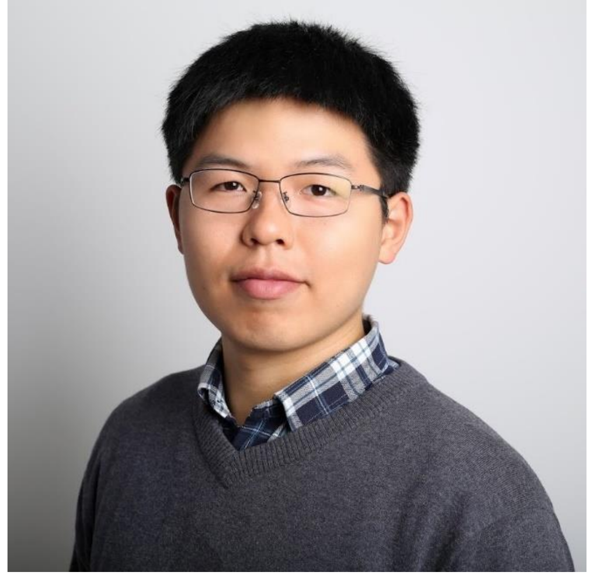
Avoid Overclaims: Summary of Complexity Bounds for Algorithms in Minimization and Minimax Optimization

Siqi Zhang
SME, NJU

Yifan Hu
EPFL & ETH Zürich



Introduction

- Review the complexity bounds of first-order methods in optimization in different convexity/smoothness scenarios
- Two problem settings
  - Minimization
  - Minimax Optimization

$$\min_{x \in \mathcal{X}} \; f(x) \qquad \min_{x \in \mathcal{X}} \left[f(x) \triangleq \max_{y \in \mathcal{Y}} \; g(x, y) \right]$$

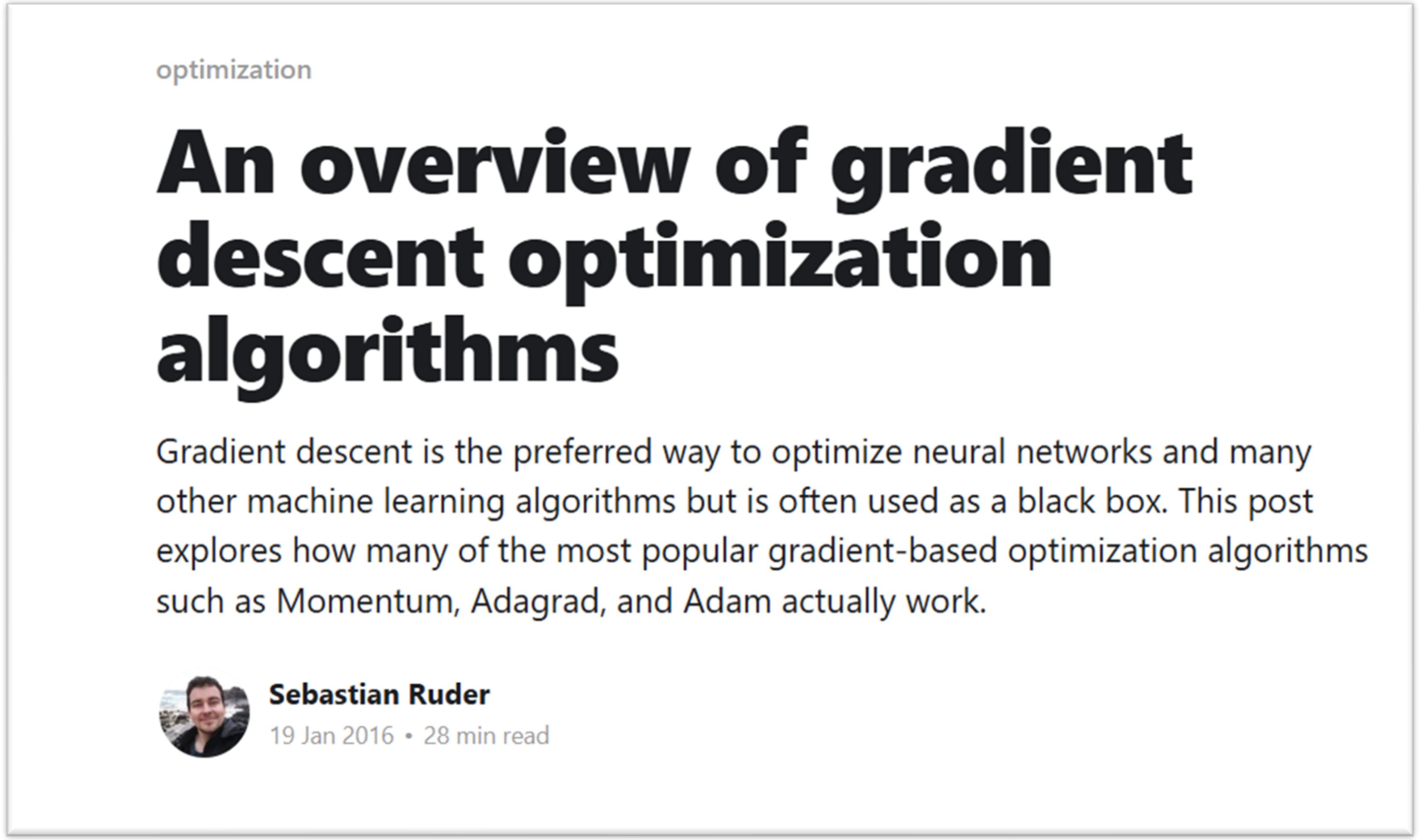
- Three stochastic settings
  - Deterministic (general)
  - Finite-sum
  - Stochastic optimization

