CS 189 Spring 2014 — Introduction to Machine Learning — Midterm

- You have 2 hours for the exam.
- The exam is closed book, closed notes except your one-page crib sheet.
- Please use non-programmable calculators only.
- Mark your answers ON THE EXAM ITSELF. If you are not sure of your answer you may wish to provide a *brief* explanation.
- For true/false questions, fill in the *True/False* bubble.
- For multiple-choice questions, fill in the bubbles for **ALL** CORRECT CHOICES (in some cases, there may be more than one). We have introduced a negative penalty for false positives for the multiple choice questions such that the expected value of randomly guessing is 0. Don't worry, for this section, your score will be the maximum of your score and 0, thus you cannot incur a negative score for this section.

First name	
Last name	
SID	
First and last name of student to your left	
First and last name of student to your right	

For staff use only:

Q1.	Hidden Markov Models: Math Review	/0
	Total	/0

Q1. [0 pts] Hidden Markov Models: Math Review

A Hidden Markov Model is a Markov Process with unobserved (hidden) states.

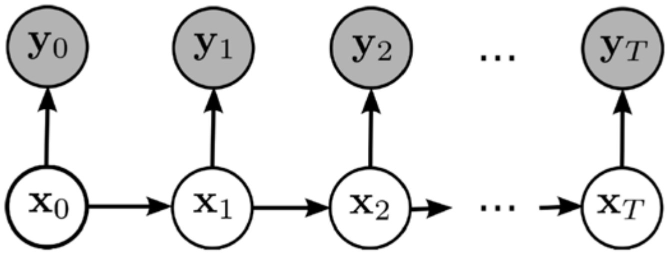

Figure 1: Example Hidden Markov Chain

Consider the following system in $\mathbb{R}^2$, where X_n is the true state at any given time n and Y_n is our observation. Given an initial state X_0, we move to future states by recursively multiplying our current state with transformation matrix A and adding i.i.d. Standard Normal Gaussian noise. When we take an observation Y_n of the true state X_n, we are also exposed to i.i.d. Standard Normal Gaussian Noise.

$$X_{n+1} = AX_n + N(0, I) \tag{1}$$
$$Y_n = X_n + N(0, I) \tag{2}$$

Where we have the $2x2$ transformation matrix A defined as follows:

$$A = \begin{bmatrix} .5 & -.25 \\ -.25 & .75 \end{bmatrix} \tag{3}$$

If we restrict the initial state X_0 to be a unit vector ($\|X_1\|_2 = 1$), determine the following

(a) What are the eigenvalues of A? Is A a positive semi-definite matrix? (Note that $\sqrt{5}\ = 2.236$)

Remember that an eigenvector is a vector $\vec{v}$ such that $A\vec{v} = \lambda\vec{v}$, where the constant λ is the eigenvalue corresponding to $\vec{v}$. We manipulate the above equation to be $(A - \lambda I)\vec{v} = 0$, which implies that $A - \lambda I$ is a singular matrix since it has an eigenvalue of 0.

$$A - \lambda I = \begin{bmatrix} \frac{1}{2} - \lambda & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} - \lambda \end{bmatrix} \tag{4}$$

We can take the determinant of the above matrix and set it to zero in order for the matrix to be singular, giving us the following characteristic polynomial:

$$0 = (\frac{1}{2} - \lambda)(\frac{3}{4} - \lambda) - (-\frac{1}{4})(-\frac{1}{4}) = \lambda^2 - \frac{5}{4}\lambda + \frac{3}{8} - \frac{1}{16} = \lambda^2 - \frac{5}{4}\lambda + \frac{5}{16} \tag{5}$$
$$\lambda = \frac{1}{2}(\frac{5}{4} \pm \sqrt{\frac{25}{16} - 4(\frac{5}{16}}) = \frac{1}{8}(5 \pm \sqrt{5}) \tag{6}$$

Since $\lambda > 0$ for all possible values, it is a positive-semidefinite matrix (in fact, it is positive definite).

—-

(b) What is the $\| E[Y_\infty] \|_2$? Prove your assertion.

Lets look at the first several expressions of the true state X

$$X_1 = AX_0 + N(0, I) \tag{7}$$
$$X_2 = A(AX_0 + N(0, I)) + N(0, I) \tag{8}$$
$$X_3 = A(A(AX_0 + N(0, I)) + N(0, I)) + N(0, I) \tag{9}$$

We note that a particular state can be defined by our original state as follows $X_n = A^n X_0 + \sum\limits_{i=0}^{n-1} A^i N(0, I)$.

