E1 260
Optimization for Machine Learning and Data Science

Sundeep Prabhakar Chepuri

Email: spchepuri@iisc.ac.in

Indian Institute of Science
भारतीय विज्ञान संस्थान

Course information

- Instructor

  Sundeep Prabhakar Chepuri
  email: *spchepuri AT iisc.ac.in*

- Class schedule

  Tuesdays and Thursdays 2.00 – 3.30 pm (*Online via MS Teams*)

  (Pls. be on mute when you don't have to ask anything)

- Attendance is mandatory (watch recorded videos only for revising!)

Course objective

- Cover optimization techniques suitable for problems that frequently appear in the areas of *data science, machine learning, communications, and signal processing*.

- Focus on the *computational*, *algorithmic*, and *implementation* aspects

Grading and course requirements

- **Prerequisite:** Basic linear algebra, probability, and knowledge of a programming language like Python (Google CoLab) to conduct simulation exercises.
- This is a 3**:1** course (*expect about 3 hours of work each week, almost*)
- **Five** assigments (problem and programming set): **10% each**
- **Two** projects**: 20% each**
- **Final Exam (take-home 24 hour exam): 10%**
- **No mid-term exam**

Grading	%
5 x HW	50
Project 1	20
Project 2	20
Final Exam	10
Total	100

Sessional (50%)	**Final (50%)**
HW 1,2,3, Project 1	HW 4,5,Project 2 and final exam

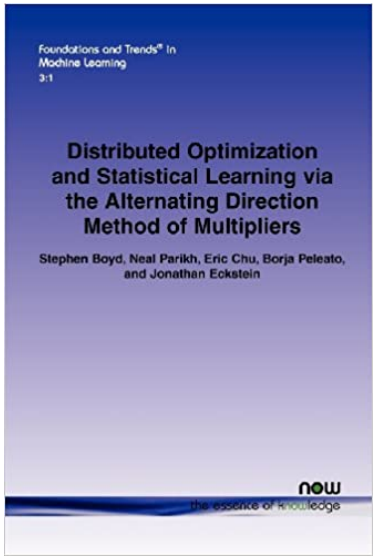
References

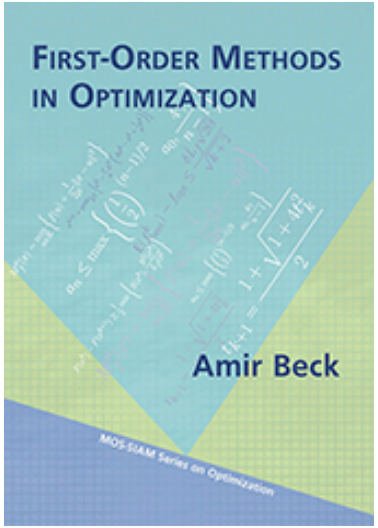
Optimization for Machine Learning

Lecture Notes CS-439, Spring 2021

Bernd Gärtner, ETH
Martin Jaggi, EPFL

May 4, 2021





and papers….

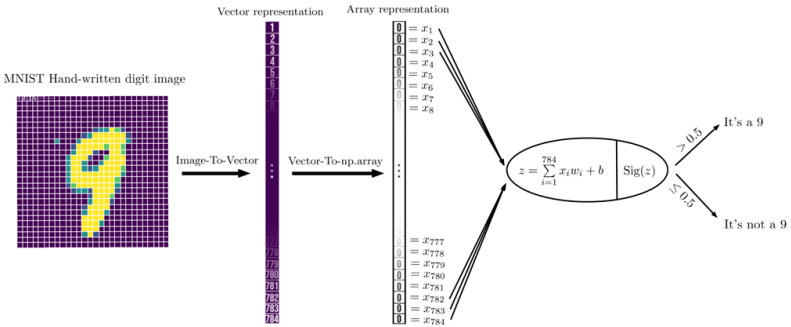
Unconstrained optimization problem

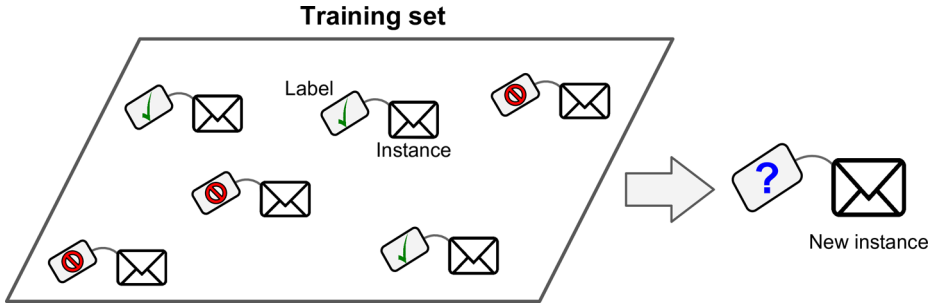
$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathbb{R}^d \end{aligned}$$

