Dorian Goldman
200 North 11th Street, 4J, 11211 NY, NY. USA

Dorian Goldman, 200 North 11th Street, 4J, 11211 NY, NY. USA

Google AI

January 8, 2018

Dear Committee,

I currently work as a Senior Data Scientist at Condé Nast, where I develop and productionalize machine learning models. I am also an Adjunct Professor of Applied Mathematics at Columbia University, where I teach *APM4990E - Introduction to Data Science in Industry*[1].

I was previously a mathematician working in the fields of the Calculus of Variations, Partial Differential Equations (PDE) and Geometric Analysis as a Herchel Smith Fellow at The University of Cambridge. I completed my PhD at the Courant Institute of Mathematical Sciences (NYU) and Université Pierre et Marie Curie (Paris 6), advised by Sylvia Serfaty. My research primarily focused on problems related to the energetic description of pattern formation in systems with competing short and long range interactions [1, 2, 3], and the geometric properties of these energies [4, 5]. Prior to this, I completed my masters degree under the supervision of Robert McCann, where I worked on problems related to dynamical systems and optimal transportation [6], which was complimented by working with Cédric Villani at École Normale Superieur de Lyon. The methods I was trained in are becoming increasingly relevant in a modern context as essential tools in understanding fundamental problems that arise in Artificial Intelligence (AI) [7, 8, 9, 10, 11]. I believe my diverse background puts me in a unique position to help advance various subsets of the field as a result.

Connections between my previous work and recent developments in AI include, but are not limited to; stability of classifications with respect to perturbations of the dependent variables [7, 13], using optimal transportation [14] for transfer learning [8, 9], and reinforcement learning for offline and online policy discovery [15, 16]. Attempts to address the issue of stability in image classification, in a computationally feasible way, can be seen in [13] for instance, however the methods do not make use of the very recent developments in [7, 10], which recasts the problem in the language

[1] `https://github.com/Columbia-Intro-Data-Science/APMAE4990-`

of nonlinear dynamical systems. Another approach to stable image classification is transfer learning, which attempts to find methods, for instance, to classify objects in the presence of various contexts, backgrounds, lighting and quality. It is only recently that methods in optimal transportation have been able to provide optimal, computationally feasible solutions to this problem in an ansatz-free way [8, 9, 11]. As a data scientist, I have developed customized Python libraries which extend the ideas in [15] for optimal ad targeting and virality prediction, with initial work in adapting them to the more general framework discussed in [16] [2]. Additionally, some of my work in [1, 2] provided meaningful insights into the asymptotics of a highly complex nonlinear functional and its corresponding PDE, by understanding the fundamental characteristics of its limit. Understanding the limiting behavior of the structures involved in neural nets, Markov processes, or any kind of discrete models, could provide meaningful insights into its qualitative characteristics. This is currently a source of frustration in the field [17, 18]. A more in depth investigation into these question is provided in the following sections.

Dynamical Systems for Stable Image Classification

The notion of *well posedness* has been studied exhaustively in the context of PDE and ODE for over a century, yet the framework used in machine learning does not currently borrow many ideas from these advances. A simple example is the following - if a model can predict accurately that a photo contains a Panda, how stable is this prediction? Can I change a subset of pixels and still get the same result? Recent attempts to answer this question in a computationally feasible way can be seen in [13] for example. However these methods all rely on studying the discrete propogation problem in neural nets, and do not take advantage of any of the theory concerning the continuous limit of these structures. In [7, 10] however, they show that in certain cases, one we can reinterpret forward propagation in a neural net as a discretization of a continuous Hamiltonian system, which has the structure

$$\dot{y}(t) = -\nabla_z H(y, z, t) \text{ and } z'(t) = -\nabla_y H(y, z, t), \ \forall t \in [0, T]. \tag{1}$$

This perspective opens the field to a wide variety of methods in dynamical systems which relate the stability of the evolution of such systems to its initial conditions (see [19] for instance).

