

## Homework 2

ECE 5412  
Fall 2019

**Yen-Lin Chen**  
yc2253@cornell.edu  
Due: 10/30/19

## Problem 37

The AR2 model

$$y_k = -a_1 y_{k-1} - a_2 y_{k-2} + w_k \quad (1)$$

has the transfer function:

$$\frac{Y(z^{-1})}{U(z^{-1})} = \frac{z}{z^2 + a_1 z + a_2} \quad (2)$$

This model is stationary if the roots of  $z^2 + a_1 z + a_2$  are in the unit circle of the complex plane, i.e.  $|z_r| < 1$  where

$$z_r = \frac{-a_1}{2} \pm \frac{1}{2}\sqrt{a_1^2 - 4a_2} \quad (3)$$

First, consider the case where  $a_1^2 < 4a_2$  and  $z_r$  is complex.

$$|z_r|^2 < 1 \iff \frac{a_1^2}{4} + \left(-\frac{a_1^2}{4} + a_2\right) < 1 \iff a_2 < 1 \quad (4)$$

Second, consider the case where  $a_1^2 > 4a_2$  and  $x_r$  is real.

$$|z_r| < 1 \iff \frac{-a_1}{2} + \frac{1}{2}\sqrt{a_1^2 - 4a_2} < 1 \quad \text{and} \quad \frac{-a_1}{2} - \frac{1}{2}\sqrt{a_1^2 - 4a_2} > -1 \quad (5)$$

$$|z_r| < 1 \iff a_1 + a_2 + 1 > 0 \quad \text{and} \quad a_1 - a_2 - 1 < 0 \quad (6)$$

The possible values of  $a_1$  and  $a_2$  are shown in Fig. 1 as the colored areas, where the blue and magenta regions correspond to complex and real poles respectively.

## Problem 38

Choose  $a_1 = -1.2$  and  $a_2 = 0.7$  and the AR2 model is now

$$y_k = 1.2y_{k-1} - 0.7y_{k-2} + w_k \quad (7)$$

with  $y_0 = y_1 = 0$ . We simulate 5000 observations of the system and compute the recursive least square (RLS) estimator for  $\theta = (a_1, a_2)'$  using recursive least square (RLS). The RLS estimator at step  $N + 1$  is given by

$$\theta_{N+1} = \theta_N + \frac{P_N \psi_{N+1}}{\rho/c + \psi'_{N+1} P_N \psi_{N+1}} (y_{N+1} - \psi'_{N+1} \theta_N) \quad (8)$$

$$P_{N+1} = \frac{1}{\rho} \left[ P_N - \frac{P_N \psi_{N+1} \psi'_{N+1} P_N}{\rho/c + \psi'_{N+1} P_N \psi_{N+1}} \right] \quad (9)$$

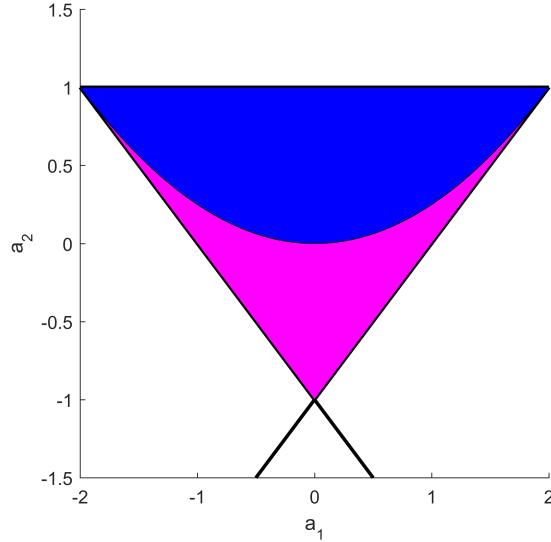


Figure 1: Problem 37: The stability triangle of the AR2 model. The blue and magenta areas correspond to complex and real poles respectively.

where  $\psi_{N+1} = (y_N, y_{N-1})'$ ,  $P_N \in \mathbb{R}^{2 \times 2}$ ,  $P_N > 0$  and the forgetting parameter  $\rho$ . We initialize the estimator and matrix using  $\theta_1 = (1, 1)'$  and  $P_1 = I$ . We choose  $\rho = 1 - 10^{-4}$  and  $c = 1$  and keep them for this problem.