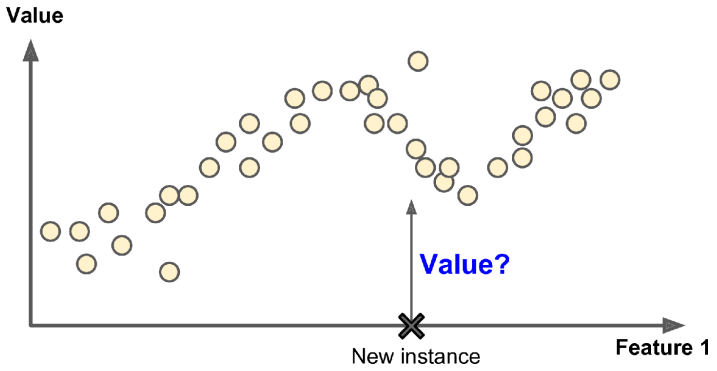
- $f : \mathbb{R}^d \to \mathbb{R}$ is the objective or the cost function. The value $f(\mathbf{x})$ is the cost of the decision $\mathbf{x}$
- $\mathbf{x}$ is the decision variable

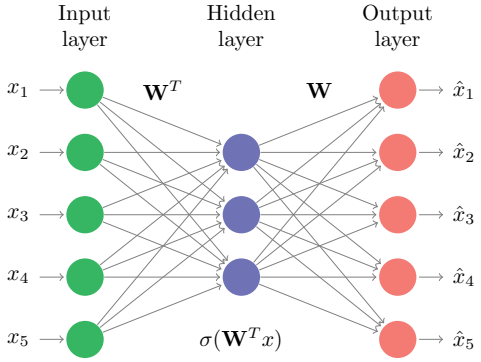
Unconstrained optimization problems

$$\operatorname*{argmin}_{\mathbf{w}\in\mathbb{R}^d} \frac{1}{m}\sum_{i=1}^{m} \log\left(1+\exp(-y_i\langle \mathbf{w}, \mathbf{x}_i\rangle)\right).$$



