EEE4114F EXAM
DIGITAL SIGNAL PROCESSING

University of Cape Town
Department of Electrical Engineering

June 2018
3 hours

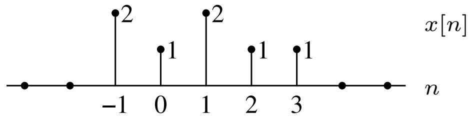
Information

- The exam is closed-book.
- There are two parts to this exam.
- **Part A** has *seven* questions totalling 65 marks. You must answer all of them.
- **Part B** has *25* questions totalling 42 marks. You must answer all of them, although there are options in some instances. Full marks for part B is 35 marks.
- Part A must be answered in an exam book. Part B must be answered on the question paper.
- A table of standard Fourier transform and z-transform pairs appears at the end of this paper.
- You have 3 hours.

PART A

Digital signal processing

1. Suppose $x[n]$ is the signal below:



   Sketch the following signals:

   (a) $y_1[n] = -2x[n+1]$

   (b) $y_2[n] = x[-2n+1]$

   (c) $y_3[n] = x[n] * \delta[2n]$

   (d) $y_4[n] = \delta[n-1] * x[-n]$

   (e) $y_5[n] = \sum_{k=-\infty}^{\infty} \delta[n-2]x[k]$.

(10 marks)

2. Consider the discrete LTI system represented by

$$y[n] = y[n-1] + x[n]$$

   where the system is causal and initially at rest, and $x[n]$ and $y[n]$ are the input and output respectively.

   (a) Determine and plot the impulse response $h[n]$. Is the system stable?

   (b) Determine and plot the step response $y[n]$ when $x[n] = u[n]$.

(10 marks)

3. An LTI system is described by the input-output relation

$$y[n] = x[n] + 2x[n-1] + x[n-2].$$

(a) Determine the impulse response $h[n]$. Is this a stable system?

(b) Show that the frequency response of the system can be written as

$$H(e^{j\omega}) = 2e^{-j\omega}(\cos(\omega) + 1).$$

(c) Plot the magnitude and phase of $H(e^{j\omega})$

(d) Now consider a new system whose frequency response is $H_1(e^{j\omega}) = H(e^{j(\omega+\pi)})$. Determine $h_1[n]$, the impulse response of the new system.

(10 marks)

---

4. Consider the following signals:

$$x_1[n] = \delta[n] + 2\delta[n-1] + 3\delta[n-2]$$

and

$$x_2[n] = 2\delta[n] + \delta[n-1] + \delta[n-3].$$

(a) Find the linear convolution of $x_1[n]$ with $x_2[n]$.

(b) Find the 4-point circular convolution of $x_1[n]$ with $x_2[n]$.

(c) Suppose $X_2[k]$ is the 4-point DFT of $x_2[n]$. Find the value $X_2[2]$.

(d) If $x_2[n]$ consists of samples of an analog signal with sampling period $T = 0.1s$, then what frequency does $X_2[2]$ in part (c) correspond to?

(e) Describe in detail how you would use an FFT to calculate the linear convolution of $x_1[n]$ with $x_2[n]$.

(10 marks)

---

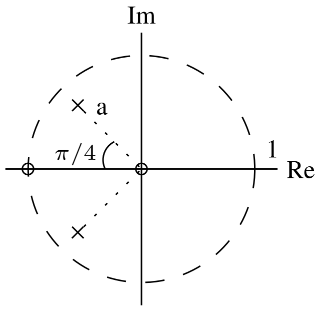
5. Find $w[n] = x[n] * y[n]$ with

$$x[n] = \cos(\pi n/2) \quad \text{and} \quad y[n] = \frac{\sin(\pi(n-5)/3)}{\pi(n-5)}.$$

(10 marks)

---

6. The following figure shows the pole-zero plot of a system with two poles and two zeros:



The poles are a distance $a = 1/\sqrt{2}$ from the origin.

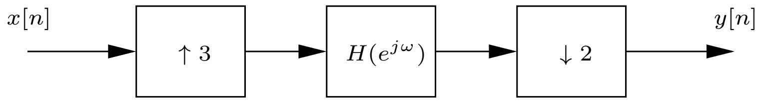
(a) Determine the transfer function $H(z)$ describing this system assuming that it has a DC gain of one.

(b) Sketch the magnitude of the frequency response of the system.

(10 marks)

---

7. Consider the system below:

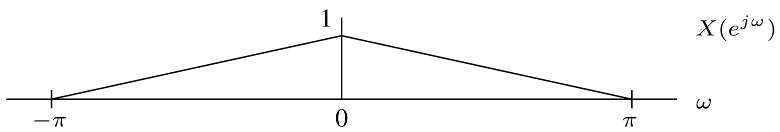


Note that the upsamping and downsampling stages do not contain any filtering.

