REPORT

January 10, 2025

Midterm Take-Home Exam

December 20, 2024 (DUE January 10, 2025)

Instructions

- Open book and open class notes are allowed (including notes taken by students during exam). No other notes are allowed.
- Each answer should be clearly written, and the solution should be developed in detail.
- Mathematical derivations need to show all steps that lead to the answer.
- Partial credit will be given for incomplete solutions.
- There is NO penalty for incorrect solutions.

Hints - equations - conventions:

- Notation
  - R represents the rate of communication in bits per channel use (bpcu),
  - ρ represents the SNR (signal to noise ratio),
  - w will denote additive noise which will be distributed as a circularly symmetric Gaussian random variable $\mathbb{C}\mathcal{N}(0, N_0)$. If N_0 is not specified, then set $N_0 = 1$,
  - Remember: for a given signal-to-noise ratio (SNR), then SNR in dB is simply $10log_{10}SNR$
  - SISO stands for single-input single-output, MISO stands for multiple-input single output, SIMO stands for single-input multiple output, MIMO stands for multiple input multiple output.
  - MU stands for multi-user.
  - CSIT stands for channel state information at the transmitter, while CSIR stands for channel state information at the receiver.
  - AWGN stands for additive white Gaussian noise.
- GOOD LUCK!!

EXAM PROBLEMS

1) ***(1 point)***. Consider a SISO setting, with no fading. Consider that the maximum possible rate (i.e., the capacity) is equal to 7 bpcu. What is the minimum SNR required to achieve this rate? Do you need CSIR?

To determine the minimum SNR required to achieve a rate of 7 bpcu (bits per channel use), we use the **Shannon capacity formula** for a single-input single-output (SISO) channel with no fading:

$C = \log_2(1 + \mathrm{SNR}),$

where: - C is the channel capacity in bits per channel use (bpcu), - SNR is the signal-to-noise ratio (linear scale).

Step 1: Solve for SNR Given that $C = 7$, we solve for SNR: $7 = \log_2(1 + \mathrm{SNR})$.

Convert to base-10 logarithms for clarity: $1 + \mathrm{SNR} = 2^7$.

Simplify: $1 + \mathrm{SNR} = 128$.

$\mathrm{SNR} = 128 - 1 = 127.$

Step 2: Minimum SNR The minimum **SNR** required is: SNR (linear) $= 127$.

Convert to decibels (dB): SNR (dB) $= 10 \cdot \log_{10}(127)$.

Using a calculator: SNR (dB) $\approx 21.03\,\mathrm{dB}$.

Step 3: Do You Need CSIR?

- **CSIR (Channel State Information at the Receiver)**: Yes, CSIR is needed to decode the signal efficiently. Without CSIR, the receiver cannot effectively interpret the transmitted data, even in a SISO setting with no fading.
- **CSIT (Channel State Information at the Transmitter)**: CSIT is not needed in this scenario because there is no fading, so the channel is constant, and the transmitter does not need to adapt its strategy dynamically.

Final Answer:

- **Minimum SNR (linear)**: 127,
- **Minimum SNR (dB)**: $21.03\,\mathrm{dB}$,
- **CSIR Requirement**: Yes, CSIR is required.

2) ***(1 point)***. Consider a SISO quasi-static fading channel with no CSIT. We wish to decrease the probability of error, from$P_{err} \approx (SNR)^1$ to $P_{err} \approx (SNR)^4$. Suggest various ways we can achieve this, based on what we have learned in class.

Ways to Improve from $P_{err} \approx (SNR)^1$to$P_{err} \approx (SNR)^4$:

1. Error-Correcting Codes:

- Use stronger codes (e.g., LDPC, Turbo, Polar codes).
- Increase codeword length.

2. Diversity Techniques:

- Time Diversity: Retransmissions (e.g., hybrid ARQ).
- Frequency Diversity: Spread spectrum or OFDM.
- Antenna Diversity: Space-time coding (e.g., Alamouti).

3. Modulation:

- Use lower-order modulation (e.g., QPSK instead of 16-QAM).
- Apply trellis-coded modulation (TCM).

