Emory University

MATH 347 Non Linear Optimization

Learning Notes

Jiuru Lyu

January 22, 2024

Contents

1 Math Preliminaries

1.1 Introduction to Optimization

Definition 1.1 (Optimization Problem). The main optimization problem can be stated as follows

$$\min_{x \in S} f(x), \tag{1}$$

where

- x is the *optimization variable*,
- S is the *feasible set*, and
- f is the *objective function*.

Remark 1.1 $\max_{x \in S} f(x) = -\min_{x \in S} -f(x)$. *Hence, we will only study minimization problems.*

Theorem 1.2 Solving an Optimization Problem

- Theoretical Analysis: analytic solution
- Numerical solution/optimization

Definition 1.3 (Solution Methods depend on the type of x, S, and f).

- When x is continuous (e.g., $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{m \times n}$, $\ldots$), then the optimization problem stated in Eq. (1) is a *continuous optimization problem*. *It will also be the focus of this class.*

  Opposite to continuous optimization problems, we have *discrete optimization problem* if x is discrete.

  If x has both types of components, then we call the problem *mixed*.

- Depending on S, we can have
  - *Unconstrained problems*: where $S = \mathbb{R}^n$, $S = \mathbb{R}^{m \times n}$, $\ldots$ $(m, n$ are fixed).
  - *Constrained problems*: where $S \subsetneq \mathbb{R}^n$, $S \subsetneq \mathbb{R}^{m \times n}$, $\ldots$.

    Both types of problems will be studied.

- Depending on f, we have
  - *Smooth optimization problems*: f has first and/or second order derivatives.

    Only smooth optimization problems will be studied.
  - *Non-smooth optimization problems*: f is not differentiable.

Definition 1.4 (Linear Optimization/Program). If f is linear and S consists of linear constrains, then the optimization problem is called a *linear problem/program.*

Example 1.5 Classification of Optimization Problems

1. Consider the following problem

$$\min_{x_1,x_2,x_3} x_1^2 - 4x_1x_2 + 3x_2x_3 + \sin x_3$$

Solution 1.

- Optimization variable: $x = (x_1, x_2, x_3) \in \mathbb{R}^3$. $\longrightarrow$ continuous.
- Feasible set: $S = \mathbb{R}^3$. $\longrightarrow$ unconstrained.
- Objective function: $f(x_1, x_2, x_3) = x_1^2 - 4x_1x_2 + 3x_2x_3 + \sin x_3$. $\longrightarrow$ smooth but non-linear.

$\square$

2. Consider the following problem

$$\max_{\substack{4x_1+7x_2+3x_3\le 1\\ x_1,x_2,x_3\ge 0}} x_1 + 2x_2 + 3x_3$$

Solution 2.

- Optimization variable: $x = (x_1, x_2, x_3) \in \mathbb{R}^3$. $\longrightarrow$ continuous.
- Feasible set: $S = \{(x_1, x_2, x_3) : x_1, x_2, x_3 \ge 0, 4x_1 + 7x_2 + 3x_3 \le 1\} \subsetneq \mathbb{R}^3$. $\longrightarrow$ constrained.
- Objective function: $f(x_1, x_2, x_3) = x_1 + 2x_2 + 3x_3$. $\longrightarrow$ smooth and linear.

$\square$

Remark 1.2 *This problem can be considered as the budget constrained optimization problem in Economics.*

3. Consider the following problem

$$\min_{x_1,x_2\ge 0} 4x_1 - 3|x_2| + \sin\left(x_1^2 - 2x_2\right)$$

Solution 3.

- Optimization variable: $x = (x_1, x_2) \in \mathbb{R}^2$. $\longrightarrow$ continuous.

- Feasible set: $S = \{(x_1, x_2) : x_1, x_2 \geq 0\} \subsetneq \mathbb{R}62$. $\longrightarrow$ constrained.
- Objective function: $f(x_1, x_2) = 4x_1 - 3|x_2| + \sin(x_1^2 - 2x_2)$. $\longrightarrow$ non-smooth and non-linear.

$\square$

Remark 1.3 *In this particular problem, $x_2 \geq 0$, and so $f(x_1, x_2) = 4x_1 - 3x_2 + \sin(x_1^2 - 2x_2)$ on the feasible set. Hence, this problem can be equivalently written as*

$$\min_{x_1, x_2 \geq 0} 4x_1 - 3x_2 + \sin\left(x_1^2 - 2x_2\right),$$

which is a smooth optimization problem.

1.2 Linear Algebra Review

Example 1.1 Why linear algebra for optimization?

Consider $\min_{x \in \mathbb{R}} f(x)$, where $f(x) = c + bx + ax^2,\ a, b, c \in \mathbb{R}$.

- $a > 0$: $x^* = -\dfrac{b}{2a}$ is a global minimum and $f(x^*) = c - \dfrac{b^2}{4a}$.
- $a < 0$: no minimum exists.
- $a = 0$: $f(x) = c + bx$.
  - $b \neq 0$: no minimum exists.
  - $b = 0$: $f(x) = c$, and every x is a minimum point.

We can approximate any smoothing function using Taylor's approximation and make them simple into the case discussed above.

Theorem 1.2 Taylor's Approximation

$$f(x) = \underbrace{f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2}_{q(x)} + \underbrace{\varepsilon(x - x_0)(x - x_0)^2}_{\text{error}},$$

where $\lim_{x \to x_0} \varepsilon(x - x_0)$.