(a) Suppose that

$$H(e^{j\omega}) = \begin{cases} 1 & |\omega| \leq \frac{\pi}{2} \\ 0 & \frac{\pi}{2} < |\omega| \leq \pi. \end{cases}$$

Find and plot $Y(e^{j\omega})$ if $X(e^{j\omega})$ is as follows:



(b) What is the overall change in sampling rate of the system?

(5 marks)

Discrete-time Fourier transform properties

Sequences $x[n], y[n]$	Transforms $X(e^{j\omega}), Y(e^{j\omega})$	Property
$ax[n] + by[n]$	$aX(e^{j\omega}) + bY(e^{j\omega})$	Linearity
$x[n - n_d]$	$e^{-j\omega n_d} X(e^{j\omega})$	Time shift
$e^{j\omega_0 n} x[n]$	$X(e^{j(\omega-\omega_0)})$	Frequency shift
$x[-n]$	$X(e^{-j\omega})$	Time reversal
$nx[n]$	$j \frac{dX(e^{j\omega})}{d\omega}$	Frequency diff.
$x[n] * y[n]$	$X(e^{-j\omega})Y(e^{-j\omega})$	Convolution
$x[n]y[n]$	$\frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\theta})Y(e^{j(\omega-\theta)}) d\theta$	Modulation

Common discrete-time Fourier transform pairs

Sequence	Fourier transform
$\delta[n]$	1
$\delta[n - n_0]$	$e^{-j\omega n_0}$
$1 \quad (-\infty < n < \infty)$	$\sum_{k=-\infty}^{\infty} 2\pi\delta(\omega + 2\pi k)$
$a^n u[n] \quad (\|a\| < 1)$	$\frac{1}{1-ae^{-j\omega}}$
$u[n]$	$\frac{1}{1-e^{-j\omega}} + \sum_{k=-\infty}^{\infty} \pi\delta(\omega + 2\pi k)$
$(n+1)a^n u[n] \quad (\|a\| < 1)$	$\frac{1}{(1-ae^{-j\omega})^2}$
$\frac{\sin(\omega_c n)}{\pi n}$	$X(e^{j\omega}) = \begin{cases} 1 & \|\omega\| < \omega_c \\ 0 & \omega_c < \|\omega\| \leq \pi \end{cases}$
$x[n] = \begin{cases} 1 & 0 \leq n \leq M \\ 0 & \text{otherwise} \end{cases}$	$\frac{\sin[\omega(M+1)/2]}{\sin(\omega/2)} e^{-j\omega M/2}$
$e^{j\omega_0 n}$	$\sum_{k=-\infty}^{\infty} 2\pi\delta(\omega - \omega_0 + 2\pi k)$

Z-transform properties

Sequences $x[n], y[n]$	Transforms $X(z), Y(z)$	ROC	Property
$ax[n] + by[n]$	$aX(z) + bY(z)$	ROC contains $R_x \cap R_y$	Linearity
$x[n - n_d]$	$z^{-n_d} X(z)$	ROC $= R_x$	Time shift
$z_0^n x[n]$	$X(z/z_0)$	ROC $= \|z_0\| R_x$	Frequency scale
$x^*[-n]$	$X^*(1/z^*)$	ROC $= \frac{1}{R_x}$	Time reversal
$nx[n]$	$-z \frac{dX(z)}{dz}$	ROC $= R_x$	Frequency diff.
$x[n] * y[n]$	$X(z)Y(z)$	ROC contains $R_x \cap R_y$	Convolution
$x^*[n]$	$X^*(z^*)$	ROC $= R_x$	Conjugation

Common z-transform pairs

Sequence	Transform	ROC
$\delta[n]$	1	All z
$u[n]$	$\frac{1}{1-z^{-1}}$	$\|z\| > 1$
$-u[-n-1]$	$\frac{1}{1-z^{-1}}$	$\|z\| < 1$
$\delta[n - m]$	z^{-m}	All z except 0 or ∞
$a^n u[n]$	$\frac{1}{1-az^{-1}}$	$\|z\| > \|a\|$
$-a^n u[-n-1]$	$\frac{1}{1-az^{-1}}$	$\|z\| < \|a\|$
$na^n u[n]$	$\frac{az^{-1}}{(1-az^{-1})^2}$	$\|z\| > \|a\|$
$-na^n u[-n-1]$	$\frac{az^{-1}}{(1-az^{-1})^2}$	$\|z\| < \|a\|$
$\begin{cases} a^n & 0 \leq n \leq N-1, \\ 0 & \text{otherwise} \end{cases}$	$\frac{1-a^N z^{-N}}{1-az^{-1}}$	$\|z\| > 0$
$\cos(\omega_0 n)u[n]$	$\frac{1-\cos(\omega_0)z^{-1}}{1-2\cos(\omega_0)z^{-1}+z^{-2}}$	$\|z\| > 1$
$r^n \cos(\omega_0 n)u[n]$	$\frac{1-r\cos(\omega_0)z^{-1}}{1-2r\cos(\omega_0)z^{-1}+r^2 z^{-2}}$	$\|z\| > \|r\|$

