Problem Sets

Some of the problems below will be part of your quizzes, assignments and take home exams. You should attempt to really understand how to solve the problems below when we cover the appropriate material in class.

Please note that these problems are not yet finalized - I will add more problems as the class progresses.

Part 1: Probabilistic Models

1. Suppose $\Theta \sim U(0, 2\pi)$ (uniform pdf) and $R \sim \lambda e^{-\lambda r}$ (exponential pdf with $\lambda = 1/2$) are indpt rvs.
   Then show that $X = \sqrt{R}\cos\Theta$ and $Y = \sqrt{R}\sin\Theta$ are indpt $N(0, 1)$ rvs. (The slides covered in class gives an outline of the derivation).

   Remark: This problem involves computing distribution of a two dimensional function of two random variables; most of the details for the soln are in the class slides – it involves computing the Jacobian matrix. This specific example is important in modeling wireless communication channels from a cell phone to a basestation – we will cover this later in the course again.

2. Suppose X and Y are independent exponentially distributed random variables. Is $Z = \max(X, Y)$ exponentially distributed? What about $W = \min(X, Y)$? Evaluate the density function for $X + Y$.

   Remark: Obviously if $Z = \min(X, Y)$, then $Z > z$ implies $X > z$ and $Y > z$. So you can compute $P(Z > z) = P(X > z)P(Y > z)$ since X and Y are indpt. Of course $P(Z \leq z)$ which is the cdf is $1 - P(Z \leq z)$. For the sum, the pdf is the convolution of the two pdfs. This problem is very important in reliability theory: suppose you have 2 machines each of which fails at a random time with an exponential distribution. Suppose that your production line stops as soon as one of the machines fails. Then the problem reduces to computing the pdf of $\min(X, Y)$. Actually there is a really cool algorithm in graph theory that uses labelling of nodes in a graph using exponential random variables to count the number of nodes in a neighborhood of a graph – it is called Cohen's algorithm [we wont cover this since it is advanced – but the point is that exponential rv and their minimum are very useful].

3. Suppose a random variable U is uniformly distributed on the interval [0,1]. Derive the density function of the random variables

   - $Y = -\ln(1 - U)$
   - $Y = U^n$ for $n \geq 1$.

   Obtain the pdf of the sum of two independent $U[0, 1]$ uniform random variables.

Remark: Start with showing the result that $Y = F(U)$ implies that Y has cdf F^{-1}. This result is crucial in stochastic simulation. It is the basis of the inverse transform method.

4. Multivariate Gaussian. An n-dimensional Gaussian random variable $X \sim N(\mu, \Sigma)$ has pdf

$$\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\big(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\big), \quad x \in \mathbb{R}^n$$

where $\mu \in \mathbb{R}^n$ and Σ is an $n \times n$ symmetric positive definite matrix.

(a) Show that if Σ is diagonal, then the individual components $x_1, x_2, \ldots, x_n$ are independent random variables.
(b) Show that the individual components x_i of x are univariate Gaussians. Therefore $(x_1, \ldots, x_n)$ jointly Gaussian implies that the marginals are Gaussian.
(c) However, x_1, x_2 being individually Gaussian does not imply that (x_1, x_2) is jointly Gaussian. Give an example.
(d) Compute the mean and covariance of AX where $A \in \mathbb{R}^{n\times n}$.

Remark: This is a standard exercise involving multivariate Gaussians - you can google it up. Multivariate Gaussians are very important - we will revisit them next semester when discussing Kalman filters.

5. Suppose $p(x)$ and $q(x)$ are probability density functions (pdfs) with cumulative distribution function $P(x)$ and $Q(x)$ respectively. Suppose $P(a) \leq Q(a)$ for all a. Prove for any increasing function $f(\cdot)$ that

$$\int_{\mathbb{R}} f(x)p(x)dx \geq \int_{\mathbb{R}} f(x)q(x)dx$$

This result is important because it gives you a way of ordering pdfs - since the expected value wrt $p(x)$ is greater than that wrt $q(x)$ for any increasing function, we can say that in some sense that the pdf $p(x)$ dominates $q(x)$. In a more advanced course, this is called first order stochastic dominance - and plays a major role in economics/finance (quantifying risk), detection theory, etc.

