Name: GEOFF CARTER

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

Notes:

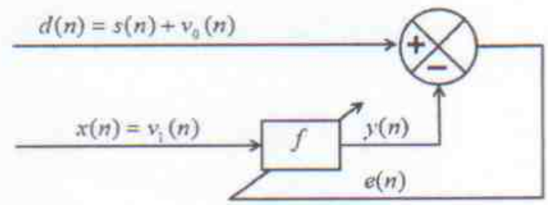
(1) The exam is closed books and notes except for two double-sided sheets of notes.

(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.

2/10

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



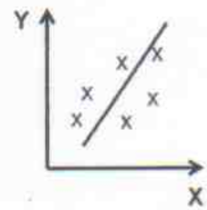
(a) Derive an expression for the optimal filter such that the energy of the noise is minimized.

(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.

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

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



(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.

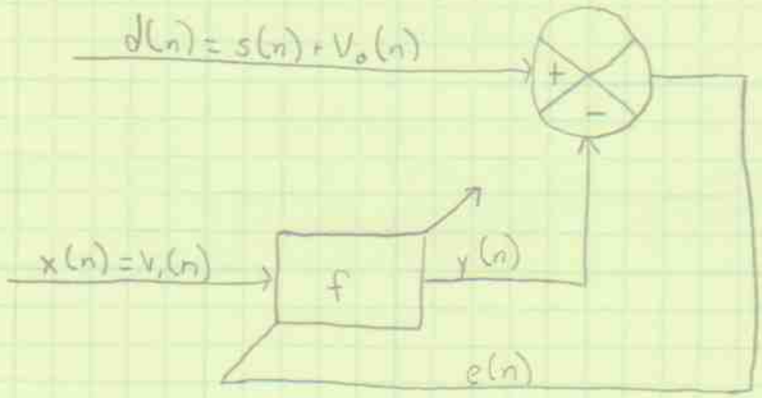
(b) Discuss (but do not derive) properties of this estimate, such as bias. Comment on the implications of the resulting equations in terms of application of this technique to a Gaussian model of common time series such as a speech or image signal.

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.

(1a)

STANDARD ADAPTIVE NOISE CANCELLER



$s(n) \equiv$ SIGNAL

$V_0(n) \equiv$ NOISE

$e(n) \equiv$ ERROR SIGNAL

$V_1(n) \equiv$ NOISE

$x(n) \equiv$ REFERENCE

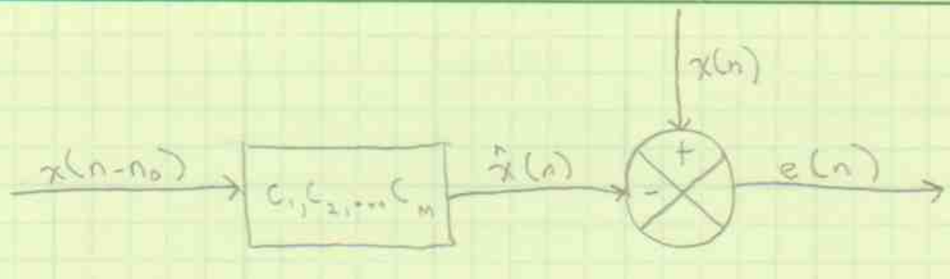
$e(n) = s(n) + V_0(n) - y(n)$

$E[e^2(n)] = E[s^2(n)] + E[(V_0(n) - y(n))^2] + \cancel{2E[s(n)(V_0(n) - y(n))]}$ $\varnothing$

$E[e^2(n)] = E[s^2(n)] + E[(V_0(n) - y(n))^2]$

$\therefore$ $E[e^2(n)]$ IS MIN WHEN $y(n) = V_0(n)$ or $e(n) = s(n)$.

(1b)



LINEAR PREDICTOR

THE ANC DOES NOT USE A PRIORI KNOWLEDGE OF THE SIGNAL, BUT DOES REQUIRE A SECOND INPUT IN ORDER TO GENERATE THE ERROR SIGNAL. THE ERROR SIGNAL IS MOST ACCURATE IF THE SECOND INPUT CONTAINS CORRELATED NOISE WITH NO PRIMARY SIGNAL

THE LINEAR PREDICTOR DOES REQUIRE PREVIOUS KNOWLEDGE OF THE SIGNAL AND THE PREDICTION IS GENERATED FROM A LINEAR COMBINATION OF THESE PREVIOUS VALUES

