Name: GERALD GODBOLD

Problem	Points	Score
1(a)	15	10
1(b)	10	9
2(a)	15	15
2(b)	10	5
3(a)	15	10
3(b)	10	10
4	25	25
Total	100	84

Notes:

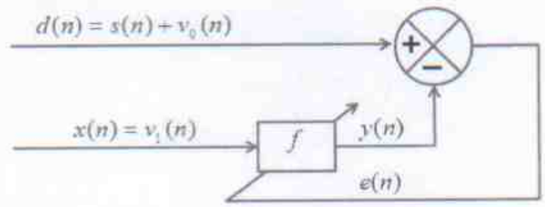
(1) The exam is closed books and notes except for two double-sided sheets of notes.

(2) Please indicate clearly your answer to the problem.

(3) The details of your solutions are more important than the answers. Please explain your solutions clearly and include as many details as possible.

600BOL10

1. The standard adaptive noise canceller is shown to the right.



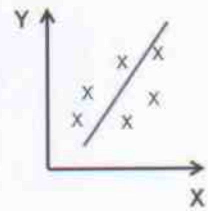
(a) Derive an expression for the optimal filter such that the energy of the noise is minimized.

(b) Suppose instead of this approach you simply computed a linear prediction model on the noisy input signal. Compare and contrast the model you would obtain to the model in (a).

2. Recall our expression for a simple linear regression model that formed the basis for maximum likelihood linear regression.

$$Y = \beta_0 + \beta_1 X$$

$$\beta_0 = \overline{Y} - \beta_1 \overline{X} \quad \beta_1 = \frac{\overline{XY} - (\overline{X})(\overline{Y})}{\overline{X^2} - (\overline{X})^2}$$



(a) Derive the optimal value of the slope.

(b) Describe, in qualitative terms, how you would apply this to the problem of adaptation of the model parameters of a Gaussian mixture model. Discuss the pros and cons of this approach.

3. In this class we introduced the concept of maximum a posteriori (MAP) adaptation.

(a) Derive the MAP estimate of the mean of a single Gaussian distribution assuming the variance is fixed.

(b) Discuss (but do not derive) properties of this estimate, such as bias. Comment on the implications of the resulting equations in terms of application of this technique to a Gaussian model of common time series such as a speech or image signal.

4. In this course, we discussed a range of adaptation topics beginning with the least mean square error (LMS) approaches and ending with approaches based on discriminative training. Describe the course in terms of a tree where the root node is labeled ECE 8423, and all other topics are arranged in a hierarchy representing their relationships with each other. Then provide a glossary: describe the essence of each term represented at each node in a small number of sentences.

Do not feel constrained by the way I presented the course – there is not only one correct answer. Your answers will be judged on their own merits based upon the amount of insight you demonstrate and the completeness of your hierarchy.

$\boxed{1}$ FROM ANC MODEL $\rightarrow$ $e(n) = d(n) - y(n)$

$$e(n) = (s(n) + V_0(n)) - y(n)$$

SIGNAL NOISE FILTER OUTPUT

MEAN ERROR ENERGY $\rightarrow$

$$E\{e^2(n)\} = E\{(s(n) + v_0(n) - y(n))^2\}$$

$$= E\{s^2(n) + (v_0(n) - y(n))^2 + 2s(n)(v_0(n) - y(n))\}$$

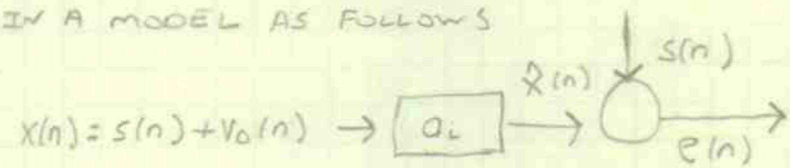
$$= 2E\{s^2(n)\} + E\{(v_0(n) - y(n))^2\} + 2E\{s(n)(v_0(n) - y(n))\}$$

UNCORRELATED

$$\therefore E\{e^2(n)\} = 2E\{s^2(n)\} + E\{(v_0(n) - y(n))^2\}$$

SO, IF $J = E\{e^2(n)\}$, min J YEILDS OPTIMAL SET OF FILTER COEFF'S RESULTING IN MINIMUM NOISE ENERGY. (i.e. $e(n)$ CONVERGES TO $s(n)$)

B) A LINEAR PREDICTOR MODEL JUST BASED ON THE NOISY INPUT SIGNAL, WOULD RESULT IN A MODEL AS FOLLOWS



$$\text{so } \hat{x}(n) = \sum_{i=1}^{p} a_i \left(s(n - n_0 - i) + v_0(n - n_0 - i) \right)$$

* THIS MODEL USES ONLY A WEIGHTED COMBINATION OF THE MEASURED SIGNAL, WITH NO INDEPENDANT MEASURE OF THE NOISE. THE ANC MODEL IS PREFFERED, BECAUSE IT NO PRIOR KNOWLEDGE OF EITHER SIGNAL.