Remark 1.4 *The hope is that the quadratic approximation will inform us on the behavior of f near x_0 and be useful for instance in referring x_0 on the subject of optimality.*

Definition 1.3 (Quadratic Approximation in Higher Dimensions). When $d > 1$, we consider $\min_{x \in \mathbb{R}^d} f(x)$. Then, the *quadratic approximation* of f is defined as

$$q(x) \coloneqq c + \langle b, x \rangle + \langle x, Ax \rangle,$$

where $c \in \mathbb{R}$, $b \in \mathbb{R}^d$, $A \in \mathbb{R}^{d \times d}$.

Remark 1.5 *Then, to know if a minimum exists, we need information on the matrix A and the vector b.*

Definition 1.4 (Vector, $\mathbb{R}^d$). We define a vector in $\mathbb{R}^d$ as a column vector.

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \in \mathbb{R}^d, \; x_i \in \mathbb{R}.$$

On $\mathbb{R}^d$, we also have the following operations defined

- Addition:
$$\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_d + y_d \end{pmatrix}, \; x_i, y_i \in \mathbb{R}$$

- Scalar multiplication:
$$\alpha \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_d \end{pmatrix}, \alpha, x_i \in \mathbb{R}$$

Definition 1.5 (Basis of $\mathbb{R}^d$). A collection of vectors $v_1 \ldots, v_d \in \mathbb{R}^d$ is a basis in $\mathbb{R}^d$ if $\forall\, x \in \mathbb{R}^d$, $\exists!\, \alpha_1, \ldots, \alpha_d \in \mathbb{R}$ *s.t.* $x = \alpha_1 v_1 + \cdots + \alpha_d v_d$.

Example 1.6 The Standard Basis

The standard basis is defines as

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix},$$

where 1 is at the i-th position for $1 \leq i \leq d$. Note that $\forall x \in \mathbb{R}^d, \; x = x_1 e_1 + \cdots + x_d e_d$.

Notation 1.7.

$$0_d = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Definition 1.8 (Inner Product). $\langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is an inner product if

- (symmetry) $\langle x, y \rangle = \langle y, x \rangle \quad \forall x, y \in \mathbb{R}^d$
- (additivity) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle \quad \forall x, y, z \in \mathbb{R}^d$
- (homogeneity) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle \quad \forall x, y \in \mathbb{R}^d,\ \lambda \in \mathbb{R}$
- (positive definiteness) $\langle x, x \rangle \geq \quad \forall x \in \mathbb{R}^d$ and $\langle x, x \rangle = 0 \iff x = 0$

Example 1.9 Examples of Inner Products

1. **Definition 1.10 (Dot Product).** The *dot product* of $x, y \in \mathbb{R}^d$ is defined as

$$\langle x, y \rangle = x_1 y_1 + \cdots + x_d y_d = \sum_{i=1}^{d} x_i y_i \quad \forall x, y \in \mathbb{R}^d.$$

   It is also referred as the *standard inner product*, and we often use the notation $x \cdot y$ to denote it.

2. **Definition 1.11 (Weighted Dot Product).** The *weighted dot product* of $x, y \in \mathbb{R}^d$ with some weight w is defined as

$$\langle x, y \rangle_w = \sum_{i=1}^{d} w_i x_i y_i,$$

   where $w_1, \ldots, w_d > 0$ are called *weights*.

Remark 1.6 *When $d = 2$, then $\langle x, y \rangle = |x||y| \cos \angle(x, y)$. Dot product measure how correlated are two vectors (with respect to their directions).*

Definition 1.12 (Vector Norm). $\|\cdot\| : \mathbb{R}^d \to \mathbb{R}$ is a norm if

- (non-negativity) $\|x\| \geq 0 \quad \forall x \in \mathbb{R}^d$ and $\|x\| = 0 \iff x = 0$
- (positive homogeneity) $\|\lambda x\| = |\lambda| \|x\| \quad \forall \lambda \in \mathbb{R},\ x \in \mathbb{R}^d$
- (triangular inequality) $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in \mathbb{R}^d$.

Remark 1.7 *Vector norm introduces the notion of length of vectors in $\mathbb{R}^d$.*

Example 1.13 Examples of Vector Norms

- If $\langle \cdot, \cdot \rangle$ is an inner product on $\mathbb{R}^d$, then

$$\|x\| = \sqrt{\langle x, x \rangle} \quad \forall x \in \mathbb{R}^d$$

  is a norm. For instance,

$$\|x\|_2 = \sqrt{x \cdot x} = \left(\sum_{i=1}^{d} x_i^2 \right)^{\frac{1}{2}}.$$

  This norm is called the *standard (Euclidean)* or ℓ_2 norm in $\mathbb{R}^d$.

- **Definition 1.14 (ℓ_p Norms).** Suppose $p \geq 1$, then

$$\|x\|_p := \left(\sum_{i=1}^{d} x_i^p \right)^{\frac{1}{p}}.$$

- **Definition 1.15 (∞-Norms).**

$$\|x\|_\infty := \max_{1 \leq i \leq d} |x_i| \quad \forall x \in \mathbb{R}^d.$$

Remark 1.8 $\lim_{p \to \infty} \|x\|_p = \|x\|_\infty$.

2 Unconstrained Optimization

3 Least Square

4 Constrained Optimization