ECE8423 | 12 DEC, 2008 | GEOFF CARTER | 5/10

(2a) $y = f(x) = \beta_0 + \beta_1 x$

$$L = \sum_{i=1}^{n} \left(Y_i - (\beta_0 + \beta_1 X_i)\right)^2$$

$$\frac{\partial L}{\partial \beta_0} = -\sum_{i=1}^{n} 2\left(Y_i - (\beta_0 + \beta_1 X_i)\right) = 0$$

$$\frac{\partial L}{\partial \beta_1} = -\sum_{i=1}^{n} 2\left(Y_i - (\beta_0 + \beta_1 X_i)\right) X_i = 0$$

KNOWING $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

$$\overline{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$$

$$\overline{Y} = \beta_0 + \beta_1 \overline{X}$$

$$\beta_0 \overline{X} + \beta_1 \overline{X^2} = \overline{XY}$$

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

$$\beta_0 = \overline{Y} - \beta_1 \overline{X}$$

(2b) Knowing that $Y = \beta_0 + \beta_1 x$

Adding Gaussian Noise ϵ we get

$\underline{Y = \beta_0 + \beta_1 x + \epsilon}$ or $Y_i = \beta_0 + \beta_1 + x_i + \epsilon_i$

$$\mathrm{VAR}(\hat{\beta}_1) = \frac{\sigma^2}{n(\bar{x^2} - (\bar{x})^2)}$$

$$\mathrm{VAR}(\hat{\beta}_0) = \left[\frac{1}{n} + \frac{(\bar{x})^2}{n(\bar{x^2} - (\bar{x})^2)}\right]\sigma^2$$

$$\mathrm{COV}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{X}\sigma^2}{n(\bar{x^2} - (\bar{x})^2)}$$

(3A) $$\mu_{MAP} = \text{ARGMAX } p(o|\mu)\, g(\mu) = \frac{1}{\sqrt{2\pi}\,\sigma_m} e^{-\frac{1}{2}\left[\frac{\mu-\mu_m}{\sigma_m}\right]^2} \longrightarrow$$

* I think that for a single Gaussian case N may be equal to 1

$$\longrightarrow \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_v} e^{-\frac{1}{2}\left[\frac{o_i-\mu}{\sigma_v}\right]^2}$$

$$\mu_{MAP} = \text{ARGMIN} \left[\frac{1}{2}\left(\frac{\mu-\mu_m}{\sigma_m}\right)^2 + \sum_{i=1}^{N} \frac{1}{2}\left(\frac{o_i-\mu}{\sigma_v}\right)^2\right]$$

$$\frac{\partial}{\partial \mu}\left[\frac{1}{2}\left(\frac{\mu-\mu_m}{\sigma_m}\right)^2 + \sum_{i=1}^{N} \frac{1}{2}\left(\frac{o_i-\mu}{\sigma_v}\right)^2\right]$$

$$= \left(\frac{1}{\sigma_m^2}\right)(\mu-\mu_m) - \sum_{i=1}^{N}\left(\frac{1}{\sigma_v^2}\right)(o_i-\mu) = 0$$

$$\mu\left(\frac{1}{\sigma_m^2} + \frac{N}{\sigma_v^2}\right) = \frac{\sum_{i=1}^{N} o_i}{\sigma_v^2} + \frac{\mu_m}{\sigma_m^2}$$

$$\mu\left(\frac{\sigma_v^2 + N\sigma_m^2}{\sigma_m^2\sigma_v^2}\right) = \frac{\sum_{i=1}^{N} o_i}{\sigma_v^2} + \frac{\mu_m}{\sigma_m^2}$$

$$\mu = \left[\frac{\sum_{i=1}^{N} o_i}{\sigma_v^2} + \frac{\mu_m}{\sigma_m^2}\right]\left[\frac{\sigma_m^2\sigma_v^2}{\sigma_v^2 + N\sigma_m^2}\right] = \frac{\sigma_m^2\sum_{i=1}^{N} o_i + \sigma_v^2\mu_m}{\sigma_v^2 + N\sigma_m^2}$$

$$\hat{\mu}_{MAP} = \frac{\sigma_m^2\sum_{i=1}^{N} o_i + \sigma_v^2\mu_m}{\sigma_v^2 + N\sigma_m^2}$$

(3b)

