SYSTEMS AND CONTROL
DEPARTMENT OF INFORMATION TECHNOLOGY
UPPSALA UNIVERSITY
http://www.it.uu.se/research/systems_and_control

Homework Assignments

Introduction

All homeworks are based on different exercises from the Spectral Analysis textbook. The exercise numbers for both the new and old version of the book are given, but the exercises in whole are also included in this document for convenience. The code for the different spectral estimators used in the homeworks can be downloaded from Studentportalen. Make sure that you use these functions correctly by making use of the "`help`"-function in MATLAB. If still in doubt how to use the functions, try inspecting the code directly. Note that the custom made functions provided to you are not necessarily compatible with built-in MATLAB functions. For example, the definition of the window function is different in MATLAB and in the provided code. Therefore, use the definitions of the window functions in the book to create your windows directly. Note that in order to solve the homework assignments, you are required to do some MATLAB coding yourself as well.

Reporting

No full report is required, only a clear description of the results with references to all figures you have chosen to include, and a discussion. All figures should have a descriptive figure text, axis labels, etc. Make sure to answer any questions asked in the homework, and to motivate your answers. Think about what needs to be displayed and show only relevant plots, motivate why you show them. For example, some figures might not show anything of interest and some might be identical to others; however, do not fully omit plots that are needed to be discussed.

Often, there is no one true solution, but you will have to make a qualified decision: which makes the best trade off for the problem at hand, in your expert opinion.

An updated list of additional advice is given in Studentportalen, which includes some potential pitfalls.

You are allowed to discuss in pairs when solving the exercises, but the homeworks should be handed in individually. Sign up for a group in Studentportalen, which is also where each homework is to be handed in electronically as a single PDF document, before the respective deadline. Furthermore, the results will be discussed in class, and each individual will be expected to explain, motivate, and answer questions regarding the solutions.

Grading and examination

The homeworks are graded based on your understanding of the problem, use of the methods, the explanations of the results, *how* the results connect to theory, and the motivation for your answers. But also, unavoidably, on your ability to clearly and efficiently present the results. Each homework is worth at most 25 points. Preliminary grades are shown below:

Points	Grade
< 40	Fail
40–59	3
60–79	4
80–100	5

1 Periodogram Methods

The first homework is based on Exercise **C2.22** (**C2.20** in the old book).
Below you find some explanations and clarifications for the corresponding parts of the exercise.

Broadband ARMA process:

(a) Monte-Carlo simulation is a means of gaining information about the statistical properties of a method by applying the algorithm to multiple simulated data sets, where the stochastic parts are redrawn from the same probability distribution for each data set. Here, this is done by using 50 different sequences of the driving noise $e(t)$ generated by `randn()`, and then using "`filter(B,A,e)`" to generate the 50 data sequences $y(t)$, where B and A are the coefficient vectors for the ARMA process.

For each method and M, plot the *sample mean* and the sample mean ± 1 *sample standard deviation* of the resulting PSD estimates (i.e. for each frequency). Use `subplot` to make the comparison easier. Also plot the sample variances of the spectral estimates for the different methods and M:s in the same plot to make the comparison easier. Use dB scales in all these plots. Discuss the differences/similarities between the methods when applied to the dataset, as well as the bias/variance trade-off.

Since **C2.21** is not a part of the homeworks nor a lab, you do not need to do the comparison with the result of this exercise. However, try to say something about what you could expect from such a comparison and why (i.e. if you had used the periodogram).

(b) How does the variance change with M in your simulations (or compared to the periodogram)? Do the simulations seem to agree with theory (see p. 38 in the book)?

(c) Try to describe why you chose this particular window and M as your "best design" for the data at hand. Plot the sample mean, the sample mean ± 1 sample standard deviation of the estimates using your "best design" in the same plot, similarly to what you did in (a).

Narrowband ARMA process:

Repeat the experiments and comparisons of the different methods, but now with the narrowband process. Compare the results from the broadband and narrowband cases. Is there a difference in the conclusions of (b), and if so why? What would be your best design for the narrowband data, and why?

