Classroom Expectations related to COVID-19

- In order to attend class meetings in person, all students are expected to abide by all behaviors indicated in [A Tartan's Responsibility], including any timely updates based on the current conditions.
- In terms of specific classroom expectations, whenever the requirement to wear a facial covering is in effect on campus, students are expected to wear a facial covering throughout class. Note: the requirement to wear a facial covering is in effect for the start of the Fall 2021 semester. If you do not wear a facial covering to class, I will ask you to put one on (and if you don't have one with you, I will direct you to a distribution location on campus, see https://www.cmu.edu/coronavirus/health-and-wellness/facial-covering.html). If you do not comply, you will be referred to the Office of Community Standards and Integrity for follow up, which could include student conduct action. Finally, please note that sanitizing wipes should be available in our classroom for those who wish to use them.

Advanced Topics in

Machine Learning and Game Theory

Lecture 1: Introduction

17599/17759

Fei Fang

feifang@cmu.edu

Course Staff

Instructor

Name	Dr. Fei Fang
Contact Info	Email: feifang@cmu.edu

No Teaching Assistant

Course Information

- Lecture modality: In person expected
- Office Hour: W 1-2pm, TCS 321 or Zoom (link on Canvas)
- Canvas: canvas.cmu.edu
  - Syllabus, Quizzes, Assignments, Lecture slides
- Website: A copy of the syllabus page on Canvas
- Prerequisites: linear algebra, probability, algorithms, and at least one course in artificial intelligence

Today

- Course Logistics

- Introduction to Convex Optimization

Course Logistics

Course Scope

- Basics of Machine Learning and Game Theory
  - Introduction to convex optimization, game theory, online learning, reinforcement learning
- Learning in Games
  - Learning rules
  - Learning game parameters
- Multiagent Reinforcement Learning (MARL)
  - Classical algorithms/Recent advances in MARL
- Strategic Behavior in Learning
  - Adversarial Machine Learning (AML)
  - Learning from strategic data sources
- Applications of Machine Learning and Game Theory
  - Security and sustainability, Transportation

Connection to Other AI Courses

- The "Basics of Machine Learning and Game Theory" part has some overlaps with other AI courses
  - 15-281: AI; 15-780: Graduate AI
  - 10-701/15-781: Machine Learning; 10-715 Advanced Introduction to Machine Learning
  - 10-725/36-725: Convex Optimization; 10-703 Deep Reinforcement Learning or 10-707 Topics in Deep Learning
- This course focuses on recent advances at the intersection of ML and game theory

Class Format

- Lectures
- Paper presentations and paper discussions led by students
- Quizzes (In-class + after-class)

Learning Objective

- Describe fundamental theoretical result in learning in games, strategic classification, and multi-agent reinforcement learning
- Describe and implement classical and recent algorithms at the intersection of machine learning and game theory
- Describe the applications of techniques integrating machine learning and game theory
- Deliver a report of course project and present the work through oral presentation

Assignments and Grading

Course Component	#	Points	Expected Workload
Class Participation	14 weeks 26 lectures	10 points	3h/week
Paper Reading Assignment	5	15 points	1h/week (≈3h/paper)
Paper Presentation	2	15 points	1h/week (7h/paper)
Programming Assignment	2	20 points	1.5h/week (10.5h/assignment)
Course Project		40 points	5 hours/week

- Final Grade: Letter graded

Remarks about Paper Presentation

- 25-30 min/paper (confirm with instructor)
- 0~3 paper presentations per lecture (start from Lec 5)

- Presenters for the same lecture should coordinate
  - E.g., choose the order of presentation and avoid repeated content

- Contact instructor 1 week before presentation for additional information (e.g., length of presentation)
- Submit slides 48h before lecture for feedback (on Canvas)
  - Combine/concatenate the slides from multiple presenters

Late Submission Policy

- Late submission will be discounted by 0.7
- Late submissions will not be peer reviewed and cannot earn the peer review points

Use of Peer Review

- For all peer-reviewed assignments, the instructor will provide the final evaluation taking into account the reviews by the peers.

Textbook and Additional Reference

- No formal textbook

- List of additional resources will be provided (check Canvas and slides)

Academic Integrity

- Be collaborative, give credits
  - If discuss with others, specify names and complete on your own
- Leverage resources
  - If use publicly available code packages, specify source
- If your complete submissions are the same, you will get zero score and the case will be reported
- Course project report should follow standard academic integrity policy. Plagiarism is not allowed.
- See CMU policy on academic integrity for general information
  - https://www.cmu.edu/student-affairs/ocsi/academic-integrity/index.html

Special Needs

- If you have a disability and require accommodations, please contact Catherine Getchell, Director of Disability Resources, 412-268-6121, getchell@cmu.edu
- If you have an accommodations letter from the Disability Resources office, discuss with me as early as possibly
- Students who may require some short-term academic accommodations related to COVID-19 should contact Disability Resources at access@andrew.cmu.edu

Student Well-Being

- Start early! Avoid last-minute panic.
- CMU services are available, and treatment does work
- http://www.cmu.edu/counseling/
- 412-268-2922

Details about Assignments and Grading

Class Participation

- Quizzes
- Asking and answering questions in class/on Canvas

Paper Reading Assignment

- Summary (2.5 points)
  - Overview of paper content (0.5 points)
  - Detailed comments (1 point)
    - Cover Strengths (0.5 points) and Weaknesses (0.5 points)
  - Questions or Discussion (0.5 points)
- Peer review (0.5 points)
  - Provide an evaluation of each other's summary
  - Discuss the strengths and weaknesses of the paper, and/or questions each other have
- Possibly useful references:
  - Guidelines for critiquing a research study
  - Sample paper review for an academic conference

Paper Presentation

- Each student present 2 papers during the semester
- Each presentation should be 25-30 minutes
- In addition to the list of paper we provide, you can propose other papers relevant to the topic
- The instructor will assign the papers by taking into account a number of factors, including the students' preference, the quality of the paper, the relevance, etc.
- Bid in the google spreadsheet (link on Canvas) by 9/6
- Assignment results will be announced on 9/8

Paper Presentation

- Peer reviewed
- Evaluation criteria: how much one has learned from the presentation
- Try engage the audience!

