

Оптимизация. Advanced Topics

1. Стохастическая оптимизация. Классическая постановка и способы сэмплирования.

Лекция на основе статей:

- 1) [Unified Optimal Analysis of the \(Stochastic\) Gradient Method](#)
- 2) [Stochastic Optimization with Importance Sampling for Regularized Loss Minimization](#)
- 3) [SGD: General Analysis and Improved Rates](#)
- 4) [A Unified Theory of SGD: Variance Reduction, Sampling, Quantization and Coordinate Descent](#)

Разбор статьи: [Random Reshuffling: Simple Analysis with Vast Improvements](#)

2. Стохастическая оптимизация. Методы редукции дисперсии.

Лекция на основе статей:

- 1) [SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives](#)
- 2) [Accelerating Stochastic Gradient Descent using Predictive Variance Reduction](#)
- 3) [Don't Jump Through Hoops and Remove Those Loops: SVRG and Katyusha are Better Without the Outer Loop](#)
- 4) [Katyusha: the first direct acceleration of stochastic gradient methods](#)
- 5) [A Unified Theory of SGD: Variance Reduction, Sampling, Quantization and Coordinate Descent](#)

Разбор статьи: [SARAH: A Novel Method for Machine Learning Problems Using Stochastic Recursive Gradient](#)

3. Стохастическая оптимизация. Координатные методы.

Лекция на основе статей:

- 1) [Efficiency of Coordinate Descent Methods on Huge-Scale Optimization Problems](#)
- 2) [On optimal probabilities in stochastic coordinate descent methods](#)
- 3) [Coordinate descent with arbitrary sampling I: algorithms and complexity](#)
- 4) [A Unified Theory of SGD: Variance Reduction, Sampling, Quantization and Coordinate Descent](#)

Разбор статьи: [SEGA: Variance Reduction via Gradient Sketching](#)

4. Распределенная оптимизация. Сжатие информации.

Лекция на основе статей:

- 1) [QSGD: Communication-Efficient SGD via Gradient Quantization and Encoding](#)

- 2) [The Error-Feedback Framework: Better Rates for SGD with Delayed Gradients and Compressed Communication](#)
- 3) [Distributed Learning with Compressed Gradient Differences](#)
- 4) [Acceleration for Compressed Gradient Descent in Distributed and Federated Optimization](#)
- 5) [MARINA: Faster Non-Convex Distributed Learning with Compression](#)

Разбор статьи: [EF21: A New, Simpler, Theoretically Better, and Practically Faster Error Feedback](#)

5. Распределенная оптимизация. Похожесть данных.

Лекция на основе статей:

- 1) [Communication-Efficient Distributed Optimization using an Approximate Newton-type Method](#)
- 2) [Communication Complexity of Distributed Convex Learning and Optimization](#)
- 3) [DiSCO: Distributed Optimization for Self-Concordant Empirical Loss](#)
- 4) [Statistically preconditioned accelerated gradient method for distributed optimization](#)

Разбор статьи: [Optimal Gradient Sliding and its Application to Optimal Distributed Optimization Under Similarity](#)

6. Распределенная оптимизация. Децентрализованные коммуникации.

Лекция на основе статей:

- 1) [Collective communication: theory, practice, and experience](#)
- 2) [Distributed average consensus with least-mean-square deviation](#)
- 3) [A Unified Theory of Decentralized SGD with Changing Topology and Local Updates](#)
- 4) [An Improved Analysis of Gradient Tracking for Decentralized Machine Learning](#)
- 5) [Optimal and Practical Algorithms for Smooth and Strongly Convex Decentralized Optimization](#)
- 6) [Lower Bounds and Optimal Algorithms for Smooth and Strongly Convex Decentralized Optimization Over Time-Varying Networks](#)

Разбор статьи: [EXTRA: An Exact First-Order Algorithm for Decentralized Consensus Optimization](#)

7. Вариационные неравенства и седловые задачи.

Лекция на основе статей:

- 1) [A Variational Inequality Perspective on Generative Adversarial Networks](#)
- 2) [On the convergence of single-call stochastic extra-gradient methods](#)
- 3) [Solving variational inequalities with stochastic mirror-prox algorithm](#)
- 4) [Stochastic Variance Reduction for Variational Inequality Methods](#)
- 5) [Revisiting Stochastic Extragradient](#)

Разбор статьи: [Extragradient Method: \$O\(1/K\)\$ Last-Iterate Convergence for Monotone Variational Inequalities and Connections With Cocoercivity](#)