Tutorial on Amortized Optimization

Other titles in Foundations and Trends® in Machine Learning

Introduction to Riemannian Geometry and Geometric Statistics: From Basic Theory to Implementation with Geomstats
Nicolas Guigui, Nina Miolane and Xavier Pennec
ISBN: 978-1-63828-154-2

Graph Neural Networks for Natural Language Processing: A Survey
Lingfei Wu, Yu Chen, Kai Shen, Xiaojie Guo, Hanning Gao, Shucheng Li, Jian Pei and Bo Long
ISBN: 978-1-63828-142-9

Model-based Reinforcement Learning: A Survey
Thomas M. Moerland, Joost Broekens, Aske Plaat and Catholijn M. Jonker
ISBN: 978-1-63828-056-9

Divided Differences, Falling Factorials, and Discrete Splines: Another Look at Trend Filtering and Related Problems
Ryan J. Tibshirani
ISBN: 978-1-63828-036-1

Risk-Sensitive Reinforcement Learning via Policy Gradient Search
Prashanth L. A. and Michael C. Fu
ISBN: 978-1-63828-026-2

A Unifying Tutorial on Approximate Message Passing
Oliver Y. Feng, Ramji Venkataramanan, Cynthia Rush and Richard J. Samworth
ISBN: 978-1-63828-004-0

Tutorial on Amortized Optimization

Brandon Amos
Meta AI
brandon.amos.cs@gmail.com

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

B. Amos. *Tutorial on Amortized Optimization*. Foundations and Trends® in Machine Learning, vol. 16, no. 5, pp. 592–732, 2023.

ISBN: 978-1-63828-209-9

Foundations and Trends® in Machine Learning
Volume 16, Issue 5, 2023

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Machine Learning publishes survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2023, Volume 16, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Contents

Tutorial on Amortized Optimization

Brandon Amos

Meta AI, USA; brandon.amos.cs@gmail.com

ABSTRACT

Optimization is a ubiquitous modeling tool and is often deployed in settings which repeatedly solve similar instances of the same problem. Amortized optimization methods use learning to predict the solutions to problems in these settings, exploiting the shared structure between similar problem instances. These methods have been crucial in variational inference and reinforcement learning and are capable of solving optimization problems many orders of magnitudes times faster than traditional optimization methods that do not use amortization. This tutorial presents an introduction to the amortized optimization foundations behind these advancements and overviews their applications in variational inference, sparse coding, gradient-based meta-learning, control, reinforcement learning, convex optimization, optimal transport, and deep equilibrium networks. The source code for this tutorial is available at https://github.com/facebookresearch/amortized-optimization-tutorial.

Brandon Amos (2023), "Tutorial on Amortized Optimization", Foundations and Trends® in Machine Learning: Vol. 16, No. 5, pp 592–732. DOI: 10.1561/2200000102.

1

Introduction

This tutorial studies the use of machine learning to improve repeated solves of parametric optimization problems of the form

$$y^\star(x) \in \arg\min_y f(y; x), \tag{1.1}$$

where the *non-convex* objective $f : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}$ takes a *context* or *parameterization* $x \in \mathcal{X}$ which can be continuous or discrete, and the *continuous, unconstrained domain* of the problem is $y \in \mathcal{Y} = \mathbb{R}^n$. Eq. (1.1) implicitly defines a *solution* $y^\star(x) \in \mathcal{Y}$. In most of the applications considered later in Section 3, $y^\star(x)$ is unique and smooth, *i.e.*, the solution continuously changes in a connected way as the context change is illustrated in Figure 1.1.

Parametric optimization problems such as Eq. (1.1) have been studied for decades (Bank *et al.*, 1982; Fiacco and Ishizuka, 1990; Shapiro, 2003; Klatte and Kummer, 2006; Bonnans and Shapiro, 2013; Still, 2018; Fiacco, 2020) with a focus on sensitivity analysis. The general formulation in Eq. (1.1) captures many tasks arising in physics, engineering, mathematics, control, inverse modeling, and machine learning. For example, when controlling a continuous robotic system, $\mathcal{X}$ is the space of *observations* or *states*, *e.g.*, angular positions and velocities describing

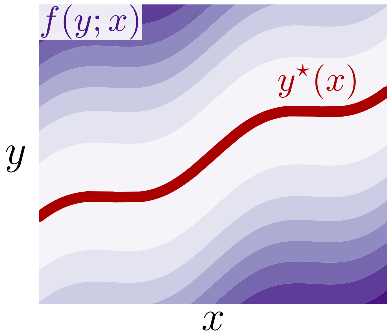

Figure 1.1: Illustration of the parametric optimization problem in Eq. (1.1). Each context x parameterizes an optimization problem that the objective $f(y;x)$ depends on. The contours show the values of the objectives where darker colors indicate higher values. The objective is then minimized over y and the resulting solution $y^\star(x)$ is shown in red. In other words, each vertical slice is an optimization problem and this visualization shows a continuum of optimization problems.

the configuration of the system, the domain $\mathcal{Y} := \mathcal{U}$ is the *control space*, *e.g.*, torques to apply to each actuated joint, and $f(u;x) := -Q(u,x)$ is the *control cost* or the negated *Q-value* of the state-action tuple (x,u), *e.g.*, to reach a goal location or to maximize the velocity. For every encountered state x, the system is controlled by solving an optimization problem in the form of Eq. (1.1). While $\mathcal{Y} = \mathbb{R}^n$ is over a deterministic real-valued space in Eq. (1.1), the formulation can also capture stochastic optimization problems as discussed in Section 2.3.1. For example, Section 3.1 optimizes over the (real-valued) parameters of a variational distribution and Section 3.6 optimizes over the (real-valued) parameters of a stochastic policy for control and reinforcement learning.

Optimization problems such as Eq. (1.1) quickly become a computational bottleneck in systems they are a part of. These problems often does not have a closed-form analytic solution and is instead solved with approximate numerical methods which iteratively search for the solution. This computational problem has led to many specialized solvers that leverage domain-specific insights to deliver fast solves. Specialized algorithms are especially prevalent in convex optimization methods for linear programming, quadratic programming, cone programming, and control and use theoretical insights of the problem structure to

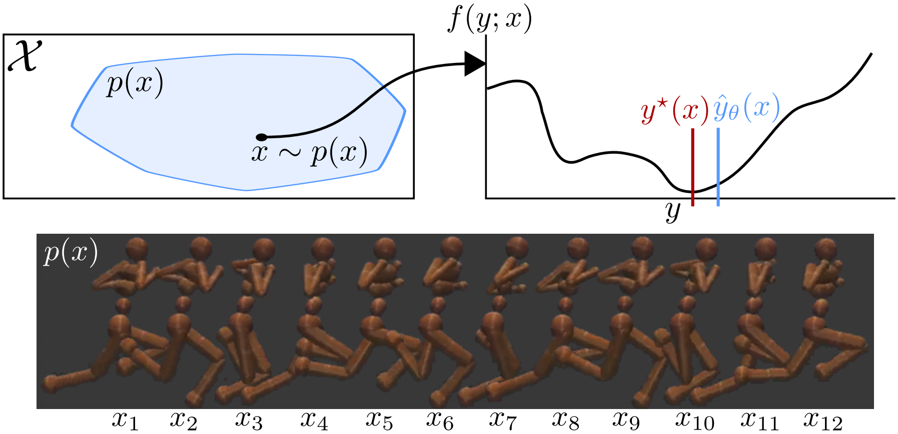

Figure 1.2: An amortized optimization method learns a model $\hat{y}_\theta$ to predict the minimum of an *objective* $f(y; x)$ to a parameterized optimization problem, as in Eq. (1.1), which depends on a *context* x. For example, in control, the context space $\mathcal{X}$ is the state space of the system, *e.g.* angular positions and velocities describing the configuration of the system, the domain $\mathcal{Y} := \mathcal{U}$ is the control space, *e.g.* torques to apply to each actuated joint, the cost (or negated value) of a state-action pair is $f(u; x) := -Q(x, u)$, and the state distribution is $p(x)$. For an encountered state x, many reinforcement learning policies $\pi_\theta(x) := \hat{y}_\theta(x)$ amortize the solution to the underlying control problem with true solution $y^\star(x)$. This humanoid policy was obtained with the model-based stochastic value gradient in Amos *et al.* (2021).

bring empirical gains of computational improvements and improved convergence (Boyd and Vandenberghe, 2004; Nocedal and Wright, 2006; Bertsekas, 2015; Bubeck, 2015; Nesterov *et al.*, 2018).

Mostly separate from optimization research and algorithmic advancements, the machine learning community has focused on developing generic function approximation methods for estimating non-trivial high-dimensional mappings from data (Murphy, 2012; Salakhutdinov, 2014; Deisenroth *et al.*, 2020). While machine learning models are often used to reconstruct mappings from data, *e.g.* for supervised classification or regression where the targets are given by human annotations. Many computational advancements on the software and hardware have been developed in recent years to make the prediction time fast: the forward pass of a neural network generating a prediction can execute in milliseconds on a graphics processing unit.

Overview. This tutorial studies the use of machine learning models to rapidly predict the solutions to the optimization problem in Eq. (1.1), which is referred to as *amortized optimization* or *learning to optimize*. Figure 1.2 illustrates a basic setup for control. Amortized optimization methods are capable of significantly improving the computational time of classical algorithms *on a focused subset of problems*. This is because the model is able to learn about the solution mapping from x to $y^\star(x)$ that classical optimization methods usually do not assume access to. My goal in writing this is to explore a unified perspective of modeling approaches of amortized optimization in Section 2 to help draw connections between the applications in Section 3, *e.g.* between amortized variational inference, meta-learning, and policy learning for control and reinforcement learning, sparse coding, convex optimization, optimal transport, and deep equilibrium networks. These topics have historically been studied in isolation without connections between their amortization components. Section 4 presents a computational tour through source code for variational inference, policy learning, and a spherical optimization problem and Section 5 concludes with a discussion of challenges, limitations, open problems, and related work.

How much does amortization help? Amortized optimization has been revolutionary to many fields, especially including variational inference and reinforcement learning. Figure 4.1 shows that the amortization component of a variational autoencoder trained on MNIST is **25000** times faster (0.4ms vs. 8 seconds!) than solving a batch of 1024 optimization problems from scratch to obtain a solution of the same quality. These optimization problems are solved in every training iteration and can become a significant bottleneck if they are inefficiently solved. If the model is being trained for millions of iterations, then the difference between solving the optimization problem in 0.4ms vs. 8 seconds makes the difference between the entire training process finishing in a few hours or a month.