The simulated trace ( $y_k$ ) is shown in the left plot of Fig. 2. The RLS estimator of  $a_1$  and  $a_2$  are shown in the middle and right panels respectively, where we show the trace(blue), true value(black) and 1000-step moving average(red).

To compare the RLS to the classical offline least square estimate, we do the estimation using 1000 observations for the estimator,  $\hat{\theta} = (\hat{a}_1, \hat{a}_2)' = (\Psi'\Psi)^{-1}\Psi'Y$ . We run 100 times of this estimation process to get the statistics of the estimators. The histograms of  $\hat{a}_1$  and  $\hat{a}_2$  are shown in the middle and right plots of Fig. 3. Using 100 simulations, we obtain  $E[\hat{a}_1] = -1.1965$  with std 0.0221 and  $E[\hat{a}_2] = 0.7001$  with std 0.0219.

To explore the effect of initialization for RLS algorithm, we use  $\theta_0 = s(a_{1,0}, a_{2,0})'$  and  $P_0 = tI$  where  $s, t$  are real numbers and  $a_{1,0}$  and  $a_{2,0}$  are random variable with standard normal distributions. We test the values of  $s, t$  ranging from  $10^{-2}$  to  $10^4$ . The 1000-step moving average for the RLS estimate of  $a_1$  and  $a_2$  are shown in blue and green in Fig. 4 where each panel represents different combinations of  $s$  and  $t$ . We observe that the RLS converges fast under reasonable initialization conditions because after enough observations, the RLS estimator forgets about the past exponentially. Here  $s$  can be thought of as how far the initial guess deviates from the real value and  $t$  is the optimization step size. Therefore, for large  $s$  (toward the right panels) and small  $t$  (toward the top panels), the RLS estimators need more steps (observations) to converge to the real values. In one of our extreme testing case where  $s = 10000$  and  $t = 0.01$  (top-right panel), the RLS estimators do not converge after 5000 observations.

## Problem 39

The Akaike information criterion (AIC) is defined as

$$\text{AIC} = -2 \log \hat{L} + 2k \quad (10)$$

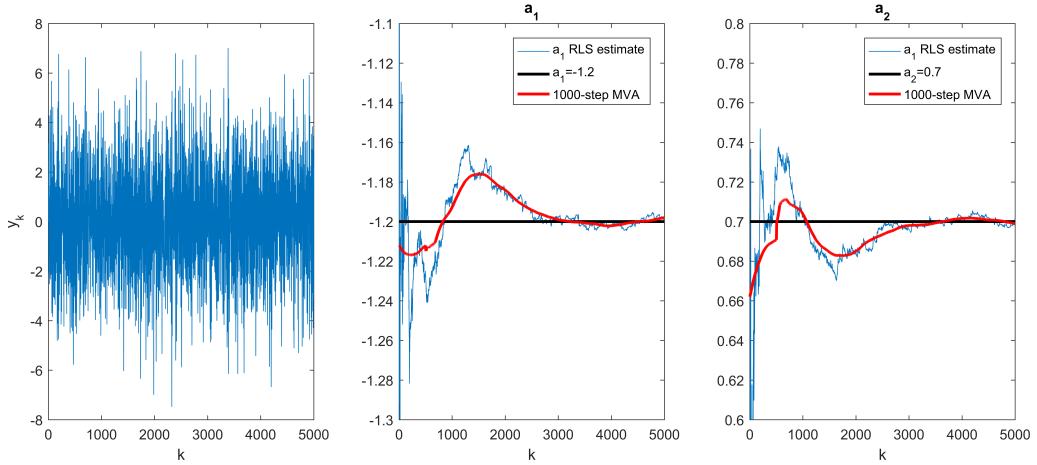


Figure 2: Problem 38: Online RLS. (left) The trace of 5000 observations using  $a_1 = -1.2$ ,  $a_2 = 0.7$  with  $y_0 = y_1 = 0$ . (middle) The RLS trace of  $\hat{a}_1$ . (right) The RLS trace of  $\hat{a}_2$ .