To summarize what needs to be done:

1. Generate 50 different realizations of the two ARMA processes.
2. Compute the BT and Welch spectral estimates for $M = N/4$ and $N/16$ for each of the realizations.
3. Plot the mean PSD estimate for each method, M, and dataset, including a $\pm\sigma$ interval showing the statistical variation for each frequency. Also plot the variance of each method, M, and dataset, together in one plot.
4. Discuss how the different methods and choices of M influence the resulting PSD estimates, both for the broadband and narrowband data. How does M relate to the bias/variance trade-off? Is the obtained variance in line with the theoretical predictions for both the broadband and narrowband data? Why/why not? How would the unmodified periodogram compare to the refined methods?
5. Based on your findings, come up with a best design for each dataset by evaluating different windows and M.
6. Motivate all your decisions/choices

Exercise C2.22: Refined Methods: Variance–Resolution Tradeoff

In this exercise we apply the Blackman–Tukey and Welch estimators to both a narrowband and broadband random process. We consider the same processes in Chapters 3 and 5 to facilitate comparison with the spectral estimation methods developed in those chapters.

Broadband ARMA Process: Generate realizations of the broadband autoregressive moving–average (ARMA) process

$$y(t) = \frac{B_1(z)}{A_1(z)}\, e(t)$$

with

$$\begin{aligned} A_1(z) &= 1 - 1.3817z^{-1} + 1.5632z^{-2} - 0.8843z^{-3} + 0.4096z^{-4} \\ B_1(z) &= 1 + 0.3544z^{-1} + 0.3508z^{-2} + 0.1736z^{-3} + 0.2401z^{-4} \end{aligned}$$

Choose the number of samples as $N = 256$.

(a) Generate 50 Monte–Carlo data realizations using different noise sequences, and compute the corresponding 50 spectral estimates using the following methods:

- The Blackman–Tukey spectral estimate using the Bartlett window $w_B(t)$. Try both $M = N/4$ and $M = N/16$.
- The Welch spectral estimate using the rectangular window $w_R(t)$, and using both $M = N/4$ and $M = N/16$ and overlap parameter $K = M/2$.

Plot the sample mean, the sample mean plus one sample standard deviation and sample mean minus one sample standard deviation spectral estimate curves. Compare with the periodogram results from Exercise C2.21, and with each other.

(b) Judging from the plots you have obtained, how has the variance decreased in the refined estimates? How does this variance decrease compare to the theoretical expectations?

(c) As discussed in the text, the value of M should be chosen to compromise between low "smearing" and low variance. For the Blackman–Tukey estimate, experiment with different values of M and different window functions to find a "best design" (in your judgment), and plot the corresponding spectral estimates.

Narrowband ARMA Process: Generate realizations of the narrowband ARMA process

$$y(t) = \frac{B_2(z)}{A_2(z)}\, e(t)$$

with

$$\begin{aligned} A_2(z) &= 1 - 1.6408z^{-1} + 2.2044z^{-2} - 1.4808z^{-3} + 0.8145z^{-4} \\ B_2(z) &= 1 + 1.5857z^{-1} + 0.9604z^{-2} \end{aligned}$$

and $N = 256$.

Repeat the experiments and comparisons in the broadband example for the narrowband process.

2 Rational Parametric Methods

This second homework is based on Exercise **C3.20**, and in parts **C2.23** (**C3.18** and **C2.21** in the old book)).
Below you find some explanations and clarifications for the corresponding parts of **C3.20**:

"Apply your favorite AR and ARMA estimator(s)" means that you can freely choose between the least squares and Yule-Walker-based approaches. You should carefully consider how to choose the required user parameters, particularly the model orders m and n. It is typically sufficient to try ARMA estimators with $m = n$ (why?). What are the risks/downsides of choosing either too low or too high order? Motivate your choices.