Give an example of two Gaussian densities that satisfy the above property.

Remark: This result follows directly using integration by parts.

---

6. Show that if Θ uniformly distributed on $(0, 2\pi)$ and $X = \cos(\Theta)$, $Y = \sin(\Theta)$, then although X and Y are dependent, $\text{Cov}(X, Y) = 0$. (Therefore independence implies uncorrelated but not vice-versa).

Soln: $\mathbf{E}\{Y\} = \frac{1}{2\pi}\int_0^{2\pi} \sin\theta d\theta = 0.$

$$\begin{aligned}\operatorname{Cov}(X,Y) = \mathbf{E}\{(X-\mathbf{E}\{X\})(Y-\mathbf{E}\{Y\})\} = \mathbf{E}(\cos\Theta\sin\Theta) &= \frac{1}{2\pi}\int_0^{2\pi}\cos\theta\sin\theta d\theta \\ &= \frac{1}{4\pi}\int_0^{2\pi}\sin(2\theta)d\theta = -\frac{1}{4\pi}\left.\frac{\cos(2\theta)}{2}\right|_{2\pi}^{0} = 0.\end{aligned}$$

7. Suppose X has a probability density

$$f(x) = ce^x, \qquad 0 < x < 1$$

Determine $\operatorname{Var}(X)$.
Soln: Clearly $c = 1/(e-1)$ for $\int_0^1 f(x)dx = 1$. Next using integration by parts $\int_0^1 xe^x dx - 1$ and $\int_0^1 x^2 e^x dx = e - 2$. So $\operatorname{Var}(X) = E\{X^2\} - (E\{X\})^2 = \frac{e^2-3e+1}{(e-1)^2}$

8. Consider the random vector $X(\omega) = [X_1(\omega), X_2(\omega)]' \in \mathbb{R}^2$. Suppose it has pdf

$$f_{X_1,X_2}(x_1,x_2) = ax_1 e^{-ax_1^2/2}\, bx_2 e^{-bx_2^2/2}, \qquad x_1 > 0, x_2 > 0, a > 0, b > 0$$

   (a) Find the joint cdf (cumulative distribution function).
   (b) Find the marginal pdfs of X_1 and X_2.

Soln: For $x_1 > 0$, $x_2 > 0$,

$$\begin{aligned} F_X(x_1,x_2) &= \int_0^{x_1}\int_0^{x_2} ax_1 e^{-ax_1^2/2} bx_2 e^{-bx_2^2/2} dx_1 dx_2 \\ &= (1-e^{-ax_1^2/2})(1-e^{-bx_2^2/2}) \\ F_{X_1}(x_1) &= \lim_{x_2\to\infty} F_{X_1,X_2}(x_1,x_2) = 1 - e^{-ax_1^2/2}, \quad x > 0 \end{aligned}$$

Therefore $f_{X_1}(x_1) = d/dx F_{X_1}(x_1) = ax_1 e^{-ax_1^2/2}$ Similarly $f_{X_2}(x_2) = bx_2 e^{-bx_2^2/2}$.

9. A system has 2 components: A and B. These components have indpt lifetimes that are exponentially distributed with parameters 2 and 3 respectively. (Recall an exponential pdf with parameter λ is $\lambda e^{-\lambda t}$). The system fails as soon as one component fails.