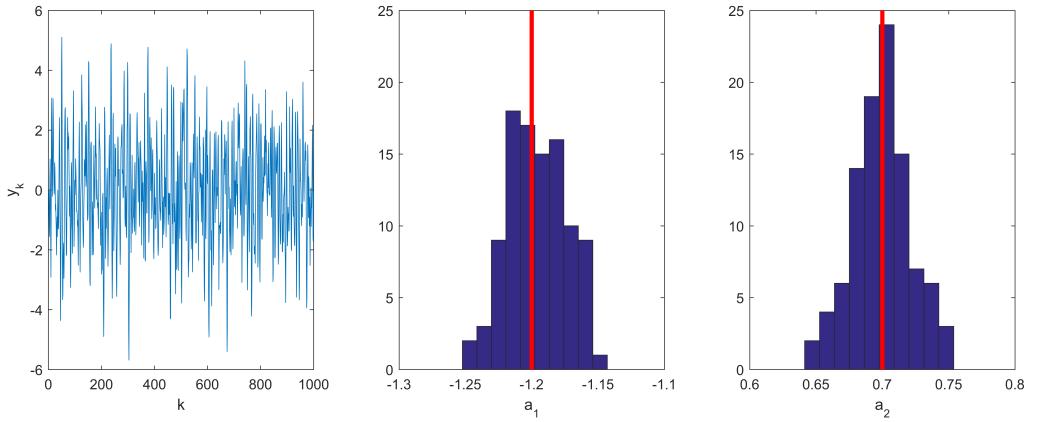


Figure 3: Problem 38: Offline Batch LS. (left) The trace of one simulation of 1000 observations using  $a_1 = -1.2$ ,  $a_2 = 0.7$  with  $y_0 = y_1 = 0$ . (middle) The histogram of  $\hat{a}_1$ . (right) The histogram of  $\hat{a}_2$  from 100 simulations.

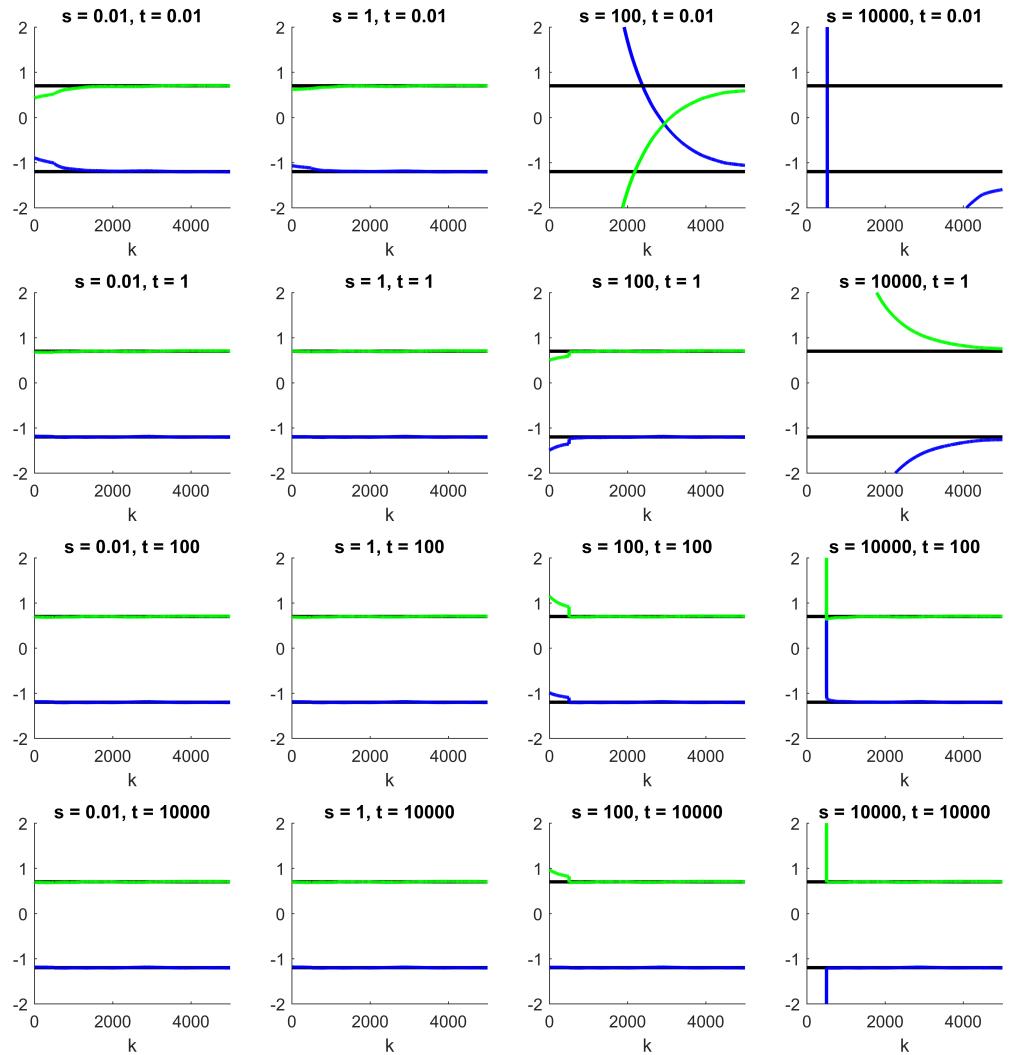


Figure 4: Problem 38: The effect of initial states.