Programming Assignment

- Two programming assignments
  - Designing an agent to play in a multi-agent particle environment
  - Defending and attacking an image classifier

Course Project

- Work in small groups (1-3 students in each group) on a project relevant to topics covered in this course
- Progress will be checked through Project Proposal, Project Progress Report, Project Presentation, and Final Project Report
- The proposal and progress report will be peer-reviewed
- Final report gets full score if at the same level as accepted papers at major AI conferences such as AAAI, IJCAI, NeurIPS, ICML, ICLR etc
- Submissions should follow AAAI format

Course Project

- Advisor is not required and will not be assigned
- Students are encouraged to reach out to faculty members/senior students/domain experts for advice
- I can provide feedback and advice during OHs

Checkpoints and Grading Criteria

- Confirm group members
  - No points; Due 9/13 (Mon)
- Project Proposal
  - 5 points; Due 9/20 (Mon); Peer-reviewed
- Project Progress Report
  - 5 points; Due 10/25 (Mon); Peer-reviewed
- Project Presentation
  - 10 points; Presentation date 11/29 & 12/1
- Full Project Report
  - 20 points; Due 12/10 (Fri)
- Each group submit one PDF file in each stage

Project Proposal (5 points)

- $\geq$300 words
- Pin down the problem (1 point)
- Briefly discuss related work (1 point)
- Describe 2~3 envisioned milestones of the proposed project, i.e., the important checkpoints that demonstrate the progress of the project (1 point)
- Describe the tentative plan of action, including the steps and the expected time needed for each step (1 point)
- Describe the tentative plan of distributing workload among team members (0.5 point)
- Provide comments and constructive feedback to the proposals assigned in peer-review (0.5 point)

Project Progress Report (5 points)

- ≥ 1 page double column (excluding references)
- Describe the problem (1 point)
- Describe the progress has been made towards the milestones (2 points)
- Provide an outline of the final report (0.5 point)
- Provide a tentative plan of next steps and distribution of workload (0.5 point)
- Provide comments and constructive feedback to the proposals assigned in peer-review (1 point)

Project Presentation (10 points)

- The presenter should be able to convey the following aspects clearly
  - Motivation (1 point)
  - Problem Description (1 point)
  - Related work / Background (1 point)
  - Contribution (3 point)
  - Evaluation/Results (2 points)
  - Future Work (1 point)
  - Q&A (1 points)

Full Project Report (20 points)

- $\geq$6 page in AAAI format (excluding references)
- Submission will be evaluated as a conference paper submission
  - Relevance to ML and Game Theory
  - Novelty and significance
  - Engagement with literature
  - Soundness of work
  - Quality of evaluation
  - Quality of presentation

Introduction to Convex Optimization

Outline

- Convex Optimization
- Gradient Descent and Projected Gradient Descent
- Linear Programming (LP)
- Dual Problem and KKT Conditions

Optimization Problem: Definition

- Optimization Problem: Determine value of optimization variable within feasible region/set to optimize optimization objective

$$\min_x f(x)$$
$$\text{s.t. } x \in \mathcal{F}$$

  - Optimization variable $x \in \mathbb{R}^n$
  - Feasible region/set $\mathcal{F} \subset \mathbb{R}^n$
  - Optimization objective $f: \mathcal{F} \rightarrow \mathbb{R}$

- Optimal solution: $x^* = \operatorname{argmin}_{x \in \mathcal{F}} f(x)$
- Optimal objective value $f^* = \min_{x \in \mathcal{F}} f(x) = f(x^*)$

Optimization Problem: How to Solve

- No general way to solve

$$\min_x f(x)$$
$$\text{s.t. } x \in \mathcal{F}$$

- Special classes
  - Convex optimization problem: f and $\mathcal{F}$ are convex
  - Linear Program (LP): f is linear, $\mathcal{F}$ is a polytope
  - Integer Linear Program (ILP): LP + variables are integer
  - Mixed Integer Linear Program (MILP)
  - Many other classes

- Existing solvers and code packages
  - Cplex, Gurobi, GLPK, Cvxopt…

Convex Optimization

- Convex Optimization Problem:

$$\min_{x} f(x)$$
$$\text{s.t. } x \in \mathcal{F}$$

  - f is a convex function
  - $\mathcal{F}$ is a convex set

Convex Set | Nonconvex Set | Convex Function | Nonconvex Function

Convex Set and Convex Function

- A set $\mathcal{F}$ is *convex* if $\forall x, y \in \mathcal{F}, \forall\theta \in [0,1]$

$$\mathrm{z} = \theta x + (1-\theta)y \in \mathcal{F}$$

  - Any convex combination of two points in the set is also in the set
- A function $f: \mathcal{F} \to \mathbb{R}$ is convex in a convex set $\mathcal{F}$ if $\forall x, y \in \mathcal{F}, \forall\theta \in [0,1]$,

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

  - Value in the middle point is lower than average value
  - If $\mathcal{F} = \mathbb{R}^n$, we simply say f is convex

Quiz 1

- The _________ of convex sets is convex
  - 1. Intersection
  - 2. Union

Convex Function

- How to determine if a function is convex?
  - If f is a twice continuously differentiable function of n variables, f is convex on $\mathcal{F}$ iff its Hessian matrix of second partial derivatives is positive semidefinite on the interior of $\mathcal{F}$