For example in [6], we study a finite dimensional Hamiltonian system arising from atmospheric fluid flow arising as the solution to an optimal transport problem. We establish the existence of a homoclinic orbit in the phase space, which as the example of the pendulum shows, are usually unstable. In this setting one can apply the method of Melnikov [48] for finite dimensional Hamiltonian systems to establish the existence of chaotic dynamics. I strongly believe that similar analysis could be done in this context which exploits this well established theory, and for which there would be immediate improvements in computational complexity and performance.

[2] Available upon request

Optimal Transport for Deep Joint Transfer Learning and Unsupervised Domain Adaptation

Transfer learning or inductive transfer is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. A common problem in industry is attempting to develop models which are trained on some *source* training data, and then being used on some *target* data, which may be different. For example, being able to classify images in the presence of varying backgrounds, lighting or quality.

The field of optimal transportation has recently been used as a tool to address this problem [8, 11]. The problem is generally stated as having a source and target distribution, $\mathbf{P}(y_s, \mathbf{x}_s)$ and $\mathbf{P}(y_t, \mathbf{x}_t)$ defined on Ω_s and Ω_t respectively. While previous work makes various assumptions about the structure and relationships between the two distributions which are unreaslistic, in [8, 11], they assume, quite generally, that the domain drift is due unknown, nonlinear transformation of the input space $\mathbf{T} : \Omega_s \mapsto \Omega_t$.

While the family of transformation is broad, in [11] they provide a proof that, for some simple affine transformations of discrete distributions, the optimal transport plan solution is exact. These results potentially only scratch the surface for understanding the optimality of transport methods for the transfer problem; utilizing the full theory from [14] could help us understand the most general conditions under which we can adapt source models to new target domains. For instance, in [8] the same methodology is extended to apply to Deep Neural Nets, which dramatically expands the class of applicable problems it can be used for, namely it could provide a powerful framework for robust, stable, deep learning image classification.

Reinforcement Learning and Offline Policy Evaluation

Unrelated to my previous academic focus, I've worked in an applied setting on problems in reinforcement learning regarding learning optimal offline policies [15, 16]. Given a collection of possible actions an agent can take (eg. sending an email) and possible rewards (eg. purchases), we can seek to infer optimal actions to maximize a reward, using only historical data. More precisely, one can decompose the expected reward of a collection of *policies* π_p which determine actions a_i at time t, given all of the history h_{t-1},

$$\mathbb{E}\left[\sum_{t=1}^{T} r_t | \pi_p\right] = \sum_{t=1}^{T} \frac{1}{m} \sum_{i=1}^{m} \frac{\prod \pi_p(a_{i,1}, \cdots, a_{i,t}|h_{t-1,i})}{\prod \pi_q(a_{i,1}, \cdots, a_{i,t}|h_{t-1,i})} r_{t,i}.$$

Decomposing the policy as

$$\pi(a_{i,1}, \cdots, a_{i,t}|h_{t-1,i}) = \prod_{t=1}^{T} \pi(a_i|h_{t-1,i}),$$

we can search for the best linear policy by, for instance, making a softmax prior on the policy distribution:

$$\pi(a_i|h_{t-1,i}) = \frac{\exp(-\theta_i^T \cdot \psi_t)}{\sum_{a_j} \exp(-\theta_j^T \cdot \psi_T)}.$$

While some research has been done in this area, the field has yet to be fully investigated. In [20], for the stationary case of $T = 1$, we extend the framework of [16] by providing a method of optimal policy discovery, which, in addition, uses the self-normalized importance sampling estimate

$$\mathbb{E}\left[\sum_{t=1}^{T} r_t|\pi_p\right] \sim \hat{V}_h(y) = \frac{\sum_{i=1}^{N} y_i \frac{\mathbf{1}(a_i=h(x_i))}{\pi_q(a_i|x_i)}}{\sum_{i=1}^{N} \frac{\mathbf{1}(a_i=h(x_i))}{\pi_q(a_i|x_i)}},$$

where $h : X \mapsto A$ is the policy and $\pi(a_i|x_i) = 1(a_i = h(x_i))$. This accounts for learning the null or negative responses in addition to the positive ones. We deal with the denominator by incorporating it as a regularization parameter into the method. I have applied this method successfully in problems related to virality prediction and online ad targeting [3].

