535520 Optimization Algorithms:
A List of Candidate Papers for Final Project

Note: The following is a list of candidate papers for the team final project. Feel free to select a paper that is beyond this list and of interest to you.

First-Order Optimization Methods:

1. Martin Jaggi, “Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization,” ICML 2013.
2. Dan Garber and Elad Hazan, “Faster Rates for the Frank-Wolfe Method over Strongly-Convex Sets,” ICML 2015.
3. Sashank J. Reddi, Suvrit Sra, Barnabas Poczos, and Alex Smola, “Stochastic Frank-Wolfe Methods for Nonconvex Optimization,” Allerton 2016.
4. Elad Hazan and Haipeng Luo, “Variance-Reduced and Projection-Free Stochastic Optimization,” ICML 2016.
5. Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar“On the Convergence of Adam and Beyond,” ICLR 2018.
6. Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M. Kakade, and Michael I. Jordan, “How to Escape Saddle Points Efficiently,” ICML 2017.
7. Cong Fang, Zhouchen Lin, and Tong Zhang, “Sharp Analysis for Nonconvex SGD Escaping From Saddle Points,” COLT 2019.
8. Jincheng Mei, Yue Gao, Bo Dai, Csaba Szepesvari, Dale Schuurmans, “Leveraging Non-uniformity in First-order Non-convex Optimization,” ICML 2021.
9. Liang Zhang, Gang Wang, Daniel Romero, and Georgios B. Giannakis, “Randomized Block Frank-Wolfe for Convergent Large-Scale Learning,” IEEE Transactions on Signal Processing, 2017.
10. Aaron Defazio, Francis Bach, and Simon Lacoste-Julien, “SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives,” NIPS 2014.
11. Lam M. Nguyen, Jie Liu, Katya Scheinberg, Martin Takáč, “SARAH: A Novel Method for Machine Learning Problems Using Stochastic Recursive Gradient,” ICML 2017.
12. Mingyi Hong, Zhi-Quan Luo, and Meisam Razaviyayn, “Convergence Analysis of Alternating Direction Method of Multipliers for a Family of Nonconvex Problems,” SIAM Journal on Optimization, 2016.
13. P. Jain, S. Kakade, R. Kidambi, P. Netrapalli, A. Sidford, “Accelerating Stochastic Gradient Descent,” COLT 2018.
14. W. Su, S. Boyd and E. J. Candes, “A Differential Equation for Modeling Nesterov’s Accelerated Gradient Method: Theory and Insights,” Journal of Machine Learning Research, 2015.
15. Cong Fang, Chris Junchi Li, Zhouchen Lin, and Tong Zhang, “SPIDER: Near-Optimal Non-Convex Optimization via Stochastic Path-Integrated Differential Estimator,” NeurIPS 2018.
16. Sébastien Bubeck, Yin Tat Lee, and Mohit Singh, “A Geometric Alternative to Nesterov’s Accelerated Gradient Descent,” arXiv 2015.
17. Song Mei, Yu Bai, and Andrea Montanari, “The Landscape of Empirical Risk for Non-convex Losses,” The Annals of Statistics, 2018.
18. H. Attouch, H. and J. Peypouquet, ”The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than $1/k^2$,” SIAM Journal on Optimization, 2016.

19. Amir Beck and Yakov Vaisbourd, “Globally Solving the Trust Region Subproblem Using Simple First-Order Methods,” SIAM Journal on Optimization, 2018.

20. Z. Zhou, P. Mertikopoulos, N. Bambos, S. Boyd, P. Glynn, “On the Convergence of Mirror Descent Beyond Stochastic Convex Programming,” 2017

21. S. Ghadimi, G. Lan, “Accelerated Gradient Methods for Nonconvex Nonlinear and Stochastic Programming,” Mathematical Programming, 2016.

22. Adrian S Lewis, Stephen J Wright, “A Proximal Method for Composite Minimization,” Mathematical Programming, 2016.

23. Mark Schmidt, Nicolas Le Roux, Francis Bach, “Minimizing Finite Sums With the Stochastic Average Gradient,” Mathematical Programming, 2017.