4. Signal Design:

- Use constellation shaping for better noise robustness.

5. Improved CSIR:

- Enhance channel estimation at the receiver for better decoding.

6. Adaptive Techniques:

- SNR-adaptive decoding or statistical power allocation.

By combining coding, diversity, and modulation optimization, the desired reduction P_{err} can be achieved.

3) ***(1 point)***. What are some of the advantages of MISO vs. SIMO, mentioned in class?

MISO (Multiple Input Single Output) has several advantages over SIMO (Single Input Multiple Output):

1. **Improved Signal Reliability**: Multiple antennas at the transmitter reduce the impact of fading.
2. **Enhanced Data Rates**: Multiple transmitting antennas can increase data rates.
3. **Diversity Gain**: Provides diversity gain to combat multi-path fading.
4. **Power Efficiency**: Distributes transmission power across multiple antennas, improving efficiency.

4) ***(1 point)***. In a single-user MIMO channel, how much diversity gain would we be able to get if we employed a transmitter with 4 transmit antennas and a receiver with 2 receive antennas, when in fact the channel between the first transmit and receive antenna, is identical always to the channel between the first transmit and second receive antenna?

With 4 transmit antennas and 2 receive antennas, if the channel between the first transmit antenna and both receive antennas is identical, the effective diversity gain is calculated as follows:

Diversity Gain $= N_t \times N_{\text{effective}} = 4 \times 1 = 4$

Thus, the diversity gain would be 4.

5) ***(1 point)***. In a single-user MISO channel, how much multiplexing gain would we be able to get if we employed a transmitter with 2 transmit antennas?

In a single-user MISO channel with 2 transmit antennas, the multiplexing gain is:

Multiplexing Gain $= \min(N_t, 1) = \min(2, 1) = 1$

Given: - $N_t = 2$ (number of transmit antennas) - Receive antennas = 1

Thus, the multiplexing gain is 1.

6) ***(1 points)***. Consider communication over a quasi-static 2×1 MISO fading channel. Assume that you must draw symbols from 16-QAM.

- Can you name a space time code, that gives full diversity in this setting, and then describe the rate (in bpcu) of such a code.

For a quasi-static 2×1 MISO (Multiple Input Single Output) fading channel, a well-known space-time code that provides full diversity is the Alamouti code. The Alamouti scheme is specifically designed for a 2×1 setup and achieves full diversity by using two transmit antennas and one receive antenna.

Space-Time Code: Alamouti Code

- **Diversity Gain**: Full diversity
- **Code Rate**: 1 bpcu (bit per channel use)

Description The Alamouti scheme transmits two symbols s_1 and s_2 over two time slots from two transmit antennas. The transmission matrix for the Alamouti code is given by:

$$\mathbf{S} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}$$

Rate Calculation Since 16-QAM is used, each symbol s_1 and s_2 carries 4 bits (since 16-QAM has $2^4 = 16$ symbols).

- In one time slot, 4 bits are transmitted from each antenna.
- Over two time slots, a total of 8 bits are transmitted.

Since the rate of the Alamouti code is 1 bpcu, and we use 16-QAM, the effective rate of the system is:

Rate $= \frac{\text{Number of bits transmitted}}{\text{Number of channel uses}} = \frac{8 \text{ bits}}{2 \text{ channel uses}} = 4$ bits per channel use (bpcu)

Thus, the Alamouti code provides full diversity with a rate of 4 bpcu when using 16-QAM.

7) ***(3 points)***. In the context of various strategies, answer if each of the following statements are true or false, justifying briefly your answers.

- In a MISO channel, we can get transmitter beamforming gain even without CSIT.

- **False** - Transmitter beamforming gain requires CSIT.

