2102531 System Identification

Term Project
Semester 1/2016

Jitkomut Songsiri

November 24, 2017

Abstract

This project aims to apply an identification technique to real-world problems. There are **four** topics selected from various applications, namely, (i) recursive estimation of solar irradiance using time-series models, (ii) subspace identification of EEG time series, (iii) identification of building temperature system, and (iv) modeling of photovalatic system. Students spend six weeks on this project where each week a progress report must be submitted.

Instruction to students

Each week students are supposed to submit the weekly progress report to CourseVille in Assessment section, where *everyone* must submit the group report in typesetting and .pdf format. Any proposed ideas, comments, or practical considerations should be stated in the report to keep track of your concerns or problem found as the work progresses. The use of LaTeXto type the document is highly encouraged. The progress reports should be named as `onlinesolar_week1.pdf` `subspace_week1.pdf` `temperature_week1.pdf` `pv_week1.pdf` where the week number refers to the week of submitting the report. We have created a shared folder `ee531_project_2016` and all the established necessary files must be kept here except your codes in progress. Put the progress reports in the folder `report` each week as well.

Student lists

1. Satayu Chunnawong
2. Tony Fang
3. Chanthawit Anuntasethakul
4. Natdanai Sontrapornpol
5. Natthapol techaphangam
6. Janenarong Klomklao

Final report format

All groups should use the following formats:

- All vector graphics (such as MATLAB figures) must be in .eps or .pdf format.

- All MATLAB codes must be in the appendix with proper comments explaining the codes. You can use `listings` package for source code listing in LaTeX. Mark the language (MATLAB) to highlight the codes. Each MATLAB file should be referred to the section where it is used, or you can explain which function is used to generate each figure.
- Use BibTeX to create reference lists.

1 Recursive Estimation of Solar Irradiance using Time-series Models

Student list:

1. Tony Fang

1.1 Progress deadlines

Wed Oct 18

- Typesetting report of problem description.
- Explain the application and define the problem explicitly; the model, model characteristics, input, and output.

Wed Oct 25: Class skipped

Wed Nov 1

- Revise the old contents.
  - Revise the report according to comments (mostly about notations, and English).
  - All contents should be supported by reliable references.
  - Explain about the characteristics of solar irradiance time series (stationarity and how to handle this.)
  - Explain about how to select appropriate candidate models for estimation.
  - All contents should be supported by reliable references.
- New list:
  - For offline estimation, explain about the estimation methods used for each candidate models in detail.
  - Provide examples of time series or other related plots to confirm about solar irradiance characteristics that you have described.
  - For online estimation, explain the principle of recursive estimation in detail. You can explain a simple method first, which is a recursive least-squares. Explain by math, not by wording.

Wed Nov 8

- Revise the old contents. Revise the report according to comments (technical details, English, math notation).
- Finish last-week contents: explain SARIMA model description and how to estimate it. Explain by diagram in each step and provide mathematical description of each step.

- Revise report structure. Problem statement includes the description of the problem, not everything you have read.
- Explain the maximum likelihood estimation of ARMA model. Derive the loglikelihood function and explain how to solve the problem.
- Explain how to obtain ML estimate of ARMA model from the libraries in MATLAB and python. What are their difference? Give some numerical examples (generate time series from a ground-truth ARMA model and estimate its parameter).

Wed Nov 15

- Revise the old contents. Revise the report according to comments (technical details, English, math notation).
- Revise report structure.
- Correct the estimation experiment and make sure you get correct results. Estimate SARIMA model from a time series you have generated from a groud-truth SARIMA model. Explain the experiment description and always attach your codes.
- Explain the prediction equation once your model is estimated. Test the prediction experiments (and always explain the experiment descriptions.)

Wed Nov 22

- Revise the old contents. Revise the report according to comments (technical details, English, math notation). YOU HAVE NOT CORRECT SOME CONTENTS FROM LAST WEEKS.
- Provide a diagram of all procedures involved in this project. Make it detailed enough and associate them with the goal and scope of the project.
- Revise the report outline (according to my comments on your senior project report). Right now, there is NO CONTENTS in the introduction, which is absurd. Figure 1 just explain only some part of this project is all about.
- Make understanding about the commands `infer` and `forecast` in MATLAB or any other commands/programming language you use in order to compute the estimation of the model, and the forecast of the responses. Are `infer` and `forecast` the same thing? Check with a simple model; for example, AR, or ARMA. Compute the fitting and forecast by yourself (according to the equations they are supposed to be) and compare with the outputs of `infer` and `forecast` commands. If things all agree, apply ARIMA model to compute the forecasts or the fitting.
- Experiments: always explain what you're going to do and what you expect first. Then discuss the results if they agree with your hypothesis. This week, you should do more experiments on real data, once you understand everything correctly.