* FOR ANC TO WORK BEST, THE REFERENCE SIGNAL AND DESIRED SIGNAL SHOULD HAVE MINIMAL CORRELATION (i.e. LOW LEAKAGE).

* IN PRACTICE, THIS IS HARD TO DO, MORE LEAKAGE DECREASED SIGNAL TO NOISE $\left(\text{i.e. } \rho_{out}(e^{j\omega}) = 1 / \rho_{ref}(e^{j\omega}) \right)$

* THE LP MODEL WOULD REQUIRE SAMPLE DATA TO BE STATIONARY WITHIN PROCESSING WINDOW. THIS MAY BE DEFFICIENT IF THE NOISE IS HIGHLY UNPREDICTABLE. THE ANC IS BETTER IN THIS RESPECT, BECAUSE NOISE IS SAMPLED AS A REFERENCE.

2) GIVEN MEASUREMENTS $X = \{x_1, x_2 \ldots x_n\}$
OBSERVATIONS $Y = \{y_1, y_2 \ldots y_n\}$
(DATA POINTS)

SLOPE OF OPTIMAL MODEL ⇒ $Y = \beta_0 + \beta_1 X$

SO MINIMIZE THE ERROR ⇒

$$J = e^2(n), \quad \min_{B_0, B_1} J = \min_{B_0, B_1} \left[\sum_{i=1}^{n} \left[y_i - B_0 + B_1 x_i \right]^2 \right]$$

TAKE PARTIALS, SET TO Ø ⇒

$$\frac{\partial J}{\partial B_0} = -\sum_{i=1}^{n} 2\left[y_i - B_0 + B_i x_i \right] = 0$$

$$\frac{\partial J}{\partial B_1} = -\sum_{i=1}^{n} 2\left[y_i - x_i (B_0 + B_i x_i) \right] = 0$$

2 EQ'S
2 UNKNOWNS

SO, $Y = B_0 + B_1 \bar{X}$ WHERE $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$ MEAN

$B_0 = \bar{Y} - B_1 \bar{X}$ WHERE $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ MEAN

$B_1 = \dfrac{\overline{XY} - (\bar{X})(\bar{Y})}{\overline{X^2} - (\bar{X}^2)}$ WHERE $\overline{XY} = \frac{1}{n} \sum_{i=1}^{n} x_i y_i$ CROSS CORR

$\overline{X^2} = \frac{1}{n} \sum_{i=1}^{n} x_i^2$ VARIANCE

B) IF MODEL HAS GAUSSIAN MIXTURE, THAT IMPLIES THAT X IS A RANDOM VARIABLE OF FORM $X \to N(\mu, \sigma^2)$.

* THE ML ESTIMATE WOULD YEILD A β_0 WHICH WOULD BE UNBIASED (i.e. $E\{\beta_0\} = \beta_0$). THE VARIANCE WOULD ALSO BE EQUIVALENT TO THE LMS ESTIMATE. SO, FOR ALL PRACTICAL PURPOSES, THE ML ESTIMATE WOULD BE IDENTICAL.

* DUE TO VARIANCE IN RV X, RESULTS MAY BE SKEWED BY WILD FLUCTUATIONS BECAUSE ALL DATA IS WEIGHTED THE SAME. (i.e. EVEN BAD DATA CONTRIBUTES EQUALLY).

(3) MAP, SINGLE GAUSSIAN, FIXED VARIANCE

X IS RV $\sim N(\mu, \sigma^2)$ $\quad X_i,\ i=1 \dots n$

A-PRIORI DENSITY IS $f(\theta)$. PDF $= f(\theta | X)$.

MAP ESTIMATE IS PEAK VALUE OF DENSITY →

$$\theta_{MAP} = \max_{\theta | x} f(\theta | x) \Rightarrow \frac{\partial}{\partial \theta} \{ f(\theta | x) \}$$

WHERE $f(\theta | x) = \dfrac{f(x|\theta) f(\theta)}{f(x)}$. USE LOG FORM ⇒

$$\theta_{MAP}: \frac{\partial}{\partial \theta} \{ \ln[f(x|\theta)] + \ln[f(\theta)] \} = 0$$

WHERE CONDITIONAL DENSITY $\ln[f(x|\theta)] = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma}(x-\theta)^2$

AND PRIOR IS $\ln[f(\theta)] = -\frac{1}{2}\ln(2\pi\sigma_\theta^2) - \frac{1}{2\sigma_\theta^2}(\theta - \bar{\theta})^2$

b) * THE MAP ESTIMATE θ_{MAP} HAS CONVERGENCE PROPERTIES SUCH THAT AS THE NUMBER OF SAMPLES INCREASES, THE ESTIMATE TRENDS TOWARDS THE MEAN OF THE TRAINING DATA, AND THE SAMPLE MEAN OF THE ADAPTATION DATA

* WITH TIME SERIES DATA, THE PRIOR ESTIMATES CAN BE UPDATED AND MIXED WITH THE ADAPTIVE DATA USING A WEIGHTING FUNCTION.