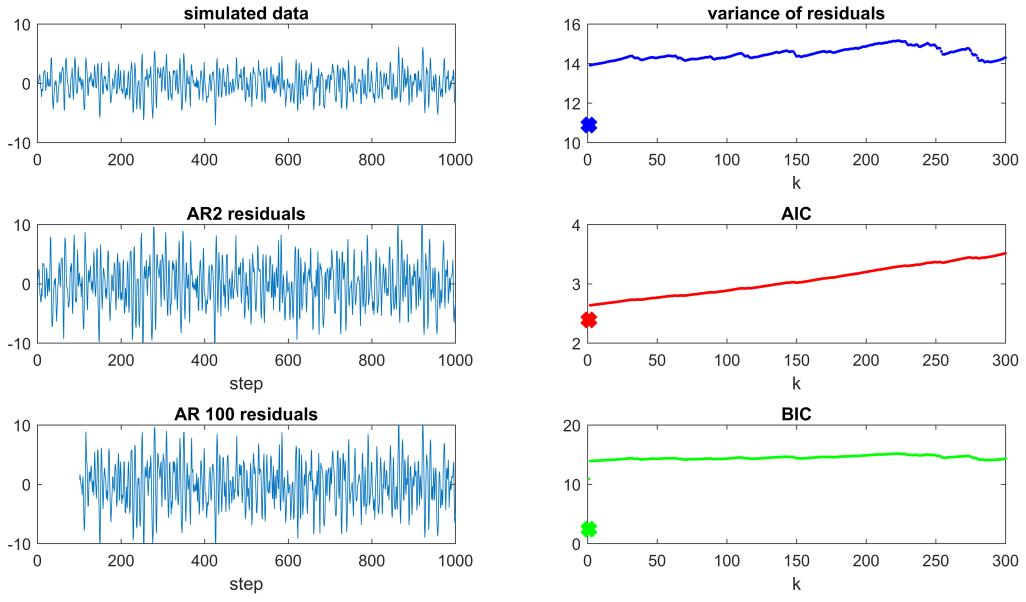


Figure 5: Problem 39: (top left) The simulated data from an AR-2 system using  $a_1 = -1.2$  and  $a_2 = 0.7$  with  $y_0 = y_1 = 0$ . The traces of the residuals (middle left) using the AR-2 model and (bottom left) AR-100 model for LS estimation. The panels on the right show the  $\hat{\sigma}^2$ , AIC and BIC of the AR- $k$  models with the lowest value indicated as "x" which is the AR-1 model.

where  $k$  is the number of estimated parameters in the model and  $\hat{L}$  is the maximum likelihood. Using the same notation, the Bayesian information criterion (BIC) is defined similarly,

$$\text{BIC} = -2 \log \hat{L} + (\log n)2k \quad (11)$$

where  $n$  is the number of data points. The AIC for the ARMA model with parameter  $p$  (order of AR, i.e. number of  $a_i$ 's),  $q$  (order of MA, i.e. number of  $b_i$ 's) is

$$\text{AIC}(arma(p, q)) = \log(\hat{\sigma}^2(p, q)) + \frac{2(p+q)}{n} \quad (12)$$

Similarly, for the BIC,

$$\text{BIC}(arma(p, q)) = \log(\hat{\sigma}^2(p, q)) + \frac{2(p+q) \log n}{n} \quad (13)$$

where  $\hat{\sigma}^2(p, q)$  is the variance of the residuals of the fitted ARMA model.

$$\hat{\sigma}^2(p, q) = \text{var}(Y - \hat{Y}) = \text{var}\left\{\left[I_{p+q} - \Psi(\Psi'\Psi)^{-1}\Psi'\right] Y\right\} \quad (14)$$

In our case of AR model,  $q = 0$ . We simulate  $n = 1002$  from an AR2 model using  $a_1 = -1.2$  and  $a_2 = 0.7$  with  $y_0 = y_1 = 0$ . The simulated data is shown in the top left panel of Fig. 5. We apply the LS estimation for AR- $k$  model where  $k \in \{1, 2, \dots, 300\}$  and compare the AIC and BIC for all the models. We show the traces of residuals of AR-2 (middle left) and AR-100 models (bottom left) in Fig. 3.

