Artificial Intelligence Methods for Social Good

Lecture 1 (Part II)

Basics of Optimization

17-537 (9-unit) and 17-737 (12-unit)
Fei Fang
feifang@cmu.edu

Outline

- Optimization Problem
- Convex Optimization Problem

Learning Objectives

- Understand the concept of
  - Optimization Problem
  - Convex Optimization Problem (CO)
- Briefly describe the following algorithms
  - Gradient Descent Algorithm
- Formulate problems as CO and use solvers to solve them

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 $f: \mathcal{F} \to \mathbb{R}$

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

Optimization Problem: Example

x_i	1.0	2.0	3.5
y_i	2.1	3.98	7.0

- Example: Linear Regression
  - Problem: Find a such that $y_i \approx a x_i, \forall i = 1..3$
  - Variable a
  - Feasible region $\mathbb{R}$
  - Objective function $f(a)$?

$$\min_a \sum_{i=1}^{3} |y_i - a x_i| \quad \text{s.t. } a \in \mathbb{R}$$

$$\min_a \sum_{i=1}^{3} (y_i - a x_i)^2 \quad \text{s.t. } a \in \mathbb{R}$$

Optimization Problem: How to Solve

- Many algorithms developed for special classes of optimization problems (i.e., when $f(x)$ and $\mathcal{F}$ satisfy certain constraints)

- We will mainly cover the following classes in this course
  - Convex optimization problem (CO)
  - Linear Programming problem (LP)
  - (Mixed) Integer Linear Programming problem (MILP)

- Many existing solvers and code packages available
  - Cplex (LP, MILP), Gurobi (LP, MILP), Cvxopt (CO)

Lazy Mode

- Formulate a problem as an optimization problem
- Identify which class the formulation belongs to
- Call the corresponding solver
- Map the solution back to the original problem
- Done!

Why Go Further?

- Learn how to identify which class the problem formulation belongs to
- Understand which formulations can be solved more efficiently
- Choose/Convert to the right formulation
- Open the black box to learn key ideas, useful for developing advanced solutions

Outline

- Optimization Problem
- Convex Optimization Problem

Convex Optimization: Definition

- Convex Optimization Problem
  - A special case of optimization problem that can be solved efficiently
  - An optimization problem whose optimization objective is a convex function and feasible region is a convex set

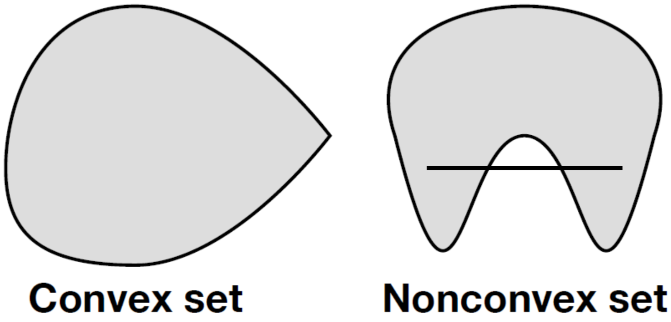
$$\min_{x} f(x)$$
$$\text{s.t. } x \in \mathcal{F}$$

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

Convex Optimization: Definition

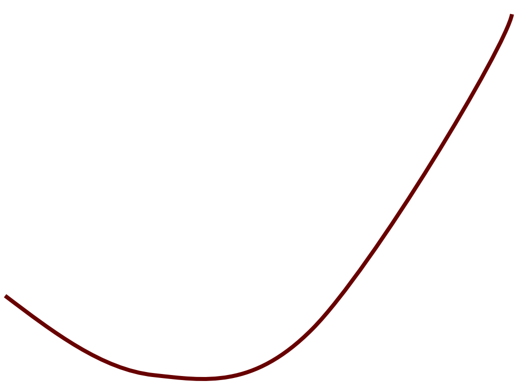
- Convex set
  - Any convex combination of two points in the set is also in the set
  - A set $\mathcal{F}$ is *convex* if $\forall x, y \in \mathcal{F}, \forall \theta \in [0,1]$,

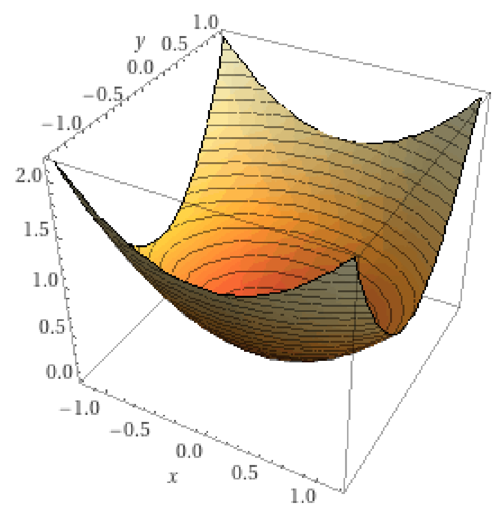
$$z = \theta x + (1 - \theta) y \in \mathcal{F}$$