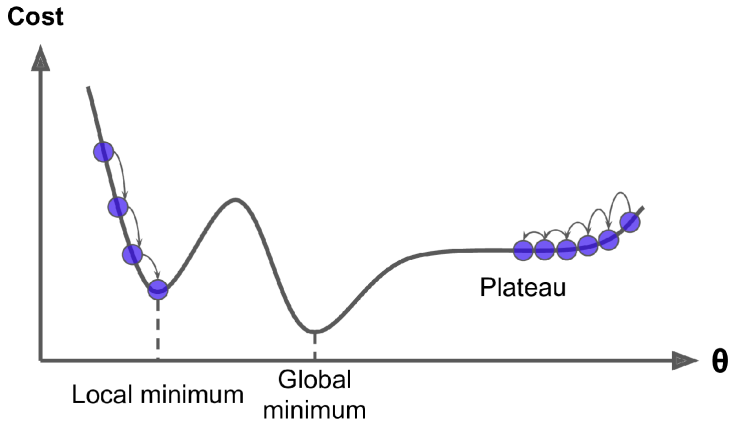


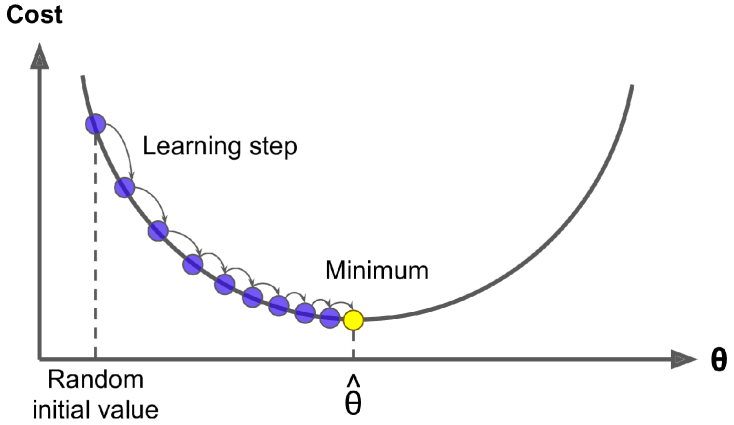




Theory of convex functions

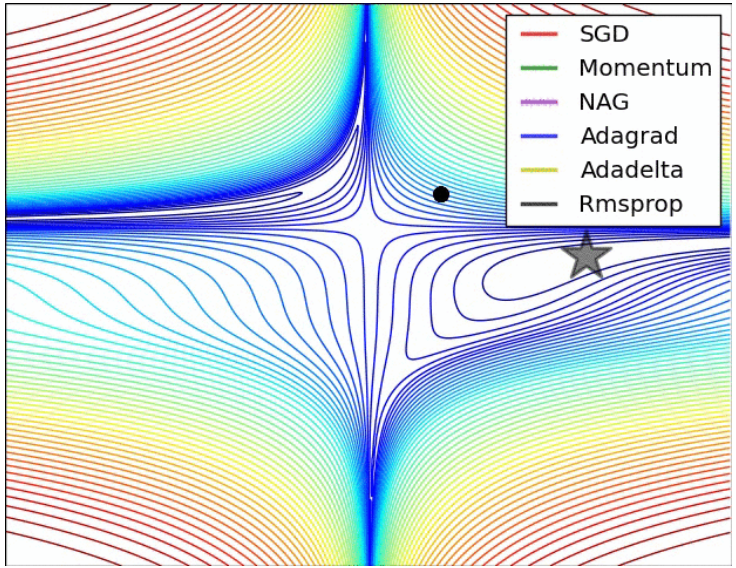
3	24/08/2021	**Theory of convex functions**
	26/08/2021	**Theory of convex functions**
4	31/08/2021	**Theory of convex functions**
	02/09/2021	**Theory of convex functions**





Unconstrained optimization

5	07/09/2021	**Gradient descent for unconstrained problems**
	09/09/2021	**Gradient descent for unconstrained problems**
6	14/09/2021	**Accelerated gradient method**
	16/09/2021	**Accelerated gradient method**
7	21/09/2021	**Proximal gradient descent**
	23/09/2021	**Mirror descent**
8	28/09/2021	**Subgradient methods**
	30/09/2021	NO CLASS
9	05/10/2021	**Stochastic Gradient Descent**
	07/10/2021	SGD and friends
10	12/10/2021	SGD and friends
	14/10/2021	**Project 1 presentations**
11	19/10/2021	HOLIDAY
	21/10/2021	SGD and friends



Many APIs, libraries are available

```
optimizer = keras.optimizers.RMSprop(lr=0.001, rho=0.9)
```

In this course:

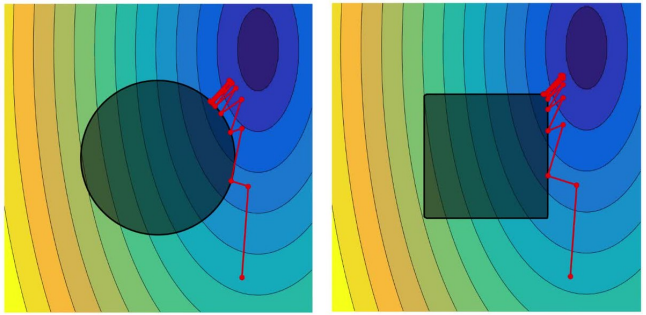
- We look inside these algorithms
- Study why, when, and how fast they work

Constrained optimization problem

12	26/10/2021	**Projected gradient descent**
	28/10/2021	**Frank-Wolfe**

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) \\ &\text{subject to} && \mathbf{x} \in \mathcal{C} \end{aligned}$$

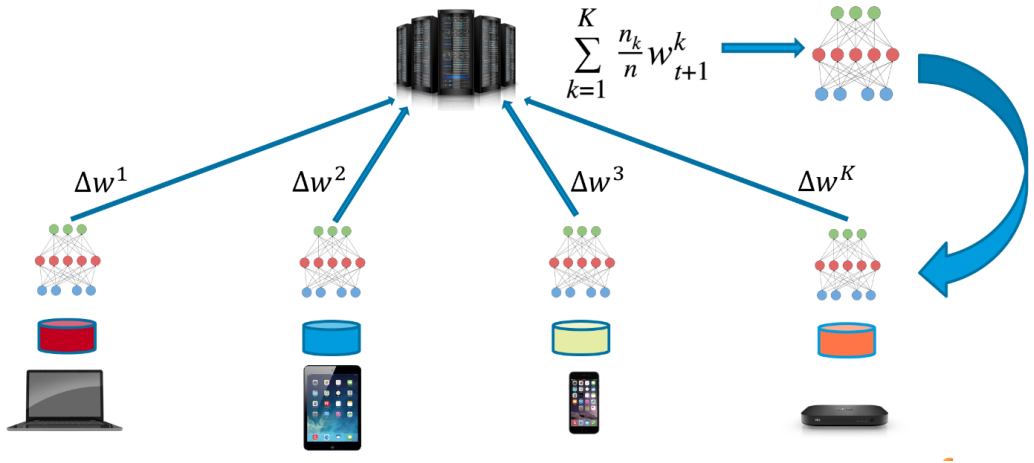
- $f : \mathbb{R}^d \to \mathbb{R}$ is the objective or the cost function. The value $f(\mathbf{x})$ is the cost of the decision $\mathbf{x}$
- $\mathbf{x}$ is the decision variable
- $\mathcal{C}$ is the feasible set



