References

--- a tentative list of papers to be mentioned in the ICML 2017 tutorial

"Recent Advances in Stochastic Convex and Non-Convex Optimization"

Disclaimer: in a quite arbitrary order.

1. **[ShalevShwartz-Zhang, 2013]: AccSDCA**
   Accelerated Proximal Stochastic Dual Coordinate Ascent for Regularized Loss Minimization
   by Shai Shalev-Shwartz, Tong Zhang

2. **[Lin-Lu-Xiao, 2014]: APCG**
   An Accelerated Proximal Coordinate Gradient Method and its Application to Regularized Empirical Risk Minimization
   by Qihang Lin, Zhaosong Lu, Lin Xiao

3. **[BenTal-Nemirovski, 2013]**
   Lectures on Modern Convex Optimization
   By Aharon Ben-Tal and Arkadi Nemirovski

4. **[Nemirovski, 2005]**
   Prox-method with rate of convergence o(1/t) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems.
   By Arkadi Nemirovski

5. **[Chambolle-Pock, 2011]**
   A first-order primal-dual algorithm for convex problems with applications to imaging
   By Antonin Chambolle and Thomas Pock

6. **[LeRoux-Schmidt-Bach, 2012]: SAG**
   A stochastic gradient method with an exponential convergence rate for finite training sets.
   By Nicolas Le Roux, Mark Schmidt, and Francis R. Bach

7. **[Johnson-Zhang, 2013]: SVRG**
   Accelerating stochastic gradient descent using predictive variance reduction.
   By Rie Johnson and Tong Zhang

8. **[Defazio-Bach-LacosteJulien, 2014]: SAGA**
   SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives.

By Defazio, A., Bach, F., & Lacoste-Julien, S.

9. **[ShalevShwartz-Zhang, 2012]: SDCA**
Stochastic Dual Coordinate Ascent Methods for Regularized Loss Minimization
By Shai Shalev-Shwartz and Tong Zhang

10. **[ShalevShwartz-Zhang, 2014]: AccSDCA**
Accelerated Proximal Stochastic Dual Coordinate As- cent for Regularized Loss Minimization
By Shai Shalev-Shwartz and Tong Zhang

11. **[Nesterov 2012]: RCDM/ACDM**
Efficiency of Coordinate Descent Methods on Huge-Scale Optimization Problems.
By Yurii Nesterov.

12. **[Lee-Sidford, 2013]: ACDM (better proof)**
Efficient accelerated coordinate descent methods and faster algorithms for solving linear systems.
By Yin Tat Lee and Aaron Sidford

13. **[AllenZhu-Qu-Richtarik-Yuan, 2016]: NUACDM**
Even Faster Accelerated Coordinate Descent Using Non-Uniform Sampling
By Zeyuan Allen-Zhu, Zheng Qu, Peter Richtarik and Yang Yuan.

14. **[Zhang-Xiao, 2015]: SPDC**
Stochastic Primal-Dual Coordinate Method for Regularized Empirical Risk Minimization
By Yuchen Zhang and Lin Xiao.

15. **[Lan-Zhou, 2015]: RPDG**
An optimal randomized incremental gradient method
By Guanghui Lan and Yi Zhou.

16. **[Frostig-Ge-Kakade-Sidford, 2015]: APPA**
Un-regularizing: approximate proximal point and faster stochastic algorithms for empirical risk minimization
By Roy Frostig, Rong Ge, Sham M. Kakade, and Aaron Sidford

17. **[Lin-Mairal-Harchaoui, 2015]: Catalyst**
A universal catalyst for first-order optimization
By Hongzhou Lin, Julien Mairal and Zaid Harchaoui

18. **[Zhang-Mahdavi-Jin, 2013]**
Linear convergence with condition number independent access of full gradients.
By Lijun Zhang, Mehrdad Mahdavi, and Rong Jin.

19. **[AllenZhu-Yuan, 2015] SVRG++**

Improved SVRG for Non-Strongly-Convex or Sum-of-Non-Convex Objectives

By Zeyuan Allen-Zhu and Yang Yuan

20. **[AllenZhu-Hazan 2016b]: reduction**

Optimal Black-Box Reductions Between Optimization Objectives

By Zeyuan Allen-Zhu and Elad Hazan

21. **[AllenZhu-Hazan 2016a]: non-convex SVRG**

Variance Reduction for Faster Non-Convex Optimization

By Zeyuan Allen-Zhu and Elad Hazan

22. **[Mahdavi-Zhang-Jin 2013]**

Mixed optimization for smooth functions

By Mehrdad Mahdavi, Lijun Zhang, and Rong Jin.

23. **[Lei-Jordan, 2017]**

Less than a Single Pass: Stochastically Controlled Stochastic Gradient

By Lihua Lei and Michael I. Jordan.

24. **[Lei-Ju-Chen-Jordan, 2017]**

Nonconvex Finite-Sum Optimization Via SCSG Methods

By Lihua Lei, Cheng Ju, Jianbo Chen, and Michael I. Jordan.

25. **[Nesterov 2005]**

Smooth minimization of non-smooth functions

By Yurii Nesterov

26. **[Su-Boyd-Candes, 2014]**

A differential equation for modeling nesterovs accelerated gradient method: Theory and insights

By Weijie Su, Stephen Boyd, Emmanuel J. Candès

27. **[Bubeck-Lee-Singh, 2015]**

A geometric alternative to Nesterov's accelerated gradient descent

By Sébastien Bubeck, Yin Tat Lee, Mohit Singh

28. **[Wibisono-Wilson-Jordan, 2016]**

A Variational Perspective on Accelerated Methods in Optimization

By Andre Wibisono, Ashia C. Wilson, Michael I. Jordan

29. **[AllenZhu-Orecchia, 2014]**

Linear Coupling: An Ultimate Unification of Gradient and Mirror Descent

By Zeyuan Allen-Zhu and Lorenzo Orecchia.