A historic note: amortization in control and statistical inference. Amortized optimization has arisen in many fields as a result to practical optimization problems being non-convex and not having easily computed, or closed-form solutions. Continuous control problems with linear dynamics and quadratic cost are convex and often easily solved with the linear quadratic regulator (LQR) and many non-convex extensions and iterative applications of LQR have been successful over the decades, but becomes increasingly infeasible on non-trivial systems and in reinforcement learning settings where the policy often needs to be rapidly executed. For this reason, the reinforcement learning community almost exclusively amortizes control optimization problems with a learned policy (Sutton and Barto, 2018). Related to this throughline in control and reinforcement learning, many statistical optimization problems have closed form solutions for known distributions such as Gaussians. For example, the original Kalman filter is defined with Gaussians and the updates take closed-form. The extended Kalman filter generalizes the distributions to non-Gaussians, but the updates are in general no longer available analytically and need to be computationally estimated. Marino *et al.* (2018a) shows how amortization helps improve this computationally challenging step. Both of these control and statistical settings start with a simple setting with analytic solutions to optimization problems, generalize to more challenging optimization problems that need to be computationally estimated, and then add back some computational tractability with amortized optimization.

References

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.* (2016). "Tensorflow: A system for large-scale machine learning". In: *12th USENIX symposium on operating systems design and implementation (OSDI 16)*. 265–283.

Absil, P.-A., R. Mahony, and R. Sepulchre. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.

Adams, R. P. and R. S. Zemel. (2011). "Ranking via sinkhorn propagation". *arXiv preprint arXiv:1106.1925*.

Adler, J., A. Ringh, O. Öktem, and J. Karlsson. (2017). "Learning to solve inverse problems using Wasserstein loss". *ArXiv preprint*. abs/1710.10898.

Agrawal, A., B. Amos, S. T. Barratt, S. P. Boyd, S. Diamond, and J. Z. Kolter. (2019a). "Differentiable Convex Optimization Layers". In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. 9558–9570. URL: https://proceedings.neurips.cc/paper/2019/hash/9ce3c52fc54362e22053399d3181c638-Abstract.html.

Agrawal, A., A. N. Modi, A. Passos, A. Lavoie, A. Agarwal, A. Shankar, I. Ganichev, J. Levenberg, M. Hong, R. Monga, and S. Cai. (2019b). "TensorFlow Eager: A multi-stage, Python-embedded DSL for machine learning". In: *Proceedings of Machine Learning and Systems 2019, MLSys 2019, Stanford, CA, USA, March 31 - April 2, 2019*. URL: https://proceedings.mlsys.org/book/286.pdf.

Aho, A. V., R. Sethi, and J. D. Ullman. (1986). "Compilers, principles, techniques". *Addison Wesley*.

Alabi, D., A. T. Kalai, K. Ligett, C. Musco, C. Tzamos, and E. Vitercik. (2019). "Learning to Prune: Speeding up Repeated Computations". In: *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*. Vol. 99. *Proceedings of Machine Learning Research*. PMLR. 30–33. URL: http://proceedings.mlr.press/v99/alabi19a.html.

Ali, A., E. Wong, and J. Z. Kolter. (2017). "A Semismooth Newton Method for Fast, Generic Convex Programming". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 70–79. URL: http://proceedings.mlr.press/v70/ali17a.html.

Allgower, E. L. and K. Georg. (2012). *Numerical continuation methods: An introduction*. Vol. 13. Springer Science & Business Media.

Amos, B. (2019). "Differentiable Optimization-Based Modeling for Machine Learning". *PhD thesis*. Carnegie Mellon University.

Amos, B. (2023). "On amortizing convex conjugates for optimal transport". In: *The Eleventh International Conference on Learning Representations, ICLR*.

Amos, B., S. Cohen, G. Luise, and I. Redko. (2023). "Meta Optimal Transport". In: *International Conference on Machine Learning, ICML 2023. Proceedings of Machine Learning Research*.

Amos, B. and J. Z. Kolter. (2017). "OptNet: Differentiable Optimization as a Layer in Neural Networks". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 136–145. URL: http://proceedings.mlr.press/v70/amos17a.html.

Amos, B., V. Koltun, and J. Z. Kolter. (2019). "The limited multi-label projection layer". *ArXiv preprint*. abs/1906.08707.

Amos, B., I. D. J. Rodriguez, J. Sacks, B. Boots, and J. Z. Kolter. (2018). "Differentiable MPC for End-to-end Planning and Control". In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. 8299–8310. URL: https://proceedings.neurips.cc/paper/2018/hash/ba6d843eb4251a4526ce65d1807a9309-Abstract.html.

Amos, B., S. Stanton, D. Yarats, and A. G. Wilson. (2021). "On the model-based stochastic value gradient for continuous reinforcement learning". In: *Proceedings of the 3rd Annual Conference on Learning for Dynamics and Control, L4DC 2021, 7-8 June 2021, Virtual Event, Switzerland*. Vol. 144. *Proceedings of Machine Learning Research*. PMLR. 6–20. URL: http://proceedings.mlr.press/v144/amos21a.html.

Amos, B., L. Xu, and J. Z. Kolter. (2017). "Input Convex Neural Networks". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 146–155. URL: http://proceedings.mlr.press/v70/amos17b.html.

Amos, B. and D. Yarats. (2020). "The Differentiable Cross-Entropy Method". In: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*. Vol. 119. *Proceedings of Machine Learning Research*. PMLR. 291–302. URL: http://proceedings.mlr.press/v119/amos20a.html.

Anderson, D. G. (1965). "Iterative procedures for nonlinear integral equations". *Journal of the ACM (JACM)*. 12(4): 547–560.

Andrychowicz, M., M. Denil, S. G. Colmenarejo, M. W. Hoffman, D. Pfau, T. Schaul, and N. de Freitas. (2016). "Learning to learn by gradient descent by gradient descent". In: *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*. 3981–3989. URL: https://proceedings.neurips.cc/paper/2016/hash/fb87582825f9d28a8d42c5e5e5e8b23d-Abstract.html.

Antoniou, A., H. Edwards, and A. J. Storkey. (2019). “How to train your MAML”. In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. URL: https://openreview.net/forum?id=HJGven05Y7.

Arbel, M. and J. Mairal. (2022). “Amortized Implicit Differentiation for Stochastic Bilevel Optimization”. In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. URL: https://openreview.net/forum?id=3PN4iyXBeF.

Arjovsky, M., S. Chintala, and L. Bottou. (2017). “Wasserstein Generative Adversarial Networks”. In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 214–223. URL: http://proceedings.mlr.press/v70/arjovsky17a.html.

Arnold, S. M., P. Mahajan, D. Datta, I. Bunner, and K. S. Zarkias. (2020). “learn2learn: A library for meta-learning research”. *ArXiv preprint*. abs/2008.12284.

Ba, J. L., J. R. Kiros, and G. E. Hinton. (2016). “Layer Normalization”. *arXiv e-prints*.

Bae, J., P. Vicol, J. Z. HaoChen, and R. B. Grosse. (2022). “Amortized proximal optimization”. *Advances in Neural Information Processing Systems*. 35: 8982–8997.

Bai, S., J. Z. Kolter, and V. Koltun. (2019). “Deep Equilibrium Models”. In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. 688–699. URL: https://proceedings.neurips.cc/paper/2019/hash/01386bd6d8e091c2ab4c7c7de644d37b-Abstract.html.

Bai, S., V. Koltun, and J. Z. Kolter. (2020). “Multiscale Deep Equilibrium Models”. In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. URL: https://proceedings.neurips.cc/paper/2020/hash/3812f9a59b634c2a9c574610eaba5bed-Abstract.html.

Bai, S., V. Koltun, and J. Z. Kolter. (2022). "Neural Deep Equilibrium Solvers". In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022.* URL: https://openreview.net/forum?id=B0oHOwT5ENL.

Baird, L. (1995). "Residual algorithms: Reinforcement learning with function approximation". In: *Machine Learning Proceedings 1995.* Elsevier. 30–37.

Baker, K. (2019). "Learning warm-start points for AC optimal power flow". In: *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP).* IEEE. 1–6.

Balcan, M.-F. (2020). "Data-driven algorithm design". *ArXiv preprint.* abs/2011.07177.

Banerjee, A. G. and N. Roy. (2015). "Efficiently solving repeated integer linear programming problems by learning solutions of similar linear programming problems using boosting trees". *Computer Science and Artificial Intelligence Laboratory Technical Report. MIT.*

Banert, S., A. Ringh, J. Adler, J. Karlsson, and O. Oktem. (2020). "Data-driven nonsmooth optimization". *SIAM Journal on Optimization.* 30(1): 102–131.

Banert, S., J. Rudzusika, O. Öktem, and J. Adler. (2021). "Accelerated Forward-Backward Optimization using Deep Learning". *ArXiv preprint.* abs/2105.05210.

Bank, B., J. Guddat, D. Klatte, B. Kummer, and K. Tammer. (1982). *Non-linear parametric optimization.* Springer.

Barratt, S. (2018). "On the differentiability of the solution to convex optimization problems". *ArXiv preprint.* abs/1804.05098.

Baxter, J. (1998). "Theoretical models of learning to learn". In: *Learning to Learn.* Springer. 71–94.

Beck, A. and M. Teboulle. (2009). "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". *SIAM Journal on Imaging Sciences.* 2(1): 183–202.

Belanger, D. (2017). "Deep energy-based models for structured prediction". *PhD thesis.* University of Massachusetts Amherst.

Belanger, D. and A. McCallum. (2016). "Structured Prediction Energy Networks". In: *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*. Vol. 48. *JMLR Workshop and Conference Proceedings*. JMLR.org. 983–992. URL: http://proceedings.mlr.press/v48/belanger16.html.

Belanger, D., B. Yang, and A. McCallum. (2017). "End-to-End Learning for Structured Prediction Energy Networks". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 429–439. URL: http://proceedings.mlr.press/v70/belanger17a.html.

Bellman, R. (1966). "Dynamic programming". *Science*. 153(3731): 34–37.

Bello, I., B. Zoph, V. Vasudevan, and Q. V. Le. (2017). "Neural Optimizer Search with Reinforcement Learning". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 459–468. URL: http://proceedings.mlr.press/v70/bello17a.html.

Bengio, S., Y. Bengio, and J. Cloutier. (1994). "Use of genetic programming for the search of a new learning rule for neural networks". In: *First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence*. IEEE. 324–327.

Bengio, Y., A. Lodi, and A. Prouvost. (2021). "Machine learning for combinatorial optimization: a methodological tour d'horizon". *European Journal of Operational Research*. 290(2): 405–421.

Bertinetto, L., J. F. Henriques, P. H. S. Torr, and A. Vedaldi. (2019). "Meta-learning with differentiable closed-form solvers". In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. URL: https://openreview.net/forum?id=HyxnZh0ct7.