Plot the spectra for AR and ARMA separately, and use e.g. `subplot()` to make the comparison easier. For the "lynx" data, also plot the corresponding spectra obtained from the logarithmically transformed data, see **C2.23b** (this data is provided). How does the two spectral estimators, AR and ARMA, differ? How does the logarithmic transform influence the lynx data and the corresponding spectrum? Use dB scale in all figures in order to simplify the interpretation of the results. Mark on each PSD plot where the most significant spectral peaks, corresponding to poles close to the unit circle, say, with a magnitude larger than 0.8, are located).

Discuss the questions and remarks raised in the exercise. For the discussion on nonparametric vs. parametric methods, you need to solve **C2.23** partly as well. Discuss the differences between nonparametric and parametric methods. How could a combination of these two approaches be used to estimate the spectral and periodic structure of the data?

To summarize what needs to be done:

1. Pick both an AR and an ARMA method (YW/LSAR and MYW/LSARMA). You can do this based on which you like most, which you think is the best, or by actually trying them out and making an informed decision.

2. Try the two chosen methods (for AR and ARMA respectively) on the three datasets (lynx, loglynx, and sunspot) for various choices of orders.

3. Try to come up with a good order for the given datasets somehow. Use the information (from 2. etc.), the methods/tools you have, and your expert opinion, to set a good order[1]. Note that the data was not generated by an AR or ARMA process, it is actual measurements of something. Based on these measurements the goal is to estimate the spectrum of the signal.

4. Plot resulting spectra, and mark locations of poles close to the unit circle (magnitude > 0.8) in these PSD-figures, for example, using vertical lines at the corresponding values of ω. Are there any periodic components? What frequencies? Does log(lynx) make the lynx data more periodic in nature?

5. What are the pros and cons with parametric/nonparametric methods, and how can you use them together (in this HW for example)? What problems can arise if the order of a parametric method (here AR/ARMA) is chosen too low/high?

6. Motivate all your decisions/choices

[1] You can also try to use the `armaorder` function that is supplied to get further guidance from a few automatic order selection methods (this is not required by any means!). However, keep in mind that your professional opinion might be better. And if you try these methods, make sure that you know what you are doing to avoid errors that can be misleading.

Exercise C3.20: AR and ARMA Estimators applied to Measured Data

Consider the data sets in the files `sunspotdata.mat` and `lynxdata.mat`. These files can be obtained from the text web site `www.prenhall.com/stoica`.

Apply your favorite AR and ARMA estimator(s) (for the `lynx` data, use both the original data and the logarithmically transformed data as in Exercise C2.23) to estimate the spectral content of these data. You will also need to determine appropriate model orders m and n (see, *e.g.*, Exercise C3.19). As in Exercise C2.23, try to answer the following questions: Are there sinusoidal components (or periodic structure) in the data? If so, how many components and at what frequencies? Discuss the relative strengths and weaknesses of parametric and nonparametric estimators for understanding the spectral content of these data. In particular, discuss how a combination of the two techniques can be used to estimate the spectral and periodic structure of the data.

Exercise C2.23: Periodogram–Based Estimators applied to Measured Data

Consider the data sets in the files `sunspotdata.mat` and `lynxdata.mat`. These files can be obtained from the text web site `www.prenhall.com/stoica`. Apply periodogram–based estimation techniques (possibly after some preprocessing; see the following) to estimate the spectral content of these data. Try to answer the following questions:

(a) Are there sinusoidal components (or periodic structure) in the data? If so, how many components and at what frequencies?

(b) Nonlinear transformations and linear or polynomial trend removal are often applied before spectral analysis of a time series. For the `lynx` data, compare your spectral analysis results from the original data, and the data transformed first by taking the logarithm of each sample and then by subtracting the sample mean of this logarithmic data. Does the logarithmic transformation make the data more sinusoidal in nature?

3 Rational Parametric Methods for Line Spectra

This third homework is based on Exercise **C3.18** (**C3.17** in the old book).
Below you find some explanations and clarifications for the corresponding parts of the exercise.

AR and ARMA Estimators for Line Spectral Estimation

(a) Verify that the expression for the true spectrum $\phi(\omega)$ is correct given the signal $y(t)$. Use the this true spectrum as a reference for the remaining parts of the exercise.