- A base station equipped with 5 antennas in the downlink, can simultaneously serve up to 5 users (single receive antenna each).
  - **True** - 5 antennas can serve 5 users with single antennas using spatial multiplexing.
- A base station equipped with 5 antennas in the downlink, can simultaneously serve up to 10 users (two receive antennas each).
  - **False** - 5 antennas cannot serve 10 users even if each user has 2 receive antennas.
- A base station equipped with 4 antennas in the downlink, can simultaneously serve up to 2 users (two receive antennas each).
  - **True** - 4 antennas can serve 2 users with 2 receive antennas each using spatial multiplexing.
- Line of sight channels are detrimental for spatial multiplexing in both single-user and multiuser MIMO.
  - **True** - Line of sight channels are detrimental for spatial multiplexing due to rank deficiency.
- For a MIMO receiver using spatial multiplexing, the complexity of ZF receiver is more than the complexity of the maximum-likelihood receiver.
  - **False** - ZF receiver is less complex than the maximum-likelihood receiver.
- CSIT is easier to obtain than CSIR.
  - **False** - CSIR is generally easier to obtain than CSIT.
- CSIT is of cardinal importance in multi-user MIMO.
  - **True** - CSIT is crucial for multi-user MIMO to perform efficient beamforming and spatial multiplexing.

8) ***(2 points)***. In a MU-MIMO channel, if I double the number of users I simultaneously serve, must I always halve the individual rate to each user? Justify your answer.

No, you do not always have to halve the individual rate to each user when you double the number of users in a MU-MIMO (Multi-User Multiple Input Multiple Output) channel. The relationship between the number of users and their individual rates depends on several factors, including the total available bandwidth, the power allocation, the channel conditions, and the spatial multiplexing capabilities of the system.

Justification:

1. **Spatial Multiplexing Gains**: MU-MIMO systems can exploit spatial multiplexing to serve multiple users simultaneously without necessarily reducing the rate per user. With efficient spatial processing (such as beamforming) and adequate channel state information at the transmitter (CSIT), the system can separate the signals intended for different users.

2. **Power Allocation**: The total transmit power can be dynamically allocated among the users. By optimizing the power allocation, it is possible to maintain or even improve the individual rates of users despite an increase in the number of users.
3. **Channel Conditions**: The individual rates are also influenced by the quality of the channels between the transmitter and the users. If the channels are favorable (e.g., high signal-to-noise ratio, low interference), the system can support more users without significantly degrading their individual rates.
4. **Advanced Techniques**: Techniques such as user scheduling, adaptive modulation and coding, and interference management can help maintain individual user rates even as the number of users increases.

Conclusion: Doubling the number of users in a MU-MIMO channel does not necessarily mean halving the individual rate for each user. Efficient use of spatial multiplexing, power allocation, favorable channel conditions, and advanced signal processing techniques can help mitigate the impact on individual rates.

9) ***(4 points)***. Consider communication over the2×1 quasi-static fading MISO channel, using a diagonal code (see below for details) such that the channel model is given by

$$\overbrace{\begin{pmatrix} y_1 & y_2 \end{pmatrix}}^{y} = \theta \overbrace{\begin{pmatrix} h_1 & h_2 \end{pmatrix}}^{h} \overbrace{\begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}}^{\mathcal{X}_{tr}} + \overbrace{\begin{pmatrix} w_1 & w_2 \end{pmatrix}}^{w}$$

where $h_i \sim \mathbb{C}N(0,1)$ and $w_i \sim \mathbb{C}N(0,1)$, and where θ is the power normalization factor that lets you regulate SNR.

- Describe the ML decoding rule for this case.
- Describe the cardinality[1] of code $\mathcal{X}_{tr}$ if you wish a rate of $R = 4$ bpcu.
- For a desired rate of $R = 8$ bpcu, and a desired SNR $= 10$ dB (where by SNR we mean the AVERAGE signal power divided by the noise unit power, under QAM) then what is the normalizing factor θ ?
- Imagine that what you transmit (x_1, x_2) are independently chosen from 16-PAM, then
  - What is the rate of your code (in bpcu)?
  - What is the slope of your probability of error, in high SNR, if you plot on the y-axis the probability of error, in log scale ($\log_{10}$(Prob)), and the x-axis is the SNR, in dB?
- Imagine now that $(x_1 x_2) = s_1 s_2 \cdot \mathbf{Q}$, where s1,s2 are independently chosen from a 64-QAM constellation, where the matrix $\mathbf{Q}$ is a randomly chosen orthogonal matrix. Then
  - What is the rate of your code?