$$\min_{x \in \mathcal{X}} \; f(x) \qquad \min_{x \in \mathcal{X}} \; f(x) \triangleq \frac{1}{n} \sum_{i=1}^{n} f_i(x)$$

$$\min_{x \in \mathcal{X}} \; f(x) \triangleq \mathbb{E}_{\xi \sim \mathcal{D}}[f(x; \xi)]$$

Prior Arts

- Sebastian Ruder’s Blog on common optimization algorithms in ML
- Several monographs reviewed optimization algorithms in various settings
- Popular repository tracking nonconvex optimization research by Ju Sun
- We focus on the SOTA upper and lower bounds in various settings

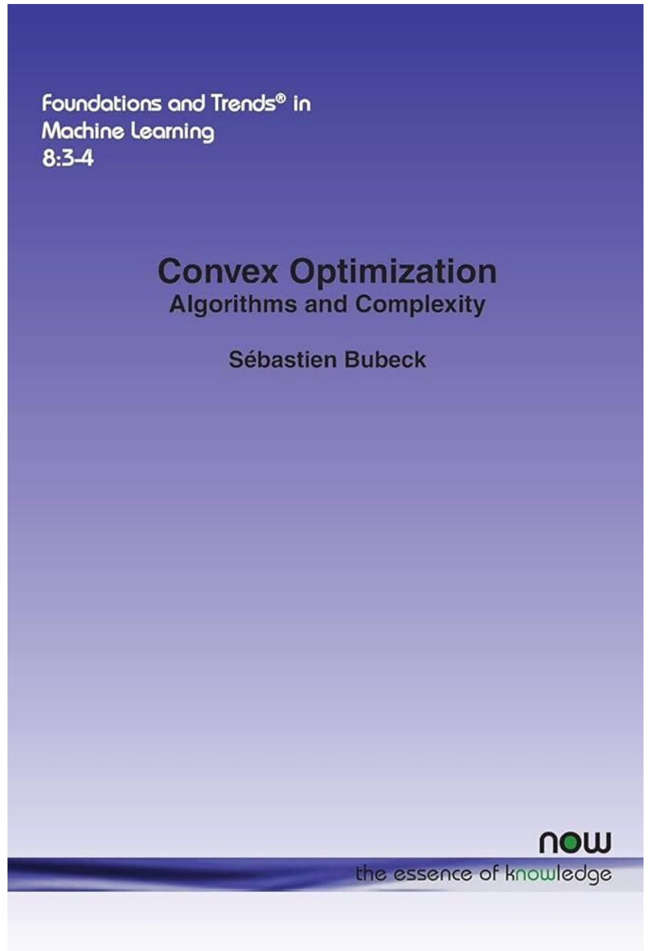


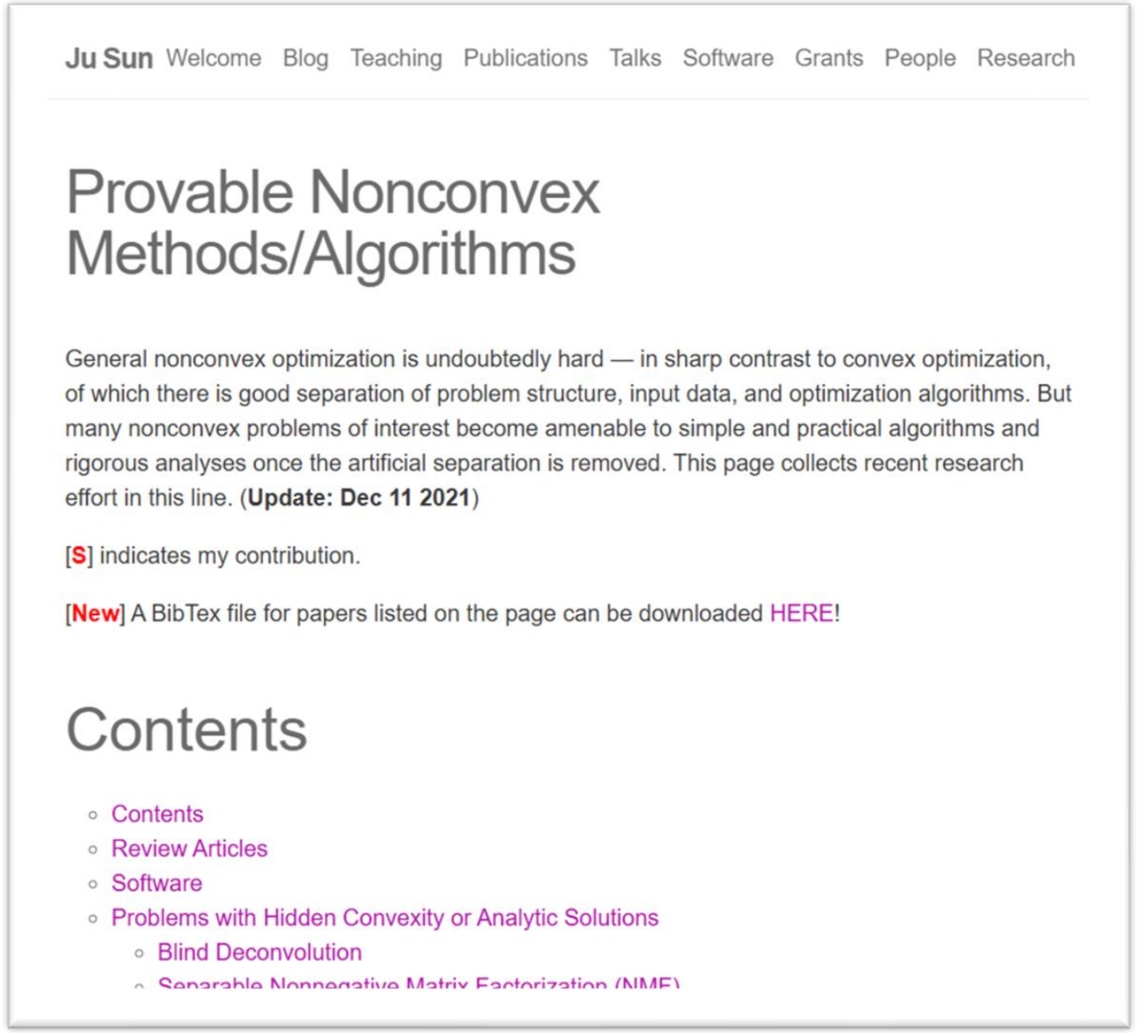
optimization

An overview of gradient descent optimization algorithms

Gradient descent is the preferred way to optimize neural networks and many other machine learning algorithms but is often used as a black box. This post explores how many of the most popular gradient-based optimization algorithms such as Momentum, Adagrad, and Adam actually work.