   (a) What is the mean time to failure for component A? *Soln*: 1/2
   (b) What is the mean time to failure for component B? *Soln*: 1/3
   (c) What is the mean time to failure for the system? *Soln*: 1/5
   Hint: Define $Z = \min(X,Y)$. Then cdf of Z is $F_Z(z) = P(Z \le z) = P(X \le z, Y \ge z) + P(X \ge z, Y \le Z) = F_X(z)(1-F_Y(z)) + F_Y(z)(1-F_X(z))$

10. A linear memoryless system (strictly speaking – affine system) gives an output $Y = aX + b$. Show that if the input X is Gaussian $N(\mu, \sigma^2)$ then Y is also Gaussian. (Moral: A Gaussian input to a linear system gives a Gaussian output. Recall for deterministic linear systems – a complex exponential input gives a complex exponential output).

    Soln:

    Let $a > 0$. Then,

    $$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq (y-b)/a) = F_X(\frac{y-b}{a})$$

    So $f_Y(y) = \frac{1}{a} f_X(\frac{y-b}{a}) = \exp\left(-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}\right)$, i.e., $Y \sim N(a\mu + b, a^2\sigma^2)$

Stochastic Simulation

11. Suppose $X \in \{-3, -1, 1, 3\}$ is a random variable with pmf $f_X(-3) = 0.3$, $f_X(-1) = 0.2$, $f_X(1) = 0.3$, $f_X(3) = 0.2$.

    (a) Suppose $Y = X^2$. Compute the pmf of Y.

    (b) Compute $\mathbf{E}\{Y\}$ from the pmf of Y. Compare this with $\mathbf{E}\{Y\}$ compute directly from the pmf of X.

    (c) Generate using the inverse transform method a 400 point iid sequence of $X[1], \ldots, X[400]$ and $Y[1], \ldots, Y[400]$.

    (d) Compute the time average of the X sequence. Compare this with $\mathbf{E}\{X\}$. Also compare the time average of the Y sequence with $\mathbf{E}\{Y\}$. (This provides a computer simulation verification of the law of large numbers).

    Soln: Clearly $Y \in \{1, 9\}$. So $P(Y = 9) = P(X = -3) + P(X = 3) = 0.5$ and $P(Y = 1) = 0.5$. $\mathbf{E}\{Y\} = 1 \times 0.5 + 9 \times 0.5 = 5$.

12. Show that the inverse transform method works. (see lecture notes).

13. Suppose X is a rv with cdf F. Define the rv $Z = F(X)$. What is the cdf of Z? (If you are unable to solve this, simulate this and check the answer. Then use your brain to solve it).

14. Use the inverse transform method to generate the following random variables. In each case verify your answer in Matlab as follows: Generate 10,000 random samples from the density. Use the `hist` command (suitably scaled) to plot the empirical density function. Check how close this empirical density is to the true density.

    (a) Random variable with density

    $$f_X(x) = e^x/(e-1), \qquad 0 \leq x \leq 1$$

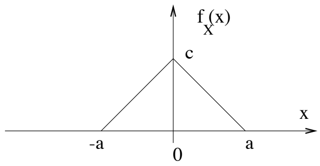
(b) Random variable with distribution

$$F(x) = (x^2 + x)/2, \qquad 0 \le x \le 1$$

(c) Random variable with Weibull distribution

$$F(x) = 1 - \exp(-\alpha x^\beta), \qquad 0 < x < \infty$$

(d) Density $f_X(x)$ is given in the following figure with $c = 2$



Soln: (a) $F(x) = \frac{1}{e-1}\int_0^x \exp(t)dt = \frac{\exp(x)-1}{e-1}, \quad x \in [0,1]$. Set $u = F(x)$ implies $x = F^{-1}(u) = \log(1 + (e-1)u)$ where $u \sim U[0,1]$.
(b): $x = F^{-1}(u) = \frac{-1\pm\sqrt{1+8u}}{2}$. The solution $x = \frac{-1-\sqrt{1+8u}}{2}$ is unacceptable, because $0 \le x \le 1$. So choose $x = F^{-1}(u) = \frac{-1+\sqrt{1+8u}}{2}$.
(c) $F^{-1}(u) = \left(-\frac{\log(1-u)}{\alpha}\right)^{1/\beta}$.
(d) Clearly $a = 1/c$ and $f(x) = \begin{cases} (a+x)/a^2, & \text{if } \; -a \le x \le 0 \\ (1-x)/a^2, & \text{if } \; 0 \le x \le a \end{cases}$.
Next evaluate $F(x)$ for $x \in [-a, 0]$ and $(0, a]$.