30. **[Nitanda, 2014]**

Stochastic proximal gradient descent with acceleration techniques.
By Atsushi Nitanda.

31. **[AllenZhu, 2016]**
Katyusha: The First Direct Acceleration of Stochastic Gradient Methods
By Zeyuan Allen-Zhu

32. **[Woodworth-Srebro, 2016]**
Tight Complexity Bounds for Optimizing Composite Objectives
By Blake Woodworth and Nathan Srebro

33. **[Frostig-Musco-Musco-Sidford, 2016]: PCR**
Principal Component Projection Without Principal Component Analysis
By Roy Frostig, Cameron Musco, Christopher Musco and Aaron Sidford.

34. **[AllenZhu-Li, 2017]: PCR**
Faster Principal Component Regression and Stable Matrix Chebyshev Approximation
By Zeyuan Allen-Zhu and Yuanzhi Li

35. **[Shibagaki-Takeuchi, 2017]: mini-batch SPDC**
Stochastic primal dual coordinate method with nonuniform sampling based on optimality violations
By Atsushi Shibagaki and Ichiro Takeuchi.

36. **[Murata-Suzuki, 2017]: DASVRDA**
Doubly accelerated stochastic variance reduced dual averaging method for regularized empirical risk minimization
By Tomoya Murata and Taiji Suzuki

37. **[Garber et al, 2015]: shift and invert (two papers about the same result)**
Faster Eigenvector Computation via Shift-and-Invert Preconditioning
By Dan Garber, Elad Hazan, Chi Jin, Sham M. Kakade, Cameron Musco, Praneeth Netrapalli, Aaron Sidford
[Garber-Hazan, 2015]
Fast and Simple PCA via Convex Optimization
By Dan Garber and Elad Hazan

38. **[AllenZhu-Li, 2016]: LazySVD**
LazySVD: Even Faster SVD Decomposition Yet Without Agonizing Pain
By Zeyuan Allen-Zhu and Yuanzhi Li.

39. **[Sun-Luo, 2014]**
Guaranteed Matrix Completion via Non-convex Factorization
By Ruoyu Sun and Zhi-Quan Luo

40. **[Arora-Ge-Ma-Moitra, 2015]**
    Simple, Efficient, and Neural Algorithms for Sparse Coding
    By Sanjeev Arora, Rong Ge, Tengyu Ma and Ankur Moitra

41. **[Chen-Candès, 2015]**
    Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems
    By Yuxin Chen, Emmanuel J. Candès

42. **[Nesterov, 2004]: textbook**
    Introductory Lectures on Convex Programming Volume: A Basic course
    By Yurii Nesterov

43. **[Li-Yuan, 2017]**
    Convergence Analysis of Two-layer Neural Networks with ReLU Activation
    By Yuanzhi Li and Yang Yuan

44. **[ShalevShwartz, 2015]**
    SDCA without Duality
    By Shai Shalev-Shwartz

45. **[Reddi et al., 2016a]**
    Stochastic variance reduction for nonconvex optimization
    By Sashank J Reddi, Ahmed Hefny, Suvrit Sra, Barnabas Poczos, and Alex Smola

46. **[Reddi et al., 2016b]**
    Fast incremental method for nonconvex optimization
    Sashank J Reddi, Suvrit Sra, Barnabás Póczos, and Alex Smola.

47. **[Reddi et al., 2016c]**
    Fast Stochastic Methods for Nonsmooth Nonconvex Optimization
    Sashank J Reddi, Suvrit Sra, Barnabás Póczos, and Alex Smola.

48. **[Lei-Ju-Chen-Jordan, 2017]**
    Nonconvex Finite-Sum Optimization Via SCSG Methods
    By Lihua Lei, Cheng Ju, Jianbo Chen, and Michael I. Jordan

49. **[Carmon-Duchi-Hinder-Sidford, 2016]**
    Accelerated Methods for Non-Convex Optimization
    By Yair Carmon, John C. Duchi, Oliver Hinder and Aaron Sidford

50. **[AllenZhu, 2017]**
    Natasha: Faster Non-Convex Stochastic Optimization via Strongly Non-Convex Parameter
    By Zeyuan Allen-Zhu

51. **[Carmon- Duchi-Hinder-Sidford, 2017]**
    "Convex Until Proven Guilty": Dimension-Free Acceleration of Gradient Descent on Non-Convex Functions
    By Yair Carmon, John C. Duchi, Oliver Hinder and Aaron Sidford

52. **[Ge-Huang-Jin-Yuan, 2015]**
    Escaping From Saddle Points --- Online Stochastic Gradient for Tensor Decomposition
    By Rong Ge, Furong Huang, Chi Jin, Yang Yuan

53. **[Jin et al., 2017]**
    How to Escape Saddle Points Efficiently
    By Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M. Kakade, Michael I. Jordan

54. **[Agarwal et al., 2016]**
    Finding Approximate Local Minima Faster Than Gradient Descent
    By Naman Agarwal, Zeyuan Allen-Zhu, Brian Bullins, Elad Hazan, Tengyu Ma