Sebastian Ruder
19 Jan 2016 • 28 min read





Ju Sun Welcome Blog Teaching Publications Talks Software Grants People Research

Provable Nonconvex Methods/Algorithms

General nonconvex optimization is undoubtedly hard — in sharp contrast to convex optimization, of which there is good separation of problem structure, input data, and optimization algorithms. But many nonconvex problems of interest become amenable to simple and practical algorithms and rigorous analyses once the artificial separation is removed. This page collects recent research effort in this line. (**Update: Dec 11 2021**)

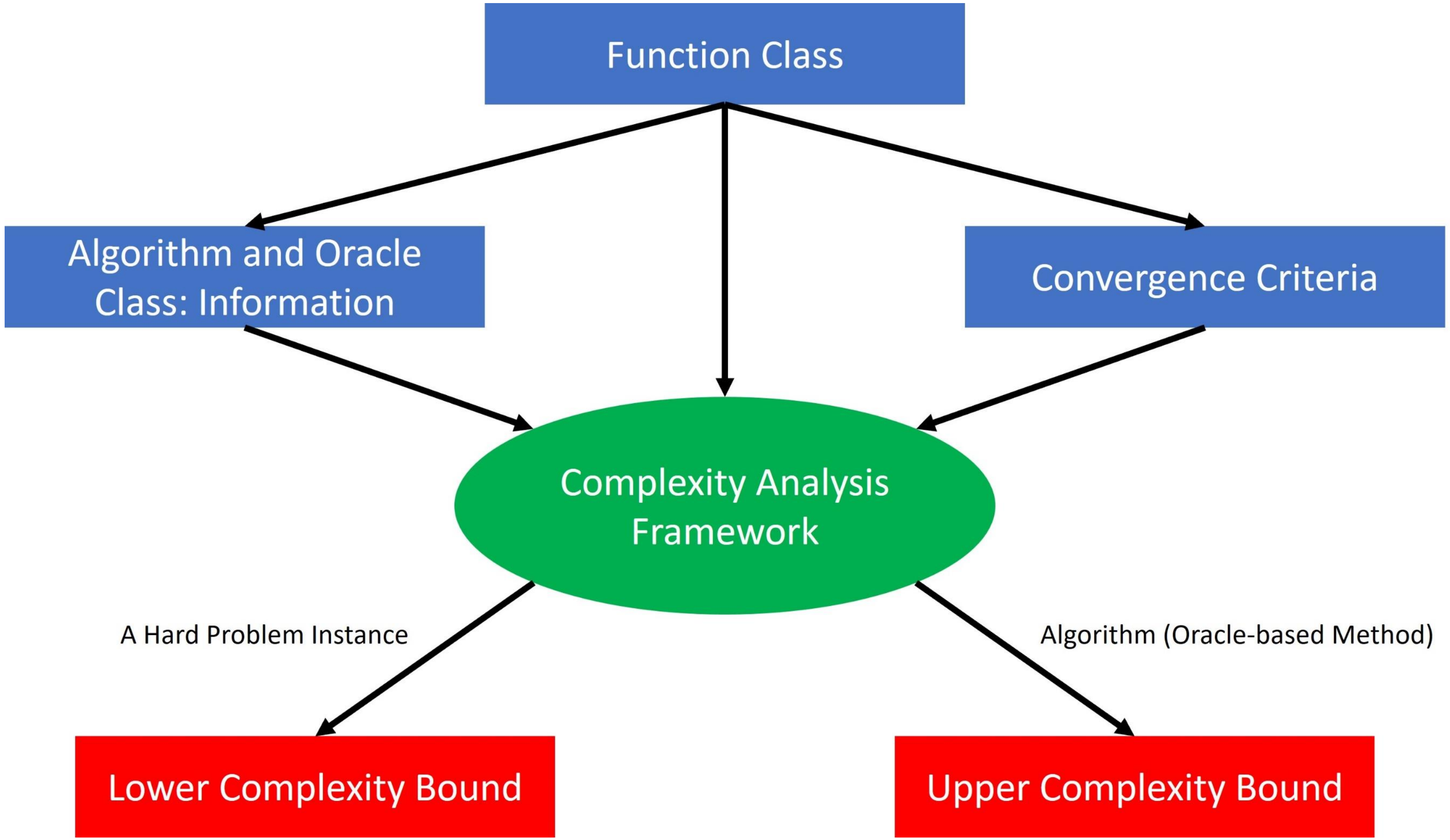
[**S**] indicates my contribution.

[**New**] A BibTex file for papers listed on the page can be downloaded HERE!