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\[2.2ex] \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\[2.2ex] \vdots & \vdots & \ddots & \vdots \\[2.2ex] \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

H is positive semidefinite in S if $\forall x \in S, \forall z \in \mathbb{R}^n, z^T H(x) z \geq 0$

H is positive semidefinite in $\mathbb{R}^n$ iff all eigenvalues of H are non-negative

Alternatively, prove $z^T H(x) z = \sum_i \left(g_i(x, z)\right)^2$

Convex Function

- Other ways to show convexity:
  - Prove by definition
  - Sum of convex functions is convex
    - If $f(x) = \sum_i w_i f_i(x)$, $w_i \geq 0$, $f_i(x)$ convex, then $f(x)$ is convex
  - Convexity is preserved under a linear transformation
    - If $f(x) = g(Ax + b)$, g convex, then $f(x)$ is convex

Concave Function

- A function f is concave if $-f$ is convex
  - Let $\mathcal{F}$ be a convex set. A function $f: \mathcal{F} \to \mathbb{R}$ is *concave* in $\mathcal{F}$ if $\forall x, y \in \mathcal{F}, \forall \theta \in [0,1]$,
$$f(\theta x + (1-\theta)y) \geq \theta f(x) + (1-\theta)f(y)$$

Q: If f is a concave function and $\mathcal{F}$ is a convex set, can the following problem converted to a convex optimization problem?

$$\max_x f(x)$$
$$\text{s.t. } x \in \mathcal{F}$$

We also view such problem as convex optimization problem

Quiz 2

- Is the following optimization problem a convex optimization problem?

$$\max_{x_1,x_2,\ldots,x_n} \sum_i a_i x_i - \sum_i x_i \log x_i$$

$$\text{s.t.} \textstyle\sum_i x_i = 1$$

$$x_i > 0$$

Affine function

- An *affine function* is a function of the form $f(x) = a^T x + b$ where $a \in \mathbb{R}^n, b \in \mathbb{R}$
  - If a function f is both convex and concave in $\mathbb{R}^n$, then f is an affine function

Affine function

- If a function f is both convex and concave in $\mathbb{R}^n$, then f is an affine function

- Proof: Let $g(x) = f(x) - f(0)$.
- Clearly $g(x)$ is convex and concave.
- By definition, $g(x)$ satisfies (i) $g(x + y) = g(x) + g(y)$; (ii) $g(\theta x) = \theta g(x)$.
- Let $\{e_j\}_1^n$ be the canonical basis of $\mathbb{R}^n$, then $x = \sum_j x_j e_j$.
- Therefore $g(x) = \sum_j x_j g(e_j) = a^T x$ where $a_j = g(e_j)$.
- So $f(x) = g(x) + f(0) = a^T x + b$

Convex Optimization: Local Optima=Global Optima

$$\min_x f(x)$$
$$\text{s.t. } x \in \mathcal{F}$$

- Theorem: For a convex optimization problem, all locally optimal points are globally optimal

x is *globally optimal* if $x \in \mathcal{F}$ and $\forall y \in \mathcal{F}, f(x) \leq f(y)$

x is *locally optimal* if $x \in \mathcal{F}$ and $\exists R > 0$ such that $\forall y: y \in \mathcal{F}$ and $\|x - y\|_2 \leq R, f(x) \leq f(y)$

Convex Optimization: Local Optima=Global Optima

- Prove by contradiction

Assume x is a local optima with optimality radius R, and $\exists y \in \mathcal{F}, f(x) > f(y)$. Let $z = \theta x + (1-\theta)y$ where $\theta = 1 - \frac{R}{2\|x-y\|_2}$.

(1) $z \in \mathcal{F}$ due to convexity of $\mathcal{F}$

(2) $\|x - z\|_2 = \|x - \theta x - (1-\theta)y\|_2 = (1-\theta)\|x-y\|_2$

$$= \frac{R}{2\|x-y\|_2}\|x-y\|_2 = \frac{R}{2} \leq R$$

(3) $f(z) \leq \theta f(x) + (1-\theta)f(y) < \theta f(x) + (1-\theta)f(x) = f(x)$

So x is not local optima. Contradiction.

Outline

- Convex Optimization
- Gradient Descent and Projected Gradient Descent
- Linear Programming (LP)
- Dual Problem and KKT Conditions

Gradient Descent (GD)

- For unconstrained optimization $\min_{x \in \mathbb{R}^n} f(x)$
- Iteratively update the value of x

Algorithm: Gradient Descent

Input: function f, initial point x_0, step size $\alpha > 0$

Initialize $x \leftarrow x_0$
Repeat
$$x \leftarrow x - \alpha \nabla_x f(x)$$
Until convergence

Gradient $\nabla_x f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n}\right]^{\mathrm{T}}$

Gradient Descent

- Theorem: For differentiable f and small enough α, for any x with $\nabla_x f(x) \neq 0, f\big(x - \alpha \nabla_x f(x)\big) < f(x)$
  - f can be convex or non-convex
  - Prove using Taylor expansion

- For convex and differentiable f and small enough α, gradient descent converges to global optimum

Gradient Descent

- Proof: $f\big(x-\alpha\nabla_x f(x)\big)=f(x)+\nabla_x f(x)^T\big(-\alpha\nabla_x f(x)\big)+O(||-\alpha\nabla_x f(x)||_2^2)$
$$=f(x)-\alpha||\nabla_x f(x)||_2^2+O(||\alpha\nabla_x f(x)||_2^2)$$
$$\le f(x)-\alpha||\nabla_x f(x)||_2^2+C\alpha^2||\nabla_x f(x)||_2^2$$
$$<f(x)$$
- Last inequality holds for $\alpha<\frac{1}{C}$ since $||\nabla_x f(x)||_2^2>0$