  - What is the aforementioned slope of your probability of error?

[1] By cardinality we mean the number of matrices that the code has.

ML Decoding Rule The Maximum Likelihood (ML) decoding rule for the given channel model is to find the transmitted symbols x_1 and x_2 that minimize the Euclidean distance between the received vector $\underline{y}$ and the expected received vector given by the channel matrix $\underline{h}$ and the transmitted symbols. Mathematically, this can be expressed as:

$\hat{x}_1, \hat{x}_2 = \arg\min_{x_1,x_2} \|\underline{y} - \theta \underline{h} \mathcal{X}_{tr}\|^2$

where $\underline{y} = (y_1 \quad y_2)$, $\underline{h} = (h_1 \quad h_2)$, $\mathcal{X}_{tr} = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}$.

Cardinality of Code $\mathcal{X}_{tr}$ for Rate $R = 4$ bpcu The rate R in bits per channel use (bpcu) determines the number of possible codewords. If $R = 4$ bpcu, the number of codewords is $2^R = 2^4 = 16$. Therefore, the cardinality of the code $\mathcal{X}_{tr}$ is 16.

Normalizing Factor θ for Rate $R = 8$ bpcu at SNR = 10 dB For an SNR of 10 dB, the average signal power to noise power ratio is 10. In terms of linear scale, this is $10^{10/10} = 10$.

Given that we are using QAM, the average power of a QAM signal is proportional to the square of the constellation size. For a rate of $R = 8$ bpcu, we need $2^8 = 256$ possible symbols. This corresponds to a 256-QAM constellation.

The normalizing factor θ ensures that the average signal power is regulated to achieve the desired SNR. The power normalization factor can be calculated as:

$\theta = \sqrt{\frac{\text{SNR}}{\text{Average Power of QAM constellation}}}$

For 256-QAM, the average power is approximately:

$P_{\text{avg}} = \frac{(M-1)}{3} = \frac{255}{3} \approx 85$

Thus,

$\theta = \sqrt{\frac{10}{85}} \approx 0.34$

Rate of Code and Slope of Probability of Error for 16-PAM

- **Rate**: If x_1 and x_2 are independently chosen from a 16-PAM constellation, each symbol carries $\log_2(16) = 4$ bits. Since there are 2 symbols, the rate is:

$R = 2 \times 4 = 8$ bpcu

- **Slope of Probability of Error**: In high SNR regimes, the slope of the probability of error (in log scale) for a system with diversity order d is:

Slope $= -d$

For a 2x1 MISO system using independent symbols, the diversity order is 2. Thus, the slope of the probability of error is:

Slope $= -2$

Rate and Slope of Probability of Error for Orthogonal Matrix Q with 64-QAM

- **Rate**: If x_1 and x_2 are chosen from a 64-QAM constellation, each symbol carries $\log_2(64) = 6$ bits. Since there are 2 symbols, the rate is:

$R = 2 \times 6 = 12$ bpcu

- **Slope of Probability of Error**: If $(x_1, x_2) = (s_1, s_2) \cdot \mathbf{Q}$ where $\mathbf{Q}$ is an orthogonal matrix, the diversity order remains 2. Therefore, the slope of the probability of error is:

Slope $= -2$

10) ***(Extra Credit: 5 points)***. Consider communication over a quasi-static 2×2 MIMO channel, utilizing the space-time code $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4\}$, where

$$\mathbf{X}_1 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \mathbf{X}_2 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \mathbf{X}_3 = \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix} \text{ and } \mathbf{X}_4 = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix},$$

- What is the average SNR?
- What is the rate of the code in bpcu?
- What is the diversity gain of this code?
- What is the approximate (in the high SNR regime) probability of error of this code, if SNR is 30dB?

