Report private PhD defense Gustavo Quintana-Carapia

VUB, Friday May 15, 2:30 pm.

The public defense is planned on Monday September 21, 2020 at 4 pm.

The jury decided that **minor changes** need to be made to the text. However, the thesis text needs to be approved by the promotors before final submission.

Below, we give a summary of the main changes requested, as well as a summary of the questions posed during the defence. The candidate is invited to use these comments and suggestions for improving the manuscript.

The candidate is required to submit a well-documented list of changes to the committee at the latest one week before the public defence.

Required changes

- The manuscript text should better position the PhD to related work. Improve the literature review and discuss (and cite) the state of the art.
- Add more clearly the motivation for the PhD research work by including motivational examples to the introduction.
- Add a discussion with comparison of existing EIV methods. Give, with the positive and negative aspects in mind, a better justification for why you use a (R)LS method, and not, e.g., (R)IV method.
- Discuss more clearly the aspects of scalability of the methods and their limitations in terms of model order, size of data sets (e.g., related to the inversion of LS matrices).
- The equations on pages 12 and 13 seemingly contain errors and should be corrected.
- The figure shown on page 66 was made with a single experiment; it is desirable to see the (averaged) effect on several experiments.

The members of the jury may provide the candidate directly with a list of detailed comments and corrections.

Questions and comments raised during defense

A summary of the questions and comments is given below:

Lyudmila Mihaylova

- You develop results for uncertainty quantification; what type of uncertainties can your approach deal with and which levels of noise can you consider? Elaborate on SNR levels; I'm also very interested in scalability, so how big the matrices can be.
- Concerning the scalability to data: eqn 3.6: how scalable is the approach in all aspects (order, data, linked to inversion for LS matrix); can you deal with hundreds, thousands of data?
- What recommendation up to scale/order can method work with. What are the limitations? It would be good to have more precise discussion about these aspects in dissertation: what do you mean by large scalability: 100, 10^4, ...? A more detailed discussion is desirable.
- On slide 25 (noisy graphs) it would be good to see a statistical analysis (more experiments than one).

Stephane Chretien

- The Taylor expansion is computed of order 2, what is the accuracy of expansion with respect to system, noise level,... Can you come up with a priori bound on error that will allow you to assess the quality of the expansion? Can you predict where the solution is going to be before computing Taylor expansion?
- Concerning other sorts of prior estimations, e.g., given by explicit function theorems. I wonder if you could get something and apply this bound in order to control the bound of the 2nd order expansion. There exist some good results that could help you devise something a priori which can control the error you are making. Neuberger's results (quite old now) but may be handy for this kind of problem. This can be investigated further.
- About trying to find a good transformation from LS to EIV model. I am wondering if you can think of other more accurate starting points instead of LS. Closer to the problem you want to study, but still preserving computational lightness? The closer you start from objective, better the Taylor expansion would be.. Something to investigate as well...
- About robustness: is it an issue in practical problems with heavy tailed noise, outliers,... Perhaps the use of *median of means* plugged into your approach

might produce interesting alternatives for outliers or heavy tailed data. But this is an intricate thing to deal with. Might be interesting to ponder this in future.
- On deep learning: could this be an alternative?

Guillaume Mercere

- In the first or second sentence: *"users cannot wait for a steady state regime"*. Do you have specific examples of real industrial systems for which it is necessary to have continuous read-out before steady-state is reached?
- Further on the industrial part and selling sensors with data-driven parts. Can you explain to experts that your solution is better? Sensors will be used for ages; why not spend one day to do system identification and get a reliable model to make a compensator, instead of using your approach?
- Chapter 2, Section 2.2.1. Can you give me some classic solutions/algorithms instead of LS in general? (Say I'm not familiar with EIV or TLS?) Usually with LS problems: multiply with K^T and compute pseudoinverse. Do you know another solution where they don't use K on the left-hand-side but something else? To get rid of noise by using a matrix on the left-hand-side, is related to EIV. There are some recursive solutions for this. It could be interesting to compare with such IV solutions.
- In Chapter 3: at beginning you say to use linear systems theory to model a sensor. Can you discuss this assumption: is it true in practice? Can you guarantee it behaves like an LTI system?
- Equation 3.1: an LTI system, described with SS representation. You only put output noise on this representation. People sometimes also add process noise. Why? How to modify your approach if there is process noise? Is it still white in this case?
- On page 12 standard approach top of page equation: there are some mistakes in the equations! there is no epsilon (noise) in the equation. But then you do an SVD of H and estimate from there. How does it work if you have noise? Have you tried implementing it and see what you get?
- On page 12: I'm not sure the equations at the end of the page are correct, e.g., Y=G ubar ... not sure there is x; same for rest; There seem to be many mistakes here and next pages. Be careful with the equations!
- On page 13, you assume persistency of excitation of order L. How do you check this? In your case the input is constant; how to guarantee that the assumption is met?
- In Chapter 4, p 17, Section 4.1.1: You assume that perturbations are IID and must be normally distributed. Why using an assumption of normally distributed? I would think that the kind of distribution is important for CRLB,