Using the same simulated data, we estimated the AR- $k$  parameters using LS method. The variance of the residuals, the AIC and BIC are plotted in the right column of Fig. 5. We label the lowest value with "x" which turns out to be from the AR-1 model. Therefore, by AIC and BIC, the order of AR model is 1.

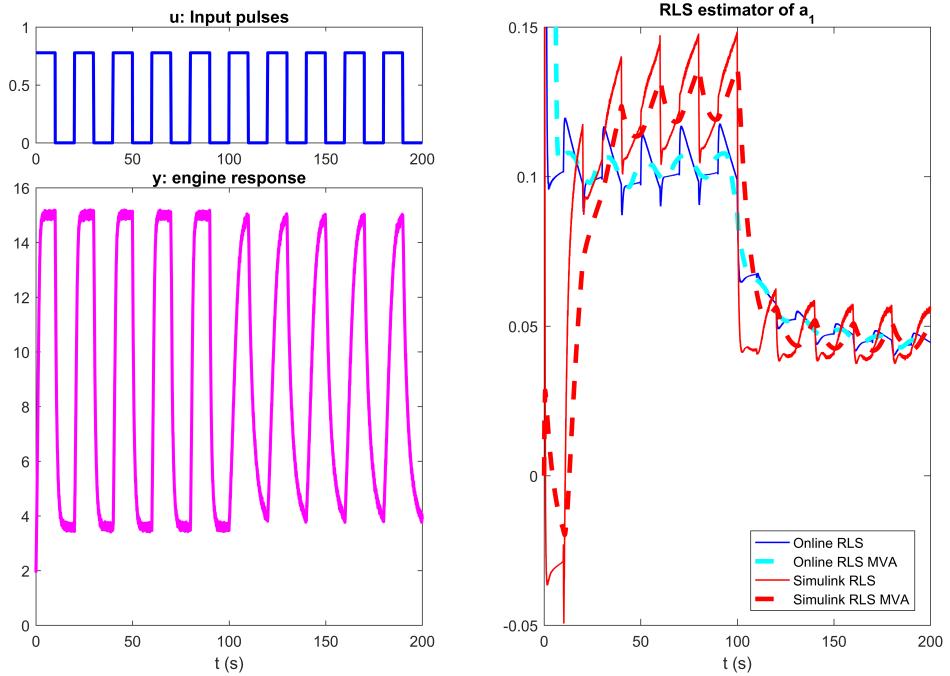


Figure 6: Problem 43: (left) The input  $u$  of the engine and the responses  $y$ . (right) The RLS estimator for  $a_1$  by implementing Eq. (8) and (9) with the 2000-step moving average (blue and cyan). The red trace is the result directly extracted from the Simulink model and the magenta trace is the moving average.

### Problem 43

The engine is modeled by a combination of non-linear input model and AR-1 response.

$$y_n = a_1 u_{n-1} + a_2 u_{n-1}^2 + a_3 y_{n-1} + w_n \quad (15)$$

Our goal is to use the vector  $u$  and  $y$  to construct a online recursive least square estimation of the parameters  $a_i$ 's. In the MATLAB Simulink, the parameter estimations are updated each step along with the simulation on the fly; however, in our attempt to reproduce the result, we extract the data from the Simulink model and do the RLS estimation using Eq. (8) and (9).

In the Simulink online RLS estimation, the forgetting parameter is set to  $\rho = 1 - 10^{-4}$ , corresponding to about 15 seconds of past data.

The input pulse and output response are shown on the left plots of Fig. 6. The online RLS estimations of  $a_1$  from our estimation and from Simulink are shown on the right panel of Fig. 6. This engine system is designed to change inertia at  $t = 100$  s, which affects the  $a_1$  value. In both our and Simulink estimations, the estimated values of  $a_1$  change abruptly at  $t = 100$  s, which signals the change in the engine inertia.