Average SNR The SNR (Signal-to-Noise Ratio) in a MIMO system is typically defined as the ratio of the average signal power to the average noise power. If we assume the transmitted signal power is normalized to 1, and the noise power is also normalized to 1 (i.e., unit variance noise), then the average SNR is given directly by the power normalization factor.

Given that the SNR is often defined per receive antenna, and assuming equal power allocation across the transmit antennas, the average SNR per receive antenna is:

$\text{SNR} = \frac{\text{Total transmit power}}{\text{Noise power per receive antenna}}$

In this case, we assume the transmit power is 1 per antenna and the noise power is 1 per receive antenna, so:

Average SNR $=$ SNR $= \frac{1}{1} = 1$ (normalized)

Rate of the Code in bpcu The rate of a space-time code is determined by the number of independent symbols transmitted per channel use. Here, we have 4 code matrices $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4\}$.

Each code matrix represents a unique codeword. Since there are 4 codewords, we can send $\log_2(4) = 2$ bits per channel use (bpcu).

Rate $= 2$ bpcu

Diversity Gain of the Code The diversity gain of a space-time code is given by the rank of the difference of any two distinct code matrices. For full diversity, the rank of all such differences should be equal to the number of transmit antennas.

Let's check the rank of the difference of any two distinct code matrices. For example, consider $\mathbf{X}_1$ and $\mathbf{X}_2$:

$\mathbf{X}_1 - \mathbf{X}_2 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$

The rank of this matrix is 2, indicating full diversity.

Diversity Gain = 2

Approximate Probability of Error at SNR = 30 dB In the high SNR regime, the probability of error P_e for a MIMO system with diversity gain d typically decreases as:

$P_e \approx \left(\frac{1}{\text{SNR}}\right)^d$

Given: - SNR = 30 dB = $10^{30/10} = 10^3$ - Diversity Gain $d = 2$

The approximate probability of error is:

$P_e \approx \left(\frac{1}{10^3}\right)^2 = 10^{-6}$

Summary

- **Average SNR**: 1 (normalized)
- **Rate**: 2 bpcu
- **Diversity Gain**: 2
- **Approximate Probability of Error at SNR = 30 dB**: 10^{-6}

11) ***(1 point)***. Consider the following distributed setup with N = 3 workers, as shown in Figure 1. There are $K = 3$ datasets, W_1, W_2, and W_3, each of size 100 MBytes and each cache has a storage capacity of 200 MBytes.

a) How does one need to distribute the datasets across the workers to ensure that the master node can recover all the datasets from any 2 workers?

b) Assume that there is a delay constraint of 10 milli-seconds (ms) allowed for the master node to receive all the information. What is the total rate R (in received bits per milli-second) with which the master node will be receiving data, in order for it to successfully recover W_1, W_2, W_3?

c) Let us now assume that the master node also has memory and can cache one dataset. What is the total rate of transmission the master needs from the workers to successfully recover W_1, W_2, W_3?

d) Comment on the role of memory on the transmission rate from the workers.

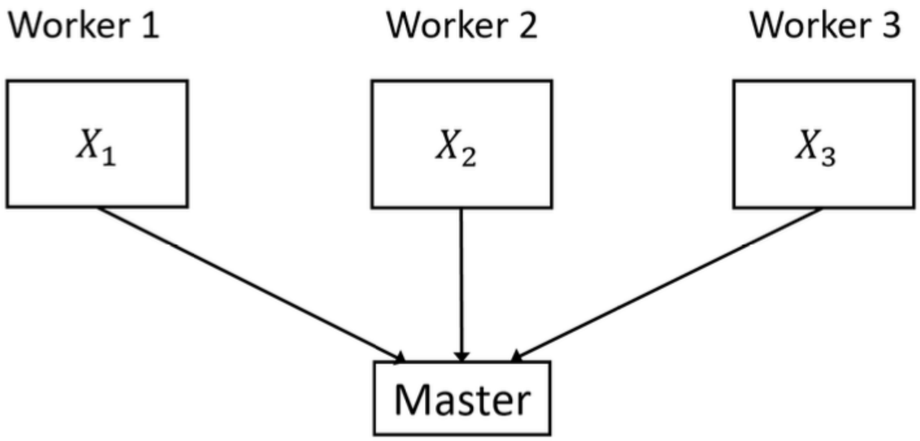

Fig. 1: A distributed computing scenario.

