

Asst. Prof. Dr. Vedran Dunjko

Leiden Institute of Advanced Computer Science
Leiden Institute of Physics

Snellius Gebouw, Niels Bohrweg 1, 2333 CA Leiden,
The Netherlands
v.dunjko@liacs.leidenuniv.nl



Review of the Doctoral Dissertation

"Application of machine learning in quantum computer science"

by **Mgr. Mateusz Ostaszewski**

The presented Dissertation "*Application of machine learning in quantum computer science*" introduces original contributions to the interdisciplinary fields of quantum machine learning and near-term intermediate-size quantum (NISQ) algorithmics. Its main aim is at capitalizing on methods classical machine learning and optimization for the benefit of quantum information processing.

In broad terms, the two main research themes discussed in this work address: 1) the applications of machine learning in the problem of optimal control -- a notoriously difficult and very relevant problem in the NISQ era; and 2) the problem of classical optimization of quantum computational parameters in so-called Variational Quantum Eigensolvers (VQE). The latter is a quantum realization of a variational method for estimating ground states and energies, where a parametrized quantum circuit serves as a parametrization of the search manifold -- a *quantum-realized Ansatz*. Although it is not presented in this way in the Dissertation, both problems are, in a broad sense, control problems -- in both case we deal with optimizing pulses as a function of time, which, in the VQE case specify the circuit. Thus, although these two areas are not typically considered closely related, I feel this thesis has an appealing underlying cohesion.

The presented Dissertation collects three first-author papers of the candidate, developed in collaboration with two distinct groups. The thesis is very short, even frugal, and after a very brief introduction to topics of quantum mechanics, quantum computing and control, and machine learning (16 pages of double spaced text for this substantial content), presents the three main Chapters: Chapter 3 & 4 dealing with machine learning for optimal control, and Chapter 5, proposing a new algorithm for VQE optimization. These Chapters contain the three published papers, essentially in their published form. The thesis finishes with a short 2 page summary of the main findings. The writing and presentation of the thesis is not the strong point of this Doctoral work, and this aspect could be much improved. More importantly, however, the research itself is innovative, timely and generally very solid.

Beyond the three papers discussed in the Dissertation, the candidate has co-authored further 7 published papers and another unpublished manuscript, totaling 11 publications. The topics of these works involve explorations of quantum walks (very nice work), and quantum machine learning and control, which are all timely and important themes. The works are published in reputable journals as well.

The overall academic output of the candidate, and research quality, evaluated on the basis of the sample presented in this Dissertation leaves no doubt that Mateusz easily meets the standard academic requirements for a PhD title.

**Consequently, I recommend that the committee grant a doctoral degree to
Mgr. Mateusz Ostaszewski
for his work in the domain of quantum computing.**

I support this recommendation with a more detailed review of the work presented next.

Detailed Review of the Doctoral Work

Theme summary

The presented thesis deals with two related, but sufficiently distinct themes that should be discussed independently. The first problem, discussed in the third and fourth Chapter of the Dissertation deals with the application of two different classes of machine learning methods for the problem of optimal control.

Specifically, the candidate considers the problem of optimal control in the presence of unknown drift fields — parts of the Hamiltonian that are not under the experimenter's control. As usual, the objective is to realize a target quantum evolution, however, instead of tackling the problem by trying to control the drifting Hamiltonian, the objective here is to learn the correlations between optimal drift-free parameters and parameters tuned to counteract the drift (with one fixed, or multiple possible drift strengths). The end result of the approaches is an algorithm which can, given a sequence specifying the target evolution for the drift-free Hamiltonian (in some versions, and a choice of target drift strength, in some for a fixed value), outputs a sequence which will counteract the drift, and realize the target evolution.

To solve these problems, the candidate experiments with bidirectional LSTMs — a well-known recurrent neural network architecture successful in certain time series problems (like machine translation) — and with a simpler combination of k-means clustering and k-nearest neighbours classification methods.

In the second theme, the candidate proposes a new algorithm for the gradient-based optimization of variational circuits in a VQE context. The algorithm is based on a nice observation characterizing the dependence of arbitrary cost functions on one free variational parameter, which leads to a very explicit and simple coordinate descent-type algorithm. Further, beyond just optimizing field strengths (“angles” or circuit

parameters), the candidate considers the optimization of directions (specified by Pauli operators), in a more general process of architecture optimization.

Relevance in the specific field

The problem of optimal control is deep research area and has been explored for decades in various context. I am not an expert on this topic, but I am aware that various variants of the question have been addressed in literature. For the last 10 years, there has been an influx of machine learning techniques to this field, often relatively complicated ones. The particular question of translation from ideal to the non-ideal case raised by the applicant is not what I commonly see in literature, but certainly makes sense to me. I believe that this question may be quite relevant as (near-)optimal pulses for drift-free cases are indeed known for many problems, so having methods to move from know to unknown would be valuable. Beyond this, the candidate also briefly discusses how the ML insights could be used to derive more general rules connecting optimal drift-free and drift counteracting sequences, which is obviously interesting.

