

# Data Analysis and System Identification

**AE566, Fall 2008**

- Course Objective:** Data is the link between theory (models) and the real world (physical systems and hardware). The objective of this course is to develop skills for using data to construct empirical models. This course is method-oriented rather than theory-oriented, and will cover a broad range of topics and algorithms, which you will learn to use and apply. Extensive use will be made of Matlab routines with applications to engineering and scientific problems.
- Prerequisites** This course is suitable for **senior-level students and graduate students** majoring in any area of engineering or science. Although the course has no specific prerequisites, some facility with Matlab is recommended.
- Credits** This is a 3-credit course.
- Instructor:** Professor D. S. Bernstein  
3020 FXB Building  
[dsbaero@umich.edu](mailto:dsbaero@umich.edu)
- Meets:** Tuesdays and Thursdays 2:30-4:00, Room FXB 1012
- Textbooks:** A combination of books will be used for this course. Students will not need to purchase any books.
- Grading:** Biweekly homework 50%  
Project 50%  
No tests

# Basic Information

## Course Content

- Analyze signals and systems
  - Signal processing techniques and systems analysis tools
- Construct models
  - System identification methods

## Philosophy

- Learn and understand by doing, failing, succeeding, and experiencing
- Some math, but “proof” and discovery mainly by Matlab
- Apply methods when legal and when *not* legal
  - What goes wrong and why?
  - Apply methods in an investigative manner
- Experimental philosophy—like a lab course
- Have fun and learn a lot!

## Work Expectations

- No tests.
- 6 required homework assignments.
- Final project in which you apply the methods of the course to an application of your choice.
- All Matlab code and all substantive work must be your own.

## Useful Background (but no specific prerequisites)

- Linear algebra and matrices
- Differential/difference equations and linear systems theory
- Optimization theory and methods
- Probability, random processes, and statistics

# Syllabus

## **Week 1      Analysis of sensors, measurements, data, and noise**

Overview of sensors used to make measurements. Sensor specifications: precision, accuracy, repeatability, calibration, nonlinearity, sensitivity, dynamic range. Basic error concepts.

## **Weeks 2,3      Frequency domain analysis of continuous-time signals and systems**

Fourier series and Fourier transform. Nyquist theorem, aliasing, and undersampling. Laplace transform, causality, convolution, and impulse response. Shannon reconstruction.

## **Week 4      Frequency domain analysis of discrete-time signals and systems**

Discrete Fourier transform and fast Fourier transform. Estimation of discrete-time transfer functions.

## **Weeks 5,6      Statistical analysis of signals in the time domain**

Probability concepts, random processes, autocorrelation, power spectral density (periodogram), and autocorrelation. Frequency response estimates using impulse, stochastic, and harmonic-response data. Markov parameters.

## **Weeks 7,8      Regression for linear time-series models**

Stochastic simulation. ARMA, ARMAX, and DARMA time-series models and their state space realizations. Mu-Markov models with explicit Markov parameters. Least squares (linear regression) techniques. Batch and recursive least squares. Bias versus consistency. Regularization. BLUE. Handling missing data. Blind and sensor-only identification. SISO versus MIMO identification.

## **Week 9      Parameter estimation**

Parameter estimation for white box models using optimization methods. Maximum likelihood and nonlinear least squares. Gradient and non-gradient search methods. Optimization for transfer function fits to frequency-response estimates.

**Week 10      Identifying linear state space models**

ERA/Ho-Kalman algorithm using Markov parameters from frequency-domain and time-domain identification. OKID. Identifying state space models in the time domain using subspace methods. Estimating initial conditions and unmodeled inputs.

**Week 11      Identifying nonlinear time series models**

Coherence as a test of nonlinearity. Function approximation and basis-function expansions of nonlinear maps. Polynomials and radial basis functions. NARMAX models. Support vector machines and Mercer kernels.

**Weeks 12,13    Identifying nonlinear models**

Grey box and black box nonlinear model structures. Hammerstein-Wiener, bilinear, linear parameter-varying (LPV), Volterra, and neural network models. Techniques for nonlinear model identification.

# References

## Books on linear system identification methods

The following books cover the basic ideas and techniques of system identification:

J.-N. Juang, *Applied System Identification*, Prentice-Hall, 1993.

C. Heij, A. Ran, and F. van Schagen, *Introduction to Mathematical Systems Theory: Linear Systems, Identification and Control*, Birkhauser, 2007.

M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*, Cambridge University Press, 2007.

The following books are references on time-domain system identification:

L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, MIT Press, 1983.

T. Soderstrom and P. Stoica, *System Identification*, Prentice Hall, 1989.

L. Ljung, *System Identification: Theory for the User*, 2<sup>nd</sup> edition, Prentice Hall, 1998.

T. Katayama, *Subspace Methods for System Identification*, Springer, 2005.

System identification in the frequency domain is covered in:

R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*, IEEE Press, 2001.