$$F(x) = \frac{1}{a^2}\int_{-a}^{x}(a+t)dt = \frac{1}{a^2}\left.(at + t^2/2)\right|_{-a}^{x} = \frac{1}{2a^2}(x+a)^2, \quad -a \le x \le 0,$$

$$F(x) = \frac{1}{a^2}\int_{-a}^{0}(a+t)dt + \frac{1}{a^2}\int_{-a}^{x}(a-t)dt = \frac{1}{2} + \frac{1}{a^2}\left.(at - t^2/2)\right|_0^x = 1 - \frac{1}{2a^2}(x-a)^2, \quad 0 \le x \le a.$$

Inverting yields

$$F^{-1}(u) = \begin{cases} -a + a\sqrt{2u}, & \text{if } \; 0 \le u \le 1/2 \\ a - a\sqrt{2(1-u)}, & \text{if } \; 1/2 \le u \le 1 \end{cases}.$$

15. (a) Describe how to simulate a random variable with Gamma density

$$f(x) = Kx^{1/2}e^{-x}, \quad x \ge 0, \text{ and } K > 0 \text{ is a constant}$$

using the acceptance rejection method, starting with an exponentially distributed random variable.

(b) Illustrate your algorithm in Matlab by plotting an empirical histogram of the cdf and comparing with the Gamma cdf.

(c) Show that for the acceptance rejection method to terminate with the minimum expected number of iterations, the best choice of the mean of the exponential distribution is the same as the mean of the Gamma distribution.

16. Give an algorithm for generating the random variable with pdf

$$f(x) = \begin{cases} e^{2x} & x < 0 \\ e^{-2x} & x \geq 0 \end{cases}$$

17. Devise and implement in Matlab an algorithm for generating a random variable with distribution function

$$F(x) = \begin{cases} \frac{1-e^{-2x}+2x}{3} & \text{if } 0 < x \leq 1 \\ \frac{3-e^{-2x}}{3} & \text{if } 1 < x < \infty \end{cases}$$

18. (a) Give an algorithm to generate a random variable having distribution function

$$F(x) = \int_0^\infty x^y e^{-y} dy, \quad 0 \leq x \leq 1$$

(b) Illustrate your algorithm in Matlab by plotting an empirical histogram of the cdf and comparing with the actual cdf.

19. Using the acceptance rejection method, show how to simulate from a Gaussian random variable starting from an exponential random variable.

20. Suppose it is easy to generate random variables from any of the distributions F_i, $i = 1, \ldots, n$. How can one generate random variables from the distribution $F(x) = \prod_{i=1}^n F_i(x)$.

21. Consider a 2-dim vector random variable with distribution

$$F_{X_1,X_2}(x_1, x_2) = \left(1 - e^{-x_1} - e^{-x_2} + e^{-(x_1+x_2)}\right) \quad x_1 \geq 0, x_2 \geq 0$$

(a) Compute the cdf $F_{X_1}(x_1)$ and pdfs $f_{X_1,X_2}(x_1, x_2)$ and $f_{X_1}(x_1)$.

(b) Devise an algorithm for simulating random numbers from this distribution $F_{X_1,X_2}(x_1, x_2)$. You need to justify why your algorithm is correct. Implement this in Matlab and show that your algorithm is correct by generating several random numbers, histograming them and comparing this histogram with the actual CDF.