* IF LARGE VARIATION IN TRAINING DATA, THE MODEL HAS LITTLE WEIGHT, AND ALL WEIGHT ENDS UP IN SAMPLE DATA.

(4)

- ECE 8423
  - CONVENTIONAL ADAPTATION
    - DIRECT METHODS
      - LEAST SQUARES
      - LINEAR PREDICTION
    - ITTERATIVE METHODS
      - LMS ADAPTIVE
        - ECHO CANCELATION
        - ANC
      - LMS VARIANTS
        - VARIBLE ADAPT. TECHNIQUES
        - LEAKY LMS
        - NORMALIZED
        - BLOCK LMS, SIGN-SIGN, ETC...
      - NEWTONS METHOD
        - RLS
        - WEIGHTED RLS
      - IIR & LATTICE
  - MODEL-BASED ADAPTATION
    - BASIC TECHNIQUES
      - MAP
      - MLLR
    - STATE-OF-THE-ART
      - DISCRIMINATIVE TECHNIQUES
        - MMI
        - MLE
        - MPE / MWE
        - SUPERVISED VS. UNSUPERVISED

(4) * CONVENTIONAL ADAPTATION - USE WELL-ESTABLISHED DIRECT AND ITERATIVE METHODS, MANY OF WHICH WERE DEVELOPED BEFORE MODERN COMPUTERS

* LEAST-SQUARES - DEVELOPS SOLUTION TO OPTIMAL FILTER USING THE NORMAL EQUATIONS TO MINIMIZE THE MEAN SQUARE ERROR.

* LINEAR PREDICTION - MODEL IN WHICH FUTURE VALUES OF SIGNAL ARE PREDICTED TO OPTIMIZE FILTER COEFFICIENTS.

* LMS - USES STEEPEST DESCENT TO OPTIMIZE FILTER COEFFICIENTS TO MINIMIZE ERROR ENERGY.

* ECHO-CANCELLATION - PROCESS OF USING A DELAY MODEL IN AN LMS FILTER SUCH THAT OPTIMAL VALUES THE FILTER ELIMINATE THE ECHO IN A TRANSMISSION PATH.

* ANC - ADAPTIVE NOISE CANCELLATION, USES LMS TO ELIMINATE BACKGROUND NOISE (OR OTHER INTERFERENCE) USING A MEASURED REFERENCE SIGNAL.

* LMS VARIANTS - MANY VARIATIONS OF THE LMS SOLUTION, EACH "TWEAKED" FOR VARYING PERFORMANCE CHARACTERISTICS.

* NEWTON'S METHOD - USE DERIVATIVE OF MEAN-SQ-ERROR TO MINIMIZE OVER A NUMBER OF STEPS. MSQE IS A QUADRATIC FUNCTION.

* RLS - RECURSIVE LEAST SQUARES, USE A DIFFERENT APPROXIMATION OF THE DERIVATIVE AND ADAPTIVE STEP SIZE.

* IIR + LATTICE - USES IIR FILTER STRUCTURE IN A RECURSIVE FORM, SUCH THAT COMPUTATIONAL PERFORMANCE IS IMPROVED.

* MAP - MAX A POSTERIOR, USING PROBABILITY MODELS, PARAMETERS ARE OPTIMIZED USING THE PEAK VALUES OF THE ASSOCIATED PROBABILITY MODELS.

* MLLR - MAX LIKELIHOOD LINEAR REGRESSION, USES THE MAX OF THE DENSITY OF DATA, BUT WITH NO PRIOR INFORMATION.

* DISCRIMINATIVE - MMI (MUTUAL INFORMATION), MCE (MIN CLASSIFICATION ERROR), MPE or MWE (MIN PHONE, WORD ERROR) - RANGE OF MODEL-BASED TECHNIQUES THAT DIRECTLY ESTIMATE PROBABILITIES OF EACH CLASS, WITHOUT HAVING ESTIMATE JOINT PROBABILITIES.

* SUPERVISED VS. UNSUPERVISED - COMPARES PERFORMANCE OF BOTH TRAINING TECHNIQUES ON LARGE AMOUNTS OF SPEECH DATA.