Contents

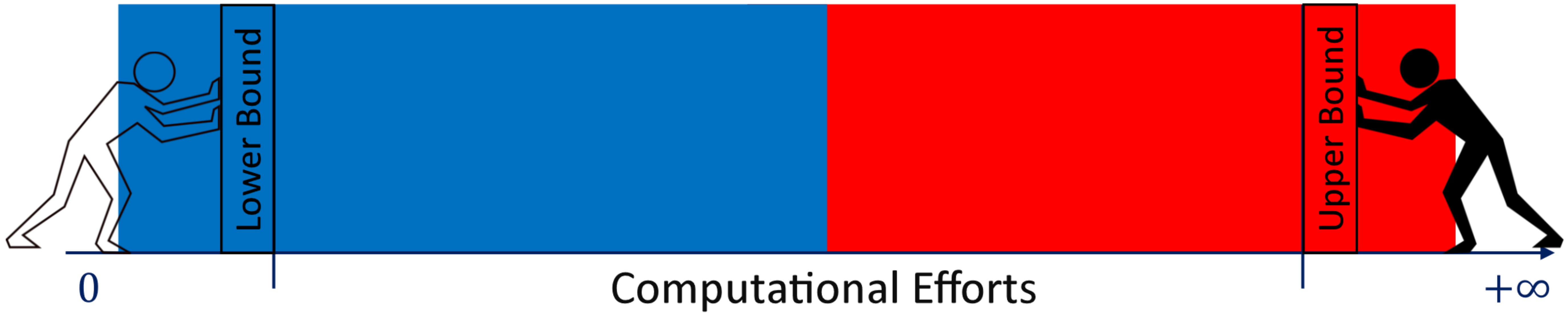
- Contents
- Review Articles
- Software
- Problems with Hidden Convexity or Analytic Solutions
  - Blind Deconvolution

Oracle Complexity Framework



Adapted from Prof. Yangyang Xu's Slides

Oracle Complexity Framework (cont.)



Main Results

- Case 1-1: Deterministic Minimization
- Case 1-2: Finite-sum and Stochastic Minimization
- Case 2-1: (S)C-(S)C Deterministic Minimax Optimization
- Case 2-2: (S)C-(S)C Finite-sum and Stochastic Minimax Optimization
- Case 2-3: NC-(S)C Deterministic Minimax Optimization
- Case 2-4: NC-(S)C Finite-sum and Stochastic Minimax Optimization