Wed Nov 29

- Revise the old contents. Revise the report according to comments (technical details, English, math notation). SOME PARTS ARE NOT REVISED ACCORDING TO MY COMMENTS FROM LAST WEEK.
- Just finish the experiment of estimating SARIMA model and computing forecasts. Compute the forecasts by yourself and compare with the results from MATLAB commands and see if they are the same. Just finish this experiment on a specified ground-truth SARIMA model.

- To make the results valid, set the experiment to follow ideal assumptions. Report the results in an average sense as well because one result from one particular run (on one data set) do not mean anything.
- Complete the report structure and contents. Have a clear and valid conclusion.
- Always put comments on MATLAB codes.

2 Subspace Identification of EEG Time Series

Student list:

1. Satayu Chunnawong

2.1 Progress deadlines

Wed Oct 18

- Typesetting report of problem description.
- Explain the application and define the problem explicitly; the model, model characteristics, input, and output.

Wed Oct 25: Class skipped

Wed Nov 1

- Revise the old contents:
  - Revise the report according to comments (mostly about notations, math symbols, and English).
  - All contents should be supported by reliable references.
  - Explain clearly which variables can be measured, which cannot. The goals are two folds: i) estimate the model and ii) used the estimated model to infer brain connectivity. Seperate these two points explicitly.
- New list:
  - Explain general characteristics of EEG time series. Explain general principle of measuring EEG signals (what signal we observe exactly.)
  - Provide example of time series plots of EEG signals (there are various kinds: resting-state, seizure, etc.) and explain if the characteristics of EEG (such as nonstationarity or any other) can be observed from the time series.
  - Study about the model structure assumption. For example, how to determine the size of state variables.
  - Explain in principle about the model estimation method for your case.

Wed Nov 8

- Revise the report according to comments (technical details, math symbols, and English).
- Revise the report structure. Problem statement is the description of the problem. The methods used to solve the problem can be separated in another section. Your problem statement are two folds: i) estimate state-space model of EEG signals and ii) conclude GC causality from the estimated model.

- Explain all the methods involved in a diagram.
- Correct the state-space represenation used to investigate GC causality.
- From the state-space, derive (by hand) how DARE will be simplified and examine some conclusion about the structure of DARE solution (P). Does P have any structure? zero in some blocks?
- Include the experimental results on verifying GC causality where you should use a groud-truth AR model of p lags. Explain and discuss the results. Give the F measure and $C_1 B^k K_2$.

Wed Nov 15

- Revise the report according to comments (technical details, math symbols, and English). My previous comments from last week have not been taken into revision.
- Reorganize the report structure. Background should contain two topics: EEG model and Granger causality (for a time series, AR model and for a general state-space model). Problem statement contains two problems: subspace identification and learning GC. Methodology explains two main topics: subspace identification and GC measure on state-space model derived for EEG model. Explain clearly that there are two measures: F and the coefficient $C\mathcal{A}_c^l K$. Give a proper reference.
- Derive the reduced DARE for AR(p) (not just p=2).
- Revise the MATLAB code. If DARE is reduced, then you can check the result computed from the reduced one. The choice of W_1 is trivial; just try with another value.
- Develope a MATLAB function used to compute two GC measures from any state-space system. Give an example of the results.
- Write MATLAB code to perform subspace identification from a time series. Use the estimated system for GC test.

Wed Nov 22

- Revise the report according to comments (technical details, math symbols, and English).
- Section 1.2 Granger causality needs a huge revision. Explain things in a right order and just present main results. Split the section to two paragraphs: GC causality for AR models and GC causaliy for state-space model. When presenting main results, always give references and express the results so that the reader know they are important.
- Problem statement needs a clearer description. What model we estimate first ? What model to learn GC causality ? Do not just explain by words but use cross referencing to refer to model equations. Always explain reasons of the problem statement.
- Methodology: Do we estimate a general SS model or SS from AR ? This is important and I'm not sure if you still understand the problem correctly. Explains the diagram from the methodology by wording as well. Do not explain just by procedure but explain reasons of each steps.
- Section 3.2 is basically a summary from Seth paper. Give the reference.
- The derivation of DARE from state-space model of AR system is NOT the main topic of this project. They are just known results (we do not prove anything new) and we can use them to double check our understanding. Therefore, move all the contents here to appendix. I suggest to use a new notation.