### Problem 46

The stochastic least square problem  $Y = \Psi\theta + \epsilon$  where  $\epsilon \sim N_N(0, \sigma^2 I_N)$  has the estimator  $\theta_*$ :

$$\theta_* = (\Psi'\Psi)^{-1}\Psi'Y \quad (16)$$

The estimator for noise variance  $\hat{\sigma}^2$  is

$$\hat{\sigma}^2 = \frac{1}{N-n}(Y - \Psi\theta_*)'(Y - \Psi\theta_*) \quad (17)$$

Plugging  $\theta_*$  using Eq. (14).

$$\hat{\sigma}^2 = \frac{1}{N-n} [(I - \Psi(\Psi'\Psi)^{-1}\Psi')Y]' [(I - \Psi(\Psi'\Psi)^{-1}\Psi')Y] = \frac{1}{N-n}(HY)'HY \quad (18)$$

where  $H = I - \Psi(\Psi'\Psi)^{-1}\Psi'$  is symmetric and idenpotent ( $H^2 = H$ ) and  $\text{rank}(H) = \text{tr}(H) = N - n$ . Define  $Z = HY$ .  $\hat{\sigma}^2 = \frac{1}{N-n}Z'Z$ .

$$\text{Cov}(Z) = H(\sigma^2 I)H' = \sigma^2 H \quad (19)$$

$$E[\hat{\sigma}^2] = \frac{1}{N-n}E[Z'Z] = \frac{1}{N-n}\text{tr}(\text{Cov}(Z)) = \frac{\sigma^2}{N-n}(N-n) = \sigma^2 \quad (20)$$

Therefore,  $\hat{\sigma}^2$  is an unbiased estimator for  $\sigma^2$ .

By imposing  $A\theta = 0$  where  $A \in \mathbb{R}^{r \times n}$ , we can rewrite the least square error with constraint by expanding the matrices as the following.

$$\begin{bmatrix} Y \\ 0_r \end{bmatrix} = \begin{bmatrix} \Psi \\ A \end{bmatrix} \theta + \begin{bmatrix} \epsilon \\ 0_r \end{bmatrix} \quad (21)$$

The LS estimator  $\hat{\theta}$  is now

$$\hat{\theta} = \left( \begin{bmatrix} \Psi' & A' \end{bmatrix} \begin{bmatrix} \Psi \\ A \end{bmatrix} \right)^{-1} \begin{bmatrix} \Psi' & A' \end{bmatrix} \begin{bmatrix} Y \\ 0_r \end{bmatrix} = (\Psi'\Psi + A'A)^{-1} \Psi'Y \quad (22)$$

We can replace  $\theta_*$  in Eq. (15) with  $\hat{\theta}$  from Eq. (20) can get the new estimation of  $\sigma^2$ :

$$\hat{\sigma}_c^2 = \frac{1}{N-n}(Y - \Psi\hat{\theta})'(Y - \Psi\hat{\theta}) \quad (23)$$

We run the simulation by varying  $r$  using  $N = 1000$ ,  $n = 10$  and  $\sigma^2 = 1$ .  $r = 0$  is the case where there is no constraint. We compute 1000 estimators of  $\sigma^2$  by using different  $\Psi$  to get good statistics of  $\hat{\sigma}^2$ . The trace and histogram are shown on the left of Fig. 6. We repeat the estimation of  $\hat{\sigma}_c^2$  using Eq. (21) for different size of  $A \in \mathbb{R}^{r \times n}$  and explore the effect of imposing the constraint. The  $\hat{\sigma}_c^2$  is increasing with  $r$  which suggests that  $\hat{\sigma}_c^2$  is no longer unbiased because it fails to taking into account the constraint by matrix  $A$ .

Let  $M = (\Psi'\Psi + A'A)^{-1}$ . The residuals from the LS estimator  $\hat{\theta}$  is

$$r = \begin{bmatrix} Y \\ 0_r \end{bmatrix} - \begin{bmatrix} \Psi M \Psi' Y \\ A M \Psi' Y \end{bmatrix} = \begin{bmatrix} (I - \Psi M \Psi') \\ -A M \Psi' \end{bmatrix} Y \quad (24)$$

$$\begin{aligned} r'r &= Y' [I - 2\Psi M \Psi' + \Psi M \Psi' \Psi M \Psi' + \Psi M A' A M \Psi'] Y \\ &= Y' [I - 2\Psi M \Psi' + \Psi M (\Psi'\Psi + A'A) M \Psi'] Y \\ &= Y' (I - \Psi M \Psi') Y \\ &= Y' [I - \Psi (\Psi'\Psi + A'A)^{-1} \Psi'] Y \end{aligned} \quad (25)$$