Berto, F., S. Massaroli, M. Poli, and J. Park. (2022). "Neural Solvers for Fast and Accurate Numerical Optimal Control". In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. URL: https://openreview.net/forum?id=m8bypnj7Yl5.

Bertsekas, D. (2015). *Convex optimization algorithms*. Athena Scientific.

Bertsekas, D. P. (1971). "Control of uncertain systems with a set-membership description of the uncertainty." *PhD thesis*. Massachusetts Institute of Technology.

Bertsekas, D. P. (2000). *Dynamic Programming and Optimal Control*. 2nd. Athena Scientific.

Bertsimas, D. and B. Stellato. (2019). "Online mixed-integer optimization in milliseconds". *ArXiv preprint*. abs/1907.02206.

Bertsimas, D. and B. Stellato. (2021). "The voice of optimization". *Machine Learning*. 110(2): 249–277.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah. (2017). "Julia: A fresh approach to numerical computing". *SIAM Review*. 59(1): 65–98.

Bhardwaj, M., B. Boots, and M. Mukadam. (2020). "Differentiable gaussian process motion planning". In: *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 10598–10604.

Blechschmidt, J. and O. G. Ernst. (2021). "Three ways to solve partial differential equations with neural networks—A review". *GAMM-Mitteilungen*: e202100006.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe. (2017). "Variational inference: A review for statisticians". *Journal of the American statistical Association*. 112(518): 859–877.

Blondel, M. (2019). "Structured Prediction with Projection Oracles". In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. 12145–12156. URL: https://proceedings.neurips.cc/paper/2019/hash/7990ec44fcf3d7a0e5a2add28362213c-Abstract.html.

Blondel, M., Q. Berthet, M. Cuturi, R. Frostig, S. Hoyer, F. Llinares-López, F. Pedregosa, and J.-P. Vert. (2022). "Efficient and modular implicit differentiation". *Advances in Neural Information Processing Systems*. 35: 5230–5242.

Blondel, M., A. F. T. Martins, and V. Niculae. (2020). "Learning with Fenchel-Young losses". *Journal of Machine Learning Research*. 21: 35:1–35:69. URL: http://jmlr.org/papers/v21/19-021.html.

Bonnans, J. F. and A. Shapiro. (2013). *Perturbation analysis of optimization problems*. Springer Science & Business Media.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein. (2011). "Distributed optimization and statistical learning via the alternating direction method of multipliers". *Foundations and Trends® in Machine Learning*. 3(1): 1–122.

Boyd, S. and L. Vandenberghe. (2004). *Convex optimization*. Cambridge University Press.

Bradbury, J., R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, and S. Wanderman-Milne. (2020). "JAX: composable transformations of Python+ NumPy programs, 2018". 4: 16.

Brockman, G., V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. (2016). "Openai gym". *ArXiv preprint*. abs/1606.01540.

Broyden, C. G. (1965). "A class of methods for solving nonlinear simultaneous equations". *Mathematics of Computation*. 19(92): 577–593.

Bubeck, S. (2015). "Convex optimization: Algorithms and complexity". *Foundations and Trends® in Machine Learning*. 8(3-4): 231–357.

Bunne, C., A. Krause, and M. Cuturi. (2022). "Supervised training of conditional monge maps". *ArXiv preprint*. abs/2206.14262.

Burgess, C. P., I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner. (2018). "Understanding disentangling in β-VAE". *ArXiv preprint*. abs/1804.03599.

Busseti, E., W. M. Moursi, and S. Boyd. (2019). "Solution refinement at regular points of conic problems". *Computational Optimization and Applications*. 74(3): 627–643.

Byravan, A., L. Hasenclever, P. Trochim, M. Mirza, A. D. Ialongo, Y. Tassa, J. T. Springenberg, A. Abdolmaleki, N. Heess, J. Merel, and M. A. Riedmiller. (2022). "Evaluating Model-Based Planning and Planner Amortization for Continuous Control". In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. URL: https://openreview.net/forum?id=SS8F6tFX3-.

Byravan, A., J. T. Springenberg, A. Abdolmaleki, R. Hafner, M. Neunert, T. Lampe, N. Siegel, N. Heess, and M. Riedmiller. (2019). "Imagined Value Gradients: Model-Based Policy Optimization with Transferable Latent Dynamics Models". *ArXiv preprint*. abs/1910.04142.

Camacho, E. F. and C. B. Alba. (2013). *Model predictive control*. Springer Science & Business Media.

Cappart, Q., D. Chételat, E. Khalil, A. Lodi, C. Morris, and P. Veličković. (2021). "Combinatorial optimization and reasoning with graph neural networks". *ArXiv preprint*. abs/2102.09544.

Carter, M. (2001). *Foundations of mathematical economics*. MIT Press.

Caruana, R. (1997). "Multitask learning". *Machine Learning*. 28(1): 41–75.

Cauligi, A., P. Culbertson, E. Schmerling, M. Schwager, B. Stellato, and M. Pavone. (2021). "CoCo: Online Mixed-Integer Control via Supervised Learning". *IEEE Robotics and Automation Letters*.

Cauligi, A., P. Culbertson, B. Stellato, D. Bertsimas, M. Schwager, and M. Pavone. (2020). "Learning mixed-integer convex optimization strategies for robot planning and control". In: *IEEE Conference on Decision and Control (CDC)*. IEEE. 1698–1705.

Chandak, Y., G. Theocharous, J. Kostas, S. M. Jordan, and P. S. Thomas. (2019). "Learning Action Representations for Reinforcement Learning". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Vol. 97. *Proceedings of Machine Learning Research*. PMLR. 941–950. URL: http://proceedings.mlr.press/v97/chandak19a.html.

Chang, J. R., C. Li, B. Póczos, and B. V. K. V. Kumar. (2017). "One Network to Solve Them All - Solving Linear Inverse Problems Using Deep Projection Models". In: *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. IEEE Computer Society. 5889–5898. DOI: 10.1109/ICCV.2017.627.

Charton, F. (2021). "Linear algebra with transformers". *ArXiv preprint*. abs/2112.01898.

Charton, F., A. Hayat, S. T. McQuade, N. J. Merrill, and B. Piccoli. (2021). "A deep language model to predict metabolic network equilibria". *ArXiv preprint*. abs/2112.03588.

Chen, J. Y., S. Silwal, A. Vakilian, and F. Zhang. (2022a). "Faster Fundamental Graph Algorithms via Learned Predictions". In: *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*. Vol. 162. *Proceedings of Machine Learning Research*. PMLR. 3583–3602. URL: https://proceedings.mlr.press/v162/chen22v.html.

Chen, S. S., D. L. Donoho, and M. A. Saunders. (2001). "Atomic decomposition by basis pursuit". *SIAM Review*. 43(1): 129–159.

Chen, S. W., T. Wang, N. Atanasov, V. Kumar, and M. Morari. (2022b). "Large scale model predictive control with neural networks and primal active sets". *Automatica*. 135: 109947.

Chen, T., X. Chen, W. Chen, H. Heaton, J. Liu, Z. Wang, and W. Yin. (2021a). "Learning to optimize: A primer and a benchmark". *ArXiv preprint*. abs/2103.12828.

Chen, T., B. Xu, C. Zhang, and C. Guestrin. (2016). "Training deep nets with sublinear memory cost". *ArXiv preprint*. abs/1604.06174.

Chen, Y., B. Hosseini, H. Owhadi, and A. M. Stuart. (2021b). "Solving and learning nonlinear PDEs with gaussian processes". *ArXiv preprint*. abs/2103.12959.

Chen, Y., A. L. Friesen, F. Behbahani, A. Doucet, D. Budden, M. Hoffman, and N. de Freitas. (2020). "Modular Meta-Learning with Shrinkage". In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. URL: https://proceedings.neurips.cc/paper/2020/hash/1e04b969bf040acd252e1faafb51f829-Abstract.html.

Chen, Y., M. W. Hoffman, S. G. Colmenarejo, M. Denil, T. P. Lillicrap, M. Botvinick, and N. de Freitas. (2017). "Learning to Learn without Gradient Descent by Gradient Descent". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 748–756. URL: http://proceedings.mlr.press/v70/chen17e.html.

Cho, K., B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. (2014). "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics. 1724–1734. DOI: 10.3115/v1/D14-1179.

Chung, J., K. Kastner, L. Dinh, K. Goel, A. C. Courville, and Y. Bengio. (2015). "A Recurrent Latent Variable Model for Sequential Data". In: *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*. 2980–2988. URL: https://proceedings.neurips.cc/paper/2015/hash/b618c3210e934362ac261db280128c22-Abstract.html.

Cohen, S., B. Amos, and Y. Lipman. (2021). "Riemannian Convex Potential Maps". In: *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*. Vol. 139. *Proceedings of Machine Learning Research*. PMLR. 2028–2038. URL: http://proceedings.mlr.press/v139/cohen21a.html.

Cover, T. M. and J. A. Thomas. (2006). "Elements of Information Theory". *Wiley Series in Telecommunications and Signal Processing*.

Cremer, C., X. Li, and D. Duvenaud. (2018). "Inference Suboptimality in Variational Autoencoders". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Vol. 80. *Proceedings of Machine Learning Research*. PMLR. 1086–1094. URL: http://proceedings.mlr.press/v80/cremer18a.html.

Cruz, R. S., B. Fernando, A. Cherian, and S. Gould. (2017). "DeepPermNet: Visual Permutation Learning". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society. 6044–6052. DOI: 10.1109/CVPR.2017.640.

Cuturi, M. (2013). "Sinkhorn Distances: Lightspeed Computation of Optimal Transport". In: *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*. 2292–2300. URL: https://proceedings.neurips.cc/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html.

Cuturi, M. and M. Blondel. (2017). "Soft-DTW: a Differentiable Loss Function for Time-Series". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 894–903. URL: http://proceedings.mlr.press/v70/cuturi17a.html.

d'Ascoli, S., P.-A. Kamienny, G. Lample, and F. Charton. (2022). "Deep Symbolic Regression for Recurrent Sequences". arXiv: 2201.04600 [cs.LG].

Dam, N., Q. Hoang, T. Le, T. D. Nguyen, H. Bui, and D. Phung. (2019). "Three-Player Wasserstein GAN via Amortised Duality". In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*. 2202–2208. DOI: 10.24963/ijcai.2019/305.

Danskin, J. M. (1966). "The theory of max-min, with applications". *SIAM Journal on Applied Mathematics*. 14(4): 641–664.

Daubechies, I., M. Defrise, and C. De Mol. (2004). "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint". *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*. 57(11): 1413–1457.

Davidson, J. W. and S. Jinturkar. (1995). "An aggressive approach to loop unrolling". *Tech. rep.* Citeseer.

