differential equation associated with the maximum likelihood estimator. (20 points)

---

a) The likelihood function is

$$L(k, \lambda) = \frac{\lambda^{nk}}{[\Gamma(k)]^n} \prod_{i=1}^{n} x_i^{k-1} e^{-\lambda x_i}.$$

Taking the logarithm, we obtain the log-likelihood function:

$$\ln L(k, \lambda) = \sum_{i=1}^{n} [k \ln \lambda + (k-1) \ln x_i - \lambda x_i - \ln \Gamma(k)]$$
$$= nk \ln \lambda - n \ln \Gamma(k) + (k-1) \sum_{i=1}^{n} \ln x_i - \lambda \sum_{i=1}^{n} x_i.$$

(b) Maximum Likelihood Estimators (MLEs)

Differentiate $\ln L(k,\lambda)$ with respect to λ:

$$\frac{\partial \ln L(k,\lambda)}{\partial \lambda} = \frac{nk}{\lambda} - \sum_{i=1}^{n} x_i.$$

Setting this equal to zero,

$$\frac{nk}{\lambda} - \sum_{i=1}^{n} x_i = 0 \quad \rightarrow \quad \hat{\lambda} = \frac{n\hat{k}}{\sum_{i=1}^{n} x_i}.$$

and replacing k by its estimator $\hat{k}$, calculated below. Differentiate $\ln L(k,\lambda)$ with respect to k:

$$\frac{\partial \ln L(k,\lambda)}{\partial k} = n\ln\lambda - n\frac{\Gamma'(k)}{\Gamma(k)} + \sum_{i=1}^{n} \ln x_i = 0,$$

Substitute $\hat{\lambda}$ to get an equation for $\hat{k}$:

$$\ln\left(\frac{n\hat{k}}{\sum_{i=1}^{n} x_i}\right) - \frac{\Gamma'(k)}{\Gamma(k)} + \frac{1}{n}\sum_{i=1}^{n} \ln x_i = 0.$$

This implicit equation determines $\hat{k}$.

16

The Gaussian quadrature has the following form,

$$\int_{-1}^{1} w(x)g(x)dx \approx \sum_{i=1}^{n} W_i g(x_i),$$

where the weighting function is $w(x) = 1$. Calculate the expression such that it is exact to the third polynomial order x^3. Find the node locations via solving the system of equations. *Hint*: you can write the above,

$$g(x) = g_0 + g_1 x + g_2 x^2 + g_3 x^3, \quad (15)$$

and,

$$\sum_i W_i g(x_i) = \sum_i (W_i g_0 + W_i g_1 x_i + W_i g_2 x_i^2 + W_i g_3 x_i^3). \quad (16)$$

Two sides must be the same for any g_0, g_1, g_2, g_3, and we must have,

$$\int_{-1}^{1} g(x)dx = \sum_i W_i g(x_i), \quad (17)$$

which gives the system of equations, e.g.

$$\int g_0 dx = \sum_i W_i g_0, \cdots. \quad (18)$$

(15 points)

Taylor expand the function $g(x)$ upto $2n-1 = 4-1$-th order polynomial,

$$\begin{aligned}
\int_{-1}^{1} dx &= 2 = W_1 + W_2 \\
\int_{-1}^{1} \mathbf{x} dx &= 0 = W_1 x_1 + W_2 x_2 \\
\int_{-1}^{1} \mathbf{x}^2 dx &= \frac{2}{3} = W_1 \mathbf{x}_1^2 + W_2 \mathbf{x}_2^2 \\
\int_{-1}^{1} \mathbf{x}^3 dx &= 0 = W_1 \mathbf{x}_1^3 + W_2 \mathbf{x}_2^3
\end{aligned} \quad (19)$$

The above system of equations gives,

$$\begin{aligned} x_1 = -x_2 &= \frac{1}{\sqrt{3}} \\ W_1 = W_2 &= 1 \end{aligned} \quad (20)$$