- SOTA upper and lower bounds comparison

Main Results

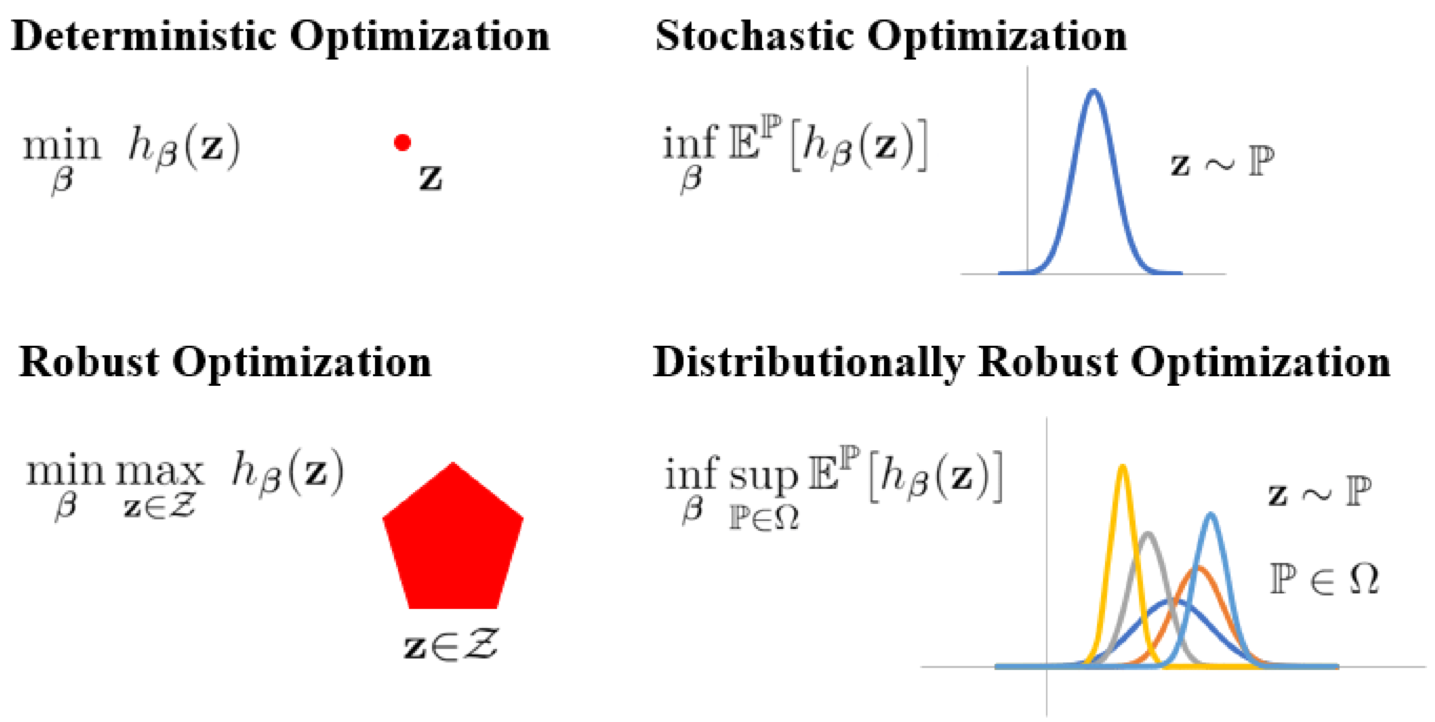
Case 1-1: Deterministic Minimization

Problem Type	Measure	Lower Bound	Upper Bound	Reference (LB)	Reference (UB) [5]
L-Smooth Convex	Optimality gap	$\Omega\left(\sqrt{L\epsilon^{-1}}\right)$	✓	[25], Theorem 2.1.7	[25], Theorem 2.2.2
L-Smooth μ-SC	Optimality gap	$\Omega\left(\sqrt{\kappa}\log\frac{1}{\epsilon}\right)$	✓	[25], Theorem 2.1.13	[25], Theorem 2.2.2
NS G-Lip Cont. Convex	Optimality gap	$\Omega(G^2\epsilon^{-2})$	✓	[25], Theorem 3.2.1	[25], Theorem 3.2.2
NS G-Lip Cont. μ-SC	Optimality gap	$\Omega(G^2(\mu\epsilon)^{-1})$	✓	[25], Theorem 3.2.5	[7], Theorem 3.9 [6]
L-Smooth Convex (function case)	Stationarity	$\Omega\left(\sqrt{\Delta L\epsilon^{-1}}\right)$	✓ (within logarithmic)	[26], Theorem 1	[26], Appendix A.1
L-Smooth Convex (domain case)	Stationarity	$\Omega\left(\sqrt{DL}\epsilon^{-\frac{1}{2}}\right)$	✓	[26], Theorem 1	[27] Section 6.5
L-Smooth NC	Stationarity	$\Omega(\Delta L\epsilon^{-2})$	✓	[20], Theorem 1	[28], Theorem 10.15
NS G-Lip Cont. ρ-WC	Near-stationarity	Unknown	$\mathcal{O}(\epsilon^{-4})$	/	[29], Corollary 2.2
L-Smooth μ-PL	Optimality gap	$\Omega\left(\kappa\log\frac{1}{\epsilon}\right)$	✓	[30], Theorem 3	[24], Theorem 1

Remark:

1. References: [25] [7] [26] [20] [27] [28] [29] [30] [24]
2. $\kappa \triangleq L/\mu \geq 1$ is called the condition number, which can be very large in many applications, e.g., the optimal regularization parameter choice in statistical learning can lead to $\kappa = \Omega(\sqrt{n})$ where n is the sample size [31].
3. The PL condition is a popular assumption in nonconvex optimization, generalizing the strong convexity condition. Based on the summary above, we can find that both smooth strongly convex and smooth PL condition optimization problems have established the optimal complexities (i.e., UB matches LB). However, the LB in the PL case is strictly larger than that of the SC case. Thus, regarding the worst-case complexity, we can say that the PL case is "strictly harder" than the strongly convex case.