Projected Gradient Descent (PGD)

- For constrained optimization problem $\min_x f(x)$ s.t. $x \in \mathcal{F}$

Algorithm: Projected Gradient Descent
Input: function f, initial point x_0, step size $\alpha > 0$
Initialize $x \leftarrow x_0$ Repeat $\quad x \leftarrow P_{\mathcal{F}}(x - \alpha \nabla_x f(x))$ Until convergence

Add a projection step to ensure feasibility

$$P_{\mathcal{F}}(x) = \underset{x' \in \mathcal{F}}{\operatorname{argmin}} \|x - x'\|_2^2$$

Again a constrained optimization problem

Projected Gradient Descent

- If $\mathcal{F}$ is a l_2 ball, i.e., $\mathcal{F} = \{x: \|x\|_2 \leq 1\}$, then
  - $P_{\mathcal{F}}(x) = x/\|x\|_2$ for $x \notin \mathcal{F}$

- If $\mathcal{F}$ is a box, i.e., $\mathcal{F} = \{x: l \leq x \leq u\}$, then
  - $[P_{\mathcal{F}}(x)]_i = l_i$ if $x_i < l_i$
  - $[P_{\mathcal{F}}(x)]_i = u_i$ if $x_i > u_i$
  - $[P_{\mathcal{F}}(x)]_i = x_i$ otherwise

Outline

- Convex Optimization
- Gradient Descent and Projected Gradient Descent
- Linear Programming (LP)
- Dual Problem and KKT Conditions

Linear Program

- Linear Program:
  - A special case of convex optimization problem
  - An optimization problem whose optimization objective is a linear function and feasible region is a polytope (defined by a set of linear constraints)

$$\max_x c^T x$$
$$\text{s.t. } Ax \leq b$$

Note: can also be minimization

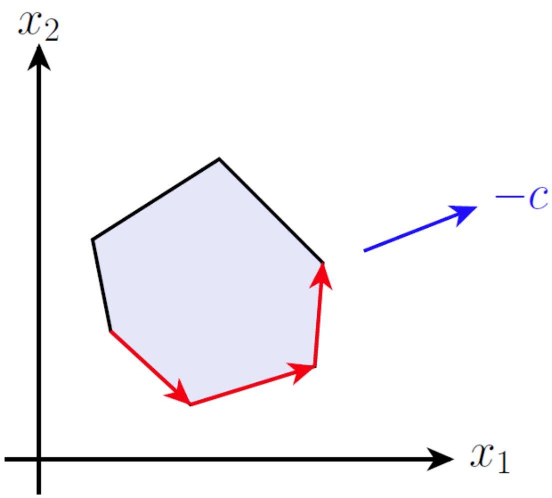
  - $c \in \mathbb{R}^n$
  - $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$

Linear Program

- An LP may be
  - Infeasible: Feasible region is empty
  - Unbounded: The objective value can be arbitrarily large (for a maximization problem)
- If an LP has an optimal solution, at least one vertex of the polytope is an optimal solution

The Simplex Algorithm

$$\min_x c^T x$$
$$\text{s.t. } Ax \le b$$



- Intuition: no need to enumerate all vertices
- Recall local search and gradient descent
- Move towards a neighbor to get reduced objective value
- Still need to enumerate almost all the vertices in the worst case, but very efficient in most cases

Q: How to find a vertex? How to find a neighboring vertex?

Quiz 3

- Consider the following two LPs (LP-L and LP-R) where $b \geq 0$

LP-L

$$\min_{x,z} 1^T z$$
$$\text{s.t. } Ax + z = b$$
$$x, z \geq 0$$

LP-R

$$\min_{x} c^T x$$
$$\text{s.t. } Ax = b$$
$$x \geq 0$$

- Applying simplex algorithm to LP-L with the initial vertex $x_0 = 0, z_0 = b$. Denote the optimal solution as (x^*, z^*). If $z^* = 0$, then which of the following claims are true about x^*?
  - A: x^* is not in the feasible region of LP-R
  - B: x^* is in the feasible region of LP-R
  - C: x^* is a vertex of the feasible region of LP-R
  - D: x^* is an optimal solution of LP-R

Solve LPs in practice

- All the solvers/algorithms for Convex Optimization problems can be applied
- Additional solvers/algorithms for LPs
  - linprog (MATLAB), linprog (in Python package SciPy)
  - Cvxpy (Python)
  - PuLP (Python)
  - Cplex, Gurobi
  - https://www.informs.org/ORMS-Today/Public-Articles/June-Volume-38-Number-3/Software-Survey-Linear-Programming

Outline

- Convex Optimization
- Gradient Descent and Projected Gradient Descent
- Linear Programming (LP)
- Dual Problem and KKT Conditions

Lagrange Multiplier

- Consider a simple optimization problem with two variables and one equality constraint

$$\min_{x_1,x_2} f(x_1,x_2)$$
$$\text{s.t. } h(x_1,x_2) = 0$$

- Define the Lagrangian

$$L(x_1,x_2,\lambda) = f(x_1,x_2) + \lambda h(x_1,x_2)$$

- Claim: The optimal solution x_1^*, x_2^* satisfies

$$\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial x_2} = \frac{\partial L}{\partial \lambda}\Big|_{x_1^*,x_2^*,\lambda} = 0$$