17

Show that the first Order Runge-Kutta Method (RK1) is the Euler method at x_0,

$$y_{n+1} = y_n + hF(x_n, y_n),$$

with the following differential equation,

$$y'(x_0) = F(x_0, y_0), \quad y(x_0) = y_0.$$

The Euler method is given by,

$$y(x_0 + h) = hy'(x_0) + \mathcal{O}(h^2).$$

(10 points)

Expanding $y(x_0+h)$ using Taylor series:

$$y(x_0+h) - y_1 = y(x_0) + hy'(x_0) + \mathcal{O}(h^2) - y_0 - hF(x_0, y_0), \tag{21}$$

which simplifies to:

$$y(x_0+h) - y_1 = y_0 + hF(x_0, y_0) - y_0 - hF(x_0, y_0) + \mathcal{O}(h^2). \tag{22}$$

Thus:

$$y(x_0+h) - y_1 = \mathcal{O}(h^2). \tag{23}$$

18

How to categorize different second order partial differential equation using discriminant? Write down the conditions, and discuss the implications. (10 points)

- Elliptic Equations describe steady-state behavior $\Delta < 0$.
- Parabolic Equations describe diffusion-like processes. $\Delta < 0$.
- Hyperbolic Equations describe wave propagation information propagation. $\Delta > 0$.

19

a) What are the most important characteristics of a chaos system? b) A classical system is fully deterministic, and why there still exists deterministic chaos? (10 points)

a) An exponential sensitivity to the initial condition. b) Local density deterministic in phase space does not imply phase space volume distribution may behave unpredictably.

20

What is optimal transport, and how does it relate to the concept of general neural network? (5 points)

The objective is the same: the minimization in the transport cost.

Extra Credit

Consider a 3-layer single neuron Multi-layer Perceptron with ReLU activation function $f = \text{ReLU}(z) = \max(0, z)$ for a regression problem.

$$z^{(l)} = \theta_1^{(l)} a^{(l-1)} + \theta_0^{(l)}, \tag{24}$$

$$a^{(l)} = f\big(z^{(l)}\big). \tag{25}$$

a) Write down the forward propagation represented by the following transport plan, $\hat{y} = T(x)$, and draw the corresponding directed graph. b) Using the chain rule, write down the backward propagate the error analytically, and draw the corresponding directed graph. (15 points)

The forward propagation equations for a 3-layer, single-neuron MLP with ReLU activation are:

$$z^{(1)} = \theta_1^{(1)} x + \theta_0^{(1)}, \tag{26}$$

$$a^{(1)} = \max(0, z^{(1)}), \tag{27}$$

$$z^{(2)} = \theta_1^{(2)} a^{(1)} + \theta_0^{(2)}, \tag{28}$$

$$a^{(2)} = \max(0, z^{(2)}), \tag{29}$$

$$z^{(3)} = \theta_1^{(3)} a^{(2)} + \theta_0^{(3)}, \tag{30}$$

$$\hat{y} = a^{(3)} = z^{(3)}. \tag{31}$$

Using the chain rule, the gradients propagate as follows:

$$\delta^{(3)} = \frac{\partial L}{\partial z^{(3)}} = \hat{y} - y, \tag{32}$$

$$\delta^{(2)} = \delta^{(3)} \theta_1^{(3)} \mathbb{1}(z^{(2)} > 0), \tag{33}$$

$$\delta^{(1)} = \delta^{(2)} \theta_1^{(2)} \mathbb{1}(z^{(1)} > 0). \tag{34}$$

The gradients for parameter updates are:

$$\frac{\partial L}{\partial \theta_1^{(3)}} = \delta^{(3)} a^{(2)}, \quad \frac{\partial L}{\partial \theta_0^{(3)}} = \delta^{(3)}, \tag{35}$$

$$\frac{\partial L}{\partial \theta_1^{(2)}} = \delta^{(2)} a^{(1)}, \quad \frac{\partial L}{\partial \theta_0^{(2)}} = \delta^{(2)}, \tag{36}$$

$$\frac{\partial L}{\partial \theta_1^{(1)}} = \delta^{(1)} x, \quad \frac{\partial L}{\partial \theta_0^{(1)}} = \delta^{(1)}. \tag{37}$$