## **Books on nonlinear system identification methods**

R. Haber and L. Keviczky, *Nonlinear System Identification—Input-Output Modeling Approach, Vol. 1: Nonlinear System Parameter Identification*, Kluwer, 1999.

R. Haber and L. Keviczky, *Nonlinear System Identification—Input-Output Modeling Approach, Vol. 2: Nonlinear System Structure Identification*, Kluwer, 1999.

O. Nelles, *Nonlinear System Identification*, Springer, 2001.

## **Books on function approximation and parameter estimation**

R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems*, Elsevier, 2005.

J. Farrell and M. M. Polycarpou, *Adaptive Approximation Based Control*, Wiley, 2006.

V. Cherkassky and F. Mulier, *Learning from Data*, 2<sup>nd</sup> Edition, Wiley, 2007.

A. van den Bos, *Parameter Estimation for Scientists and Engineers*, Wiley, 2007.

## Books on system theory and signal processing

A good textbook on signals and systems is:

C. D. McGillem and G. R. Cooper, *Continuous and Discrete Signal and System Analysis*, 3<sup>rd</sup> edition, Saunders, 1991.

There are many books on probability, random processes, and statistics. Good, modern treatments are:

J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*, Cambridge, 2006.

J. A. Rice, *Mathematical Statistics and Data Analysis*, 3<sup>rd</sup> Edition, Thomson, 2007.

Some advanced topics relevant to signal processing are covered in:

R. W. Hamming, *Digital Filters*, 3<sup>rd</sup> edition, Dover, 1998.

W. A. Gardner, *Statistical Spectral Analysis: A Nonprobabilistic Theory*, Prentice Hall, 1988.

S. M. Kay, *Fundamentals of Statistical Signal Processing: Vol. 1, Estimation Theory*, Prentice Hall, 1993.

# AE566 Fall 2008 Course Project

Version December 4, 2008

**Due:** Tuesday, December 23, 2008, by 4:00 PM in my mailbox in FXB and by email

## 1 Overview

The purpose of this project is to have you apply the methods learned in this course to data.

In order to do the project, you therefore must have data. In particular, the data that you use must satisfy the following criteria:

1. The data must be either generated by a physical process or it can be generated by a sophisticated computer simulation.
2. Several “officially approved” data sets are available for you to use. If you use data generated by another student, then I ask that you make your report available to that student after the course is over. You need not do this if you do not wish to do so. If you do agree to do this, then the student who sees your report will be asked to agree to acknowledge your work if it has any impact on their future research.
3. If you use your own data, then I must approve your project. Therefore, be sure that you have emailed me a description of your project before Tuesday, November 25 so that I can approve it in advance.
4. You must have at least 3 signals. Ideally at least one of the signals is an input that causally affects at least one other variable. However, you can do the project if all 3 signals are outputs (that is, using sensor-only identification).
5. The process from which the data comes must have some dynamics that can be modeled.
6. You must have a sufficient amount of data to do interesting identification. I recommend 5,000-10,000 data points for all inputs and outputs. The absolute minimum is probably around 2,000. With this amount of data you can do both model fitting and prediction.
7. Either the input data can be commanded by you (such as a sine sweep or impulse that you or someone else implemented) or it can be measured (for example, solar wind data, where we have no control over the Sun). In either case, you need to make sure that the data are sufficiently persistent to ensure the ability to do useful identification of the system’s dynamics.



## 2 Project Elements

1. First analyze your data in terms of its sample rate, spectral content, noise content and statistics, input-output or signal-signal correlation, and other features. Discuss whether your signals are inputs or outputs, and how you expect them to be related causally or how they might be correlated based on your knowledge of the relevant physics.
2. The main identification portions of the project include 1) linear identification, and 2) nonlinear identification or parameter estimation.

1) Linear identification. First, perform coherence analysis to check for degree of linearity. Then, perform frequency domain identification on one SISO transfer function as a preliminary step. Now you will obtain least-squares  $\mu$ -Markov SISO/MISO/SIMO/MIMO fits. Although SISO fits can be useful, you *must* consider fits that involve more than one input or output (which is why your data must involve at least 3 signals). **You will need to program a multivariable least squares algorithm for this portion of the project.** Note that we did not do multivariable identification for homework, but I explained in class how it can be done. **You need to provide one “toy” example to demonstrate that your code is working. Be sure this demonstration example is clearly presented in a short subsection.** The goal of the linear identification is to estimate Markov parameters. You can do this either by 1) identifying an ARMA model and then impulsing it to obtain the Markov parameters, or 2) identifying a  $\mu$ -Markov model and then extracting the Markov parameters directly from it. Note that if you do 1) and you have more than 1 input signal, then you will need to be sure to impulse the system properly in order to extract the Markov parameters. For your data set, be sure to vary the order, number of Markov parameters, amount of data, data set, and anything else that will help you assess the validity and accuracy of your linear fits. Also, consider the use of input/disturbance identification through input augmentation. Look at the frequency content of the fit error to determine whether you might have missed some dynamics. Use ERA (eigensystem realization algorithm) analysis with your estimated Markov parameters to check on the order of the system and construct a state space model.