The second problem of VQE optimization is an absolutely trending topic nowadays, as it is vital in both quantum-computing-for-quantum-chemistry applications, and in quantum machine learning. Many conventional methods have been put forward, and we are far from a clear winner. The insights introduced in this thesis, and the proposed algorithm will I am sure prompt substantial follow up work.

In summary, I am convinced this thesis addresses timely questions, presents highly relevant work, and contributes substantially to our knowledge-base in these topics.

Methods, results, discussions

I review the methods results and discussions for the two themes separately.

Theme 1: in Chapters 3 and 4, the thesis explores two main ML approaches for the learning of drift-correcting sequences, and the methods are adequate. The models are evaluated via random set cross-validation, and the experiments are done on a collection of relatively natural instances.

One of the main questions, of whether the method can generalize to larger control systems remains mostly open as most work is done on 1 and 2 qubit, relatively simple examples, but I do think this is justified by virtue of the presented methods being novel, and exploratory. Similarly, the problem setting has other scalings (time, number of intervals, etc.), and I hope follow up work will shed more light on the actual value of the methods.

Both methods achieve solid results, with fidelities in the high 90's, which looks impressive. From the presented thesis, it is not entirely clear, however, what should be considered a breakthrough result in this. Nonetheless it is clear that the results are promising.

The discussions are short, and mostly consist in a result summary, and some interesting possibilities were left unexplored in the presented exposition (I give some

examples later). In general, the entire thesis could have capitalized a lot more on establishing connections between the individual results presented, and connecting the novel work to other state-of-the-art.

Theme 2 deals with VQE optimization. The candidate introduces a novel, coordinate-descent-based algorithm, and a method to find the optima exactly using very few expectation evaluations. The algorithm is extended to allow for also optimization of rotation axes, as an aspect of architecture optimization. I note that the latter is only tractable due to the coordinate descent approach, and I find this very interesting.

The algorithm is empirically evaluated in a solid fashion, and benchmarked against common methods (Adam, SPSA), and the results look very impressive indeed.

I am looking forward to follow up work where the evaluation is done on broader classes of Ansätze, more depths, and for different problem Hamiltonians. I expect that this method may become one of the standard benchmarks in future works on VQE optimization. As in the previous Chapters, the discussion is very brief, and I have many follow-up questions I am hoping will be answered in the future.

While the thesis could certainly have been made stronger by exploring some of these relatively obvious questions, the mere fact that it is clear that these questions deserve follow up research is a confirmation of the quality of the questions, and of this preliminary work.

Thesis (Presentation, writing, language)

The presentation of the work is, in my view, significantly below the level of the work presented itself. The thesis is not a conventional monograph, but more of a collection of papers. Although introduction sections are added, they are absolutely minimal.

The thesis is 70-odd pages of content (certainly the shortest PhD thesis I have ever reviewed or read), and it collects 3 full papers, with generous spacing — so not much room for anything else remains. Further, since the Chapters are essentially the papers, some topics are introduced twice; I personally am not overly bothered by this, but it further evidences that not much space is dedicated to the background. Since the thesis deals with very different topics, all of which are quite deep, a more generous intro and background would easily cover another 30-60 pages, which with additional discussions would bring us to the lengths of thesis I am accustomed with.

I personally did not have problems following the text (the text being so concise helps), but I believe readers less informed about the presented topics would find it a challenge to fully grasp the main points without venturing to other literature.

To exemplify, I simply do not think readers outside of ML can grasp what bi-directional LSTMs are, or have any intuition about them based on the given description.

Finally, the thesis could be improved in structuring, and certainly in terms of language and typos/stylistic problems. The thesis could certainly be improved by a focused proofreading and perhaps re-structuring (e.g. the LSTM is split over two Chapters, even though the 4th Chapter is explicitly about geometric interpretation,

which LSTM is not, and 3rd is about LSTM exclusively. This makes little sense to me).

Comments, suggestions and points of clarification on the research

In the remainder of this review, I will provide (hopefully) constructive criticisms for the presented research. Note that per definition I now focus on the short-comings, and I will be critical. However, this part of the evaluation should not be taken out of context to be representative of my opinion on the Dissertation work in general. To avoid any chance of misunderstanding, as I have clarified, I believe this constitutes very good work, which will prompt a lot of exciting research, and that I unreservedly believe the presented work more than suffices for a PhD thesis.

The thesis is a collection of papers, with very concise introductions, and the papers themselves follow a minimal and “fast” paradigm: question - method - result.

What this is missing out on are the following. On the “question” part, what I find wanting is *context and related work*; the questions the author explores are very important, and consequently have been studied in literature extensively. However, the presented work does the bare minimum of discussing the mainstream methods, trends, strengths and weaknesses of current methods. This is missing out also on the didactic aspects as well, very much needed for a reader not versed in the presented topics.

Closely related to this is the framing of the work by *appropriate motivation and background*. Especially for topics of Theme 1, it remains undiscussed why study the particular questions, what are the benefits over direct methods, what are the shortcomings?