22. Consider the following bivariate density:

$$f_{X,Y}(x,y) = \begin{cases} 2e^{-x}e^{-y} & 0 \le y \le x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

   (a) Compute the marginals and show that X and Y are dependent.
   (b) Devise an algorithm for simulating random numbers from this bivariate density.
   (c) Use the simulation algorithm to compute by Monte-Carlo simulation the following integral:

$$\int_0^\infty \int_0^x \sin(x+y)\,\cos(x-y)\,e^{-x}\,e^{-y}\,dy\,dx$$

23. Use Monte-Carlo integration to evaluate $\int_0^1 \frac{4}{1+x^2}dx$

24. *Randomized algorithm for multiplying two vectors.* Consider evaluating the inner product of two vectors $x \in \mathbb{R}^N$ and $p \in \mathbb{R}^N$:

$$S = \sum_{i=1}^N x_i p_i = x'p = \mathbf{E}_p\{x\}$$

   Assume p is a probability vector, i.e $p_i \ge 0$ and $\sum_i p_i = 1$.

   - Show that computing S requires N multiplications.
   - Simulate the following randomized algorithm for estimating S. Below $\hat{S}_n$ denotes the estimate of S using n random samples:
     Step 1: For $k = 1 : n$, simulate integers $i_k^* \sim p$
     Step 2: Compute the estimate $\hat{S}_n = \frac{1}{n}\sum_{k=1}^n x_{i_k^*}$
     Note that this randomized algorithm only needs n additions to evaluate $\hat{S}_n$. Illustrate via Matlab simulations the performance of this algorithm for various choices of n and N, where N is large.
   - Describe and illustrate via Matlab simulations how you can extend this algorithm to estimate matrix vector products, i.e., computing Ap where $A \in \mathbb{R}^{N\times N}$ and p is a probability vector.

Random processes and Markov chains.

25. A random process $x(t)$ is defined by

$$x(t) = 2\,\cos(2\,\pi\,t + y)$$

   where y is a discrete random variable with $P(y=0) = 1/2$ and $P(y = \pi/2) = 1/2$. Find the mean $m(t)$ at $t = 1$ and the auto-correlation $R(0,1)$.

26. A second order S-state Markov chain is one where $P(X_{k+1}|X_k, X_{k-1} \ldots) = P(X_{k+1}|X_k, X_{k-1})$ where $X_k \in \{q_1, \ldots, q_S\}$. Show that a second order Markov chain can be expressed as a first order Markov chain by enlarging the state space.

27. The following question is on Markov chains.

    (a) Using the inverse transform method write a Matlab program to generate 500 sample points of a 3-state Markov chain with initial prob vector and transition probability matrix

    $$\pi[0] = \begin{bmatrix} 0.1 \\ 0.9 \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.1 & 0.9 & 0 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

    (b) Use the acceptance rejection method to simulate the above Markov chain using Matlab.

    (c) Given the sample path of 500 points, estimate the transition probabilities of the Markov chain.

    (d) Compute the stationary distribution $\pi[\infty]$ of the Markov chain. Verify from the 500 point sample path generated above that the time the Markov chain spends in the three states is proportional to $\pi[\infty]$.

28. For a 2-state Markov chain show mathematically that the state probabilities $\pi[k]$ converge to the stationary distribution geometrically fast in terms of the second largest eigenvalue.

    For a 3-state Markov chain, verify this numerically.

29. Suppose a transition matrix A satisfies $A_{ij}v_i = A_{ji}v_j$. Show that v is the stationary distribution of the Markov chain.

    Using this relationship, given a stationary distribution v, give a procedure for constructing a nontrivial transition matrix (an iid matrix is trivial).

30. Consider a Markov chain with transition matrix P. Let T_{ij} denote the time the process takes for its first entrance to state j given that it started at state i at time 0. We are interested in computing the probability mass function for T_{ij}, namely,

    $$P(x_n = j, x_{n-1} \neq j, \ldots, x_1 \neq j | x_0 = i)$$

    for any time n. Derive an expression for this in terms of the transition matrix.