but I'm not sure you need it for eqns 4.8, 4.9, 4.10,... IID is important, normality perhaps not.

- In equation 4.11 you have sigma_eps^2 + sigma_e^2; how to get in practice sigma_e? Do you have to do this each time? Recursively?
- Switch to 4.2, simulation results: first part dedicated to K, randomly generated; not sure if still possible, but in all your results it would be nice to have somewhere explicitly K = ... (even if randomly generated). You use only one realization also for matrices A, B, C, D in the text, so it can be added explicitly.
- in Fig 4.1 the bottom figure also discusses the structured case. Why do you talk about structured results in a section about the unstructured case?
- In 4.1 uses a constant K, but in 4.2.2 you use A, B, C, D, without any links to K. You compare the two curves generated from different data sets… It is strange to compare them. I don't see the link between both.
- To show comparison between structured and unstructured case, use unstructured results for structured case; then use unstructured results for structured case and cover all the possibilities. Now you compare curves related to two different simulations.
- I was surprised to read in Chapter 5 something very close to what I read in chapter 4. In both cases generated A, B, C. In 5.1 again oscillating 5th order system, similar conclusions, ... What is the added value of 5.1 wrt 4.2.2?
- On the impact of selected order on results. You were talking about a related bias/variance trade off. So, do you know in the literature some solutions to avoid overfitting? Can you cite something? Can you use something in your case to test for a good order of system?
- Do you know of cross-validation: split data into two parts or several parts; some for estimation, some for validation? Have you tried this? Can this be used here to select good orders? Better results? Influence of system order on bias/variance; if I only use one dataset it is complicated and usually you can increase complexity to mimic a data set, but to avoid overfitting (too high order) use a test/validation data set. with this i should see what is the best value... Have you tried this?
- In Section 6.2.2. Why do you call it a max likelihood approach? To me it is not one. Why do you call it like this? Where is the likelihood? Depending on which likelihood (eg. Gaussian, ...) it leads to LS. Don't' call it ML if you dont give distributions on ... In this case it is better to call it a nonlinear minimization procedure. Not clear why you call it ML.
- Concerning the ML approach in simulations, you say that it takes 30 seconds to complete. What is done exactly in these 30 seconds; how many samples, realizations,... I am a bit surprised you require 30 sec for 2 parameters.

Nikos Deligiannis

- When solving algorithms. Do you opt for closed-form solutions, or do you use optimization based algorithms? How to solve LS problems? Do you solve them using pseudo-inverse or do you do gradient descent?
- Today there is an overwhelming amount of data sensed and communicated; How would you apply your method in a scenario where dimensionality increases significantly? This relates to the questions of previous jury members; what is the implication of using numerical versus closed form solutions?
- What if the $K^T K$ matrix is not invertible? What if you have fewer samples than unknowns; what can you do? Solutions that assume structure like in compressed sensing, regularization,... can you use them, what would be the impact of that?
- You mention that you improve complexity with your approach, but I missed a complexity analysis. Did you do this? Anything in processing time, flops,...?
- You mentioned deep learning could be applied; could you show how? An issue is that neural nets are trained in high SNR data, but applied in engineering with lower SNR. This clogs performance; any ideas? Is this a limitation?
- You mention that you consider Kalman filters; have you seen approaches where they write Deep Learning in form of Kalman Filter; Could this be applicable?
- Recommendation: if you revise the thesis: maybe change "I propose" to "we propose".