2) Nonlinear Identification/parameter estimation. For this objective you may use any nonlinear model structure and identification methodology you wish. Most of you will use NARMAX, but you can use any other model structure (such as Wiener-Hammerstein, LPV, bilinear, neural network, or a physics-based white box model) discussed in class or in the literature. For fit optimization, you must use `fminsearch`. Be sure to check a variety of model structures, orders, and other features that might impact the fits. Consider the use of input identification through input augmentation. Check the frequency content of the fit error. Use some of the data for prediction as a form of validation. **Note: You must do fits for EITHER two SISO relations or one MISO/SIMO relation. I prefer the latter.**

3. Inevitably, some things will not work, and you will get puzzling results. By using toy examples, you can test conjectures about the source of the problem.

### 3 Report Outline

1. Section 1 is an introduction that describes the goals of the report. **If you use one of the data sets provided to the class, then you must state whether or not you agree to allow the person who provided the data to read your report.**
2. Section 2 describes the application from which the data are taken. Show diagrams and figures if appropriate explaining the source of the data and its physical meaning.
3. Section 3 analyzes the data. You can include relevant information about how the data were measured (precision, noise, resolution in the sensors and actuators if this information is known). Perform statistical and spectral analysis of the data to get insight into its properties before you begin identification.
4. Section 4 is on linear time-domain identification, including frequency-domain identification and coherence analysis.
5. Section 5 is on nonlinear identification (or parameter estimation).
6. Section 6 includes discussion and conclusions. This section must provide substantial and detailed discussion of what your investigations have revealed about the data and system. Summarize what you learned from the experience.
7. Bibliography. List any references that you used or are relevant to any aspect of your project and report.
8. Appendix. See below.

## 4 Report Requirements

1. Your report must be typed, neat, well organized, and well written. You must explain clearly what you are doing in each section of the report.
2. All discussions are written with good grammar and paragraph structure. Please review my writeup “Precision Writing.”
3. You do not need to explain how the identification algorithms work, but you must include enough discussion of each algorithm so that it is clear what is being plotted.
4. All figures must appear in order and must be mentioned in the text. All plots must be clear in terms of axes, units, labels, legends, and must have explanatory captions. All curves in all figures are labeled.
5. All code that you use must be your own. Except for `fminsearch` and basic Matlab functions, you cannot use code from any other source, including my homework solutions. **In particular, you cannot use identification routines from other sources such as Matlab toolboxes.** If in doubt, please email me.
6. Except for very limited conceptual discussions with other students, no collaboration of any kind is permitted on this project. Everything you do must be your own work, and all programs and text must be original. You may consult books and papers, and those should be cited as appropriate.
7. Include equations that are needed to explain the application if they are available.
8. Give all information in a logical order. **Do not use appendices except as noted below.**
9. List some references at the end of your report that are relevant to your application or data analysis and system identification.
10. Provide an appendix that *lists the titles* of all of your m-files along with a brief description of what each file does. Then, zip up all relevant files and email them to me when you submit your report. Note: Do not print out and attach all of your Matlab code to the report.
11. Hand in one paper copy by the specified deadline. Use one-sided printing only.
12. Be creative in producing an interesting report. Strive to do more than the minimum required.
13. The thought process is important. For example, if you believe that your system is nonlinear because the linear methods do not work well, do a sufficient number of cases to support this view. In other words, provide a convincing case that the system is not linear. Think about diagnostics in analyzing the data.

14. I do not expect all methods to be successful. In fact, I expect many methods to fail in many cases. However, I will be looking for a clear logical flow in terms of the thought process. What did you learn from each test? How did the result of one test suggest another test or variation? What are the novel or creative elements? Did you preprocess data? Did you exploit any knowledge of physics? What insights into the system that generated the data does your data analysis reveal?

## 5 Grading

1. 50% points will be devoted to the mechanics of the report, including the quality of writing, figures, organization, clarity, and all other aspects of presentation. Any aspects that are not clear or are poorly presented will detract from the grade.
2. 50% points will be devoted to the technical quality, including the correct application of the the methods that are specified above. Any elements that are missing or are not properly executed will detract from the grade.
3. I will add extra credit for items that go beyond the minimum or are especially innovative.