Dayan, P., G. E. Hinton, R. M. Neal, and R. S. Zemel. (1995). "The helmholtz machine". *Neural Computation*. 7(5): 889–904.

De Boer, P.-T., D. P. Kroese, S. Mannor, and R. Y. Rubinstein. (2005). "A tutorial on the cross-entropy method". *Annals of Operations Research*. 134(1): 19–67.

Deisenroth, M. P., A. A. Faisal, and C. S. Ong. (2020). *Mathematics for machine learning*. Cambridge University Press.

Deisenroth, M. P. and C. E. Rasmussen. (2011). "PILCO: A Model-Based and Data-Efficient Approach to Policy Search". In: *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*. Omnipress. 465–472. URL: https://icml.cc/2011/papers/323%5C_icmlpaper.pdf.

Deleu, T., T. Würfl, M. Samiei, J. P. Cohen, and Y. Bengio. (2019). "Torchmeta: A meta-learning library for pytorch". *ArXiv preprint*. abs/1909.06576.

Deshpande, I., Y. Hu, R. Sun, A. Pyrros, N. Siddiqui, S. Koyejo, Z. Zhao, D. A. Forsyth, and A. G. Schwing. (2019). "Max-Sliced Wasserstein Distance and Its Use for GANs". In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE. 10648–10656. DOI: 10.1109/CVPR.2019.01090.

Diamond, S. and S. Boyd. (2016). "CVXPY: A Python-embedded modeling language for convex optimization". *The Journal of Machine Learning Research*. 17(1): 2909–2913.

Dini, U. (1878). *Analisi infinitesimale*. Lithografia Gorani.

Dinitz, M., S. Im, T. Lavastida, B. Moseley, and S. Vassilvitskii. (2021). "Faster Matchings via Learned Duals". In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. 10393–10406. URL: https://proceedings.neurips.cc/paper/2021/hash/5616060fb8ae85d93f334e7267307664-Abstract.html.

Doersch, C. (2016). "Tutorial on variational autoencoders". *ArXiv preprint*. abs/1606.05908.

Domke, J. (2012). “Generic methods for optimization-based modeling”. In: *AISTATS*. 318–326.

Dong, W., Z. Xie, G. Kestor, and D. Li. (2020). “Smart-PGSim: using neural network to accelerate AC-OPF power grid simulation”. In: *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE. 1–15.

Donoho, D. L. and M. Elad. (2003). “Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell 1$ minimization”. *Proceedings of the National Academy of Sciences*. 100(5): 2197–2202.

Dontchev, A. L. and R. T. Rockafellar. (2009). *Implicit functions and solution mappings*. Springer.

Donti, P. L., D. Rolnick, and J. Z. Kolter. (2021). “DC3: A learning method for optimization with hard constraints”. In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. URL: https://openreview.net/forum?id=V1ZHVxJ6dSS.

Drori, I., S. Tran, R. Wang, N. Cheng, K. Liu, L. Tang, E. Ke, N. Singh, T. L. Patti, J. Lynch, A. Shporer, N. Verma, E. Wu, and G. Strang. (2021). “A Neural Network Solves and Generates Mathematics Problems by Program Synthesis: Calculus, Differential Equations, Linear Algebra, and More”. arXiv: 2112.15594 [cs.LG].

Duchi, J. C., E. Hazan, and Y. Singer. (2010). “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization”. In: *COLT 2010 - The 23rd Conference on Learning Theory, Haifa, Israel, June 27-29, 2010*. Omnipress. 257–269. URL: http://colt2010.haifa.il.ibm.com/papers/COLT2010proceedings.pdf%5C#page=265.

Dunning, I., J. Huchette, and M. Lubin. (2017). “JuMP: A modeling language for mathematical optimization”. *SIAM Review*. 59(2): 295–320.

Dupont, E. (2018). “Learning Disentangled Joint Continuous and Discrete Representations”. In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. 708–718. URL: https://proceedings.neurips.cc/paper/2018/hash/b9228e0962a78b84f3d5d92f4faa000b-Abstract.html.

Duruisseaux, V. and M. Leok. (2022). "Accelerated Optimization on Riemannian Manifolds via Projected Variational Integrators". arXiv: 2201.02904 [math.OC].

Ernst, D., P. Geurts, and L. Wehenkel. (2005). "Tree-based batch mode reinforcement learning". *Journal of Machine Learning Research*. 6.

Fiacco, A. V. (2020). *Mathematical programming with data perturbations*. CRC Press.

Fiacco, A. V. and Y. Ishizuka. (1990). "Sensitivity and stability analysis for nonlinear programming". *Annals of Operations Research*. 27(1): 215–235.

Fickinger, A., H. Hu, B. Amos, S. J. Russell, and N. Brown. (2021). "Scalable Online Planning via Reinforcement Learning Fine-Tuning". In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. 16951–16963. URL: https://proceedings.neurips.cc/paper/2021/hash/8ce8b102d40392a688f8c04b3cd6cae0-Abstract.html.

Finn, C., P. Abbeel, and S. Levine. (2017). "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 1126–1135. URL: http://proceedings.mlr.press/v70/finn17a.html.

Fleiss, J. (1993). "Review papers: The statistical basis of meta-analysis". *Statistical Methods in Medical Research*. 2(2): 121–145.

Foerster, J. N., R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch. (2017). "Learning with opponent-learning awareness". *ArXiv preprint*. abs/1709.04326.

Franceschi, L., M. Donini, P. Frasconi, and M. Pontil. (2017). "Forward and Reverse Gradient-Based Hyperparameter Optimization". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 1165–1173. URL: http://proceedings.mlr.press/v70/franceschi17a.html.

Franceschi, L., P. Frasconi, S. Salzo, R. Grazzi, and M. Pontil. (2018). "Bilevel Programming for Hyperparameter Optimization and Meta-Learning". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Vol. 80. *Proceedings of Machine Learning Research*. PMLR. 1563–1572. URL: http://proceedings.mlr.press/v80/franceschi18a.html.

Fujimoto, S., H. van Hoof, and D. Meger. (2018). "Addressing Function Approximation Error in Actor-Critic Methods". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Vol. 80. *Proceedings of Machine Learning Research*. PMLR. 1582–1591. URL: http://proceedings.mlr.press/v80/fujimoto18a.html.

Fujimoto, S., D. Meger, D. Precup, O. Nachum, and S. S. Gu. (2022). "Why Should I Trust You, Bellman? The Bellman Error is a Poor Replacement for Value Error". In: *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*. Vol. 162. *Proceedings of Machine Learning Research*. PMLR. 6918–6943. URL: https://proceedings.mlr.press/v162/fujimoto22a.html.

Gao, Z., Y. Wu, Y. Jia, and M. Harandi. (2020). "Learning to Optimize on SPD Manifolds". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. IEEE. 7697–7706. DOI: 10.1109/CVPR42600.2020.00772.

Garcia, J. R., F. Freddi, S. Fotiadis, M. Li, S. Vakili, A. Bernacchia, and G. Hennequin. (2023). "Fisher-Legendre (FishLeg) optimization of deep neural networks". In: *The Eleventh International Conference on Learning Representations, ICLR*.

Garnelo, M., D. Rosenbaum, C. Maddison, T. Ramalho, D. Saxton, M. Shanahan, Y. W. Teh, D. J. Rezende, and S. M. A. Eslami. (2018). "Conditional Neural Processes". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Vol. 80. *Proceedings of Machine Learning Research*. PMLR. 1690–1699. URL: http://proceedings.mlr.press/v80/garnelo18a.html.

Geist, M., B. Piot, and O. Pietquin. (2017). "Is the Bellman residual a bad proxy?" In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. 3205–3214. URL: https://proceedings.neurips.cc/paper/2017/hash/e0ab531ec312161511493b002f9be2ee-Abstract.html.

Gershman, S. and N. Goodman. (2014). "Amortized inference in probabilistic reasoning". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 36.

Goodfellow, I. J., J. Shlens, and C. Szegedy. (2015). "Explaining and Harnessing Adversarial Examples". In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. URL: http://arxiv.org/abs/1412.6572.

Gordon, J., J. Bronskill, M. Bauer, S. Nowozin, and R. E. Turner. (2019). "Meta-Learning Probabilistic Inference for Prediction". In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. URL: https://openreview.net/forum?id=HkxStoC5F7.

Gould, S., B. Fernando, A. Cherian, P. Anderson, R. S. Cruz, and E. Guo. (2016). "On differentiating parameterized argmin and argmax problems with application to bi-level optimization". *ArXiv preprint*. abs/1607.05447.

Grazzi, R., L. Franceschi, M. Pontil, and S. Salzo. (2020). "On the Iteration Complexity of Hypergradient Computation". In: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*. Vol. 119. *Proceedings of Machine Learning Research*. PMLR. 3748–3758. URL: http://proceedings.mlr.press/v119/grazzi20a.html.

Grefenstette, E., B. Amos, D. Yarats, P. M. Htut, A. Molchanov, F. Meier, D. Kiela, K. Cho, and S. Chintala. (2019). "Generalized Inner Loop Meta-Learning". *ArXiv preprint*. abs/1910.01727.

Gregor, K. and Y. LeCun. (2010). "Learning Fast Approximations of Sparse Coding". In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel.* Omnipress. 399–406. URL: https://icml.cc/Conferences/2010/papers/449.pdf.

Gruslys, A., R. Munos, I. Danihelka, M. Lanctot, and A. Graves. (2016). "Memory-Efficient Backpropagation Through Time". In: *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain.* 4125–4133. URL: https://proceedings.neurips.cc/paper/2016/hash/a501bebf79d570651ff601788ea9d16d-Abstract.html.

Grzeszczuk, R., D. Terzopoulos, and G. Hinton. (1998). "Neuroanimator: Fast neural network emulation and control of physics-based models". In: *25th Annual Conference on Computer Graphics and Interactive Techniques.* 9–20.

Guiasu, S. and A. Shenitzer. (1985). "The principle of maximum entropy". *The Mathematical Intelligencer.* 7(1): 42–48.

Gurumurthy, S., S. Bai, Z. Manchester, and J. Z. Kolter. (2021). "Joint inference and input optimization in equilibrium networks". In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual.* 16818–16832. URL: https://proceedings.neurips.cc/paper/2021/hash/8c3c27ac7d298331a1bdfd0a5e8703d3-Abstract.html.

Ha, D., A. M. Dai, and Q. V. Le. (2017). "HyperNetworks". In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings.* URL: https://openreview.net/forum?id=rkpACe1lx.

Haarnoja, T., A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, *et al.* (2018). "Soft actor-critic algorithms and applications". *ArXiv preprint.* abs/1812.05905.