Thus, our observation of that is $Y_n = A^n X_0 + N(0, I) + \sum\limits_{i=0}^{n-1} A^i N(0, I)$.

Remember that since matrix A is a real symmetric matrix, we can use spectral decomposition to prove that $A^N = (UDU^\top)^N = UD^N U^\top$, where U is a unitary matrix and D is a diagonal matrix of eigenvalues. Note that our eigenvalues are such that $0 < \lambda < 1$. Therefore, $D^N = 0 \Rightarrow A^N = 0$

Thus, when we take expectations and norm, we see that

$$\| \lim_{n\to\infty} E[Y_n]\|_2 = \|E[A^n X_0 + N(0, I) + \sum_{i=0}^{n-1} A^i N(0, I)]\|_2 \tag{10}$$
$$= \|E[N(0, I) + \sum_{i=0}^{n-1} A^i N(0, I)]\|_2 \tag{11}$$
$$= \|0\|_2 \tag{12}$$
$$= 0 \tag{13}$$

(c) Consider the Frobenius Norm of an arbitrary M x N matrix Q, defined as $\|Q\|_F = \sqrt{\sum\limits_i \sum\limits_j |Q_{i,j}|^2}$, which indicates the magnitude or largeness of a matrix. Is $\| Var[Y_\infty] \|_F$ finite or infinite? Prove your assertion.

You may find the following facts to be useful:

(i) Triangle Inequality: $\|X + Y\| \leq \|X\|+\|Y\|$

(ii) Cauchy Schwarz: $\|XY\| \leq \|X\|\|Y\|$

(iii) Geometric Sum: $\sum_{i=0}^{\infty} ar^i = \frac{1}{1-r} \qquad \forall r$ s.t. $0 < r < 1$; $a, r \in \mathbb{R}$

We will approach this part in the same way as part b). Remember from discussion that for multidimensional i.i.d. random variables X,Y with variance I, and constant matrix B:

$$BVar[X] = BIB^\top = BB^\top \qquad Var[B + X] = Var[X] \qquad Var[X + Y] = Var[X] + Var[Y] = I + I = 2I$$

Therefore, if we examine $\lim_{n\to\infty} Var[Y_n]$, where we define N(0,I) = Q and note that A is symmetric ($A = A^\top$), we see that:

$$\lim_{n\to\infty} Var[Y_n] = Var[A^n X_0 + \sum_{i=0}^{n-1} A^i Q + Q] \tag{14}$$

$$= Var[\sum_{i=0}^{n-1} A^i Q] + Var[Q] = \sum_{i=0}^{n-1} Var[A^i Q] + I \tag{15}$$

$$= \sum_{i=0}^{n-1} A^i I (A^\top)^i + I = \sum_{i=0}^{n-1} (AA^\top)^i + I \tag{16}$$

$$= \sum_{i=0}^{n-1} (A)^{2i} + I = \sum_{i=0}^{n-1} (UDU^\top)^{2i} + I \tag{17}$$

$$= \sum_{i=0}^{n-1} UD^{2i}U^\top + I = U(\sum_{i=0}^{n-1} D^{2i})U^\top + I \tag{18}$$

$$\tag{19}$$

We could stop here and note that $\sum_{i=0}^{n-1} D^{2i}$ is finite since $0 < D_{1,1}, D_{2,2} < 1$. Thus, since D is a diagonal matrix and D^n is also diagonal we can apply the geometric sum formula for each term $\sum_{i=0}^{n-1} (D_{1,1})^{2i}$ and $\sum_{i=0}^{n-1} (D_{2,2})^{2i}$. We then note that the sum is finite, that U and $U^\top$ will preserve magnitude, and I is finite. Therefore, the above limit is finite, which means that the Frobenius Norm is also finite.

If we want to decompose further, we can use the Triangle Inequality and Cauchy Schwarz Inequality:

$$\|\lim_{n\to\infty} Var[Y_n]\|_F = \|U(\sum_{i=0}^{n-1} D^{2i})U^\top + I\|_F \tag{20}$$

$$\leq \|I\|_F + \|U(\sum_{i=0}^{n-1} D^{2i})U^\top\|_F \tag{21}$$

$$\leq \|I\|_F + \|U\|_F \, \|(\sum_{i=0}^{n-1} D^{2i})\|_F \, \|U^\top\|_F \tag{22}$$

$$\tag{23}$$

We then use the same argument as before to show that the sum of diagonal matrices is a geometric series, and note that I and U are finite matrices. Therefore, both have finite norms and the sum must be finite.