---

a) Distributing the datasets To ensure the master can recover all datasets W_1, W_2, W_3 from **any 2 workers**: 1. **Storage Capacity**: Each worker has a storage capacity of 200 MB, and the total size of all datasets is 300 MB. Therefore, each dataset must be partially stored at multiple workers to enable recovery.

2. **Data Distribution**:
   - Split each dataset W_1, W_2, W_3 into **two equal parts**: $W_{i,1}$ and $W_{i,2}$ (e.g., $W_1 = W_{1,1} + W_{1,2}$, each part 50 MB).
   - Assign the parts such that **any two workers have all parts of all datasets**. For instance:
     - Worker 1: $W_{1,1}, W_{2,1}, W_{3,1}$ (150 MB total).
     - Worker 2: $W_{1,2}, W_{2,1}, W_{3,2}$ (150 MB total).
     - Worker 3: $W_{1,1}, W_{2,2}, W_{3,2}$ (150 MB total).

   This ensures that the master can recover W_1, W_2, W_3 from any two workers since all parts of each dataset are distributed across at least two workers.

b) Total rate R required (in bits/ms)

- **Delay Constraint**: The master must recover all W_1, W_2, W_3 (300 MB total) within 10 ms.
- **Total Rate** R: $R = \frac{\text{Total Size of Datasets (in bits)}}{\text{Delay Constraint (in ms)}}$.
  - Total size of datasets = 300 MB = $300 \times 8 \times 10^6$ bits = 2.4×10^9 bits.
  - Delay constraint = 10 ms = 10^{-2} s.

  Substituting: $R = \frac{2.4 \times 10^9}{10} = 2.4 \times 10^8$ bits/ms.

  Answer: $R = 240$ Mbps.

c) Total rate if the master can cache one dataset

- If the master caches one dataset, say W_1, then it only needs to recover W_2 and W_3 from the workers.
- Size of data to be transmitted: Total Size $= W_2 + W_3 = 200\,\text{MB} = 200 \times 8 \times 10^6\,\text{bits} = 1.6 \times 10^9\,\text{bits}$.
- **Total Rate** R: $R = \frac{1.6\times10^9}{10} = 1.6 \times 10^8\,\text{bits/ms}$.

  Answer: $R = 160\,\text{Mbps}$.

d) Role of memory in transmission rate

- **Without Master Caching**: The master must recover all datasets (W_1, W_2, W_3) from the workers, resulting in a higher transmission rate ($240\,\text{Mbps}$).
- **With Master Caching**: The master pre-storing one dataset reduces the amount of data it needs to recover, thereby lowering the transmission rate ($160\,\text{Mbps}$).

Conclusion: - Increasing the memory at the master node significantly reduces the transmission rate required from the workers. - Memory acts as a **trade-off between storage and communication cost**, with more memory reducing the load on the communication channel.

12) ***(1 point)***. Name two applications that in your opinion popularized Reinforcement Learning in scientific communities working on Artificial Intelligence?

Applications That Popularized Reinforcement Learning (RL):

1. **Game Playing (e.g., AlphaGo, AlphaZero)**:
   - **Description**: RL gained significant attention with the development of DeepMind's **AlphaGo** and later **AlphaZero**, which used RL to achieve superhuman performance in complex games like Go, chess, and shogi.
   - **Impact**: These breakthroughs demonstrated RL's capability to handle high-dimensional decision-making problems, inspiring its use in other domains.
2. **Autonomous Robotics**:
   - **Description**: RL has been widely adopted in training robots to perform tasks such as walking, grasping, and manipulation. Robots learn optimal policies through interaction with their environments.
   - **Impact**: Showcased RL's potential in real-world, dynamic, and uncertain environments, leading to advancements in robotics and control systems.