- In experiments: explain first what you're going to do, and what your assumptions are. What are expected results from theory and compare it with experimental results. Always discuss the results. At the moment, your experimental results are wrong. I can suggest two possible causes: 1) you may try using subspace identification toolbox because their code is more optimized and consider many parameter settings 2) did you check if P and P^R are always positive? This relates to assumption for solving DARE. Study the theory and relate with the experiments.

Wed Nov 29

- Revise the report according to comments (technical details, math symbols, and English).
- Complete the experiment on verifying GC on state-space model. To make the results valid, set all the experiment to follow ideal assumptions. See if the F measures should be the value it's supposed to be. Report the results in an average sense as well because one result from one particular run (on one data set) do not mean anything.
- Complete the report structure and contents. Have a clear and valid conclusion.

3 Identification of Building Temperature System

Student list:

1. Chanthawit Anuntasethakul
2. Natdanai Sontrapornpol
3. Natthapol techaphangam

3.1 Progress deadlines

Wed Oct 18

- Typesetting report of problem description.
- Explain the application and define the problem explicitly; the model, model characteristics, input, and output.

Wed Oct 25: Class skipped

Wed Nov 1

- Revise the old contents:
  - Revise the report according to comments (mostly about notations, math symbols, some assumptions on the parameters, and English).
  - All contents should be supported by reliable references.
  - Explain clearly what variables and parameters are. Which variable/parameter can be measured or known.
  - The system description lacks of dynamical equations. It's not clear what the input signal is and can we manipulate this variable or not (if you let it be the energy used by air.)
- New list:
  - Make a better and clearer description of system. The dynamical models are in continuous-time or discrete-time ? Simplify the notation and try to use a vector form.

- Explain how one can sample the continous-time system to discrete-time system. If we can identify the discrete-time system, how can be retrieve the parameters in the continous-time ? Explain mathematically.
- Explain in principle, what estimation method will be used in your case.

Wed Nov 8

- Revise the report according to comments (technical details, math symbols, and English).
- Revise the report structure. Problem statement is the description of the problem. The methods used to solve the problem can be separated in another section. Background section can be used to explain about the model first.
- Revise the variable notations so that they are consistent (T_c, T_h, T_i) and others.
- From the first law of thermodynamics, explain more what Q and W mean. Derive the complete dynamical equations of each room (may derive with notation subscript i) and see if you have obtained nonlinear or linear ODE. Explain what standard techniques we can use to approximate nonlinear ODE to linear ODE. State clearly what state variables, input, and disturbance are.
- Derive state-space equations from the approximated linear ODE you obtained (continuous-time) and see if you have a pre-defined structure in (A, B, C, D). If you have a structure, just examine how that structure will be carried over when the model is transformed into the discrete-time model
- Explain what estimation method you will use in order to estimate the model. Can you do subspace identification?
- Start to explore about the measurement data. Give some example of plots.

Wed Nov 15

- Revise the report according to comments (technical details, math symbols, and English).
- Distinguish between the equations that can be explained by some principle, and the equations that are proposed. If an equation is derived from some principle, you can just give a proper reference.
- Reorganize the report structure. Background contains AC system, Building structure, Heat transfer. Modeling contains the deriviation of (nonlinear) dynamical system equations, approximated model, linearized model, discrete-time model. The approximated model is the one you assume COP is a constant. These models should be explained in a standard notation $\dot{x} = f(x, u, w)$. For nonlinear model, you can treate $u, \dot{u}$ as two inputs (which are known but they are independent which is okay for subspace identification purpose.). Problem statement explains what you are going to do with the approximated and linearized models.
- In methodology section, explain what estimation method you're going to use and which model you will apply on.
- In Data section, explain the variables, parameters, pre-processing, etc.
- Run preliminary estimation experiments.