Habets, P. (2010). "Stabilite Asyptotique Pour des Problemes de Perturbations Singulieres". In: *Stability Problems.* Springer. 2–18.

Hafner, D., T. P. Lillicrap, J. Ba, and M. Norouzi. (2020). "Dream to Control: Learning Behaviors by Latent Imagination". In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. URL: https://openreview.net/forum?id=S1lOTC4tDS.

Han, T., Y. Lu, S. Zhu, and Y. N. Wu. (2017). "Alternating Back-Propagation for Generator Network". In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*. AAAI Press. 1976–1984. URL: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14784.

Harlow, H. F. (1949). "The formation of learning sets." *Psychological Review*. 56(1): 51.

Harrison, J., L. Metz, and J. Sohl-Dickstein. (2022). "A Closer Look at Learned Optimization: Stability, Robustness, and Inductive Biases". *ArXiv preprint*. abs/2209.11208.

He, H. and R. Zou. (2021). "functorch: JAX-like composable function transforms for PyTorch". Github.

He, K., X. Zhang, S. Ren, and J. Sun. (2016). "Identity mappings in deep residual networks". In: *ECCV*. Springer. 630–645.

He, S., Y. Li, Y. Feng, S. Ho, S. Ravanbakhsh, W. Chen, and B. Póczos. (2019). "Learning to predict the cosmological structure formation". *Proceedings of the National Academy of Sciences*. 116(28): 13825–13832.

Heess, N., G. Wayne, D. Silver, T. P. Lillicrap, T. Erez, and Y. Tassa. (2015). "Learning Continuous Control Policies by Stochastic Value Gradients". In: *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*. 2944–2952. URL: https://proceedings.neurips.cc/paper/2015/hash/148510031349642de5ca0c544f31b2ef-Abstract.html.

Henaff, M., A. Canziani, and Y. LeCun. (2019). "Model-Predictive Policy Learning with Uncertainty Regularization for Driving in Dense Traffic". In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. URL: https://openreview.net/forum?id=HygQBn0cYm.

Higgins, I., L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner. (2017). "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. URL: https://openreview.net/forum?id=Sy2fzU9gl.

Hochreiter, S. and J. Schmidhuber. (1997). "Long short-term memory". *Neural Computation*. 9(8): 1735–1780.

Hochreiter, S., A. S. Younger, and P. R. Conwell. (2001). "Learning to learn using gradient descent". In: *International Conference on Artificial Neural Networks*. Springer. 87–94.

Hoffman, M. D., D. M. Blei, C. Wang, and J. Paisley. (2013). "Stochastic variational inference." *Journal of Machine Learning Research*. 14(5).

Hospedales, T., A. Antoniou, P. Micaelli, and A. Storkey. (2020). "Meta-learning in neural networks: A survey". *ArXiv preprint*. abs/2004.05439.

Hoyer, S., J. Sohl-Dickstein, and S. Greydanus. (2019). "Neural reparameterization improves structural optimization". *ArXiv preprint*. abs/1909.04240.

Hu, J., X. Liu, Z. Wen, and Y. Yuan. (2019). "A Brief Introduction to Manifold Optimization". *Journal of the Operations Research Society of China*. 8: 199–248.

Huang, K., N. D. Sidiropoulos, and A. P. Liavas. (2016). "A flexible and efficient algorithmic framework for constrained matrix and tensor factorization". *IEEE Transactions on Signal Processing*. 64(19): 5052–5065.

Huang, T., T. Chen, S. Liu, S. Chang, L. Amini, and Z. Wang. (2022). "Optimizer Amalgamation". In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. URL: https://openreview.net/forum?id=VqzXzA9hjaX.

Huszár, F. (2019). "Notes on iMAML: Meta-Learning with Implicit Gradients". URL: http://inference.vc.

Ichnowski, J., P. Jain, B. Stellato, G. Banjac, M. Luo, F. Borrelli, J. E. Gonzalez, I. Stoica, and K. Goldberg. (2021). "Accelerating Quadratic Optimization with Reinforcement Learning". In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual.* 21043–21055. URL: https://proceedings.neurips.cc/paper/2021/hash/afdec7005cc9f14302cd0474fd0f3c96-Abstract.html.

Jaeger, H. (2002). *Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach.* Vol. 5. GMD-Forschungszentrum Informationstechnik Bonn.

Jeong, Y. and H. O. Song. (2019). "Learning Discrete and Continuous Factors of Data via Alternating Disentanglement". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Vol. 97. *Proceedings of Machine Learning Research.* PMLR. 3091–3099. URL: http://proceedings.mlr.press/v97/jeong19d.html.

Jordan, M. I., Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. (1999). "An introduction to variational methods for graphical models". *Machine Learning.* 37(2): 183–233.

Karniadakis, G. E., I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. (2021). "Physics-informed machine learning". *Nature Reviews Physics.* 3(6): 422–440.

Kavukcuoglu, K., M. Ranzato, and Y. LeCun. (2010). "Fast inference in sparse coding algorithms with applications to object recognition". *arXiv preprint arXiv:1010.3467.*

Kehoe, E. J. (1988). "A layered network model of associative learning: learning to learn and configuration." *Psychological Review.* 95(4): 411.

Khalil, E. B., H. Dai, Y. Zhang, B. Dilkina, and L. Song. (2017). "Learning Combinatorial Optimization Algorithms over Graphs". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. 6348–6358. URL: https://proceedings.neurips.cc/paper/2017/hash/d9896106ca98d3d05b8cbdf4fd8b13a1-Abstract.html.

Khalil, E. B., P. L. Bodic, L. Song, G. L. Nemhauser, and B. Dilkina. (2016). "Learning to Branch in Mixed Integer Programming". In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*. AAAI Press. 724–731. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12514.

Khodak, M., M.-F. F. Balcan, A. Talwalkar, and S. Vassilvitskii. (2022). "Learning predictions for algorithms with predictions". *Advances in Neural Information Processing Systems*. 35: 3542–3555.

Kim, Y., S. Wiseman, A. C. Miller, D. A. Sontag, and A. M. Rush. (2018). "Semi-Amortized Variational Autoencoders". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Vol. 80. *Proceedings of Machine Learning Research*. PMLR. 2683–2692. URL: http://proceedings.mlr.press/v80/kim18e.html.

Kim, Y. H. (2020). "Deep latent variable models of natural language". *PhD thesis*. Harvard University.

Kingma, D. P. and M. Welling. (2019). "An Introduction to Variational Autoencoders". *Foundations and Trends® in Machine Learning*. 12(4): 307–392.

Kingma, D. P. and J. Ba. (2015). "Adam: A Method for Stochastic Optimization". In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. URL: http://arxiv.org/abs/1412.6980.

Kingma, D. P. and M. Welling. (2014). "Auto-Encoding Variational Bayes". In: *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. URL: http://arxiv.org/abs/1312.6114.

Kirk, D. E. (2004). *Optimal control theory: an introduction*. Courier Corporation.

Klatte, D. and B. Kummer. (2006). *Nonsmooth equations in optimization: regularity, calculus, methods and applications*. Vol. 60. Springer Science & Business Media.

Knyazev, B., M. Drozdzal, G. W. Taylor, and A. Romero-Soriano. (2021). "Parameter Prediction for Unseen Deep Architectures". In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. 29433–29448. URL: https://proceedings.neurips.cc/paper/2021/hash/f6185f0ef02dcaec414a3171cd01c697-Abstract.html.

Konda, V. and J. Tsitsiklis. (1999). "Actor-critic algorithms". *Advances in Neural Information Processing Systems*. 12.

Korotin, A., V. Egiazarian, A. Asadulaev, A. Safin, and E. Burnaev. (2021). "Wasserstein-2 Generative Networks". In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. URL: https://openreview.net/forum?id=bEoxzW%5C_EXsa.

Kotary, J., F. Fioretto, P. Van Hentenryck, and B. Wilder. (2021). "End-to-end constrained optimization learning: A survey". *ArXiv preprint*. abs/2103.16378.

Kovachki, N., S. Lanthaler, and S. Mishra. (2021). "On universal approximation and error bounds for fourier neural operators". *The Journal of Machine Learning Research*. 22(1): 13237–13312.

Kriváchy, T., Y. Cai, J. Bowles, D. Cavalcanti, and N. Brunner. (2020). "Fast semidefinite programming with feedforward neural networks". *ArXiv preprint*. abs/2011.05785.

Ladickỳ, L., S. Jeong, B. Solenthaler, M. Pollefeys, and M. Gross. (2015). "Data-driven fluid simulations using regression forests". *ACM Transactions on Graphics (TOG)*. 34(6): 1–9.

Lake, B. M., T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. (2017). "Building machines that learn and think like people". *Behavioral and Brain Sciences*. 40.

Lambert, N., B. Amos, O. Yadan, and R. Calandra. (2020). "Objective mismatch in model-based reinforcement learning". *ArXiv preprint*. abs/2002.04523.

Lample, G. and F. Charton. (2020). "Deep Learning For Symbolic Mathematics". In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. URL: https://openreview.net/forum?id=S1eZYeHFDS.

Le, H. M., C. Voloshin, and Y. Yue. (2019). "Batch Policy Learning under Constraints". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Vol. 97. *Proceedings of Machine Learning Research*. PMLR. 3703–3712. URL: http://proceedings.mlr.press/v97/le19a.html.

LeCun, Y. (1998). "The MNIST database of handwritten digits". URL: http://yann.lecun.com/exdb/mnist/.

Lee, J., Y. Lee, J. Kim, A. R. Kosiorek, S. Choi, and Y. W. Teh. (2019a). "Set Transformer: A Framework for Attention-based Permutation-Invariant Neural Networks". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Vol. 97. *Proceedings of Machine Learning Research*. PMLR. 3744–3753. URL: http://proceedings.mlr.press/v97/lee19d.html.

Lee, K., S. Maji, A. Ravichandran, and S. Soatto. (2019b). "Meta-Learning With Differentiable Convex Optimization". In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. 10657–10665. DOI: 10.1109/CVPR.2019.01091.

Levine, S. and P. Abbeel. (2014). "Learning Neural Network Policies with Guided Policy Search under Unknown Dynamics". In: *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*. 1071–1079. URL: https://proceedings.neurips.cc/paper/2014/hash/6766aa2750c19aad2fa1b32f36ed4aee-Abstract.html.

Levine, S., C. Finn, T. Darrell, and P. Abbeel. (2016). "End-to-end training of deep visuomotor policies". *The Journal of Machine Learning Research*. 17(1): 1334–1373.

Levine, S. and V. Koltun. (2013). "Guided Policy Search". In: *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*. Vol. 28. *JMLR Workshop and Conference Proceedings*. JMLR.org. 1–9. URL: http://proceedings.mlr.press/v28/levine13.html.