24. Simon Lacoste-Julien, Mark Schmidt, Francis Bach, “A Simpler Approach to Obtaining an $O(1/t)$ Convergence Rate for the Projected Stochastic Subgradient Method,” arXiv 2012.

25. Sharan Vaswani, Aaron Mishkin, Issam Laradji, Mark Schmidt, Gauthier Gidel, and Simon Lacoste-Julien, ”Painless Stochastic Gradient: Interpolation, Line-Search, and Convergence Rates,” NeurIPS 2019.

26. Sharan Vaswani, Francis Bach, and Mark Schmidt, ”Fast and Faster Convergence of SGD for Over-Parameterized Models and an Accelerated Perceptron,” AISTATS 2019.

27. Alexandre Défossez, Léon Bottou, Francis Bach, Nicolas Usunier, “A Simple Convergence Proof of Adam and Adagrad,” arXiv 2020.

28. Ayoub El Hanchi, David Stephens, and Chris Maddison, “Stochastic Reweighted Gradient Descent,” ICML 2021.

29. Nicolas Loizou, Sharan Vaswani, Issam Laradji, and Simon Lacoste-Julien, “Stochastic Polyak Step-size for SGD: An Adaptive Learning Rate for Fast Convergence,” AISTATS 2021.

30. C. Jin, P. Netrapalli, and M. I. Jordan, ”Accelerated Gradient Descent Escapes Saddle Points Faster Than Gradient Descent,” COLT 2018.

31. H. Li and Z. Lin, ”Accelerated Proximal Gradient Methods for Nonconvex Programming,” NeurIPS 2015.

32. Othmane Sebbouh, Robert M Gower, and Aaron Defazio, “Almost Sure Convergence Rates for Stochastic Gradient Descent and Stochastic Heavy Ball,” COLT 2021.

33. Jun Liu, Ye Yuan, “On Almost Sure Convergence Rates of Stochastic Gradient Methods,” COLT 2022.

34. Jincheng Mei, Zixin Zhong, Bo Dai, Alekh Agarwal, Csaba Szepesvari, and Dale Schuurmans, “Stochastic Gradient Succeeds for Bandits,” ICML 2023.

Optimization of Neural Networks:

35. Brandon Amos and J. Zico Kolter, “OptNet: Differentiable Optimization as a Layer in Neural Networks,” ICML 2017.

36. Arthur Jacot, Franck Gabriel, Clement Hongler, ”Neural Tangent Kernel: Convergence and Generalization in Neural Networks,” NeurIPS 2018.

37. Boris Hanin and David Rolnick, “How to Start Training: The Effect of Initialization and Architecture, NeurIPS 2018.

38. Lénaïc Chizat and Francis Bach, “On the Global Convergence of Gradient Descent for Over-parameterized Models using Optimal Transport, ”

39. Z. Allen-Zhu and Y. Li, “What Can ResNet Learn Efficiently, Going Beyond Kernels?,” NeurIPS 2019.

40. Z. Allen-Zhu and Y. Li, “Can SGD Learn Recurrent Neural Networks with Provable Generalization?,” NeurIPS 2019.

41. Sanjeev Arora, Nadav Cohen, and Elad Hazan, “On the Optimization of Deep Networks: Implicit Acceleration by Overparameterization,” ICML 2018.

42. Rong Ge, Jason D. Lee, and Tengyu Ma, "Learning One-hidden-layer Neural Networks with Landscape Design," ICLR 2018.

43. Quentin Berthet, Mathieu Blondel, Olivier Teboul, Marco Cuturi, Jean-Philippe Vert, and Francis Bach, "Learning with Differentiable Perturbed Optimizers," NeurIPS 2020.

Minimax Problems:

44. Tianyi Lin, Chi Jin, and Michael I. Jordan, "On Gradient Descent Ascent for Nonconvex-Concave Minimax Problems," ICML 2020.

45. Tengyuan Liang and James Stokes, "Interaction Matters: A Note on Non-asymptotic Local Convergence of Generative Adversarial Networks," AISTATS 2020.