The next aspect is the *self-reflection aspect* of the work, which overlaps “questions” and “methods”. While naturally it is the scientist’s prerogative to choose their method to tackle a problem, it is a characteristic of a good scientist to have good motivating reasons to prefer one method over another. Is the method well suited? Does it have known shortcomings? Does it rely on some assumptions which may fail to be justified in my context? How difficult is this problem to learn at all? While I am sure the candidate had good reasons to choose one method over another, one set of hyperparameters over a differing one, unfortunately this is not reflected in the thesis. Building on a solid understanding of the problem and the method, rather than blindly choosing a method and “seeing if it works” contributes a lot more to research as it allows us to refine our intuitions and assumptions. I can summarize the above comments by a frequent thought I had in my head while reading the work: “Ok, you are doing this and this in this way... but **why** in this way?” — I believe the candidate probably has good reasons for the made choices, and I look forward to the Q&A session to gain some insight.

In a similar vein, the question - method - result method is minimal in the potentially valuable reflection of the overall result. While much of ML is about “getting the thing to work”, there are many, for some more interesting questions beyond the

engineering aspect. The thesis would have been stronger if this was more present. In Chapters 3-4, there is an obvious interplay between the structure of space of sequences and Hamiltonian perturbations (the candidate even explores this explicitly to some extent), and the applicability of the methods; what is required for the k-NN method to work? If it does work, what does this tell us about control problems? What does the fact that we need 500 clusters for optimal results tell us? Is this connected to the dimension of the problem. Let me give an example: in the 2 qubit problem, the thesis deals with training sets north of 8000. Now, if I equidistantly spread 8000 unitaries (w.r.t. the preferred distance measure, such as to minimize the distance between a point and its closest neighbour), what is the maximal error I get? If this is below or close to the performance of the ML, this can shed a lot of light on what is actually happening. E.g. understanding this would tell us if this is already a training set which allows heavy overfitting, in which case we are not actually learning the problem at all. This could have been not just discussed, but various interesting hypotheses could have been tested.

Finally, the discussion section is unfortunately very abridged, and due to this many possibly exciting connections between the findings of the sections/chapters have not been discussed. Further, this is the place where the reader could see a bit more about where the imagination and curiosity are driving the candidate, and have a glimpse of the candidate's thought process. I would have appreciated more hypotheses, conjectures, criticisms and evaluations of the candidate's own presented ideas.

I finish this review with a small selection of other questions that I could not find answered in the thesis, yet I find very interesting. Perhaps they may help the candidate to decide on his future research directions as well, but I also I hope I will be given the chance to fulfill my curiosity about some of them in the Dissertation defense/presentation.

Question 1 (Theme 1): Machine learning models always allow for a set of hyperparameters which can be tuned to obtain better results. In the presented work, it is not clear whether the presented results constitute optimized performance (in which case it would have been nice to see how robust this all is), or if this was not needed. What do we learn about the underlying physics from this?

Question 2 (Theme 1): It is not entirely clear what should be considered excellent results in the "sequence translation problem". Ideally, some experiments could be done in the future on cases where optimal achievable results are known (e.g. by extensive brute-force methods or from theory, if possible), to have an absolute benchmark. Is beating GRAPE possible?

Question 3 (Theme 1): What is the best application of the translation algorithm?

Question/Comment 4 (Theme 1): Chapter 3 and Chapter 4 have a common aspect which I could be better explored/connected: in Chapter 3 the question of the effects of perturbations on “input” drift-free sequences on the output “drift-compensating” sequences is studied. In Chapter 4, the candidate explores k-NN algorithms that rely on k-means-based clusters, which are distance-based. The the performance of this method is very much dependent on the nature of this perturbation mapping. The k-NN method used vitally depends on certain locality assumptions, and sufficient smoothness of the perturbation (which, if true also makes the problem much easier). I would have enjoyed reading some reflections on this, and it would provide more cohesion to the overall research directions. I hope this too will be pursued in the future.

Question/Comment 5 (Theme 1/2): Could methods of Theme 2 be applied in Theme 1, and vice-versa? Why, why not?

Question 6 (Theme 2): The candidate presents a comparison of Adam and SPSA against the algorithm with axes selection. This to me suggests that rotoselect has the capacity to change the rotation axes whereas Adam and SPSA, in their basic applications do not do this. This feels like an unfair comparison. I would like to hear this elaborated.

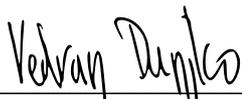
Question 7 (Theme 2): It is known that coordinate descent achieves global optima for certain classes of non-convex functions (the form is explicitly known). Does, for any Ansatz and cost Hamiltonian VQE optimization attain something close to this form? Can subsequent global optimization sweeps help attain true minima? What are we learning about the problem if coordinate search works?

Question 8 (Theme 2): How can the algorithm be made noise-robust?

Question 9 (Theme 2): Is this method avoiding barren plateaus? Can it be used as a good initialization method?

I will of course save some of the questions for the defense itself.

In Leiden, 5. April 2020,



Asst. Prof. Vedran Dunjko,
LIACS/LION,
University of Leiden,
The Netherlands