Name: Chad Williams

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	

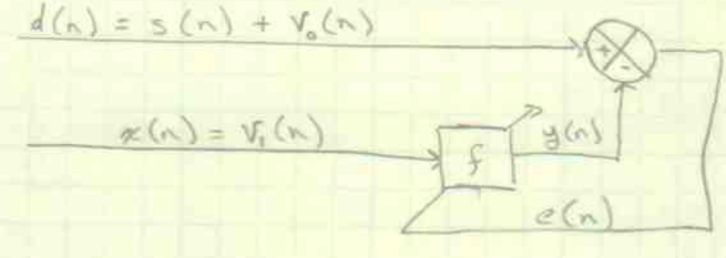
Notes:

(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.

1a)



- We can write the error as:

$$e(n) = s(n) + v_0(n) - y(n)$$

- Square and take the expectation:

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

- This filter assumes that $s(n)$ is uncorrelated with $v_0(n)$ and $v_1(n)$. Also, $y(n) = v_1(n) * f(n)$, so $s(n)$ and $y(n)$ are also uncorrelated. This means the last term in the previous expression is zero, leaving:

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

- clearly $E\{e^2(n)\}$ is minimized when $y(n) = v_0(n)$, resulting in $e(n) = s(n)$.

- In practice the perfect "reference" measurement, $v_1(n)$, is not possible, but it is assumed that the $v_1(n)$ is correlated with the noise component, $v_0(n)$, of the "primary" input signal, $d(n)$.

- So to minimize $e(n)$ we define the objective function as such:

$$J = E\{e^2(n)\}$$

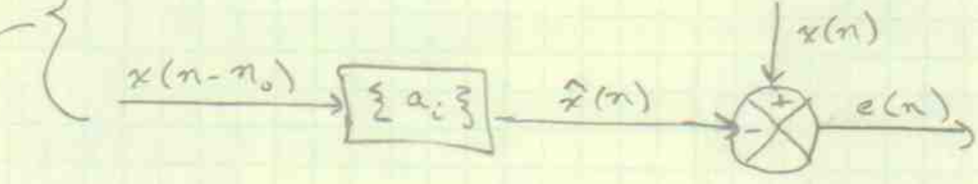
- Then we differentiate J w.r.t. the parameters of f and equate to zero in order to minimize.

- In practice signal leakage can occur between the primary and reference measurements. This is handled assuming a model:

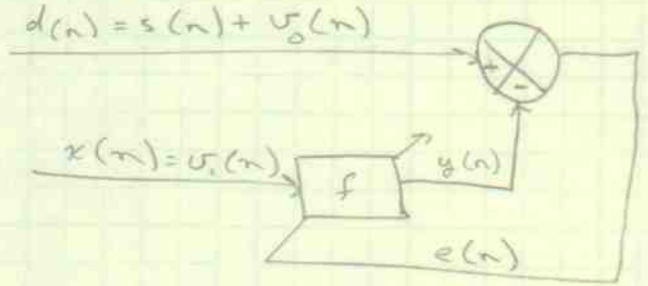
$$d(n) = s(n) + v_0(n)$$
$$x(n) = g(n) * s(n) + v_1(n)$$
$$y(n) = s(n) + v(n)$$

1b)

Linear Prediction Model



- In laymans terms Adaptive Noise Cancellation (ANC) uses a reference measurement of the system noise and adaptively filters that reference to produce a replica of the actual system noise in order to subtract it from the main measurement in the system. No prior knowledge of the signal or noise is required for this method to be applied.



The filter, f, is adapted based on feedback from the system output $e(n)$.

$d(n)$ is the primary measurement, $s(n)$ is the actual signal of interest, $v_0(n)$ is the noise associated with the measurement of $s(n)$, $x(n)$ is the reference measurement that contains noise that is correlated with $v_0(n)$ but not with $s(n)$.

- The linear prediction model uses previous knowledge of the signal to predict future values of the signal. A linear predictor requires inputs to be stationary or at least stationary within the processing frame. For the LP filter coefficients must be recalculated with each processing frame, and frame size must be selected according to the rate at which the signal statistical properties change. This can be a problem for impulsive noise or signals.

- ANC, however, has a "direct" measure of the noise at each sample instance for the signal. As long as f is adapted appropriately and a good reference measurement can be obtained, ANC is superior to LP.

2a) Observations

$(X_1, Y_1), \ldots, (X_n, Y_n)$

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

- Minimize the mean square error w.r.t. parameters β_0 & β_1

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

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

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

Define:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \;;\quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \;;\quad \overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \;;\quad \overline{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \;;$$

① $\bar{Y} = \beta_0 + \beta_1 \bar{X}$; ② $\beta_0 \bar{X} + \beta_1 \overline{X^2} = \overline{XY}$

① & ② can be solved for the optimum slope β_1

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

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

$\beta_1 \left(-(\bar{X})^2 + \overline{X^2}\right) + (\bar{X})(\bar{Y}) = \overline{XY}$

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

2b) Assuming a probabilistic model Y is still a function of X with some additive measurement noise:

$Y = f(X) + \varepsilon \Rightarrow \varepsilon$ is drawn from $N(0, \sigma^2)$

Assume a Linear Regressive model for our sequence:

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

For each particular X_i, β_0, β_1, and σ are unknown and Y_i has a distribution $N(f(X_i), \sigma^2)$

The values for β_0 and β_1 are found using the optimal equations from the Linear Regression model.