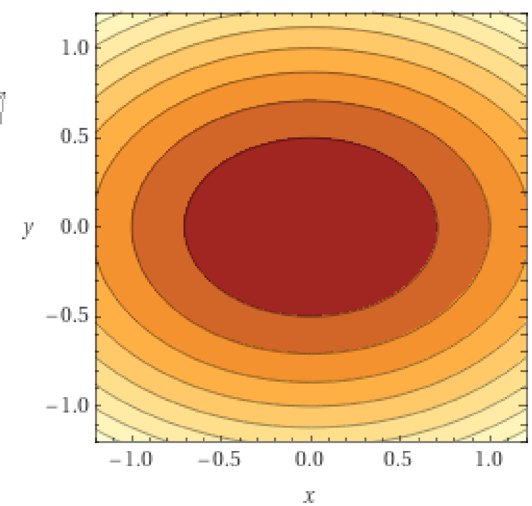


Convex Optimization: Definition

- Convex function
  - Value in the middle point is lower than average value
  - Let $\mathcal{F}$ be a convex set. A function $f: \mathcal{F} \to \mathbb{R}$ is convex in $\mathcal{F}$ if $\forall x, y \in \mathcal{F}, \forall \theta \in [0,1]$,
  $$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$
  - If $\mathcal{F} = \mathbb{R}^n$, we simply say f is convex







Convex Optimization: Which

- How to determine if a function is convex?
  - Prove by definition
  - Use properties
    - 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
    - If f is a twice differentiable function of one variable, f is convex on an interval $[a, b] \subset \mathbb{R}$ iff (if and only if) its second derivative $f''(x) \geq 0$ in $[a, b]$

Poll 1

- Which of the following functions are convex functions of a?
  - A: Only f_1
  - B: Only f_2
  - C: Both f_1 and f_2
  - D: Neither f_1 nor f_2
  - E: I don't know

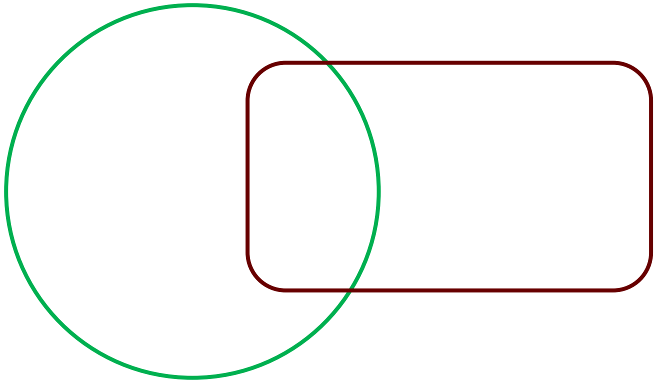
x_i	1.0	2.0	3.5
y_i	2.1	3.98	7.0

$$f_1(a) = \sum_{i=1}^{3} |y_i - ax_i|$$

$$f_2(a) = \sum_{i=1}^{3} (y_i - ax_i)^2$$

Convex Optimization Problem: Which

- Which sets are convex?
  - Prove by definition
  - Check known convex sets
    - Region defined by linear inequalities, Unit ball
  - Use properties
    - Intersection of convex sets is convex: $\mathcal{F} = \cap_i \mathcal{F}_i$ is convex if $\mathcal{F}_i$ is convex, $\forall i$



Poll 2

- Which of the following are convex optimization problems?
  - A: Only P_1
  - B: Only P_2
  - C: Both P_1 and P_2
  - D: Neither P_1 nor P_2
  - E: I don't know