EEE4114F EXAM
DIGITAL SIGNAL PROCESSING

University of Cape Town
Department of Electrical Engineering
June 2018

PART B: Computational intelligence

All the questions may be attempted. Full marks can be obtained on about 80% of the paper. Answer questions directly on the question paper. If you need more space, use the opposing blank page. Answers must be specific, but can be very concise, often only a sentence, phrase or mathematical expression. Algebraic notation can be augmented by MATLAB-style notation, where this improves clarity.

Available marks 42, full marks 35

Peoplesoft ID:

Student number:

1. Typically, Euclidean distance is used as a measure of similarity/difference between instances represented in a vector space, but in certain cases this fails to capture the required notion of 'nearness', and an alternative measure is preferred. Explain, with an example, and state what measure is used.

(1 mark)

2. In an inductive learning system intended for weather prediction, two of the feature variables are: (i) a character symbol $\{\mathbf{b}, \mathbf{m}, \mathbf{c}, \mathbf{o}\}$, representing 4 mutually-exclusive states $\{$**b**lue-sky, **m**isty, **c**loudy, **o**vercast$\}$, and (ii) a number (0–359) representing wind direction in degrees. Show how these two variables can be transformed into a form suitable for use as inputs to the system.

(2 marks)

3. In the case of a proximity-based classifier such as k-nearest neighbour (kNN), the predictive accuracy can be estimated rapidly, without any need for repeated train/test cycles. Explain how this is done.

(1 mark)

4. In the 'adaptive metric' version of the 1-NN classifier, the distance between the query point $\mathbf{q}$ and a training-data point $\mathbf{x}_i$ is modified in such a way that "typical" or "reliable" data points have greater influence than more marginal ones in predicting the class of the query point. Explain how this is done.

(1 mark)

5. A data subset in a tree-based classifier contains n_A instances from class A, and n_B from class B. Write down an expression for the entropy H of the subset. State how H can be used in the construction of a classification tree.

(1 mark)

6. Suppose a particular data partition, with n instances has entropy H. It is split into two sub-partitions with n_1 and n_2 instances, and corresponding entropy values H_1 and H_2. Write down an expression for the resultant Information Gain (IG). State the purpose for which IG is used in the construction of a classification tree.

(1 mark)

7. Splitting the data on successive variables will eventually result in a situation in which classes are perfectly separated. Explain why this is not usually a desirable outcome, and state how it can be avoided.

(1 mark)

8. In the random forest classifier, generalisation is improved by averaging classification results over multiple trees. How can we ensure diversity in the individual trees?

(1 mark)

9. An advantage of tree-based classification is that it is possible to obtain a ’confidence factor’ or estimation of posterior probability of class membership. Explain how this is done.

(1 mark)

10. Show how Bayes' Theorem follows from the axioms of probability for dependent events.

(1 mark)

11. A collection of 100 objects contains 60 from class A and 40 from class B. A certain identifying feature f occurs in 15% of those from class A but in only 5% of those from class B.

   (a) What is the probability that an object selected at random displays the feature f?

   (b) What is the probability that an object selected at random, and found to display the identifying feature, belongs to class A?

(2 marks)

12. Certain words, w_1, w_2, w_3 are known to occur in spam emails with a frequency greater than that of their general occurrence. Examination of a large random sample of N emails yields counts of n_1, n_2, n_3 respectively for these three words. S of the emails are then identified as spam emails, and a word-count on those yields counts s_1, s_2, s_3 of those same words. Write down expressions in $N, S, n_1, n_2, n_3, s_1, s_2$ and s_3 for

   (a) the estimated probability that, in a randomly-selected email, all three words will appear.

   (b) the estimated probability that an email containing all three words is spam.

   (c) What is the assumption on which the calculations in a) and b) above are based?

   (d) Describe how a work-around (sometimes called 'Laplace smoothing') is used to allow an estimate of $P(w_1\ w_2\ w_3)$ to be obtained in the event that a particular word (say w_1) happens not to occur in a particular sampling.