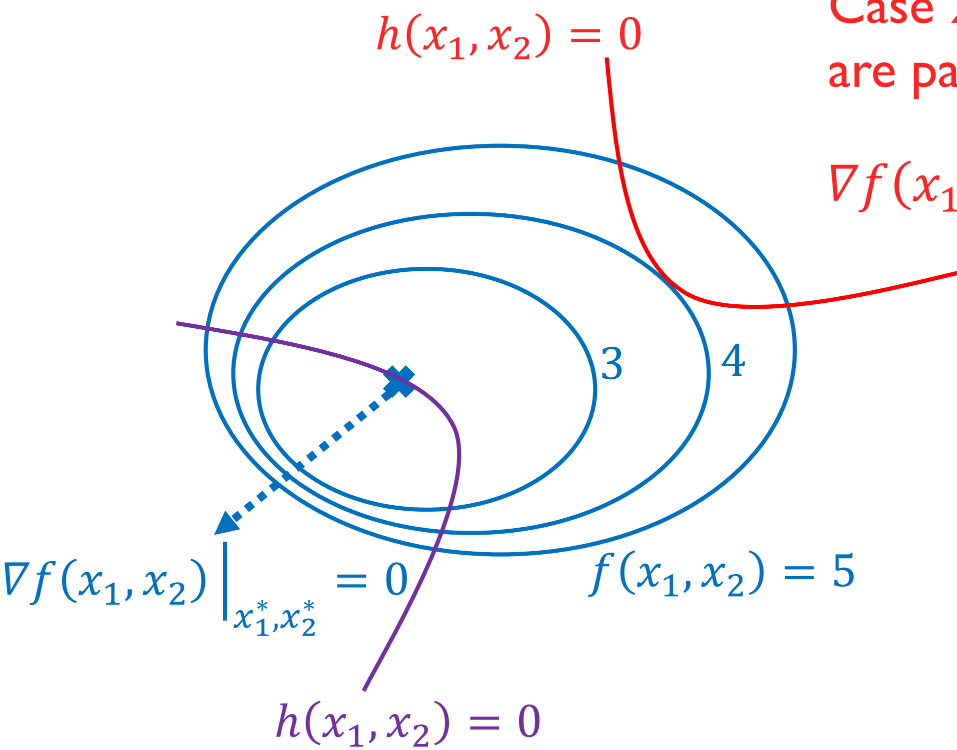
For some λ

Lagrange Multiplier

$$\min_{x_1,x_2} f(x_1,x_2)$$
$$\text{s.t. } h(x_1,x_2)=0$$

Case 2: Gradient vectors of h and f are parallel at x_1^*, x_2^*

$$\nabla f(x_1,x_2)\Big|_{x_1^*,x_2^*} = \lambda \nabla h(x_1,x_2)\Big|_{x_1^*,x_2^*}, \lambda \neq 0$$



Case 1: $h(x_1^*,x_2^*)=0, \nabla f(x_1,x_2)|_{x_1^*,x_2^*}=0$

$$L(x_1,x_2,\lambda)=f(x_1,x_2)+\lambda h(x_1,x_2)$$

$$\frac{\partial L}{\partial x_1}=\frac{\partial L}{\partial x_2}=\frac{\partial L}{\partial \lambda}\Big|_{x_1^*,x_2^*,\lambda}=0$$

Combines the two cases

Dual Problem

$$\min_{x\in\mathbb{R}^n} f(x)$$
$$\text{s.t. } g_i(x) \le 0, i = 1 \dots m$$
$$h_j(x) = 0, j = 1 \dots l$$

x: primal variables
μ and λ: dual variables (Lagrange multipliers)

- Define the Lagrangian

$$L(x,\mu,\lambda) = f(x) + \sum_i \mu_i g_i(x) + \sum_j \lambda_j h_j(x)$$

- Let $d(\mu,\lambda) = \inf_x L(x,\mu,\lambda)$. Define the dual problem as

$$\max_{\mu,\lambda} d(\mu,\lambda)$$
$$\text{s.t. } \mu_i \ge 0, i = 1 \dots m$$

- Whether or not f is convex, we can show $f^* \ge d^*$ (weak duality)
- When f is convex, we can prove $f^* = d^*$ (strong duality)

Quiz 4

Recall

$$\min_{x\in\mathbb{R}^n} f(x)$$
s.t. $g_i(x) \le 0, i = 1 \dots m$
$h_j(x) = 0, j = 1 \dots l$

$$L(x,\mu,\lambda) = f(x) + \sum_i \mu_i g_i(x) + \sum_j \lambda_j h_j(x)$$
$$d(\mu,\lambda) = \inf_x L(x,\mu,\lambda)$$
$$\max_{\mu,\lambda} d(\mu,\lambda)$$
s.t. $\mu_i \ge 0, i = 1 \dots m$

- For the following problem

$$\min_{x\in\mathbb{R}} 3x + 4$$
s.t. $2x - 1 \le 0$
$-x + 5 \le 0$

- The dual problem is equivalent to

A: $$\max_{\mu_1,\mu_2} 2\mu_1 + \mu_2$$
s.t. $2\mu_1 - \mu_2 - 3 \ge 0$
$\mu_1 \ge 0, \mu_2 \ge 0$

B: $$\max_{\mu_1,\mu_2} -\mu_1 + 5\mu_2$$
s.t. $2\mu_1 - \mu_2 = -3$
$\mu_1 \ge 0, \mu_2 \ge 0$

