ERCIM "ALAIN BENSOUSSAN"
FELLOWSHIP PROGRAMME

Scientific Report

First name / Family name	Nikita Kopylov
Nationality	Russian
Name of the *Host Organisation*	NTNU
First Name / family name of the *Scientific Coordinator*	Elena Celledoni Kurusch Ebrahimi-Fard
Period of the fellowship	01/03/2019 to 29/02/2020

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

Deep learning as an optimal control problem

Our goal was to research how optimal control theory can be applied to the learning procedure of deep neural networks. In this case, the network is seen as a function $y=f(x,w)$(where x is the training set, and w represents weights), and the trained parameters w are interpreted as external controls. The weights are to minimize an objective functional$L(y)=L(w)$–traditionally called loss– that represents the quality of training of the network.

Usually, the training problem is solved numerically using, for example, error back-propagation with stochastic gradient descent. However, this approach may suffer from gradient explosion or vanishing, and provides no explanation of the analytical behaviour of the network. For this reason, we researched into the analytical approach to the construction of controls.

Control theory is most developed for continuous control-affine systems, which can be represented as $\dot{y}(t)=g_0(x)+\sum g_i(x)w_i(t)$. In this case many techniques can be used, for instance, the Chen–Fliess series. Prof. Ebrahimi-Fard is an expert on this subject, and we, including the exchange master student Andrea Leone, investigated how the CF series could be used for the problem at hand.

It was shown that certain architectures of neural networks (like ResNet) have a continuous limit and can be seen as a discretization [1,2] of a nonlinear differential equation of a form $\dot{y}(t)=f(x,w(t))$. To apply analytical tools like the Chen–Fliess series, one needs to affinize in control this nonlinear equation. To this end we explored methods suggested in, e.g. [3,4]. The former suggests a rather straightforward affinization procedure which, however, does not produce acceptable results for the problem at hand. The latter describes a similar approach that, however, additionally requires knowing or imposing target higher-order dynamics of the system, which in general is not known for neural networks.

Therefore, we decided to use a more direct empirical approach similar to [5]. Using simple finite-width feed-forward and ResNet neural networks, we ran experiments to compare these original neural networks with their affinization w.r.t the weights by calculating the corresponding Jacobian matrices. Firstly, we trained the original networks and used the parameters in the linearized model. In this case the linear model showed a notable performance drop. In this setting, we tested the hypothesis that linearization improves if the width of the layers is increased (that is, when we increase the dimension of the system, allowing it to evolve in a 'larger' space). We found that the decrease in performance was less significant when the width of the network increased (which is similar to the results in [5]). Afterwards, we ran the following experiment: affinize the original model, train this affine representation, use the weights in the original one to assess its accuracy. If affinization was good enough, one might expect high accuracy of the original model. We observed an acceptable accuracy in the affine model but the same weights used in the original network led to a performance that was worse than random guessing. Therefore, it is necessary to research other methods, and we evaluated the possibility to use control on principal bundles and truncated power series algebras.

Predicting stress of slender structures

The estimation of the fatigue of various slender structures under repeated loads due to the environment is an important engineering problem that should be solved to guarantee performance and safety. Frequently, data which describe the state of the structure are scarce because the possibility to install sensors is limited. Moreover, the measurements are often noisy as it is not possible to ensure that the sensors stay well-oriented with respect to the structure. For these reasons, we investigated the applicability of machine learning to this task. We considered two applications: estimating the maximal bending moment of an Euler–Bernoulli beam and the fatigue of a marine drilling riser given few sensors. Specifically, drilling risers were considered due to the interest from the oil and gas company TechnipFMC.

To this end, we solved the equations corresponding to the beam and the riser, thus obtaining the fatigue data for machine learning. Afterwards, we used the data for training simple linear regression and an advanced recurrent neural network. Our hypothesis was that linear regression would be enough for the linear beam but would fail on the nonlinear riser, and it was proven true. By generating additional data sets, we investigated how the position of sensors affects the accuracy of prediction. We concluded that ML is a promising tool for such problems and it is promising to investigate advanced models. The results were presented at the GSI 2019 conference. They, along with twisted rod simulations we did with master students, may be of use in the THREAD (Numerical Modelling of Highly Flexible Structures) project that NTNU is starting.