Li, K. and J. Malik. (2017a). “Learning to Optimize”. In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. URL: https://openreview.net/forum?id=ry4Vrt5gl.

Li, K. and J. Malik. (2017b). “Learning to optimize neural nets”. *ArXiv preprint*. abs/1703.00441.

Li, Z., N. B. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. M. Stuart, and A. Anandkumar. (2021a). “Fourier Neural Operator for Parametric Partial Differential Equations”. In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. URL: https://openreview.net/forum?id=c8P9NQVtmnO.

Li, Z., H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar. (2021b). “Physics-informed neural operator for learning partial differential equations”. *ArXiv preprint*. abs/2111.03794.

Liao, R., Y. Xiong, E. Fetaya, L. Zhang, K. Yoon, X. Pitkow, R. Urtasun, and R. S. Zemel. (2018). “Reviving and Improving Recurrent Back-Propagation”. In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Vol. 80. *Proceedings of Machine Learning Research*. PMLR. 3088–3097. URL: http://proceedings.mlr.press/v80/liao18c.html.

Lillicrap, T. P., J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. (2016). “Continuous control with deep reinforcement learning”. In: *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*. URL: http://arxiv.org/abs/1509.02971.

Liu, X., Y. Lu, A. Abbasi, M. Li, J. Mohammadi, and S. Kolouri. (2022). “Teaching Networks to Solve Optimization Problems”. arXiv: 2202.04104 [cs.LG].

Lodi, A. and G. Zarpellon. (2017). “On learning and branching: a survey”. *Top*. 25(2): 207–236.

Long, J., E. Shelhamer, and T. Darrell. (2015). “Fully convolutional networks for semantic segmentation”. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*. IEEE Computer Society. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

Lorraine, J., P. Vicol, and D. Duvenaud. (2020). “Optimizing Millions of Hyperparameters by Implicit Differentiation”. In: *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020, Online [Palermo, Sicily, Italy]*. Vol. 108. *Proceedings of Machine Learning Research*. PMLR. 1540–1552. URL: http://proceedings.mlr.press/v108/lorraine20a.html.

Lowrey, K., A. Rajeswaran, S. M. Kakade, E. Todorov, and I. Mordatch. (2019). “Plan Online, Learn Offline: Efficient Learning and Exploration via Model-Based Control”. In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. URL: https://openreview.net/forum?id=Byey7n05FQ.

Luo, R., F. Tian, T. Qin, E. Chen, and T. Liu. (2018). “Neural Architecture Optimization”. In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. 7827–7838. URL: https://proceedings.neurips.cc/paper/2018/hash/933670f1ac8ba969f32989c312faba75-Abstract.html.

Lv, K., S. Jiang, and J. Li. (2017). “Learning Gradient Descent: Better Generalization and Longer Horizons”. In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 2247–2255. URL: http://proceedings.mlr.press/v70/lv17a.html.

Maclaurin, D. (2016). “Modeling, inference and optimization with composable differentiable procedures”. *PhD thesis*. Harvard University.

Maclaurin, D., D. Duvenaud, and R. P. Adams. (2015a). “Autograd: Effortless gradients in numpy”. In: *ICML 2015 AutoML Workshop*. Vol. 238. 5.

Maclaurin, D., D. Duvenaud, and R. P. Adams. (2015b). "Gradient-based Hyperparameter Optimization through Reversible Learning". In: *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*. Vol. 37. *JMLR Workshop and Conference Proceedings*. JMLR.org. 2113–2122. URL: http://proceedings.mlr.press/v37/maclaurin15.html.

Maheswaranathan, N., D. Sussillo, L. Metz, R. Sun, and J. Sohl-Dickstein. (2021). "Reverse engineering learned optimizers reveals known and novel mechanisms". In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. 19910–19922. URL: https://proceedings.neurips.cc/paper/2021/hash/a57ecd54d4df7d999bd9c5e3b973ec75-Abstract.html.

Maillard, O.-A., R. Munos, A. Lazaric, and M. Ghavamzadeh. (2010). "Finite-sample analysis of Bellman residual minimization". In: *2nd Asian Conference on Machine Learning*. JMLR Workshop and Conference Proceedings. 299–314.

Makkuva, A. V., A. Taghvaei, S. Oh, and J. D. Lee. (2020). "Optimal transport mapping via input convex neural networks". In: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*. Vol. 119. *Proceedings of Machine Learning Research*. PMLR. 6672–6681. URL: http://proceedings.mlr.press/v119/makkuva20a.html.

Marino, J., M. Cvitkovic, and Y. Yue. (2018a). "A General Method for Amortizing Variational Filtering". In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. 7868–7879. URL: https://proceedings.neurips.cc/paper/2018/hash/060afc8a563aaccd288f98b7c8723b61-Abstract.html.

Marino, J., A. Piché, A. D. Ialongo, and Y. Yue. (2021). "Iterative Amortized Policy Optimization". In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. 15667–15681. URL: https://proceedings.neurips.cc/paper/2021/hash/83fa5a432ae55c253d0e60dbfa716723-Abstract.html.

Marino, J., Y. Yue, and S. Mandt. (2018b). "Iterative Amortized Inference". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Vol. 80. *Proceedings of Machine Learning Research*. PMLR. 3400–3409. URL: http://proceedings.mlr.press/v80/marino18a.html.

Marino, J. L. (2021). "Learned Feedback & Feedforward Perception & Control". *PhD thesis*. California Institute of Technology.

Marwah, T., Z. C. Lipton, and A. Risteski. (2021). "Parametric Complexity Bounds for Approximating PDEs with Neural Networks". In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. 15044–15055. URL: https://proceedings.neurips.cc/paper/2021/hash/7edccc661418aeb5761dbcdc06ad490c-Abstract.html.

Meinhardt, T., M. Möller, C. Hazirbas, and D. Cremers. (2017). "Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems". In: *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. IEEE Computer Society. 1799–1808. DOI: 10.1109/ICCV.2017.198.

Mena, G. E., D. Belanger, S. W. Linderman, and J. Snoek. (2018). "Learning Latent Permutations with Gumbel-Sinkhorn Networks". In: *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. URL: https://openreview.net/forum?id=Byt3oJ-0W.

Merchant, A., L. Metz, S. S. Schoenholz, and E. D. Cubuk. (2021). "Learn2Hop: Learned Optimization on Rough Landscapes". In: *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*. Vol. 139. *Proceedings of Machine Learning Research*. PMLR. 7643–7653. URL: http://proceedings.mlr.press/v139/merchant21a.html.

Metz, L., C. D. Freeman, S. S. Schoenholz, and T. Kachman. (2021). "Gradients are Not All You Need". *ArXiv preprint*. abs/2111.05803.

Metz, L., J. Harrison, C. D. Freeman, A. Merchant, L. Beyer, J. Bradbury, N. Agrawal, B. Poole, I. Mordatch, A. Roberts, *et al.* (2022). "VeLO: Training Versatile Learned Optimizers by Scaling Up". *ArXiv preprint*. abs/2211.09760.

Metz, L., N. Maheswaranathan, J. Nixon, C. D. Freeman, and J. Sohl-Dickstein. (2019a). "Understanding and correcting pathologies in the training of learned optimizers". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Vol. 97. *Proceedings of Machine Learning Research*. PMLR. 4556–4565. URL: http://proceedings.mlr.press/v97/metz19a.html.

Metz, L., N. Maheswaranathan, J. Shlens, J. Sohl-Dickstein, and E. D. Cubuk. (2019b). "Using learned optimizers to make models robust to input noise". *ArXiv preprint*. abs/1906.03367.

Metz, L., B. Poole, D. Pfau, and J. Sohl-Dickstein. (2017). "Unrolled Generative Adversarial Networks". In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. URL: https://openreview.net/forum?id=BydrOIcle.

Milgrom, P. and I. Segal. (2002). "Envelope theorems for arbitrary choice sets". *Econometrica*. 70(2): 583–601.

Misra, S., L. Roald, and Y. Ng. (2021). "Learning for constrained optimization: Identifying optimal active constraint sets". *INFORMS Journal on Computing*.

Mnih, A. and K. Gregor. (2014). "Neural Variational Inference and Learning in Belief Networks". In: *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*. Vol. 32. *JMLR Workshop and Conference Proceedings*. JMLR.org. 1791–1799. URL: http://proceedings.mlr.press/v32/mnih14.html.

Mohamed, S., M. Rosca, M. Figurnov, and A. Mnih. (2020). "Monte Carlo Gradient Estimation in Machine Learning". *Journal of Machine Learning Research*. 21: 132:1–132:62. URL: http://jmlr.org/papers/v21/19-346.html.

Monga, V., Y. Li, and Y. C. Eldar. (2021). "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing". *IEEE Signal Processing Magazine*. 38(2): 18–44.

Montgomery, W. H. and S. Levine. (2016). "Guided Policy Search via Approximate Mirror Descent". In: *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain.* 4008–4016. URL: https://proceedings.neurips.cc/paper/2016/hash/a00e5eb0973d24649a4a920fc53d9564-Abstract.html.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT Press.

Nesterov, Y. (1983). "A method for unconstrained convex minimization problem with the rate of convergence O (1/kˆ 2)". In: *Doklady an ussr*. Vol. 269. 543–547.

Nesterov, Y. *et al.* (2018). *Lectures on convex optimization*. Vol. 137. Springer.

Nguyen, K. and N. Ho. (2022). "Amortized Projection Optimization for Sliced Wasserstein Generative Models". *ArXiv preprint*. abs/2203.13417.

Nichol, A., J. Achiam, and J. Schulman. (2018). "On first-order meta-learning algorithms". *ArXiv preprint*. abs/1803.02999.

Nocedal, J. and S. Wright. (2006). *Numerical optimization*. Springer Science & Business Media.

O'Donoghue, B., E. Chu, N. Parikh, and S. Boyd. (2016). "Conic optimization via operator splitting and homogeneous self-dual embedding". *Journal of Optimization Theory and Applications*. 169(3): 1042–1068.

Olshausen, B. A. and D. J. Field. (1996). "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". *Nature*. 381(6583): 607–609.

Osa, T., J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel, and J. Peters. (2018). "An algorithmic perspective on imitation learning". *Foundations and Trends® in Robotics*. 7(1-2): 1–179.

Pan, X., M. Chen, T. Zhao, and S. H. Low. (2020). "DeepOPF: A feasibility-optimized deep neural network approach for AC optimal power flow problems". *ArXiv preprint*. abs/2007.01002.

Parikh, N. and S. Boyd. (2014). "Proximal algorithms". *Foundations and Trends® in Optimization*. 1(3): 127–239.