Quiz 4

$$L(x,\mu,\lambda) = f(x) + \sum_i \mu_i g_i(x) + \sum_j \lambda_j h_j(x)$$

$$d(\mu,\lambda) = \inf_x L(x,\mu,\lambda)$$

$$\max_{\mu,\lambda} d(\mu,\lambda)$$

s.t. $\mu_i \geq 0, i = 1 \dots m$

$$\max_{\mu_1,\mu_2} \inf_x 3x + 4 + \mu_1(2x-1) + \mu_2(-x+5)$$

s.t. $\mu_1 \geq 0, \mu_2 \geq 0$

Quiz 4

$$L(x,\mu,\lambda) = f(x) + \sum_i \mu_i g_i(x) + \sum_j \lambda_j h_j(x)$$

$$d(\mu,\lambda) = \inf_x L(x,\mu,\lambda)$$

$$\max_{\mu,\lambda} d(\mu,\lambda)$$

s.t. $\mu_i \geq 0, i = 1 \dots m$

$$\max_{\mu_1,\mu_2} \inf_x 3x + 4 + \mu_1(2x-1) + \mu_2(-x+5)$$

s.t. $\mu_1 \geq 0, \mu_2 \geq 0$

$$\max_{\mu_1,\mu_2} \inf_x (3 + 2\mu_1 - \mu_2)x + 4 - \mu_1 + 5\mu_2$$

s.t. $\mu_1 \geq 0, \mu_2 \geq 0$

$$\max_{\mu_1,\mu_2} -\mu_1 + 5\mu_2$$

$$2\mu_1 - \mu_2 = -3$$

$$\mu_1 \geq 0, \mu_2 \geq 0$$

Linear Program Duality

- Dual problem of an LP: also a linear program
  - Each dual variable corresponds to a constraint in primal LP

Primal LP	Dual LP
$\max_x c^T x$ s.t. $Ax \leq b$	$\min_y b^T y$ s.t. $A^T y = c$ $y \geq 0$
$\boxed{x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, m < n}$	$\boxed{y \in \mathbb{R}^m}$

Linear Program Duality

- Strong duality holds (if feasible and bounded)
  - Primal and dual have the same optimal objective value
- The dual of the dual of a problem is itself

Primal LP	Dual LP
$\max_x c^T x$ s.t. $Ax \leq b$	$\min_y b^T y$ s.t. $A^T y = c$ $y \geq 0$
$\boxed{x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, m < n}$	$\boxed{y \in \mathbb{R}^m}$

Weak duality: $c^T x^* \leq b^T y^*$

Strong duality: $c^T x^* = b^T y^*$

Linear Program Duality

- Prove weak duality: $c^T x^* \leq b^T y^*$

Primal LP

$$\max_x c^T x$$
$$\text{s.t. } Ax \leq b$$

$$\boxed{x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, m < n}$$

Dual LP

$$\min_y b^T y$$
$$\text{s.t. } A^T y = c$$
$$y \geq 0$$

$$\boxed{y \in \mathbb{R}^m}$$

Linear Program Duality

- Prove weak duality: $c^T x^* \leq b^T y^*$

$$c^T x^* = (A^T y^*)^T x^* = y^{*T} A x^* = y^{*T}(Ax^*) \leq y^{*T} b$$

Primal LP

$$\max_x c^T x$$
$$\text{s.t. } Ax \leq b$$

$$\boxed{x \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}, m < n}$$

Dual LP

$$\min_y b^T y$$
$$\text{s.t. } A^T y = c$$
$$y \geq 0$$

$$\boxed{y \in \mathbb{R}^m}$$

Write the Dual of an LP

Maximize	Minimize
ith constraint ≤	ith variable ≥ 0
ith constraint ≥	ith variable ≤ 0
ith constraint =	ith variable unrestricted
jth variable ≥ 0	jth constraint ≥
jth variable ≤ 0	jth constraint ≤
jth variable unrestricted	jth constraint =

Linear Program Duality

- Let LP-1 denote the original LP, LP-2 denote the dual of LP-1, and LP-3 denote the dual of LP-2. Then LP-1 and LP-3 are the same (or can be converted to each other with variable substitution)

Lp-1

$$\min_x c^T x$$
$$\text{s.t. } Ax = b$$
$$x \geq 0$$

Lp-2

$$\max_y b^T y$$
$$\text{s.t. } A^T y \leq c$$

$y = y^+ - y^-$

Lp-2 (Standard form)

$$\min_{y^+, y^-, z} \; b^T y^+ - b^T y^-$$
$$\text{s.t. } A^T y^+ - A^T y^- + z = c$$
$$y^+, y^-, z \geq 0$$

dual

LP-3

$$\max_w c^T w$$
$$\text{s.t. } Aw \leq b^T$$
$$Aw \leq -b^T$$
$$w \leq 0$$

$x = w$

Proof of strong duality theorem

- Farkas' lemma: Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then exactly one of the following two statements is true
  - I. There exists an $x \in \mathbb{R}^n$ such that $Ax = b$ and $x \geq 0$
  - II. There exists a $y \in \mathbb{R}^m$ such that $A^T y \geq 0$ and $b^T y < 0$
  - Proof:
    - If (I) is true, i.e., $Ax = b$ holds for some x. If $A^T y \geq 0$ for some y, then $b^T y = (Ax)^T y = x^T (A^T y) \geq x^T \mathbf{0} = 0$. So (I)(II) cannot both be true.
    - If (I) is false, then define $C = \{q \in \mathbb{R}^m : \exists x \geq 0, Ax = q\}$. We know $b \neq C$. Notice that C is convex. From separating hyperplane theorem, we know for some $y \in \mathbb{R}^m \backslash \mathbf{0}$ s.t. $q^T y \geq 0\ \forall q \in C$ and $b^T y < 0$. Then we can show that for this y, $A^T y \geq 0$. If not, i.e., if $A^T y < 0$, then choose any $q \in C$, and choose any $x \geq 0$ such that $Ax = q$, we have $0 \leq q^T y = (Ax)^T y = x^T A^T y = x^T (A^T y) < x^T \mathbf{0} = 0$. Contradiction. So this y satisfies $A^T y \geq 0$ and $b^T y < 0$. Therefore (II) is true.
    - So exactly one of (I) and (II) is true

