Optimization Theory and Algorithm II **September 5, 2022**

Lecture 1: Review

Lecturer:Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Xiangyu Chang

1 Review

1.1 What is Optimization?

Optimization is a special field that is built on the three intertwined pillars (footsones):

- **Model**: gives rise to optimization problems.
- **Algorithm**: solves optimization problems.
- **Theory**: supports algorithms and models.

We need to remember that

Optimization = Modeling + Algorithm + Theory.

1.2 General Form of Optimization

Definition 1 *(General Form of Optimization Modeling)*

Suppose that $f : \mathcal{X} \subset \mathbb{R}^n \to \mathbb{R}$ *is a well-defined function. Then*

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

$$s.t.\ \mathbf{x} \in \mathcal{X}, \tag{2}$$

where f *is called as an objective function,* $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top \in \mathcal{X}$ *is a decision variable, and* $\mathcal{X}$ *is the so-called feasible set. For the feasible set* $\mathcal{X}$*, it is commonly denoted as*

$$\mathcal{X} = \{\mathbf{x} : c_i(\mathbf{x}) \leq 0, i = 1, \ldots, l \text{ and } c_j(\mathbf{x}) = 0, j = l+1, \ldots, l+m\},$$

where $c_i(\mathbf{x}) \leq 0, i = 1, \ldots, l$ *are* l *inequality constrains, and* $c_j(\mathbf{x}) = 0, j = l+1, \ldots, l+m$ *are* m *equality constrains.*

Definition 2 *(Global Minimum)*

Point $\mathbf{x}^* \in \mathcal{X}$ *is the global minimum of* (1) *if for any* $\mathbf{x} \in \mathcal{X}$, $f(\mathbf{x}) \geq f(\mathbf{x}^*) = f^*$.

Definition 3 *(Local Minimum)*

Point $\mathbf{x}^* \in \mathcal{X}$ *is a local minimum of* (1) *if there exists a neighborhood of* $\mathbf{x}^*$, $N(\mathbf{x}^*, \epsilon) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}^*\| \leq \epsilon\}$, *such that for any* $\mathbf{x} \in N(\mathbf{x}^*, \epsilon)$, $f(\mathbf{x}) \geq f(\mathbf{x}^*)$.

For an optimization problem, we may have many local minimum points and global minimum points. Draw an example by yourself!

Q: Give us an optimization example you have learnt with the general optimization formulation in Definition 1.

1.3 Modeling in Optimization

Example 1 *(Transportation Problem in the Operational Management)*

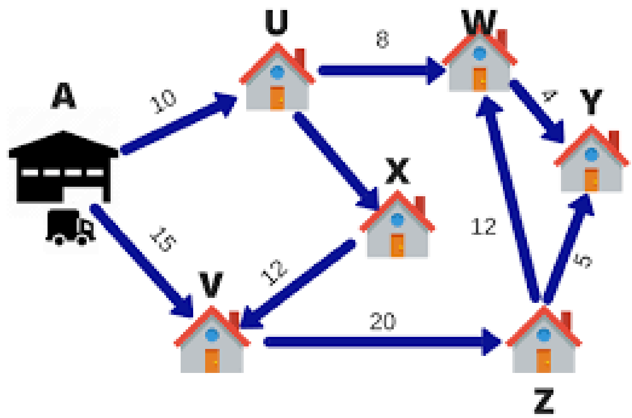

Figure 1: An example of Transportation Problem

Transportation problem (see Figure 1) is a typical problem of operational management where the objective is to minimize the cost of distributing a product form a number of sources or origins to a number of destinations.

Modeling:

- *Origin:* $O_1, O_2, \ldots, O_m$, *and each origin* $O_i, i = 1, \ldots, m$ *has a supply of* a_i *units.*
- *Destination:* $D_1, D_2, \ldots, D_n$, *and each* D_j *has a demand for* $b_j, j = 1, \ldots, n$ *to be delivered from the origins.*
- c_{ij} *is the cost per unit distributed from the origin* O_i *to the destination* D_j.
- *Aim: Finding a set of* x_{ij}*'s* $i = 1, \ldots, m; j = 1, \ldots, n$ *to meet supply and demand requirements at a minimum distribution cost.*

Optimization Formulation:

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} c_{ij}, \tag{3}$$

$$s.t.\ x_{ij} \geq 0, i = 1, \ldots, m; j = 1, \ldots, n, \tag{4}$$

$$\sum_{j=1}^{n} x_{ij} = a_i, \tag{5}$$

$$\sum_{i=1}^{m} x_{ij} = b_j, \tag{6}$$

where (4) are the inequality constrains and (5) and (6) are the equality constrains.

Q: Is this the general form of optimization (1)?

Q: Why called it as a *Linear Program*?

From the managerial perspective, optimization is also a general quantitive decision-making problem which focus on how to distribute and control the limited resource for achieving the optimal value.