13) ***(1 point)***. AlphaZero is one of the most exciting real-world applications that works using Reinforcement Learning. Name the two algorithms that are used by AlphaZero based on which this software achieves superior performance compared to humans.

The two key algorithms used by **AlphaZero** that enable its superior performance are:

1. **Monte Carlo Tree Search (MCTS)**:
   - **Description**: MCTS is a search algorithm used to explore possible moves in a game by simulating future outcomes and using the results to guide the search process.
   - **Role in AlphaZero**: MCTS evaluates potential actions efficiently, focusing on the most promising moves, which helps AlphaZero make better decisions.

2. **Deep Reinforcement Learning**:
   - **Description**: AlphaZero combines reinforcement learning with deep neural networks to approximate the value function (estimating the likelihood of winning) and the policy (deciding the best move).
   - **Role in AlphaZero**: The neural network predicts move probabilities and board evaluations, guiding MCTS and improving its performance over time through self-play.

These two algorithms work together synergistically, with MCTS leveraging the neural network's outputs and the network improving its predictions through reinforcement learning.

14) ***(1 point)***. Name two classes based on which one can distinguish exact dynamic programming algorithms?

Two classes used to distinguish exact dynamic programming algorithms are:

1. **Value Iteration**:
   - Iteratively updates the value function for each state until convergence, focusing on finding the optimal value function.
2. **Policy Iteration**:
   - Alternates between policy evaluation (calculating the value of a given policy) and policy improvement (updating the policy based on the value function) until an optimal policy is found.

15) ***(1 point)***. What is the difference between deterministic and stochastic dynamic programming algorithms?

The key differences between **deterministic** and **stochastic dynamic programming algorithms** are:

1. **Nature of the Environment**:
   - **Deterministic**: Assumes the system evolves in a fully predictable manner, where the next state is uniquely determined by the current state and action.
   - **Stochastic**: Accounts for uncertainty, where the next state depends on probabilistic transitions.
2. **Transition Model**:
   - **Deterministic**: Uses fixed transition rules (e.g., $s_{t+1} = f(s_t, a_t)$).
   - **Stochastic**: Uses a probability distribution over possible transitions (e.g., $P(s_{t+1}|s_t, a_t)$).
3. **Computation Complexity**:
   - **Deterministic**: Simpler and computationally less intensive due to the absence of randomness.
   - **Stochastic**: More complex, as it requires handling probabilities and expectations over all possible transitions.

16) ***(5 points)***. Quantum. The two components of the state $|\theta> = \frac{\sqrt(3)}{2\sqrt(2)}|11> + \frac{\sqrt(5)}{2\sqrt(2)}|00>$ are shared amidst Alice and Bob. In other words, Alice has the first component and Bob has the second. Alice performs a measurement with two outcomes $\{+1, 1\}$. Outcome $+1$ is associated with operator $|+><+|$ and Outcome 1 is associated with $|><|$. Bob performs a measurement with two outcomes $\{+1, 1\}$. Outcome $+1$ is associated with operator $|0><0|$ and Outcome 1 is associated with $|1><1|$.

a) Before the measurement is performed, is the joint state entangled or separable?

b) Compute the probability of Alice observing -1 and Bob observing $+1$.

c) Compute the probability of Alice and Bob BOTH observing 1.

d) Compute the probability of Alice and Bob BOTH observing $+1$.

e) Identify the post measurement state when Alice observes $+1$ and Bob observes 1.

---

a) Is the joint state entangled or separable? The state is:

$|\theta\rangle = \frac{\sqrt{3}}{2\sqrt{2}}|11\rangle + \frac{\sqrt{5}}{2\sqrt{2}}|00\rangle$.

- A state is **entangled** if it cannot be written as a tensor product of individual states, $|\theta\rangle \neq |\phi_A\rangle \otimes |\phi_B\rangle$.
- Here, $|\theta\rangle$ is a superposition of $|11\rangle$ and $|00\rangle$, and it cannot be decomposed into separable parts for Alice and Bob.