x_i	1.0	2.0	3.5
y_i	2.1	3.98	7.0

P_1

$$\min_a \sum_{i=1}^{3} |y_i - ax_i|$$
$$\text{s.t. } 2a + 1 \leq 3$$

P_2

$$\min_a \sum_{i=1}^{3} (y_i - ax_i)^2$$
$$\text{s.t. } |a| \geq 3$$

Convex Optimization: Local Optima=Global Optima

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

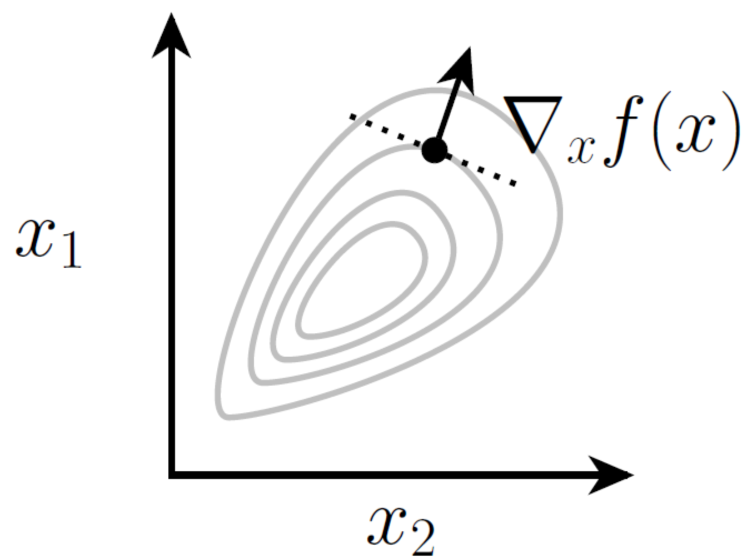
- Given an optimization problem, a point $x \in \mathbb{R}^n$ is *globally optimal* if $x \in \mathcal{F}$ and $\forall y \in \mathcal{F}, f(x) \leq f(y)$
- Given an optimization problem, a point $x \in \mathbb{R}^n$ is *locally optimal* if $x \in \mathcal{F}$ and $\exists R > 0$ such that $\forall y: y \in \mathcal{F}$ and $\|x - y\|_2 \leq R, f(x) \leq f(y)$

- Theorem 1: For a convex optimization problem, all locally optimal points are globally optimal

Convex Optimization: How to Solve

- For $f: \mathbb{R}^n \rightarrow \mathbb{R}$, *gradient* is the vector of partial derivatives
  - A multi-variable generalization of the derivative
  - Point in the direction of steepest increase in f

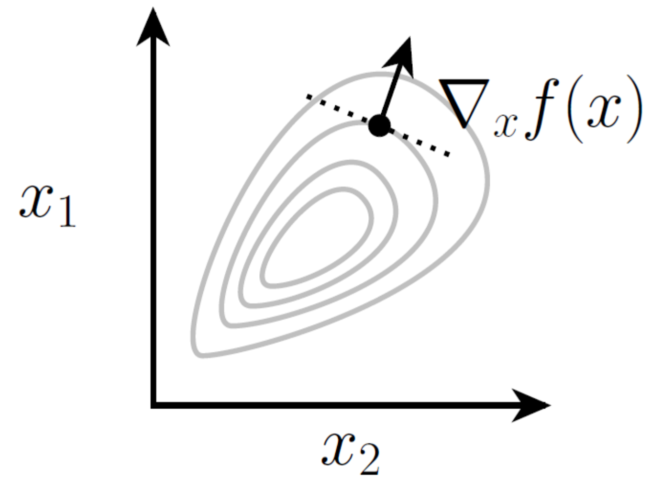
$$\nabla_x f(x) \in \mathbb{R}^n = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$



Gradient Descent

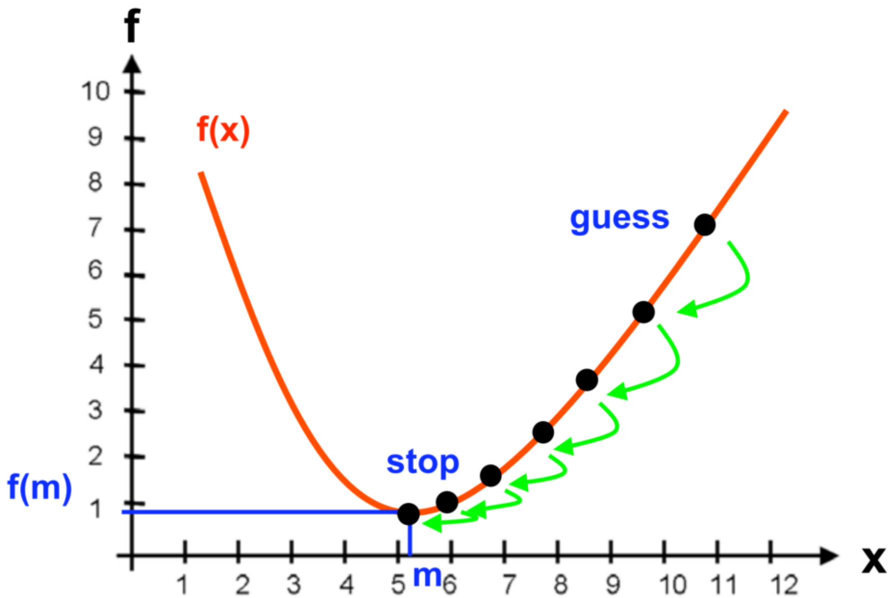
- Gradient descent: iteratively update the value of x
  - A simple algorithm for unconstrained optimization $\min_{x \in \mathbb{R}^n} f(x)$
  - Move towards a promising direction