46. Balamurugan Palaniappan, Francis Bach, "Stochastic Variance Reduction Methods for Saddle-Point Problems," NIPS 2016.

47. Aryan Mokhtari, Asuman E. Ozdaglar, and Sarath Pattathil, "Convergence Rate of for Optimistic Gradient and Extragradient Methods in Smooth Convex-Concave Saddle Point Problems," SIAM Journal on Optimization, 2020

48. Noah Golowich, Sarath Pattathil, Constantinos Daskalakis, and Asuman Ozdaglar, "Last Iterate is Slower than Averaged Iterate in Smooth Convex-Concave Saddle Point Problems," COLT 2020.

49. Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng, "Training GANs With Optimism," ICLR 2018.

Zeroth-Order Methods:

50. John C. Duchi, Michael I. Jordan, Martin J. Wainwright, and Andre Wibisono, "Optimal Rates for Zero-Order Convex Optimization: The Power of Two Function Evaluations," IEEE Transactions on Information Theory, 2015.

51. Sijia Liu, Bhavya Kailkhura, Pin-Yu Chen, Paishun Ting, Shiyu Chang, and Lisa Amini, "Zeroth-Order Stochastic Variance Reduction for Nonconvex Optimization," NeurIPS 2018.

52. Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger, "Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design," ICML 2010.

53. Sayak Ray Chowdhury and Aditya Gopalan, "On Kernelized Multi-armed Bandits," ICML 2017.

54. Felix Berkenkamp, Angela P. Schoellig, and Andreas Krause, "No-Regret Bayesian Optimization with Unknown Hyperparameters," JMLR 2019.

55. Sattar Vakili, Nacime Bouziani, Sepehr Jalali, Alberto Bernacchia, and Da-shan Shiu, "Optimal Order Simple Regret for Gaussian Process Bandits," NeurIPS 2021.

56. Yu. Nesterov, "Efficiency of Coordinate Descent Methods on Huge-Scale Optimization Problems," SIAM Journal on Optimization, 2012.

57. Amir Beck and Luba Tetruashvili, "On the Convergence of Block Coordinate Descent Type Methods," SIAM Journal on Optimization, 2013.

58. Jinshan Zeng, Tim Tsz-Kit Lau, Shaobo Lin, and Yuan Yao, "Global Convergence of Block Coordinate Descent in Deep Learning," ICML 2019

59. Mert Gurbuzbalaban, Asuman Ozdaglar, Pablo A. Parrilo, and Nuri Vanli, "When Cyclic Coordinate Descent Outperforms Randomized Coordinate Descent," NIPS 2017.

Others (e.g., Meta Learning and Federated Learning):

60. Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar, "On the Convergence Theory of Gradient-Based Model-Agnostic Meta-Learning Algorithms," AISTATS 2020.

61. Hadrien Hendrikx, Francis Bach, Laurent Massoulie, “An Optimal Algorithm for Decentralized Finite-Sum Optimization,” SIAM Journal on Optimization, 2021.

62. Liam Collins, Aryan Mokhtari, and Sanjay Shakkottai, “Task-Robust Model-Agnostic Meta-Learning,” NeurIPS 2020.

63. Fallah et al., “On the Convergence Theory of Debiased Model-Agnostic Meta-Reinforcement Learning,” NeurIPS 2021

64. Xiang Li, Kaixuan Huang, Wenhao Yang, Shusen Wang, Zhihua Zhang, “On the Convergence of FedAvg on Non-IID Data,” ICLR 2020.

65. Honglin Yuan, Tengyu Ma, “Federated Accelerated Stochastic Gradient Descent,” NeurIPS 2020.

66. Alexander Gasnikov, Anton Novitskii, Vasilii Novitskii, Farshed Abdukhakimov, Dmitry Kamzolov, Aleksandr Beznosikov, Martin Takac, Pavel Dvurechenskii, Bin Gu, “The Power of First-Order Smooth Optimization for Black-Box Non-Smooth Problems,” ICML 2022.