Wed Nov 22

- Revise the report according to comments (technical details, math symbols, and English).
- All experimental figures should be plotted in MATLAB and exported to .eps. Use

  ```
  print -depsc filename.eps
  ```

- The dynamic equation should be completely provided after (10), and the equation should be stated clearly as T_1 and T_2, not just T_i. You can explain into two cases: complete equation with the derivative of COP and the other case is to assume COP as a constant. Give a number of equation for referencing. Then when you explain state-space model in 2.2 you can refer to the part of dynamic equation. Explain the structure of (A, B, Γ) if any. Explain that w can be measured, so you can treat it as input term in the state-space.
- Problem statement part should be MORE clearly explain in detail. It seems that your state variables can be all measured, then you just apply the least-squares method, not the subspace identification.
- Think about how to present the experimental results. Printing the values on the screen is pointless. Always discuss the results.

Wed Nov 29

- Revise the report according to comments (technical details, math symbols, and English).
- Correct the method of solving least-squares with linear constraints (zero constraints, in deed). If you transform all the matrices to vectors, you can use some command in MATLAB. However, if you would like to keep the matrix variables, you can use `CVX` http://cvxr.com/cvx/
- Consider the model obtained by substituting parameter values from the manufacturer datasheet. Compare the fitting results with your model.
- Reorganize the way you present the results.
- Complete the report structure and contents. Have a clear and valid conclusion.
- Always put comments on MATLAB codes.

4 Modeling of Photovalatic System

Student list:

1. Janenarong Klomklao

4.1 Progress deadlines

Wed Oct 18

- Typesetting report of problem description.
- Explain the application and define the problem explicitly; the model, model characteristics, input, and output.

Wed Oct 25: Class skipped

Wed Nov 1

- Revise the old contents:
  - Revise the report according to comments (mostly about notations, math symbols, some assumptions on the parameters, and English).
  - All contents should be supported by reliable references.
  - Explain clearly what variables and parameters are. Which variable/parameter can be measured or known.
- New list:
  - Do a literature review about a simple model on PV system. You can start from ajarn Surachai's previous work. Provide the list of models and reference. You can pick a simple one.
  - After you simplify about notations, explain how the generated power is a function of T and I. Explain what that function is (how nonlinear) and explain what estimation methods used in the literature.

Wed Nov 8

- Revise the report according to comments (technical details, math symbols, and English).
- Revise the report structure. Problem statement is the description of the problem. The methods used to solve the problem can be separated in another section. You should have a section "Background on PV model" that explains the models in the literature and the one you're considering.
- For models from literature, try to adjust the variable notation to be consistent with other sections. Express the problem of estimating parameter in an optimization problem. That is we should see the cost objective and optimization variables clearly. Then summarize what optimization technique used in each model estimation.
- For your model, specify the model estimation method you will use and express it as an optimization problem (with cost objective and optimization variables.)
- You can come up with another simple model (for example, polynomial model that used in one of the papers you used to read).
- Start to look at measurement data from CUBEMS. Give some examples of the time series plots and see if some data values are missing.

Wed Nov 15

- Revise the report according to comments (technical details, math symbols, and English).
- Distinguish between the equations that can be explained by some principle, and the equations that are proposed. If an equation is derived from some principle, you can just give a proper reference. I have specific questions in the comments, please read and explain more in the report.
- You should consider two models: nonlinear and linear model (polynomial if you like) and state the numerical methods used to solve the two model estimation problems. Explain how to choose an initial guess of the parameter in the numerical method. Provide preliminary results of estimation from the real data.
- Explain about data preprocessing and how to split the data sets.

Wed Nov 22

- Revise the report according to comments (technical details, math symbols, and English).
- All experimental figures should be plotted in MATLAB and exported to .eps. Use

  ```
  print -depsc filename.eps
  ```

- Check the maximum power equation (35) again and see the attached paper.
- In the experiments, explain the estimation experiments: how much data you use for training, estimated parameter values, fitting results, and compare these between nonlinear and polynomial models. Is polynomial model a nested model of nonlinear one? If yes, what can you say about the fitting comparison theoretically?
- Check if you have similar constraints on the parameter sign in the polynomial model as well. If yes, the estimation problem is least-squares with inequality constraints. This can be solved in MATLAB too (as it's a special case of quadratic programming).
- Attach example of MATLAB codes in estimation part to see how you put the constraints and cost objective.
- It will be interesting to apply ANN (neural network) to train as a substitute for nonlinear model as well (this has been used in the literature.) Ask Poomipat for help.

Wed Nov 29

- Revise the report according to comments (technical details, math symbols, and English).
- Compare all the models: nonlinear model, polynomial model 1, polynomial model 2 (with constant term), ANN models. Make the results clear to make a comparison.
- Always put comments on MATLAB codes.
- Complete the report structure and contents. Have a clear and valid conclusion.