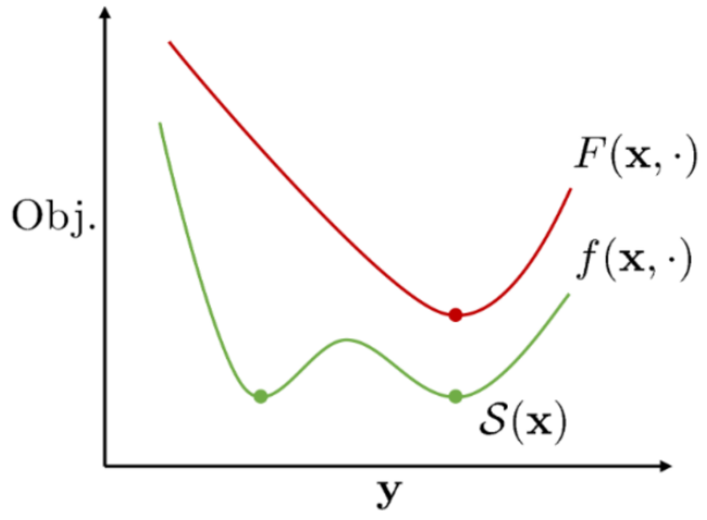
Future Directions



- Richer Problem Structure
  - Bilevel Optimization
  - Compositional Stochastic Optimization
  - Conditional Stochastic Optimization
  - Performative Prediction
  - Contextual Stochastic Optimization
  - Distributionally Robust Optimization
  - ……

$$\min_{\mathbf{x}\in\mathcal{X}} F(\mathbf{x},\mathbf{y}), \;\; s.t. \; \mathbf{y} \in \mathcal{S}(\mathbf{x})$$

$$\mathcal{S}(\mathbf{x}) := \arg\min_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{x},\mathbf{y})$$

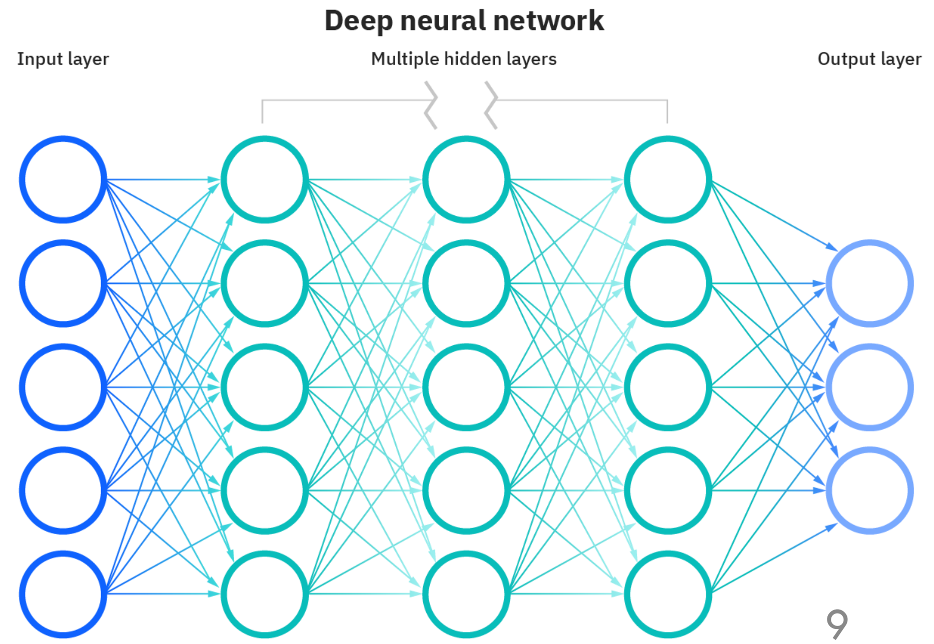


- Various optimization problems arising from ML & OR, which come with more involved problem structure and complicated landscape characteristics

Source: arXiv: 2101.11517, 2108.08993

Future Directions (cont.)

- Landscape Analysis
  - Hidden Convexity
  - PL/KL Conditions
  - Relaxed Smoothness
  - Low Rank Structure
- Beyond Classical Oracle Model
  - Average-case complexity
  - Arithmetic complexity
  - Communication complexity in distributed optimization
  - Long stepsize in first-order methods and achieve a faster convergence rate
- Unified Lower Bounds
  - Lower bound valid for any given dimension
  - Information theoretic-based lower bounds



Feedback Appreciated!

- Possibility of overlooking certain relevant works, subtle technical conditions, or potential inaccuracies in interpreting the literature
- Don’t hesitate to send emails to bring them to our attention!
- Constructive feedback, corrections, and suggestions are highly appreciated.