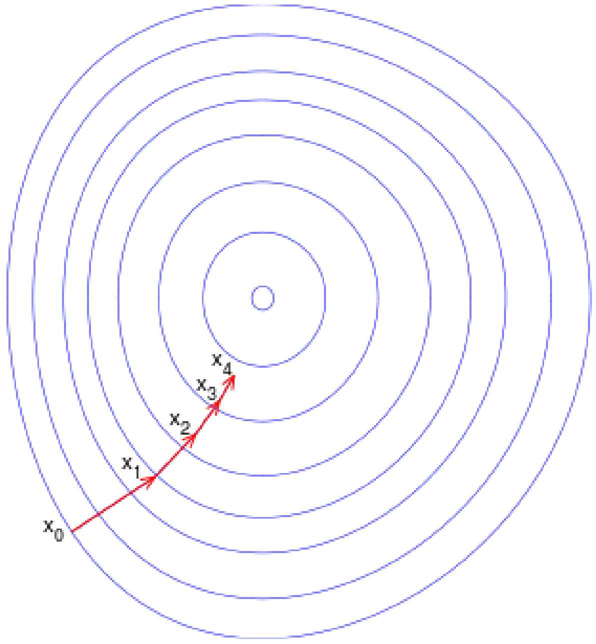
Algorithm: Gradient Descent
Input: function f, initial point x_0, step size $\alpha > 0$
Initialize $x \leftarrow x_0$ Repeat $\quad x \leftarrow x - \alpha \nabla_x f(x)$ Until convergence



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

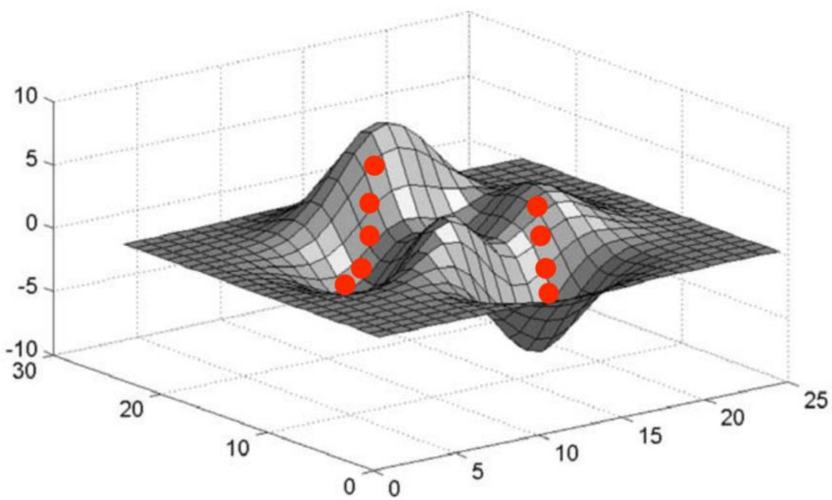
Gradient Descent: Example

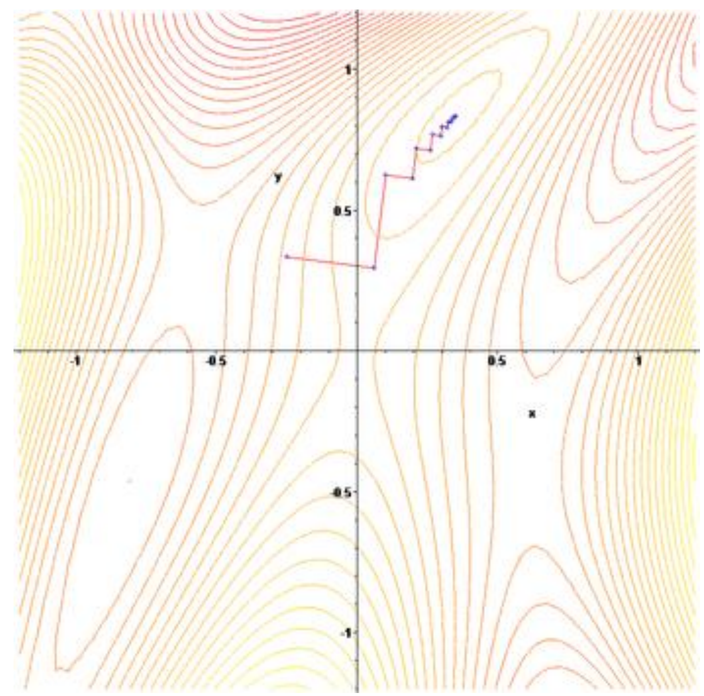




Gradient Descent: Remark

- Can also be used for nonconvex but continuous optimization
  - Lead to local optima