Answer: The state is **entangled**.

b) Probability of Alice observing -1 and Bob observing $+1$:

1. Measurement Operators:

- Alice's measurement outcomes are:
  - $+1$: $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$,
  - -1: $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.
- Bob's measurement outcomes are:
  - $+1$: $|0\rangle$,
  - -1: $|1\rangle$.

2. Compute $|\theta\rangle$ in the measurement basis: $|\theta\rangle = \frac{\sqrt{3}}{2\sqrt{2}}|11\rangle + \frac{\sqrt{5}}{2\sqrt{2}}|00\rangle$.

Express $|1\rangle$ and $|0\rangle$ in terms of $|+\rangle$ and $|-\rangle$ for Alice: $|0\rangle = \frac{|+\rangle+|-\rangle}{\sqrt{2}}, \quad |1\rangle = \frac{|+\rangle-|-\rangle}{\sqrt{2}}$.

Substituting for Alice's basis: $|11\rangle = \frac{1}{\sqrt{2}}(|+\rangle - |-\rangle)|1\rangle, \quad |00\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)|0\rangle$.

Expand: $|\theta\rangle = \frac{\sqrt{3}}{2\sqrt{2}} \cdot \frac{1}{\sqrt{2}}(|+\rangle - |-\rangle)|1\rangle + \frac{\sqrt{5}}{2\sqrt{2}} \cdot \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)|0\rangle$. $|\theta\rangle = \frac{\sqrt{3}}{4}(|+\rangle - |-\rangle)|1\rangle + \frac{\sqrt{5}}{4}(|+\rangle + |-\rangle)|0\rangle$.

3. Amplitude for $|-\rangle|0\rangle$ (Alice observes -1, Bob observes $+1$): The coefficient of $|-\rangle|0\rangle$ is: $\frac{\sqrt{5}}{4}$.

4. Probability: $P(-1,+1) = \left|\frac{\sqrt{5}}{4}\right|^2 = \frac{5}{16}$.

c) Probability of Alice and Bob BOTH observing -1:

1. Amplitude for $|-\rangle|1\rangle$: The coefficient of $|-\rangle|1\rangle$ is: $-\frac{\sqrt{3}}{4}$.

2. Probability: $P(-1,-1) = \left|-\frac{\sqrt{3}}{4}\right|^2 = \frac{3}{16}$.

d) Probability of Alice and Bob BOTH observing $+1$:

1. Amplitude for $|+\rangle|0\rangle$: The coefficient of $|+\rangle|0\rangle$ is: $\frac{\sqrt{5}}{4}$.

2. Probability: $P(+1,+1) = \left|\frac{\sqrt{5}}{4}\right|^2 = \frac{5}{16}$.

e) Post-measurement state when Alice observes $+1$ and Bob observes -1:

1. Projection: The projection operator for this outcome is: $P_{+1,-1} = (|+\rangle\langle+|) \otimes (|1\rangle\langle 1|)$.

Apply this to $|\theta\rangle$: $|\theta'\rangle = P_{+1,-1}|\theta\rangle$.

From the expanded $|\theta\rangle$, the term contributing to $|+\rangle|1\rangle$ has coefficient: $\frac{\sqrt{3}}{4}$.

2. Normalize: $|\theta'\rangle = \frac{|+\rangle|1\rangle}{\sqrt{P(+1,-1)}}$.

Probability $P(+1,-1)$: $P(+1,-1) = \left|\frac{\sqrt{3}}{4}\right|^2 = \frac{3}{16}$.

Normalized state: $|\theta'\rangle = \frac{|+\rangle|1\rangle}{\sqrt{\frac{3}{16}}} = \frac{4}{\sqrt{3}}|+\rangle|1\rangle$.

Summary of Answers:

a) The state is **entangled**.

b) $P(-1,+1) = \frac{5}{16}$.

c) $P(-1,-1) = \frac{3}{16}$.

d) $P(+1,+1) = \frac{5}{16}$.

e) Post-measurement state: $|\theta'\rangle = \frac{4}{\sqrt{3}}|+\rangle|1\rangle$.

```
[ ]:
```