Philippe Dreesen

- Your main assumption is that a sensor outputs a steady state value that is proportional to the to-be-measured quantity. This means that you assume that the sensor is a linear system. What if this assumption is not met, and the sensor has a saturation for high values? Can you still use your methods?
- Sensors are physical devices and deliver a continuous-time output. The models that you use are in discrete-time. Are there any risks that you miss information by not sampling fast enough?
- Related to earlier questions about compressed sensing. There are some recent results on the connections between sparsity and low-rank Hankel matrix completion and approximation. Say that you have an issue with missing data, e.g., from sensor outages. Would it be possible to plug in a

procedure that does structured matrix completion and approximation into your methods?

Roger Vounckx

- Two practical questions. On page 35-49 you consider a weight of 138.x grams. You sample relatively slowly (4 kHz) and after 500 samples we can consider that we have correct estimation. If I have a container that fills up and weight is constantly changing. Can we tackle this?
- Consider an optical fiber link with a detector capable of 100 million measurements per second. Can I speed up the link by using your method? You have to do calculations: how high can you go in speed before running into troubles with DSP?

Rik Pintelon

- Returning to a question by Guillaume Mercere: on page 17; do you need normality? You made 2nd order taylor series expansion. Which moment do you need to estimate? Look at the equations page 16. There are 3rd, 4th order moments,... you need to know something more about the noise; this is one of the reasons to make Gaussian assumption: 3rd order moment is 0 (symmetry); 4th order for Gaussian is easy to have an explicit relationship.
- Returning to a question by Philippe Dreesen on missing data. If you can deal with missing data, you can deal with saturation in sensors: you just remove the saturations. What is the rank of the Hankel matrix? Is it full rank or not? The question is if matrix completion can help? It works in low-rank approximations; not if the matrix is of full rank.

Ivan Markovsky

- About scalability of algorithms: You said in the very beginning that it can be very fast compared with system identification. This is the point: can you think of a good answer for these questions on scalability? Nikos Deligiannis mentioned computational analysis. You can directly say for (R)LS what is the computational complexity. Order three if full matrix inversion, now per iteration only linear. So you can scale to very large data sizes.

Detailed comments provided by jury members

Guillaume Mercere

- It is important for me to know why these new developments are generated. That is the reason why I suggest adding sections or paragraphs dedicated to
  - the reasons why, in the industry, we need such a new technique. As I said during the defense, do you have a list of sensors for which waiting for the steady regime for getting the measurement is not possible? Why do you need a fast algorithm? What are the reasons why industry people need fast algorithms? In the end, it seems that, for you, it is more important to have a fast biased result than an unbiased estimate. Why? What do you mean by fast?
  - the literature on EIV estimation problems. As shortly mentioned in Chapter 2, there are solutions in the literature (by the way, you never mention the recent book written by Torsten Soderstrom in your document which is strange for me). But you do not describe them (even shortly) and, more importantly, you do not criticize them. Not in general but, at least, explain why you do not use them, what are their advantages and drawbacks and, finally, w.r.t. this list of positive and negative points, why you choose a (R)LS solution. E.g., why not using a recursive instrumental variable solution instead?
- In the document, you always talk about recursive implementations of least squares solution but, in the end, you never give access to any recursive algorithm description (as you do for what you call the ML method). Why? Do it as well for the recursive weighted least squares solution.
- Double check your equations page 12 and 13. By the way, can you explain where Eq. (3.6) comes from?
- In Section 4.1.1 and 4.1.2, discuss the normality assumption. Please restructure Section 4.1.3 which is not clear for me.
- Why do you always study the statistical properties of your least squares solution with high SNR only? What do you get for SNR = 10 dB for instance?
- Give the values of the matrices you use for simulating data sets so that the reader can test what you have carried out.
- Please remove Fig. 4.1 b from Section 4.2.1 because the top and bottom curves are obtained with different systems and simulation condition. Or explain why you can compare both. It could be interesting to test each time (i.e., in Section 4.2.1 and 4.2.2) the bias and covariance formulas obtained by considering the structured and unstructured EIV problems.
- When you generate different simulation examples, explain why the new ones are interesting or, more importantly, what they bring w.r.t. the former

simulations results. This is the case for instance when I compare Section 4.2.2 and Section 5.1.