Convex Optimization: How to Solve

- Example solvers: fmincon (MATLAB), cvxpy (Python), cvxopt (Python), cvx (MATLAB)

With Python 3.7 and Cvxpy 1.1

x_i	1.0	2.0	3.5
y_i	2.1	3.98	7.0

$$\min_a \sum_{i=1}^{3} (y_i - a x_i)^2$$
$$\text{s.t. } a \in \mathbb{R}$$

```python
import cvxpy as cp
import numpy as np

x = np.array([1.0, 2.0, 3.5])
y = np.array([2.1, 3.98, 7.0])

# Define and solve the CVXPY problem.
a = cp.Variable(1)
cost = cp.sum_squares(y-a * x)
prob = cp.Problem(cp.Minimize(cost))
prob.solve()

# Print result.
print("\nThe optimal value is", prob.value)
print("The optimal a is", a.value)
```


https://www.cvxpy.org/examples/basic/least_squares.html

Methods for Convex Optimization

- Unconstrained and differentiable
  - Gradient Descent
  - Find saddle point (set gradient to be 0)
    - Closed form solution
    - Newton's Method (if twice differentiable)
- Constrained and differentiable
  - Interior Point Method
  - Karush–Kuhn–Tucker (KKT) conditions are sufficient and necessary. Apply Newton's method or other methods.
- Non-differentiable
  - ϵ-Subgradient Method
  - Cutting Plane Method

References and Additional Resources

Convex Optimization: Additional Resources

- Textbook
  - *Convex Optimization, Chapters 1-4*
    Stephen Boyd and Lieven Vandenberghe
    Cambridge University Press
    https://web.stanford.edu/~boyd/cvxbook/
- Online course
  - Stanford University, Convex Optimization I (EE 364A), taught by Stephen Boyd
    - http://ee364a.stanford.edu/courseinfo.html
    - https://youtu.be/McLq1hEq3UY

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
- Online course
  - https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-251j-introduction-to-mathematical-programming-fall-2009/index.htm
- Survey of existing software: https://www.informs.org/ORMS-Today/Public-Articles/June-Volume-38-Number-3/Software-Survey-Linear-Programming
  -

Backup Slides

Optimization Problem: How to Solve

- Many algorithms developed for special classes of optimization problems (i.e., when $f(x)$ and $\mathcal{F}$ satisfy certain constraints)
  - Convex optimization problem (CO)
  - Linear Programming problem (LP)
  - (Mixed) Integer Linear Programming problem (MILP)
  - Quadratic programming (QP), (Mixed) Integer Quadratic programming (MIQP), Semidefinite programming (SDP), Second-order cone programming (SOCP), …

- Existing solvers and code packages for these problems
  - Cplex (LP, MILP, QP), Gurobi (LP, MILP, MIQP), GLPK (LP, MILP), Cvxopt (CO), DSDP5 (SDP), MOSEK (QP, SOCP), Yalmip (SDP), …

Convex Optimization: Which

- 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 $\mathcal{F}$

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

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 $z^T H(x) z = \sum_i \big(g_i(x,z)\big)^2$

Gradient Descent: Advanced Material

- Gradient descent: iteratively update the value of x
  - A simple algorithm for unconstrained optimization $\min_{x \in \mathbb{R}^n} f(x)$

Algorithm: Gradient Descent

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

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

- Variants
  - How to choose x_0, e.g., $x_0 = 0$
  - How to update α, e.g., $\alpha^{i+1} = \frac{\left(x^{i+1}-x^{i}\right)^T \left(\nabla_x f\left(x^{i+1}\right)-\nabla_x f\left(x^{i}\right)\right)}{\left\|\nabla_x f\left(x^{i+1}\right)-\nabla_x f\left(x^{i}\right)\right\|_2^2}$
  - How to define "convergence", e.g., $\left\|x^{i+1} - x^{i}\right\|_2 \leq \epsilon$