(4 marks)

13. Describe a method that can be used, in the case of continuous high-dimensional data, to estimate empirically the relative class-related distribution of continuous data in the form of a prior-weighted probability density function (i.e. to estimate $P_A \cdot PDF_A(\mathbf{x})$ where A is a class label and $\mathbf{x}$ an input vector).

(1 mark)

14. Given $P_A \cdot PDF_A(\mathbf{x})$ and $P_B \cdot PDF_B(\mathbf{x})$ for class A and B respectively, write down an expression for $P(A|\mathbf{q})$ (i.e. the posterior probability of membership of class A if $\mathbf{x} = \mathbf{q}$.

(1 mark)

15. Give an expression for a Gaussian kernel $k(x_i, x_j)$ as a function of an input vector x_i, centred at x_j, and with a 'width' σ.

(1 mark)

16. For Gaussian Kernel Machines such as the Probabilistic Neural Network (PNN) and the Generalised Regression Neural Network (GRNN), describe a simple way of determining the leave-one-out (LOO) error without iterative testing and training.

(1 mark)

17. Given training data expressed in the usual way as $\{X, \mathbf{t}\}$ we can fit a linear model $H\mathbf{w} = \mathbf{t}$ where $H = [\mathbf{1}\ X]$.

    (a) Show how $\mathbf{w}$ can be calculated by singular value decomposition (SVD) of H (unregularised).

    (b) Write down an expression for the 'test error' or vector of residuals $\mathbf{r} = \mathbf{t} - \mathbf{t}_{\text{pred}}$ in terms of $\mathbf{w}$, H and $\mathbf{t}$.

    (c) The above process minimises $\sum r^2$. How is this changed if regularisation is used (with ridge factor λ)?

    (d) Show how the use of SVD facilitates the regularisation using multiple values for λ.

(4 marks)

18. Give an expression for the 'hat' matrix, H_{hat} and show how it can be used for rapid evaluation of leave-one-out (LOO) error.

(2 marks)

19. (a) Give an expression for the design matrix H in the case of the Extreme Learning Machine (ELM), where $\mathbf{1}_p$ is a unit column vector of length p (the number of training samples), and R is a random matrix with elements in the range -1 to 1. Indicate the dimensionality of R, assuming that the number of hidden neurons is h.

    (b) Why does the ELM lend itself to use in an ensemble or 'committee machine'?

(2 marks)

20. Averaging over the outputs of a full ensemble of learning machines may not necessarily yield the best possible final result. State the general properties required of a 'good' candidate machine for inclusion in an ensemble.

(1 mark)

21. (a) Suppose we perform multiple runs of ELM with different sets of random weights. The ith run produces, on the training data, a mean-square error $e^{(i)}$ and a vector of predicted training targets $t^{(i)}$. Give an expression for quantifying the relative diversity $d^{(i)}$ of the ith ELM's output.

(b) Show how $r^{(i)}$ and $d^{(i)}$ can be used to select promising candidate ELMs from a larger pool for inclusion in a committee machine.

(2 marks)

22. Do only **EITHER** part (a) **OR** part (b):

(a) Given a current trial solution $\mathbf{x}_i$, and three other randomly-selected trial vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$, indicate how a new trial solution is generated in Differential Evolution. How does this differ from the creation of new trial solutions in other real-valued stochastic search algorithms (such as the Breeder Genetic Algorithm)? Explain why this difference often results in improved search performance.

(b) The Population Vector (PV) in PBIL is initialised so that all elements have the value 0.5. Trial solutions are generated by randomly sampling PV to generate a population of binary masks $\mathbf{m}$, which are evaluated in the context of the optimisation task in hand. In each population a 'best' trial solution $\mathbf{m}_{\text{best}}$ is identified. Show how the PV is updated in the next generation to favour trial solutions that resemble $\mathbf{m}_{\text{best}}$, and how a further adjustment of PV is needed to prevent premature convergence.

(3 marks)

23. Explain how Singular Value Decomposition (SVD) can sometimes be used to allow the visualisation of the class structure in a high dimensional data set $\{X, \mathbf{t}\}$.

(1 mark)

24. Do only **EITHER** part (a) **OR** part (b):

(a) Describe the structure of the ELM autoencoder and show how it can be used to produce an enhanced feature set for supervised learning or clustering.

(b) Explain with the aid of a diagram how a contractive auto-encoder can be used for data dimensionality reduction and novelty detection.

(3 marks)

25. Do only **EITHER** part (a) **OR** part (b):

(a) Describe briefly the k-means clustering algorithm, and suggest either a simple way of initialising it which might outperform random initialisation.

(b) Describe briefly the k-means clustering algorithm, and describe briefly how it can be part of a more comprehensive 'global' clustering strategy.

(3 marks)