Proof of strong duality theorem

- Second variant of Farkas' lemma: Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then system $Ax \leq b$ has a solution if and only if $\lambda^T b \geq 0$ holds for all λ that satisfies $\lambda \geq 0$ and $\lambda^T A = 0$
  - Proof:
    - If $Ax \leq b$ has a solution, denote the solution as x^*. If $\lambda \geq 0$ and $\lambda^T A = 0$, then $\lambda^T b \geq \lambda^T (Ax^*) = (\lambda^T A) x^* = 0$
    - If $Ax \leq b$ does not have a solution, then $Ax^+ - Ax^- + z = b, x^+, x^-, z \geq 0$ does not have a solution (otherwise you can easily construct a solution for $Ax \leq b$). By Farkas' lemma, there exists a λ such that $[A \;\; -A \;\; I]^T \lambda \geq 0$ and $b^T \lambda < 0$. Then for this λ, we know $A^T \lambda = 0$ (and therefore $\lambda^T A = 0$) and $\lambda \geq 0$

Proof of strong duality theorem

- Suppose the primal has an optimal solution x^*, leading to optimal value $\mathrm{f}^* = c^T x^*$, $(y^*, g^* = b^T y^*)$ is the optimal solution and the optimal value of the dual, and $f^* > g^*$. Then for any $\epsilon > 0$, we know that $\nexists y, b^T y \geq g^* + \epsilon, A^T y \leq c$, i.e., $\begin{bmatrix} A^T \\ -b^T \end{bmatrix} y \leq \begin{bmatrix} c \\ -g^* - \epsilon \end{bmatrix}$ does not have a solution. Based on the variant of the Farkas' lemma, there exists a $\lambda \in \mathbb{R}^{n+1}$ satisfying $\lambda \geq 0$, $\lambda^T \begin{bmatrix} A^T \\ -b^T \end{bmatrix} = 0$, and $\lambda^T \begin{bmatrix} c \\ -g^* - \epsilon \end{bmatrix} < 0$. Write this λ as $\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$ where $\lambda_1 \in \mathbb{R}^n, \lambda_2 \in \mathbb{R}, \lambda_1 \geq 0, \lambda_2 \geq 0$.
- If $\lambda_2 = 0$, then $\lambda_1^T A^T = 0, \lambda_1^T c < 0, \lambda_1 \geq 0$. According to the variant of the Farkas' lemma, $A^T y \leq c$ should not have a solution. But y^* is a solution of the dual and therefore $A^T y^* \leq c$. Contradiction.
- If $\lambda_2 > 0$, then we can scale every the parameters in the problem so that $\lambda_2 = 1$. Then $\lambda_1^T A^T = b^T$ and $\lambda_1^T c < g^* + \epsilon$. Therefore λ_1 is a feasible solution of the primal and has a corresponding objective value lower than $g^* + \epsilon$. Since primal is minimization, we have $f^* \leq c^T \lambda_1 < g^* + \epsilon$. According to weak duality theorem, $f^* \geq g^*$. So $g^* \leq f^* < g^* + \epsilon$ for any $\epsilon > 0$. Then the only possibility is $f^* = g^*$.

Karush–Kuhn–Tucker (KKT) conditions

$$\min_{x\in\mathbb{R}^n} f(x)$$
$$\text{s.t. } g_i(x) \le 0, i = 1 \dots m$$
$$h_j(x) = 0, j = 1 \dots l$$

- Let x^* be the globally optimal point
- The KKT conditions are

$$\nabla f(x^*) + \sum_i \mu_i \nabla g_i(x^*) + \sum_j \lambda_j \nabla h_j(x^*) = 0$$

$$\frac{\partial L}{\partial x} = 0$$

$$g_i(x^*) \le 0, i = 1 \dots m$$
$$h_j(x^*) = 0, j = 1 \dots l$$

Primal feasibility

$$\mu_i \ge 0, i = 1 \dots m$$

Dual feasibility

$$\mu_i g_i(x^*) = 0, i = 1 \dots m$$

Complementary Slackness

Karush–Kuhn–Tucker (KKT) conditions

- For the following convex optimization problem

$$\min_{x\in\mathbb{R}^n} f(x)$$
$$\text{s.t. } g_i(x) \le 0, i = 1 \dots m$$
$$h_j(x) = 0, j = 1 \dots l$$

where g_i are differentiable convex functions, h_i are affine function, f is convex

- The KKT conditions are necessary and sufficient conditions for global optima

Quiz 5

Recall $\nabla f(x^*) + \sum_i \mu_i \nabla g_i(x^*) + \sum_j \lambda_j \nabla h_j(x^*) = 0$

$g_i(x^*) \leq 0, i = 1 \dots m$
$h_j(x^*) = 0, j = 1 \dots l$

$\mu_i \geq 0, i = 1 \dots m$

$\mu_i g_i(x^*) = 0, i = 1 \dots m$

- For the following problem

$$\min_{x \in \mathbb{R}^n} 3x + 4$$
$$\text{s.t. } 2x - 1 \leq 0$$
$$-x + 5 \leq 0$$

- The KKT conditions are (choose all that apply)
  - A: $2x - 1 \leq 0$
  - B: $-x + 5 \leq 0$
  - C: $\mu_1 \geq 0$
  - D: $\mu_2 \geq 0$
  - E: $\mu_1(2x - 1) = 0$
  - F: $\mu_2(-x + 5) = 0$
  - G: $3x + 4 \leq 0$
  - H: $\mu_1 + 2\mu_2 = 4$
  - I: $2\mu_1 - \mu_2 = -3$

Quiz 5

$$\min_{x \in \mathbb{R}^n} f(x) = 3x + 4$$

s.t. $g_1(x) = 2x - 1 \leq 0$

$$g_2(x) = -x + 5 \leq 0$$

$$\nabla f(x^*) + \sum_i \mu_i \nabla g_i(x^*) + \sum_j \lambda_j \nabla h_j(x^*) = 0$$

$$\Downarrow$$

$$3 + \mu_1 \cdot 2 + \mu_2 \cdot (-1) = 0$$

KKT Conditions for LP

$$\max_x c^T x$$
$$\text{s.t. } Ax \le b$$

$$\frac{\partial L}{\partial x} = 0$$

Primal feasibility

Dual feasibility

Complementary Slackness

Necessary and sufficient conditions for optimality!