In summary, I believe my previous research experience in the above relevant fields puts me in a unique position to make meaningful contributions in various fields of machine learning, and I hope to be given the opportunity to do so.

References

[1] Dorian Goldman, Sylvia Serfaty, and Cyrill Muratov. The gamma-limit of the two-dimensional ohta-kawasaki energy. i. droplet density. *Arch. Ration. Mech. Anal., 210:581–613, 2013*, 2013.

[2] Dorian Goldman, Sylvia Serfaty, and Cyrill Muratov. The gamma-limit of the two-dimensional ohta-kawasaki energy. ii. derivation of the renormalized energy. *Arch. Ration. Mech. Anal., 212:445–501*, 2014.

[3] Dorian Goldman. Asymptotics of non-minimizing critical points of the ohta-kawasaki energy. *Preprint: https://arxiv.org/pdf/1410.7047.pdf*, 2014.

[4] Dorian Goldman and Alexander Volkmann. On the regularity of stationary points of a non-local isoperimetric problem. *Interfaces and Free Boundaries, Volume 17, Issue 4, 2015, pp. 539–553*, 2015.

[5] Dorian Goldman. Uniqueness results for critical points of a non-local isoperimetric problem via curve shortening. *Preprint: https://arxiv.org/abs/1206.5984*, 2012.

[6] Dorian Goldman and Robert McCann. Chaotic response of the 2d semi-geostrophic and 3d quasi-geostrophic equations to gentle periodic forcing. *Nonlinearity 21, 1455-1470 doi 10.1088/0951–7715/21/7/005*, 2008.

[7] Eldad Haber and Lars Ruthotto. Stable architectures for deep neural networks. *Inverse Problems, Volume 34, Number 1*, 2017.

[8] Ying Lu, Liming Chen, and Alexandre Saidi. Optimal transport for deep joint transfer learning. *Preprint: https://arxiv.org/abs/1709.02995*, 2017.

[3] details available upon request

[9] Vivien Seguy, Bharath Bhushan Damodaran, Rmi Flamary, Nicolas Courty, Antoine Rolet, and Mathieu Blondel. Large-scale optimal transport and mapping estimation. *Preprint*, 2017.

[10] Eldad Haber, Lars Ruthotto, Elliot Holtham, and Seong-Hwan Jun. Learning across scales - multiscale methods for convolution neural networks. *Preprint*, 2017.

[11] Nicolas Courty, Remi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *CoRR*, abs/1507.00504, 2015. URL `http://arxiv.org/abs/1507.00504`.

[12] Nicolas Hunke, Zia Yusuf, Michael Rubmann, Florian Schmieg, Akash Bhatia, and Nipun Kalra. Winning in internet of things: It?s all about the business processes. *Boston Consulting Group*, 2017.

[13] Stephan Zheng, Yang Song, Thomas Leungand, and Ian Goodfellow. Improving the robustness of deep neural networks via stability training. *Conference: 2016 IEEE Conference on Computer Vision and Pattern*, 2016.

[14] Cédric Villani. *Optimal Transport: Old and New*. Springer Science, 2008. ISBN 978-3-540-71050-9. An optional note.

[15] Travis Mandel, Yun-En Liu, Sergey Levine, Emma Brunskill, and Zoran Popovic. Offline policy evaluation across representations with applications to educational games. *Proceeding AAMAS 14 Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems, 1077–1084*, 2014.

[16] Alex Strehl andJohn Langford, ham Kakade, and Lihong Li. Learning from logged implicit exploration data. *NIPS'10 Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 2 Pages 2217–2225*, 2010.

[17] Will Knight. The dark secret at the heart of ai. *MIT Technology Review, Intelligent Machines, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/*, 2017.

[18] Natalie Wolchover. New theory cracks open the black box of deep neural networks. *Wired Magazine, Science, https://www.wired.com/story/new-theory-deep-learning/*, 2017.

[19] J. P. Lasalle. The stability of dynamical systems. *Society for Industrial and Applied Mathematics*, 1976.

[20] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. URL `http://www.scipy.org/`. [Online; accessed ¡today¿].