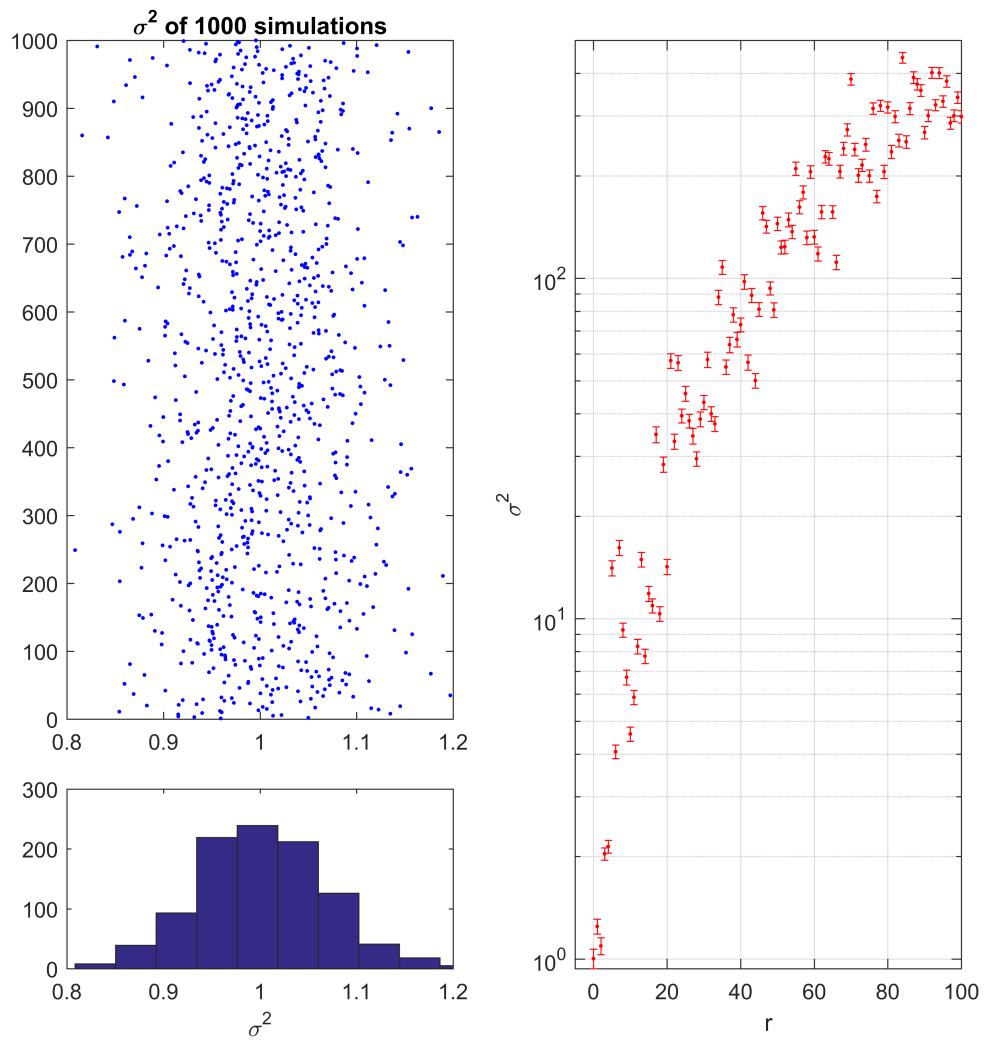


Figure 7: Problem 46: (left) The trace of 1000 simulated  $\hat{\sigma}^2$  without the constraint and the histogram. (right) The  $\hat{\sigma}_c^2$  using Eq. (21) with respect to  $r$ . The error bar is the standard deviation of 1000 simulations of  $\hat{\sigma}_c^2$ .



Figure 8: Problem 48: Snippet of the downloaded face data set.