(b) The Yule-Walker and modified Yule-Walker methods use the ACS sequence to compute the parameters. Usually the ACS sequence is estimated from the data. Here you have to use the *true* ACS. Use (4.1.6) to compute the true ACS for the given signal. Now modify the m-files to use this true ACS sequence instead of an estimated sequence. For example, you can modify the function such that the true ACS sequence is passed as input argument, rather than the data. In that case the computation of $r(k)$ inside the m-file is unnecessary. Using the true ACS helps to eliminate the effects of random estimation errors, and makes it easy to study the theoretical resolution properties of various methods. For the MYW technique, use $M = n$. Plot the locations of the roots of $A(z)$ in a separate pole-zero plot for each example (manually, or using the `zplane(B,A)` function where `B` and `A` need to be row vectors). Make sure the unit circle is drawn, and use the `axis equal` command if it is displayed as an ellipse. Note that running MYW(12,12) gives a singularity warning. Also try using the pseudo inverse when computing the AR coefficients, i.e. `a=-pinv(R1)*r1;`, instead of the backslash operator, is there any difference in the result/estimation performance?

Users of the old book: Note that the last sentence in Exercise C3.17b belongs to Exercise C3.17c (see the errata).

(c) Generate 50 data sequences (realization of the model) with random phases ϕ_1 and ϕ_2 (but no noise). Then try all four methods (two AR and two ARMA) with the two different choices of orders from (b). For the ARMA methods, also try two different values of the user parameters. This gives in total 12 plots of $1/|A(\omega)|^2$ with 50 estimates in each (one for each realization). Then also plot the corresponding pole(/zero) plot for each of the 12 $1/|A(\omega)|^2$-plots, again with all 50 pole estimates in the same plot. For example, use subplot to display the $1/|A(\omega)|^2$-plot and the pole-plot side by side for each method (and setting).

Note that in this noise-free case ($\sigma^2 = 0$) the covariance matrix R will have rank $= 4$ (the number of complex sinusoids in the data). This means that for model orders > 4 there will be a rank deficiency which shows up as warnings in MATLAB. Does this fact seem to cause any problems? Again, try using the pseudo inverse instead of backslash when computing the AR coefficients to see if there is any difference.

(d) Experimenting with the model orders, SNR, K and M is highly recommended in order to understand their impact on the results. Try at least orders 4, 8, and 12 ($m = n$ for ARMA), $K = M = \{n,\ 2n,\ 3n\}$, and $\sigma^2 = \{1,\ 0.125\}$. Show the plots you think are most relevant to motivate your answers and conclusions.

(e) This should be straight forward.

Exercise C3.18: AR and ARMA Estimators for Line Spectral Estimation

The ARMA methods can also be used to estimate line spectra (estimation of line spectra by other methods is the topic of Chapter 4). In this application, AR(MA) techniques are often said to provide *super–resolution* capabilities because they are able to resolve sinusoids too closely spaced in frequency to be resolved by periodogram–based methods.

We again consider the four AR and ARMA estimators described above.

(a) Generate realizations of the signal

$$y(t) = 10\sin(0.24\pi t + \varphi_1) + 5\sin(0.26\pi t + \varphi_2) + e(t), \qquad t = 1, \ldots, N$$

where $e(t)$ is (real) white Gaussian noise with variance σ^2, and where φ_1, φ_2 are independent random variables each uniformly distributed on $[0, 2\pi]$. From the results in Chapter 4, we find the spectrum of $y(t)$ to be

$$\begin{aligned}\phi(\omega) = {} & 50\pi\,[\delta(\omega - 0.24\pi) + \delta(\omega + 0.24\pi)] \\ & +12.5\pi\,[\delta(\omega - 0.26\pi) + \delta(\omega + 0.26\pi)] + \sigma^2\end{aligned}$$

(b) Compute the "true" AR polynomial (using the true ACS sequence; see equation (4.1.6)) using the Yule–Walker equations for both AR(4), AR(12), ARMA(4,4) and ARMA(12,12) models when $\sigma^2 = 1$. This experiment corresponds to estimates obtained as $N \to \infty$. Plot $1/|A(\omega)|^2$ for each case, and find the roots of $A(z)$. Which method(s) are able to resolve the two sinusoids?