The ML estimate of σ is found by minimizing over σ using β_0 and β_1.

The pros of this approach are that the ML estimate of the regression function are unbiased.

The cons are that outliers in the data can greatly skew the results because all data is weighted equally.

3a) Let: $O = \{O_1, O_2, \ldots, O_N\}$ be a random variable drawn from $N(\mu_v, \sigma_v)$

Let μ have an a priori distribution given by $N(\mu_m, \sigma_v)$

variance is fixed as stated in the problem

$$\hat{\mu}_{MAP} = \arg\max_{\mu}\, p(O|\mu)\, g(\mu) = \frac{1}{\sqrt{2\pi}\,\sigma_v} e^{-\frac{1}{2}\left(\frac{\mu-\mu_m}{\sigma_v}\right)^2} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_v} e^{-\frac{1}{2}\left(\frac{o_i-\mu}{\sigma_v}\right)^2}$$

This maximization equates to minimizing:

$$\hat{\mu}_{MAP} = \arg\min_{\mu} \left[\frac{1}{2}\left(\frac{\mu-\mu_m}{\sigma_v}\right)^2 + \sum_{i=1}^{N} \frac{1}{2}\left(\frac{o_i-\mu}{\sigma_v}\right)^2 \right]$$

differentiate w.r.t. μ:

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

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

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

3b) - MAP converges to the sample mean of the adaptation data

- MAP converges to the prior mean for the training data

- MAP mixes the prior mean and the sample mean using a weighting scheme based on the number of samples in the adaptation data set

- If the distribution of of the prior has large variance ($\sigma_m \to \infty$) then this data contains no real information, so MAP converges to the ML estimate (sample mean). Basically the training data has no weight.

4)

ECE 8423

- Optimal Estimation
  - MAP
  - ML
  - MMSE
- Adaptive Filters
  - Linear Prediction
  - LMS
  - ANC
  - Adaptive line Enhancement
- Time Delay Estimation
  - LMS
  - ADPCM
- Recursive Least Squares
  - RLS
  - Newton's Method
  - Exponentially Weighted RLS
- IIR Adaptive Filters
  - IIR LMS
  - Lattice filters
- Spectral Estimation
  - Periodograms
  - Parametric Estimation
    - AR
    - MA
    - ARMA
- Descriminative Training
  - Supperuised
    - MAP
    - MLLR
    - MMIE
  - Unsuppervised
    - DMT
    - DLT
  - State of the Art

- MAP - (Maximum A Posteriori) Bayesian estimator that finds the peak value of a parameters density

- ML - (Maximum Likelyhood) - deterministic or classical estimator that maximizes the density of the observed data

- MMSE - (Minimum Mean-Squared Error) - Bayesian because the estimate is a function of the posterior density. The parameter estimate is the mean value.

- Linear Prediction - adaptive filter that can be viewed as a digital filter that changes with time to track signals in real-time.

- LMS - Iterative adaptive filter. The filter impulse response at each sample is the previous value plus a term related to the projection of the error onto the signal.

- ANC - (Adaptive Noise Cancellation) uses a reference measurement to try to cancel noise from a desired signal.

- Adaptive Line Enhancement - good for very narrow band signals

- Least Mean-Squared Time-Delay Estimator - uses LMS approach to estimate time delays for example in array processing.

- ADPCM - (Adaptive Differential Pulse Code Modulation uses the LMS filter to reduce dynamic range for compression.

- RLS - uses an adaptive step size based on the error from the old filter applied to the new data
- Newton's Method - uses a quadratic model for adaptation, slope and curvature information influences iteration steps.

- Exponentially Weighted RLS - most recent errors carry the most weight.

- IIR LMS - implements a recursive LMS algorithm for a filter with poles and zeros.

- Lattice filters –
- Periodograms – inconsistant estimator because the variance doesn't decrease with sample size.
- AR – (Autoregressive Spectral Estimation).
- MA – (Moving Average Estimation
- ARMA – Autoregressive Moving Average

- MAP – As a trained model it an estimate that maximizes the posterior probability based on the observed data and the training data
- MLLR – (Maximum Likelihood Linear Regression) uses a ML approach to estimate parameters related to feature vectors.
- MMIE – (Maximum Mutual Information Estimation) maximizes the mutual information between observations and labels.
- DMT – (Discriminative Mapping Transform)
- DLT – (Discriminative Linear Transform)