Parmas, P., C. E. Rasmussen, J. Peters, and K. Doya. (2018). "PIPPS: Flexible Model-Based Policy Search Robust to the Curse of Chaos". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Vol. 80. *Proceedings of Machine Learning Research*. PMLR. 4062–4071. URL: http://proceedings.mlr.press/v80/parmas18a.html.

Parmas, P. and M. Sugiyama. (2021). "A unified view of likelihood ratio and reparameterization gradients". In: *The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021, April 13-15, 2021, Virtual Event*. Vol. 130. *Proceedings of Machine Learning Research*. PMLR. 4078–4086. URL: http://proceedings.mlr.press/v130/parmas21a.html.

Pascanu, R., T. Mikolov, and Y. Bengio. (2013). "On the difficulty of training recurrent neural networks". In: *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*. Vol. 28. *JMLR Workshop and Conference Proceedings*. JMLR.org. 1310–1318. URL: http://proceedings.mlr.press/v28/pascanu13.html.

Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. (2019). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. 8024–8035. URL: https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html.

Pearlmutter, B. A. (1994). "Fast exact multiplication by the Hessian". *Neural Computation*. 6(1): 147–160.

Pearlmutter, B. A. (1996). "An investigation of the gradient descent process in neural networks". *PhD thesis*. Carnegie Mellon University.

Pearlmutter, B. A. and J. M. Siskind. (2008). "Reverse-mode AD in a functional framework: Lambda the ultimate backpropagator". *ACM Transactions on Programming Languages and Systems (TOPLAS)*. 30(2): 1–36.

Pennec, X. (2006). "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements". *Journal of Mathematical Imaging and Vision*. 25(1): 127–154.

Peyré, G. and M. Cuturi. (2019). "Computational optimal transport: With applications to data science". *Foundations and Trends® in Machine Learning*. 11(5-6): 355–607.

Poli, M., S. Massaroli, A. Yamashita, H. Asama, and J. Park. (2020). "Hypersolvers: Toward Fast Continuous-Depth Models". In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. URL: https://proceedings.neurips.cc/paper/2020/hash/f1686b4badcf28d33ed632036c7ab0b8-Abstract.html.

Prémont-Schwarz, I., J. Vitku, and J. Feyereisl. (2022). "A Simple Guard for Learned Optimizers". In: *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*. Vol. 162. *Proceedings of Machine Learning Research*. PMLR. 17910–17925. URL: https://proceedings.mlr.press/v162/premont-schwarz22a.html.

Raghu, A., M. Raghu, S. Bengio, and O. Vinyals. (2020). "Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML". In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. URL: https://openreview.net/forum?id=rkgMkCEtPB.

Rajeswaran, A., C. Finn, S. M. Kakade, and S. Levine. (2019). "Meta-Learning with Implicit Gradients". In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. 113–124. URL: https://proceedings.neurips.cc/paper/2019/hash/072b030ba126b2f4b2374f342be9ed44-Abstract.html.

Ravi, S. and A. Beatson. (2019). "Amortized Bayesian Meta-Learning". In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. URL: https://openreview.net/forum?id=rkgpy3C5tX.

Ravi, S. and H. Larochelle. (2017). "Optimization as a Model for Few-Shot Learning". In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. URL: https://openreview.net/forum?id=rJY0-Kcll.

Real, E., C. Liang, D. R. So, and Q. V. Le. (2020). "AutoML-Zero: Evolving Machine Learning Algorithms From Scratch". In: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*. Vol. 119. *Proceedings of Machine Learning Research*. PMLR. 8007–8019. URL: http://proceedings.mlr.press/v119/real20a.html.

Rezende, D. J. and S. Mohamed. (2015). "Variational Inference with Normalizing Flows". In: *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*. Vol. 37. *JMLR Workshop and Conference Proceedings*. JMLR.org. 1530–1538. URL: http://proceedings.mlr.press/v37/rezende15.html.

Rezende, D. J., S. Mohamed, and D. Wierstra. (2014). "Stochastic Backpropagation and Approximate Inference in Deep Generative Models". In: *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*. Vol. 32. *JMLR Workshop and Conference Proceedings*. JMLR.org. 1278–1286. URL: http://proceedings.mlr.press/v32/rezende14.html.

Rezende, D. J. and F. Viola. (2018). "Taming vaes". *ArXiv preprint*. abs/1810.00597.

Al-Rfou, R., G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky, *et al.* (2016). "Theano: A Python framework for fast computation of mathematical expressions". *arXiv e-prints*: arXiv–1605.

Richter, S. L. and R. A. Decarlo. (1983). "Continuation methods: Theory and applications". *IEEE Transactions on Systems, Man, and Cybernetics*. SMC-13(4): 459–464.

Ronneberger, O., P. Fischer, and T. Brox. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer. 234–241.

Ruder, S. (2017). "An overview of multi-task learning in deep neural networks". *ArXiv preprint*. abs/1706.05098.

Runarsson, T. P. and M. T. Jonsson. (2000). "Evolution and design of distributed learning rules". In: *2000 IEEE Symposium on Combinations of Evolutionary Computation and Neural Networks*. IEEE. 59–63.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein. (2015). "Imagenet large scale visual recognition challenge". *International Journal of Computer Vision*. 115(3): 211–252.

Rusu, A. A., D. Rao, J. Sygnowski, O. Vinyals, R. Pascanu, S. Osindero, and R. Hadsell. (2019). "Meta-Learning with Latent Embedding Optimization". In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. URL: https://openreview.net/forum?id=BJgklhAcK7.

Ryu, M., Y. Chow, R. Anderson, C. Tjandraatmadja, and C. Boutilier. (2020). "CAQL: Continuous Action Q-Learning". In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. URL: https://openreview.net/forum?id=BkxXe0Etwr.

Sacks, J. and B. Boots. (2022). "Learning to Optimize in Model Predictive Control". In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE. 10549–10556.

Sakaue, S. and T. Oki. (2022). "Discrete-Convex-Analysis-Based Framework for Warm-Starting Algorithms with Predictions". In: *Advances in Neural Information Processing Systems*. Ed. by A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho. URL: https://openreview.net/forum?id=-GgDBzwZ-e7.

Salakhutdinov, R. (2014). "Deep learning". In: *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*. ACM. 1973. DOI: 10.1145/2623330.2630809.

Salimans, T. and J. Ho. (2022). “Progressive Distillation for Fast Sampling of Diffusion Models”. In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. URL: https://openreview.net/forum?id=TIdIXIpzhoI.

Sanchez-Gonzalez, A., J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. W. Battaglia. (2020). “Learning to Simulate Complex Physics with Graph Networks”. In: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*. Vol. 119. *Proceedings of Machine Learning Research*. PMLR. 8459–8468. URL: http://proceedings.mlr.press/v119/sanchez-gonzalez20a.html.

Santambrogio, F. (2015). “Optimal transport for applied mathematicians”. *Birkäuser Cham*.

Scherrer, B. (2010). “Should one compute the Temporal Difference fix point or minimize the Bellman Residual? The unified oblique projection view”. In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel*. Omnipress. 959–966. URL: https://icml.cc/Conferences/2010/papers/654.pdf.

Schmidhuber, J. (1987). “Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-... hook”. *PhD thesis*. Technische Universität München.

Schmidhuber, J. (1995). “On learning how to learn learning strategies”. *Tech. rep.* TU Munchen.

Schwarzschild, A., E. Borgnia, A. Gupta, F. Huang, U. Vishkin, M. Goldblum, and T. Goldstein. (2021). “Can You Learn an Algorithm? Generalizing from Easy to Hard Problems with Recurrent Networks”. In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. 6695–6706. URL: https://proceedings.neurips.cc/paper/2021/hash/3501672ebc68a5524629080e3ef60aef-Abstract.html.

Sercu, T., R. Verkuil, J. Meier, B. Amos, Z. Lin, C. Chen, J. Liu, Y. LeCun, and A. Rives. (2021). “Neural Potts Model”. *bioRxiv*.

Shaban, A., C. Cheng, N. Hatch, and B. Boots. (2019). "Truncated Back-propagation for Bilevel Optimization". In: *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan*. Vol. 89. *Proceedings of Machine Learning Research*. PMLR. 1723–1732. URL: http://proceedings.mlr.press/v89/shaban19a.html.

Shao, Z., J. Yang, C. Shen, and S. Ren. (2021). "Learning for Robust Combinatorial Optimization: Algorithm and Application". *ArXiv preprint*. abs/2112.10377.

Shapiro, A. (2003). "Sensitivity Analysis of Generalized Equations." *Journal of Mathematical Sciences*. 115(4).

Shu, R. (2017). "Amortized Optimization". URL: https://ruishu.io/2017/11/07/amortized-optimization/.

Silver, D., A. Goyal, I. Danihelka, M. Hessel, and H. van Hasselt. (2022). "Learning by Directional Gradient Descent". In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. URL: https://openreview.net/forum?id=5i7lJLuhTm.

Silver, D., G. Lever, N. Heess, T. Degris, D. Wierstra, and M. A. Riedmiller. (2014). "Deterministic Policy Gradient Algorithms". In: *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*. Vol. 32. *JMLR Workshop and Conference Proceedings*. JMLR.org. 387–395. URL: http://proceedings.mlr.press/v32/silver14.html.

Sjölund, J. (2023). "A Tutorial on Parametric Variational Inference". *ArXiv preprint*. abs/2301.01236.

Sjölund, J. and M. Bånkestad. (2022). "Graph-based Neural Acceleration for Nonnegative Matrix Factorization". arXiv: 2202.00264 [cs.LG].

Smola, A. J., S. V. N. Vishwanathan, and Q. V. Le. (2007). "Bundle Methods for Machine Learning". In: *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3-6, 2007*. Curran Associates, Inc. 1377–1384. URL: https://proceedings.neurips.cc/paper/2007/hash/26337353b7962f533d78c762373b3318-Abstract.html.

Sønderby, C. K., T. Raiko, L. Maaløe, S. K. Sønderby, and O. Winther. (2016). "Ladder Variational Autoencoders". In: *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*. 3738–3746. URL: https://proceedings.neurips.cc/paper/2016/hash/6ae07dcb33ec3b7c814df797cbda0f87-Abstract.html.

Stanley, K. O., D. B. D'Ambrosio, and J. Gauci. (2009). "A hypercube-based encoding for evolving large-scale neural networks". *Artificial Life*. 15(2): 185–212.

Stellato, B., G. Banjac, P. Goulart, A. Bemporad, and S. Boyd. (2018). "OSQP: An operator splitting solver for quadratic programs". In: *UKACC 12th International Conference on Control (CONTROL)*. IEEE. 339–339.