(c) Consider now $N = 64$, and set $\sigma^2 = 0$; this corresponds to the finite data length but infinite SNR case. Compute estimated AR polynomials using the four spectral estimators and the AR and ARMA model orders described above; for the MYW technique consider both $M = n$ and $M = 2n$, and for the LS ARMA technique use both $K = n$ and $K = 2n$. Plot $1/|A(\omega)|^2$, overlaid, for 50 different Monte–Carlo simulations (using different values of φ_1 and φ_2 for each). Also plot the zeroes of $A(z)$, overlaid, for these 50 simulations. Which method(s) are reliably able to resolve the sinusoids? Explain why. Note that as $\sigma^2 \to 0$, $y(t)$ corresponds to a (limiting) AR(4) process. How does the choice of M or K in the ARMA methods affect resolution or accuracy of the frequency estimates?

(d) Obtain spectral estimates ($\hat{\sigma}^2|\hat{B}(\omega)|^2/|\hat{A}(\omega)|^2$ for the ARMA estimators and $\hat{\sigma}^2/|\hat{A}(\omega)|^2$ for the AR estimators) for the four methods when $N = 64$ and $\sigma^2 = 1$. Plot ten overlaid spectral estimates and overlaid polynomial zeroes of the $\hat{A}(z)$ estimates. Experiment with different AR and ARMA model orders to see if the true frequencies are estimated more accurately; note also the appearance and severity of "spurious" sinusoids in the estimates for higher model orders. Which method(s) give reliable "super–resolution" estimation of the sinusoids? How does the model order influence the resolution properties? Which method appears to have the best resolution?

You may want to experiment further by changing the SNR and the relative amplitudes of the sinusoids to gain a better understanding of the relative differences between the methods. Also, experiment with different model orders and parameters K and M to understand their impact on estimation accuracy.

(e) Compare the estimation results with periodogram–based estimates obtained from the same signals. Discuss differences in resolution, bias, and variance of the techniques.

4 Parametric Methods for Line Spectra

This fourth homework is based on Exercise **C4.14**, and **C2.23** + **C3.20** in parts (**C4.10**, and **C2.21** + **C3.18** in the old book).
Below you find some explanations and clarifications for the corresponding parts of **C4.14**:

The following approach (to be executed for each method) is recommended to simplify the interpretation of the results:

(a)
1. Remove the mean ($\bar{y}$) of the data (y) giving a new dataset $y_m = y - \bar{y}$ (as zero mean is assumed by the model/methods)
2. Compute frequency estimates for even orders between 4 and 40 (corresponding to $n_s = 2$ to 20 sinusodial components). For HOYW, explore the methods performance for $L = M \in \{n, \ldots, 3N/4\}$; and for MUSIC, Min-Norm, and ESPRIT, explore $m = \{n+1, \ldots, N/2\}$.
3. Compute the least-squares (LS) amplitude and phase estimates $\hat{\beta}$, by using Eq. (4.3.8) in the book (implemented in the `lsa()` function), which takes the current dataset and the frequencies $\{\hat{\omega}_j\}_{j=1}^{n}$ given by the estimation methods above, as input.
4. Reconstruct the signal $\hat{y}$ using the estimated parameters, i.e. simulate the model.
5. Compute and plot the relative mean-squared errors (MSE) between the data and reconstructed signals, versus both the model order and the method parameter (i.e. a 2D plot). Here, we use the following definition which ignores the mean[1]

$$\mathrm{MSE_{rel}} = \frac{\frac{1}{N}\sum_t (y(t) - \hat{y}(t))^2}{\frac{1}{N}\sum_t (y(t) - \bar{y})^2} = \frac{1}{N}\frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\mathrm{var_b}(\mathbf{y})} = \frac{\|\mathbf{y}_m - \hat{\mathbf{y}}_m\|^2}{\|\mathbf{y}_m\|^2} = 1 - \frac{\|\hat{\mathbf{y}}_m\|^2}{\|\mathbf{y}_m\|^2},$$