31. Let X_k be a two state Markov chain with states $-1, 1$ and transition probability matrix

$$A = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}, \quad 0 \leq p \leq 1$$

Derive an expression for the correlation $\mathbf{E}\{X_k X_{k+\Delta}\}$ as $k \to \infty$ where Δ is a positive integer. Your answer should be in terms of p and Δ.

32. A robot moves from cell 1 according to a Markov chain with transition probabilities $1/M$ where M is the number of adjoining cells. There is an object in cell 9 that the robot seeks to pick up. However, if the robot enters cell 7 then it gets stuck and can never exit. Let x_n denote the position of the robot at time n. Compute the transition matrix of the process x_n. How can you use the Problem 32 to determine the probability that the robot reaches the target at time 10?

1 (robot starts)	2	3
4	5	6
7 (absorbing state)	8	9 (object)

33. This problem shows how stochastic simulation can be used for efficient matrix vector multiplication. Recall the Chapman Kolmogorov equation $\pi[k+1] = A'\pi[k]$ where A is the transition matrix. For a N state Markov chain, evaluating the matrix vector product requires N^2 multiplications which can be expensive for large N.

Consider instead the following simulation based estimator for $A'\pi[k]$.

For $m = 1 : M$,

  Simulate integer i_m from pdf $\pi[k]$.

end for

Then simply estimate the product $A'\pi[k]$ as

$$\hat{\pi}_{k+1} = \frac{1}{M} \sum_{m=1}^{M} A'(i_m, :)$$

where $A(i_m, :)$ denotes the i_mth row of the matrix A.

Illustrate using Matlab how this estimator performs for $N = 1000$ by choosing various values of M in the range $M = 100$ to $M = 10000$. Plot a graph of variance of the estimator versus M vs run-time for the program. Compare vs the run-time for multiplying $A'\pi[k]$

34. Collect stock market data to show that the log return over a fixed period of time is normal; implying that the return is lognormal. Use

one of the "frequentist tests" in `https://en.wikipedia.org/wiki/Normality_test` to verify how close to normal the log return is.

State Space Models and Least Squares Inference

35. Consider the discrete time system with transfer function

$$\frac{Y(z^{-1})}{U(z^{-1})} = \frac{0.5z^{-1} + 08z^{-2}}{1 + 0.8z^{-1} + z^{-2}}$$

Suppose the input u_k is zero mean iid noise with variance 1.

Express the above model in the controller canonical state space form.

Compute the predicted covariance of the state at time infinity. How would you solve for this asymptotic covariance algebraically?

36. Consider the scalar stochastic system

$$x_{k+1} = a(s_k)x_k + w_k$$

where w_k is zero mean iid noise with unit variance, and s_k is a 2 -state Markov chain with transition probability matrix A. Note $a(1)$ and $a(2)$ are constants depending on the state of the Markov chain (state 1 or state 2).

Given the initial mean m_0 of x_0 and initial probability distribution $\pi[0]$ of the Markov chain, compute the mean of the x_k at time k. Simulate this and examine the effect of the parameter $a(1), a(2)$ on the evolution of the mean. In particular, is it possible to choose $a(1) > 1$ and still have the mean converge to zero as $k \to \infty$? Check this via Matlab.

37. Consider an AR 2 model

$$y_k = -a_1 y_{k-1} - a_2 y_{k-2} + w_k$$

where w_k is iid zero mean Gaussian noise with variance 1.

For what values of a_1, a_2 is the above model asymptotically stationary? Plot these values on graph with horizontal axis a_1 and vertical axis a_2. (You should obtain the so called stability triangle).

38. For the above model, simulate the observations for such a model in Matlab and simulate the recursive least square (RLS) estimator for a_1 and a_2 in Matlab. You are free to choose a_1 and a_2 as you like. Examine the effect of choosing different initializations for your RLS algorithm.