Notice that  $I - \Psi(\Psi' \Psi + A'A)^{-1} \Psi'$  is not an idenpotent matrix and might have large singular values,  $\sigma_i$ , depending on matrix  $A$ . This is the effect we observe in Fig. 7 where the estimator  $\hat{\sigma}_c^2$  can be roughly written as

$$\hat{\sigma}_c^2 \approx \sigma^2 C(r) \sum_{i=1}^n \sigma_i^2 \quad (26)$$

where  $C(r)$  is a constant that depends on  $r$ . If  $r = 0$ ,  $\sigma_i = 1 \forall i \in \{1, 2, \dots, N-n\}$ . Therefore,

$$\hat{\sigma}^2 = \sigma^2 C(0) \sum_{i=1}^{N-n} \sigma_i^2 = \sigma^2 (N-n) C(0) \quad (27)$$

We can set  $C(0) = 1/(N-n)$  to make  $\hat{\sigma}^2$  an unbiased estimator.

## Problem 47

Given a set of  $n$  face images with dimension  $m$  as the database matrix  $A$ , we can do the PCA of  $A = UDV' = \sum_{i=1}^m \sigma_i u_i v_i'$  using rank  $r \leq m$ . The PCA approximation of the database is

$$Z = \sum_{i=1}^r \sigma_i u_i v_i' = U_r D_r V_r' \quad (28)$$

Given the unknown face  $a^{new}$ , in order to determine which face is the most similar in the database, we compute the deviation  $\|a_i - a^{new}\|$ . The lowest deviation should be the most similar face and we can identify this unknown face. Using PCA, the deviation is

$$\|a_i - a^{new}\| = \|U_r D_r V_r' e_i - a^{new}\| = \|D_r V_r' e_i - U_r' a^{new}\| \quad (29)$$

Therefore, instead of computing  $\|a_i - a^{new}\|$  elementwise in the database, we can project  $a^{new}$  to the principal axes and compare  $r$  components with the feature vectors  $D_r V_r' e_i$  in the database.

We download the face data of the celebrities from UMass Amherst (<http://vis-www.cs.umass.edu/lfw/>). We select the people with 10 images for this problem. We show 40 of the faces in Fig. 8. All the

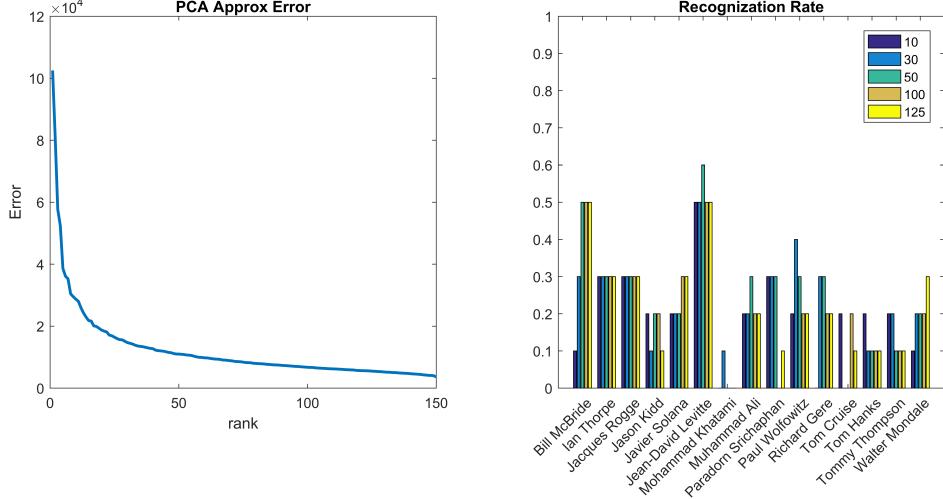


Figure 9: Problem 48: (left) The PCA approximation error v.s. rank. (right) The recognition rates of all the people using different ranks as in the legend.

image data is flattened and stored in a  $m \times n$  matrix,  $A$ . In this small demonstration,  $m = 250^2$  and  $n = 150$ . The PCA is computed and the PCA approximation error is plotted against the rank in the left panel of Fig. 9.

We test the face recognition using PCA algorithm via leave-one-out cross validation. We remove one image  $a^{new}$  from the data base  $A$  and finding the following

$$\hat{i} = \operatorname{argmin}_i \|D_r V'_r e_i - U'_r a^{new}\| \quad (30)$$

If  $\hat{i}$  is the index of the correct person, the image  $a^{new}$  is recognized. We run this procedure and show the recognition rates of all the people we include. Different PCA ranks: 10, 30, 50, 100, 125, are also tested. The recognition rates for all the people are shown in the right panel of Fig. 9. The recognition rates are not optimal because of the lack of pre-processing of the background and the small size of the data set.

## Attachments

*p37.m, ar2simulate.m, arestimate.m, rls.m*