Supervision and teaching

During my fellowship, I co-supervised (with Prof. Celledoni) two master students which successfully defended their theses and secured PhD positions.

I assisted Geir Bogfjellmo with preparing course material for the TMA4205 Numerical Linear Algebra.

1. E. Haber and L. Ruthotto, "Stable architectures for deep neural networks," Inverse Problems, vol. 34, no. 1, p. 14004, 2017.
2. M. Benning, E. Celledoni, M.J. Ehrhardt, B. Owren, C.-B. Schönlieb, "Deep learning as optimal control problems: Models and numerical methods," Journal of Computational Dynamics, vol. 6, no. 2, pp. 171–198, 2019.
3. J. Stefanovski, "Feedback affinization of nonlinear control systems," Systems & Control Letters, vol. 46, no. 3, pp. 207–217, 2002.
4. J. D. Boskovic, L. Chen, and R. K. Mehra, "Adaptive Control Design for Nonaffine Models Arising in Flight Control," Journal of Guidance, Control, and Dynamics, vol. 27, no. 2, pp. 209–217, 2004.
5. R. Novak, L. Xiao, J. Hron, J. Lee, A. A. Alemi, J. Sohl-Dickstein, S. S. Schoenholz. "Neural Tangents: Fast and Easy Infinite Neural Networks in Python", arXiv preprint 1912.02803, 2019.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

Celledoni E., Gustad H.S., Kopylov N., Sundklakk H.S. (2019) Predicting Bending Moments with Machine Learning. In: Nielsen F., Barbaresco F. (eds) Geometric Science of Information. GSI 2019. Lecture Notes in Computer Science, vol 11712. Springer, Cham. https://doi.org/10.1007/978-3-030-26980-7_19

Celledoni E., Gustad H.S., Kopylov N., Sundklakk H.S. NTNU and industry research machine learning for slender structures, European Consortium for Mathematics in Industry Blog, 13th of November 2019. (Dissemination.)

III – ATTENDED SEMINARS, WORKSHOPS, CONFERENCES

- Seminar "Working with Norwegians". 20th of March 2019, NTNU, Trondheim, Norway.
- DNA seminar. Nikita Kopylov, "Magnus-based geometric integrators for dynamical systems with time-dependent potentials", 27th of March 2019, NTNU, Trondheim, Norway.
- Seminar "The NTNU Challenge. Marie Skłodowska-Curie Actions", 22nd – 24th of May 2019, NTNU, Trondheim, Norway.
- BMS Summer School "Mathematics of deep learning", 19th – 26th of August 2019, Zuse Institute, Berlin, Germany.
- Conference "Geometric science of information", 27th – 29th of August 2019, ENAC, Toulouse, France.
- Biweekly seminars at the Norwegian Open AI Lab, NTNU, Trondheim, Norway.
- Various IMF and DNA seminars, NTNU, Trondheim, Norway.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

My REP was hosted by Erwan Faou and Philippe Chartier at INRIA Rennes Bretagne Atlantique, team Mingus and the IRMAR, mathematical laboratory of the University of Rennes 1, Rennes, France from 9 January to 17 January 2020.

Following suggestion of Dr. Faou, I researched into application of machine learning techniques to solving partial differential equations. More specifically, whether it is possible, given a data set generated from a certain set of functions (for example, harmonic or polynomials) and corresponding PDE solutions, to approximate the operator of the PDE. For linear PDEs, I looked into representing operators as products of matrices of a predefined shape, like tridiagonal or block matrices. These representations are equivalent to feed-forward neural networks without activation functions. For non-linear PDEs I used classical feed-forward neural networks with non-linear activation functions.