   where $\mathrm{var_b}(\mathbf{y})$ is the biased sample variance. This measure can be seen as the MSE of your estimator in relation to the MSE obtained when fitting a constant (the mean), a ratio which should be less than unity for any reasonable estimator.
6. The relative-MSE plot will indicate how stable each method is with respect to the user parameters (including the model order), comment on their respective performances. Are there any methods that can reliably capture the variations of the data? Does it seem like the method parameters (L, M or m) can be chosen beforehand by some rule of thumb or similar?

(b) Plot the relative MSE corresponding best choice of $L = M$ and m, for each order. Choose a reasonable model order, e.g. such that no *significant* improvement in relative MSE can be achieved by increasing the order further. Also, select the model order using the BIC criterion and compare to your personal choice.

(c) Plot the spectral estimates corresponding to the choices in (b). Are the obtained frequencies fairly similar for all methods? In particular, look at the most significant periodicities (i.e. frequencies with significant amplitudes). Are any of the frequencies reliably estimated with all methods? Which frequencies in that case?

For convenience, some MATLAB skeleton code is provided to help with the implementation. Present only the relevant plots (needed to make your point), and use `subplot()` to make the comparison easier. For a complete discussion on line spectral methods versus non-parametric and ARMA methods, you need to solve **C2.23** and **C3.20** partly as well. If you have completed **homework 2**, this should only require minor work, and it can be interesting to see how the results compare.

[1] Since we are only interested in the *variations* of the data, this is a suitable choice.

Exercise C4.14: Line Spectral Methods applied to Measured Data

Apply the Min–Norm, MUSIC, ESPRIT, and HOYW frequency estimators to the data in the files `sunspotdata.mat` and `lynxdata.mat` (use both the original `lynx` data and the logarithmically transformed data as in Exercise C2.23). These files can be obtained from the text web site `www.prenhall.com/stoica`. Try to answer the following questions:

(a) Is the sinusoidal model appropriate for the data sets under study?

(b) Suggest how to choose the number of sinusoids in the model (see Exercise C4.13).

(c) What periodicities can you find in the two data sets?

Compare the results you obtain here to the AR(MA) and nonparametric spectral estimation results you obtained in Exercises C2.23 and C3.20.

Exercise C3.20: AR and ARMA Estimators applied to Measured Data

Consider the data sets in the files `sunspotdata.mat` and `lynxdata.mat`. These files can be obtained from the text web site `www.prenhall.com/stoica`.

Apply your favorite AR and ARMA estimator(s) (for the `lynx` data, use both the original data and the logarithmically transformed data as in Exercise C2.23) to estimate the spectral content of these data. You will also need to determine appropriate model orders m and n (see, *e.g.*, Exercise C3.19). As in Exercise C2.23, try to answer the following questions: Are there sinusoidal components (or periodic structure) in the data? If so, how many components and at what frequencies? Discuss the relative strengths and weaknesses of parametric and nonparametric estimators for understanding the spectral content of these data. In particular, discuss how a combination of the two techniques can be used to estimate the spectral and periodic structure of the data.

Exercise C2.23: Periodogram–Based Estimators applied to Measured Data

Consider the data sets in the files `sunspotdata.mat` and `lynxdata.mat`. These files can be obtained from the text web site `www.prenhall.com/stoica`. Apply periodogram–based estimation techniques (possibly after some preprocessing; see the following) to estimate the spectral content of these data. Try to answer the following questions:

(a) Are there sinusoidal components (or periodic structure) in the data? If so, how many components and at what frequencies?

(b) Nonlinear transformations and linear or polynomial trend removal are often applied before spectral analysis of a time series. For the `lynx` data, compare your spectral analysis results from the original data, and the data transformed first by taking the logarithm of each sample and then by subtracting the sample mean of this logarithmic data. Does the logarithmic transformation make the data more sinusoidal in nature?