Theorem: The Frank-Wolfe method using standard step sizes $\gamma_k = 2/(k+1)$, $k = 1, 2, 3, \ldots$ satisfies

$$f(x^{(k)}) - f^\star \leq \frac{2M}{k+2}$$

Distributed optimization: ADMM and federated learning



14	09/11/2021	ADMM
	11/11/2021	ADMM
15	16/11/2021	Federated Learning
	18/11/2021	Federated Learning

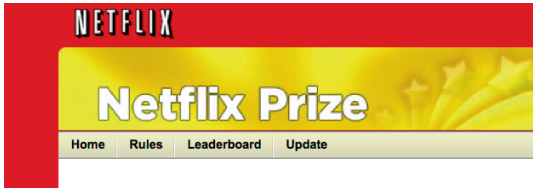
Submodular optimization

- Optimization of a set function

$$\begin{aligned} &\text{minimize} \quad F(X) \\ &\text{subject to} \quad X \subset \{1, 2, \cdots, n\} \quad \text{and} \quad |X| \leq K \end{aligned}$$

16	23/11/2021	Submodular optimization
	25/11/2021	Submodular optimization

Non-convex optimization



$$\text{minimize } \operatorname{rank}(\mathbf{M}) \quad \text{subject to } m_{ij} = z_{ij} \text{ for all } (i,j) \in \Omega$$

	Dirty Dancing	Meet the Parents	Top Gun	The Sixth Sense	Catch Me If You Can	The Royal Tenenbaums	Con Air	Big Fish	The Matrix	A Few Good Men
Customer 1					4					
Customer 2			3				3			3
Customer 3		2		4					2	
Customer 4	3									
Customer 5	5	5			4					
Customer 6						2	4			
Customer 7			5					3		
Customer 8						2				3
Customer 9	3				5			5		
Customer 10										

17	30/11/2021	Non-convex optimization
	02/12/2021	Non-convex optimization

Week	Date		Project/HW announcement	Due dates
1	10/08/2021	**Introduction to the course**: logistics, optimization in ML and data science, applications		
	12/08/2021	**Mathematical background**		
2	17/08/2021	**Mathematical background**		
	19/08/2021	HOLIDAY		
3	24/08/2021	**Theory of convex functions**		
	26/08/2021	**Theory of convex functions**		
4	31/08/2021	**Theory of convex functions**	HW1	
	02/09/2021	**Theory of convex functions**		
5	07/09/2021	**Gradient descent for unconstrained problems**		
	09/09/2021	**Gradient descent for unconstrained problems**	**Project 1**	
6	14/09/2021	**Accelerated gradient method**	HW2	HW1
	16/09/2021	**Accelerated gradient method**		
7	21/09/2021	**Proximal gradient descent**		
	23/09/2021	**Mirror descent**		
8	28/09/2021	**Subgradient methods**		
	30/09/2021	NO CLASS		
9	05/10/2021	**Stochastic Gradient Descent**		HW2
	07/10/2021	SGD and friends		**Project 1**
10	12/10/2021	SGD and friends	HW3	
	14/10/2021	**Project 1 presentations**		
11	19/10/2021	HOLIDAY		
	21/10/2021	SGD and friends		
12	26/10/2021	**Projected gradient descent**		HW3
	28/10/2021	**Frank-Wolfe**	HW4	
13	02/11/2021	NO CLASS	**Project 2**	
	04/11/2021	HOLIDAY		
14	09/11/2021	ADMM		
	11/11/2021	ADMM		
15	16/11/2021	Federated Learning		
	18/11/2021	Federated Learning	HW5	HW4
16	23/11/2021	Submodular optimization		
	25/11/2021	Submodular optimization		
17	30/11/2021	Non-convex optimization		
	02/12/2021	Non-convex optimization		HW5
18	07/12/2021	NO CLASS		**Project 2**
	09/12/2021	**Project 2 presentations**		