Still, G. (2018). "Lectures on parametric optimization: An introduction". *Optimization Online*.

Stuhlmüller, A., J. Taylor, and N. D. Goodman. (2013). "Learning Stochastic Inverses". In: *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*. 3048–3056. URL: https://proceedings.neurips.cc/paper/2013/hash/7f53f8c6c730af6aeb52e66eb74d8507-Abstract.html.

Sutton, R. S. and A. G. Barto. (2018). *Reinforcement learning: An introduction*. MIT Press.

Swersky, K., Y. Rubanova, D. Dohan, and K. Murphy. (2020). "Amortized Bayesian Optimization over Discrete Spaces". In: *Proceedings of the Thirty-Sixth Conference on Uncertainty in Artificial Intelligence, UAI 2020, virtual online, August 3-6, 2020*. Vol. 124. *Proceedings of Machine Learning Research*. AUAI Press. 769–778. URL: http://proceedings.mlr.press/v124/swersky20a.html.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. (2016). "Rethinking the Inception Architecture for Computer Vision". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. IEEE Computer Society. 2818–2826. DOI: 10.1109/CVPR.2016.308.

Taghvaei, A. and A. Jalali. (2019). "2-wasserstein approximation via restricted convex potentials with application to improved training for gans". *ArXiv preprint*. abs/1902.07197.

Tallec, C. and Y. Ollivier. (2017). "Unbiasing truncated backpropagation through time". *ArXiv preprint*. abs/1705.08209.

Tallec, C. and Y. Ollivier. (2018). "Unbiased Online Recurrent Optimization". In: *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. URL: https://openreview.net/forum?id=rJQDjk-0b.

Tennenholtz, G. and S. Mannor. (2019). "The Natural Language of Actions". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Vol. 97. *Proceedings of Machine Learning Research*. PMLR. 6196–6205. URL: http://proceedings.mlr.press/v97/tennenholtz19a.html.

Thornton, J. and M. Cuturi. (2022). "Rethinking Initialization of the Sinkhorn Algorithm". *ArXiv preprint*. abs/2206.07630.

Thrun, S. and L. Pratt. (1998). "Learning to learn: Introduction and overview". In: *Learning to Learn*. Springer. 3–17.

Usman, A., M. Rafiq, M. Saeed, A. Nauman, A. Almqvist, and M. Liwicki. (2021). "Machine Learning Computational Fluid Dynamics". In: *2021 Swedish Artificial Intelligence Society Workshop (SAIS)*. IEEE. 1–4.

Van de Wiele, T., D. Warde-Farley, A. Mnih, and V. Mnih. (2020). "Q-learning in enormous action spaces via amortized approximate maximization". *ArXiv preprint*. abs/2001.08116.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. 5998–6008. URL: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Venkatakrishnan, S. V., C. A. Bouman, and B. Wohlberg. (2013). "Plug-and-play priors for model based reconstruction". In: *IEEE Global Conference on Signal and Information Processing*. IEEE. 945–948.

Venkataraman, S. and B. Amos. (2021). "Neural Fixed-Point Acceleration for Convex Optimization". *ArXiv preprint*. abs/2107.10254.

Vicol, P., L. Metz, and J. Sohl-Dickstein. (2021). "Unbiased Gradient Estimation in Unrolled Computation Graphs with Persistent Evolution Strategies". In: *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*. Vol. 139. *Proceedings of Machine Learning Research*. PMLR. 10553–10563. URL: http://proceedings.mlr.press/v139/vicol21a.html.

Vilalta, R. and Y. Drissi. (2002). "A perspective view and survey of meta-learning". *Artificial Intelligence Review*. 18(2): 77–95.

Villani, C. (2009). *Optimal transport: old and new*. Vol. 338. Springer.

Vinuesa, R. and S. L. Brunton. (2021). "The potential of machine learning to enhance computational fluid dynamics". *ArXiv preprint*. abs/2110.02085.

Wainwright, M. J. and M. I. Jordan. (2008). "Graphical models, exponential families, and variational inference". *Foundations and Trends® in Machine Learning*. 1(1-2): 1–305.

Walker, H. F. and P. Ni. (2011). "Anderson acceleration for fixed-point iterations". *SIAM Journal on Numerical Analysis*. 49(4): 1715–1735.

Wang, H., H. Zhao, and B. Li. (2021). "Bridging Multi-Task Learning and Meta-Learning: Towards Efficient Training and Effective Adaptation". In: *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*. Vol. 139. *Proceedings of Machine Learning Research*. PMLR. 10991–11002. URL: http://proceedings.mlr.press/v139/wang21ad.html.

Wang, T. and J. Ba. (2020). "Exploring Model-based Planning with Policy Networks". In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. URL: https://openreview.net/forum?id=H1exf64KwH.

Ward, L. B. (1937). "Reminiscence and rote learning." *Psychological Monographs*. 49(4): i.

Watkins, C. J. and P. Dayan. (1992). "Q-learning". *Machine Learning*. 8(3-4): 279–292.

Watson, L. T. and R. T. Haftka. (1989). "Modern homotopy methods in optimization". *Computer Methods in Applied Mechanics and Engineering*. 74(3): 289–305.

Webb, S., A. Golinski, R. Zinkov, S. Narayanaswamy, T. Rainforth, Y. W. Teh, and F. Wood. (2018). "Faithful Inversion of Generative Models for Effective Amortized Inference". In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. 3074–3084. URL: https://proceedings.neurips.cc/paper/2018/hash/894b77f805bd94d292574c38c5d628d5-Abstract.html.

Weng, L. (2018). "Meta-Learning: Learning to Learn Fast". URL: http://lilianweng.github.io/lil-log.

Werbos, P. J. (1990). "Backpropagation through time: what it does and how to do it". *Proceedings of the IEEE*. 78(10): 1550–1560.

Wichrowska, O., N. Maheswaranathan, M. W. Hoffman, S. G. Colmenarejo, M. Denil, N. de Freitas, and J. Sohl-Dickstein. (2017). "Learned Optimizers that Scale and Generalize". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 3751–3760. URL: http://proceedings.mlr.press/v70/wichrowska17a.html.

Wiewel, S., M. Becher, and N. Thuerey. (2019). "Latent space physics: Towards learning the temporal evolution of fluid flow". In: *Computer Graphics Forum*. Vol. 38. Wiley Online Library. 71–82.

Williams, R. J. (1992). "Simple statistical gradient-following algorithms for connectionist reinforcement learning". *Reinforcement Learning*: 5–32.

Williams, R. J. and D. Zipser. (1989). "A learning algorithm for continually running fully recurrent neural networks". *Neural Computation*. 1(2): 270–280.

Wu, M., K. Choi, N. D. Goodman, and S. Ermon. (2020). "Meta-Amortized Variational Inference and Learning". In: *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press. 6404–6412. URL: https://aaai.org/ojs/index.php/AAAI/article/view/6111.

Wu, Y., M. Ren, R. Liao, and R. B. Grosse. (2018). "Understanding Short-Horizon Bias in Stochastic Meta-Optimization". In: *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. URL: https://openreview.net/forum?id=H1MczcgR-.

Xiao, Y., E. P. Xing, and W. Neiswanger. (2021). "Amortized Auto-Tuning: Cost-Efficient Transfer Optimization for Hyperparameter Recommendation". *ArXiv preprint*. abs/2106.09179.

Xie, K., H. Bharadhwaj, D. Hafner, A. Garg, and F. Shkurti. (2021). "Latent Skill Planning for Exploration and Transfer". In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. URL: https://openreview.net/forum?id=jXe91kq3jAq.

Xue, T., A. Beatson, S. Adriaenssens, and R. P. Adams. (2020). "Amortized Finite Element Analysis for Fast PDE-Constrained Optimization". In: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*. Vol. 119. *Proceedings of Machine Learning Research*. PMLR. 10638–10647. URL: http://proceedings.mlr.press/v119/xue20a.html.

You, Y., Y. Cao, T. Chen, Z. Wang, and Y. Shen. (2022). "Bayesian Modeling and Uncertainty Quantification for Learning to Optimize: What, Why, and How". In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. URL: https://openreview.net/forum?id=EVVadRFRgL7.

Yu, F. and V. Koltun. (2016). "Multi-Scale Context Aggregation by Dilated Convolutions". In: *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*. URL: http://arxiv.org/abs/1511.07122.

Zaheer, M., S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov, and A. J. Smola. (2017). "Deep Sets". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. 3391–3401. URL: https://proceedings.neurips.cc/paper/2017/hash/f22e4747da1aa27e363d86d40ff442fe-Abstract.html.

Zamzam, A. S. and K. Baker. (2020). "Learning optimal solutions for extremely fast AC optimal power flow". In: *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE. 1–6.

Zeiler, M. D. (2012). "Adadelta: an adaptive learning rate method". *arXiv preprint arXiv:1212.5701*.

Zhang, C. and V. R. Lesser. (2010). "Multi-Agent Learning with Policy Prediction". In: *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*. AAAI Press. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1885.

Zhang, C., M. Ren, and R. Urtasun. (2019a). "Graph HyperNetworks for Neural Architecture Search". In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. URL: https://openreview.net/forum?id=rkgW0oA9FX.

Zhang, J., B. O'Donoghue, and S. Boyd. (2020). "Globally convergent type-I Anderson acceleration for nonsmooth fixed-point iterations". *SIAM Journal on Optimization*. 30(4): 3170–3197.

Zhang, K., W. Zuo, S. Gu, and L. Zhang. (2017). "Learning Deep CNN Denoiser Prior for Image Restoration". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society. 2808–2817. DOI: 10.1109/CVPR.2017.300.

Zhang, X., M. Bujarbaruah, and F. Borrelli. (2019b). "Safe and near-optimal policy learning for model predictive control using primal-dual neural networks". In: *American Control Conference (ACC)*. IEEE. 354–359.

Zheng, W., T. Chen, T. Hu, and Z. Wang. (2022). "Symbolic Learning to Optimize: Towards Interpretability and Scalability". In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. URL: https://openreview.net/forum?id=ef0nInZHKIC.

Zhmoginov, A., M. Sandler, and M. Vladymyrov. (2022). "HyperTransformer: Model Generation for Supervised and Semi-Supervised Few-Shot Learning". In: *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*. Vol. 162. *Proceedings of Machine Learning Research*. PMLR. 27075–27098. URL: https://proceedings.mlr.press/v162/zhmoginov22a.html.

Zintgraf, L. M., K. Shiarlis, V. Kurin, K. Hofmann, and S. Whiteson. (2019). "Fast Context Adaptation via Meta-Learning". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Vol. 97. *Proceedings of Machine Learning Research*. PMLR. 7693–7702. URL: http://proceedings.mlr.press/v97/zintgraf19a.html.