39. Consider the above model but suppose you did not know that the dimension of the AR model was 2. Use the Akaike information criterion to estimate the model order. Illustrate this using Matlab simulations.

Also illustrate the performance of the Bayesian information criterion (BIC) for model order estimation. (BIC was not covered in class, but you can read about it from several sources).

40. Consider the deterministic least absolute value estimator (instead of a least squares estimator):

$$\min \sum_{k=1}^{50} |y_k - b_1 u_{k-1} - b_2 u_{k-2}|$$

for the parameters b_1 and b_2. Devise and simulate in Matlab a scheme for estimating b_1 and b_2. You can choose b_1 and b_2 as you like (but obviously you are not allowed to choose $b_1 = b_2 = 0$!).

41. Suppose you are fitting the model

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + w_k$$

where w_k $N(0, \sigma^2)$ to a data set of 100 points. Derive the least squares estimate subject to the constraint that $a_1 + a_2 = 1$. Simulate this constrained least squares estimate in Matlab and show how the noise variance σ_w^2 affects the estimates.

42. Illustrate the derivation of the recursive least squares algorithm for a scalar parameter.

43. **Internal Combustion Engine Model Estimation**. Read and implement the example in `https://www.mathworks.com/help/ident/examples/online-recursive-least-squares-estimation.html` Using recursive least squares, estimate the parameters of the nonlinear model for an internal combustion engine and also detect changes in the engine inertia.

44. **Echo cancellation**: Consider the setup where the signal input to a microphone is

$$y_k = \psi_k' \theta + v_k$$

where v_k is iid noise (in the absence of near end speech),

$$\psi_k = [x_{k-1}, \ldots, x_{k-L}]'$$

is the vector containing the L most recent time samples and θ is the impulse response from loudspeaker to microphone. The vector θ is not known and needs to be estimated to cancel the echo signal $\psi_k' \theta$. Simulate such as system in Matlab and compare RLS with LMS in performance.

45. Consider the perturbed linear system of equations

$$(A + \Delta A)(\Delta x + x) = b + \Delta b$$

It can be shown that if $|A^{-1}\| \|\Delta A\| < 1$, then the sensitivity of the solution is

$$\frac{\|\Delta x\|}{\|x\|} \leq \frac{c(A)}{1 - c(A)\frac{\|\Delta A\|}{\|A\|}} \left[\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|}{\|b\|} \right]$$

where $c(A) = \max \sigma(A) / \min \sigma(A)$ is the condition number of the matrix. Using the above result, illustrate via Matlab simulation the sensitivity of the constrained least squares estimator as the number of constraints is increased.

46. Consider the stochastic least squares problem $Y = \Psi\theta + \epsilon$ where ϵ is zero mean noise vector with covariance $\sigma^2 I$, $\theta \in \mathbb{R}^n$ and $Y \in \mathbb{R}^N$. Let θ_* denote the least squares estimator. Show that the following estimate for the noise variance σ^2 is unbiased:

$$\hat{\sigma^2} = \frac{1}{N - n}(Y - \Psi\theta_*)'(Y - \Psi\theta_*)$$

Illustrate via Matlab simulation how this variance estimate is affected when constraints of the form $A\theta = 0$ are added to the least squares formulation.

47. Describe how the PCA algorithm can be used for face classification. Your description should include actual face data downloaded from the internet along with Matlab simulations illustrating how well the algorithm performs.

48. Describe how the PCA algorithm can be used for image compression. Your description should include actual images (e.g. .jpg file) downloaded from the internet along with Matlab simulations illustrating how well the algorithm performs.

```
A=imread('image.jpg');
X=double(rgb2gray(A)); % convert RBG to gray, 256 bit to double.
nx = size(X,1); ny = size(X,2);
```

Problems for Part 3 and 4 of the course will be added at a later date.