- Except if you can convince me, do not use the name "maximum likelihood" when you do not introduce a likelihood you want to maximize. In Chapter 6, you introduce a nonlinear least squares algorithm, nothing else (see the attached document, Numerical Optimization, Chapter 10).

Nikos Deligiannis

Comments related to the structure of the thesis:

In the introduction chapter it is best to describe a set of applications and settings that motivate your research and proposed solutions. This can make your thesis more accessible. Furthermore, you can add a subsection that describes in a high-level manner the remainder and structure of the thesis (describing briefly the contributions of each chapter). You can also mention your contributions there but in a less technical way since the problem statement only comes in the Chapter preliminaries.

In each main chapter (Chapter 4, Chapter 5, Chapter 6) you can add an introduction subsection, a state-of-the-art subsection and then elaborate your contributions and methods.

I believe that this structure will improve the understanding of the thesis and light up your contributions.

Further comments:

Please discuss the complexity and scalability issues related to your approach in relation to the state of the art.

Position your work with regard to the state of the art as well as regarding solutions based on low-rank matrix completion and compressed sensing.

I recommend to use the formulation "we propose" than "I propose"

You might wish to add a list of notation in the beginning of the thesis.

Philippe Dreesen

General remarks:

It would be good to include some layman's examples that motivate the work of the thesis already in the introduction. In the current text, there is a nice example on page 52, almost at the end of the thesis. These kinds of examples should appear in the introduction to motivate and position the work.

Detailed remarks:
- Citations that are used in-line should have brackets only around the year, not around the authors' names. For instance on page 3, line -2 is written "reviewed in [Hack and ten Caten, 2012]". This should be displayed as "reviewed in Hack and ten Caten [2012]. In LaTeX this can typically be done by making the proper distinction between the commands \citet and \citep. The former is used for references in text, the latter for parenthetical references.
- On page 1 of the PDF (front of thesis), I believe also N. Deligiannis should be mentioned with the title "prof.".
- Page 1 of thesis: typo on line 6: "and on the sensor initial conditions" => "sensor's initial conditions"
- Page 3, line -2: "uncertainty analysis are reviewed" => "is reviewed"
- Page 3, line -1: remove comma before "that"
- Page 3, line -2: fix the in-text reference (see above)
- Page 4, lines 11-13: fix the in-text references (see above)
- Page 5, line 5: Remove "The" at the beginning of the sentence (The Hankel…)
- Page 6, line 10: fix the in-text references (see above)
- Page 7, lines 10-11: the sentence on heart and breathing physiological monitoring sounds a bit awkward. Please reformulate.
- Page 9, line -1: fix the in-line reference (see above)
- Page 10, lines 1-3: in the enumeration of the types of sensors, remove the article 'the' before the sensor name. So: "like three-axis sensors [ref], radio-frequency intruder sensors [ref], and radar sensors [ref].
- Page 10, equation 3.10: The symbol D was not bold in the equations above. Be consistent!
- Page 10, caption of figure: replace "reverts" by "inverts"
- Page 11, line 9, put "poles" between parentheses instead of between commas.

- Page 12, line -6: "that in matrix form is" => "which in matrix form is"
- Page 13, line -9: "in the practice" => "in practice"
- Page 14, line -14: missing space between "problem" and "(3.11)"
- Page 16, halfway page: the sentence "Applying a second-order…" is not a proper sentence (perhaps fix this by removing "that").
- Page 16 introduces seemingly some new notation, e.g., b(.), mu(.) and C(.). While their meaning is clear from the context, it is desirable to give proper definitions when introducing new mathematical tools.
- Page 17, line -1: fix in-text references (see above)
- Page 20 (and beyond): The abbreviation CRLB is introduced, but sometimes CRB is used instead.
- Page 21, line -10: fix in-text reference (See above)
- Page 35, second sentence "The sensor is a dynamic…" is not a proper sentence.
- Page 50: Try to reorganize this section into two or more paragraphs.
- Page 59, line 2: fix in-text reference (See above)
- Page 74, line 11: "The learning obtained from…" this is an awkward sentence; please reformulate
- Page 81, all three journal publication references do not have (or have wrong?) page numbers.