KKT Conditions for LP

$$\max_x c^T x$$
$$\text{s.t. } Ax \leq b$$

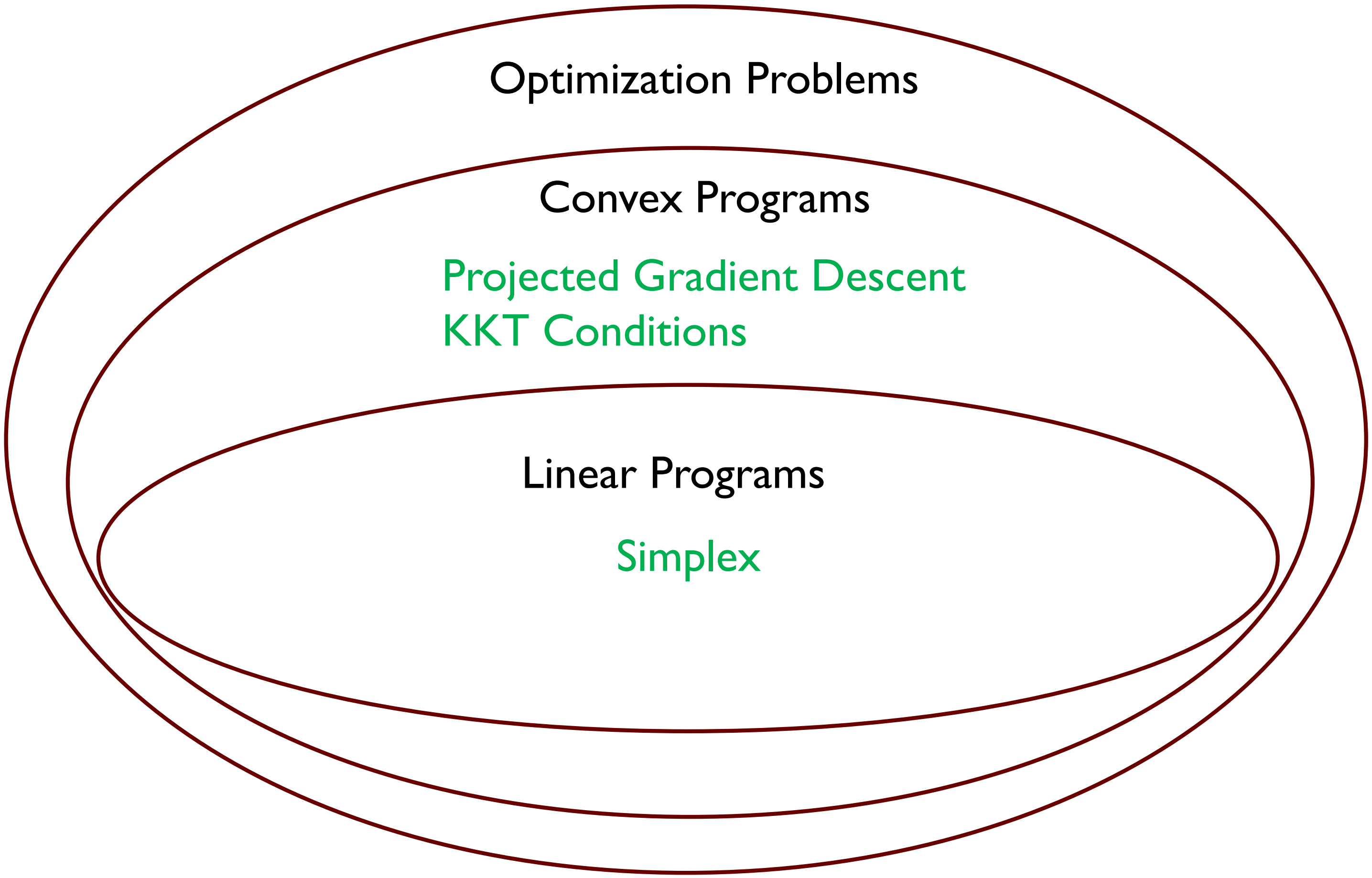
$A^T y = c$	$\dfrac{\partial L}{\partial x} = 0$
$Ax \leq b$	Primal feasibility
$y \geq 0$	Dual feasibility
$y^T(Ax - b) = 0$	Complementary Slackness

Necessary and sufficient conditions for optimality!

Dual Problem and KKT Conditions

- Sometimes the dual problem is easier to solve than the primal problem
- Sometimes the KKT conditions are easier to solve than the primal problem

Summary



Optimization Problems
Convex Programs
Projected Gradient Descent
KKT Conditions
Linear Programs
Simplex

Linear Program: Additional Resources

- Textbook
  - *Applied Mathematical Programming, Chapters 2-4*

    By Bradley, Hax, and Magnanti (Addison-Wesley, 1977)

    http://web.mit.edu/15.053/www/AMP.htm
  - *Convex Optimization, Chapters 1-4*

    Stephen Boyd and Lieven Vandenberghe

    Cambridge University Press

    https://web.stanford.edu/~boyd/cvxbook/

- Online course
  - https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-251j-introduction-to-mathematical-programming-fall-2009/index.htm
  - http://ee364a.stanford.edu/courseinfo.html
  - https://youtu.be/McLq1hEq3UY

Acknowledgment

- Some slides are borrowed from previous slides made by J. Zico Kolter