(1) As $N \to \infty$ $\hat{\mu}_{MAP} = \frac{1}{N} \sum_{i=1}^{N} o_i$

(2) As $N \to 0$ $\hat{\mu}_{MAP} = \mu_m$

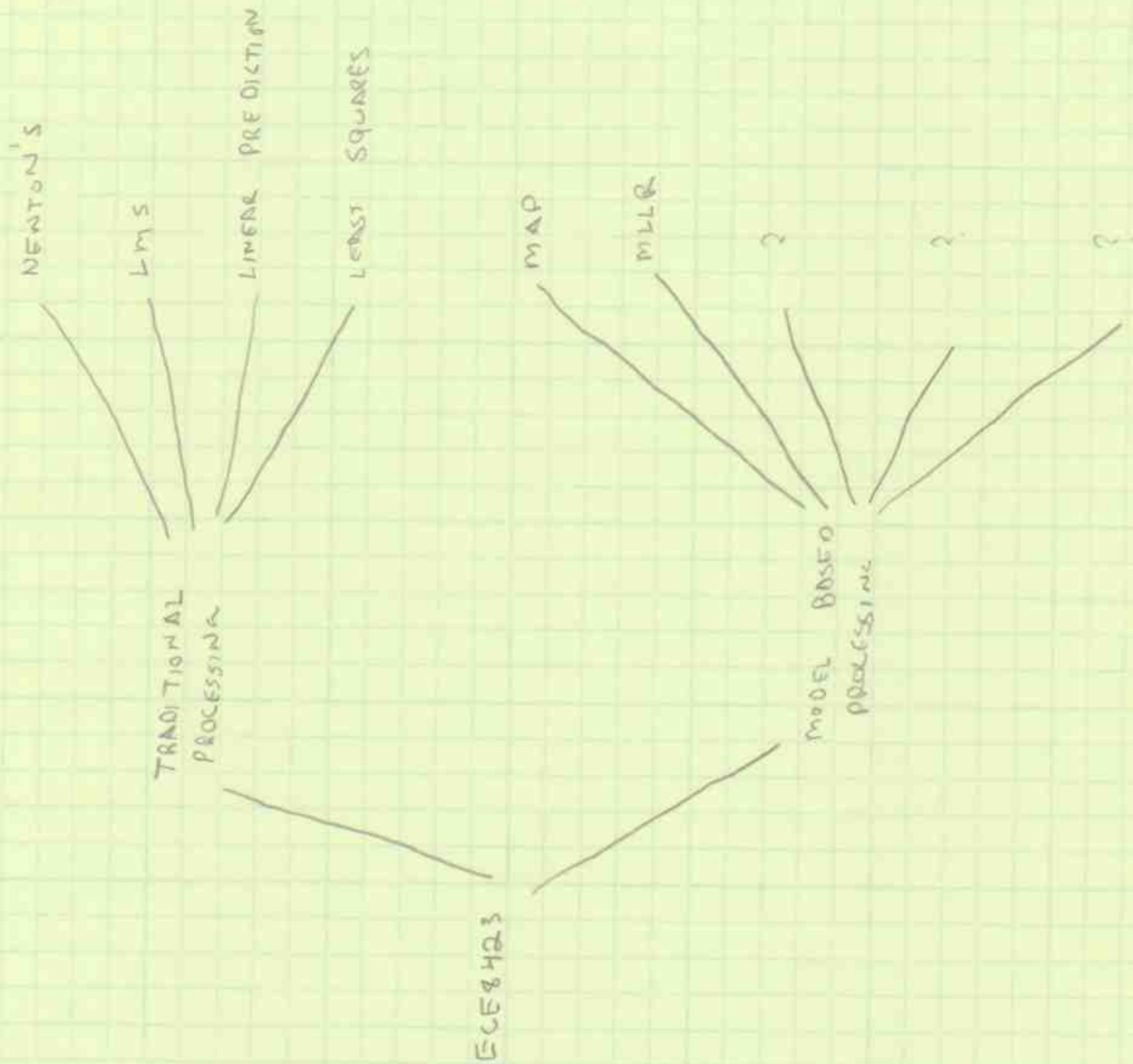
(3) THE ADAPTED MEAN MIXES THE PRIOR & SAMPLE MEAN AND IS WEIGHTED BY THE NUMBER OF SAMPLES IN THE DATA SET.

(4) NON-INFORMATIVE PRIOR $\rightarrow$ $\sigma_m \rightarrow \infty$

ALL VALUES OF THE MEAN EQUALLY WEIGHTED

ECE8423	12 DEC 2008	GEOFF CARTER	9/10

4A



(4B)

NEWTON'S METHOD - SOLUTION IS MINIMIZED USING MSE OVER A FINITE NUMBER OF STEPS.

LMS - SOLUTION IS MINIMIZED USING STEEPEST DESCENT TO OPTIMIZE FILTER PARAMETERS

LINEAR PREDICTION - USES KNOWLEDGE OF ALL PREVIOUS DATA TO PREDICT THE FILTER PARAMETERS

LEAST SQUARES - SOLUTION IS OBTAINED BY MINIMIZING THE MEAN SQUARED ERROR OF THE NORMAL EQUATIONS

MAP - PARAMETERS ARE CALCULATED USING THE OPTIMIZED VALUES OF THE PROBABILITY MODELS.

MLLR - MAXIMUM LIKELIHOOD LINEAR REGRESSION