Example 2 *(Portfolio Management)*

Portfolio Management (see Figure 2) is the art and science of making decisions about investment mix and policy, matching investments to objectives and balancing risk against performance.

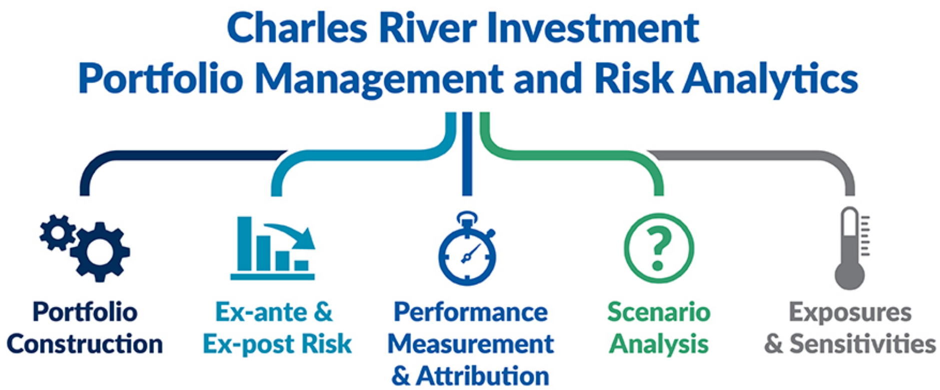

Figure 2: An example of Portfolio Management

Modeling:

- *n assets or stocks that are hold over a period of time.*
- *x_i denotes the amount of asset i, the final period.*
- *original price p_{i0} for asset i, the final price p_{it} at time t, then the return on asset i is $r_i = \frac{p_{it}-p_{i0}}{p_{i0}}$.*
- *the overall return is $R = \sum_{i=1}^{n} r_i x_i$.*
- *Suppose that $\mathbf{r} = (r_1, \ldots, r_n)^\top$ is a random vector with expectation of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, and covariance of Σ.*
- *Aim: Finding a set of asset $\mathbf{x} = (x_1, \ldots, x_n)^\top$ to maximize the expected overall return and balancing the risk perform.*

Optimization Formulation:

$$\max_{\mathbf{x}} \ \mathbb{E}(R) - \lambda Var(R), \tag{7}$$
$$s.t. \ x_i \geq 0, i = 1 \ldots, n, \tag{8}$$
$$\sum_{i=1}^{n} x_i = 1, \tag{9}$$

where $\mathbb{E}(R)$ is the expectation of R, $Var(R)$ is the variance of R and $\lambda > 0$ is called risk aversion parameter for balancing the investment risk and expected return. Finally, we have that

$$\max_{\mathbf{x}} \ \boldsymbol{\mu}^\top \mathbf{x} - \lambda \mathbf{x}^\top \Sigma \mathbf{x}, \tag{10}$$
$$s.t. \ x_i \geq 0, i = 1 \ldots, n, \tag{11}$$
$$\sum_{i=1}^{n} x_i = 1, \tag{12}$$

Q: How to compute $\mathbb{E}(R)$ and $Var(R)$?

Q: Why not $x_i < 0$?

Remark 1 *This example is significantly important. Because*

- *This is called a* **nonlinear program** *due to the nonlinear objective function.*
- *It is also called a***quadratic program**. *Why???*
- *Harry Markowiz proposed this model called* **Modern Portfolio Theory or Mean-Variance Analysis** *and obtain the* **Nobel Prize in 1990**.

Example 3 *Generalized Linear Model (GLM). Let us consider the following three management problems.*

- $b = $ *House Price* $= F(a_1 = $ *number of rooms*, $a_2 = $ *school distriction*, $a_3, \dots)$
- $b = $ *Credit Rate* $= F(a_1 = $ *education*, $a_2 = $ *salary*, $a_3, \dots)$
- $b = $ *Number of Visit this month* $= F(a_1 = $ *number of visit last month*, $a_2 = RFM, a_3, \dots)$

In this example, we introduced three classic regression models, linear regression(house price), Poisson regression (number of visit this month) and logistic regression (credit rate) derived from GLM. We parameterized the parameters in the statistic models as a linear function of covariant variables $\mathbf{a}$, *and formed the optimization problem from the likelihood.*

Consider the input-output pairs $\{\mathbf{a}_i, b_i\}_{i=1}^m$ *as the data. The procedure can be summarized as following recipe,*

1. *write down a probabilistic model for* b_i
2. *link model parameter* $\mathbf{x}$ *with* $\mathbf{a}_i$
3. *formed the optimization problem using maximum likelihood that aim to discover* $\mathbf{x}$ *with all data* $\{\mathbf{a}_i, b_i\}_{i=1}^m$

Next we instantiate this recipe by three examples.

(i) *Linear Regression: Given training data* $\{\mathbf{a}_i, b_i\}_{i=1}^m$ *with* $\mathbf{a}_i \in \mathbb{R}^p$ *and* $b_i \in \mathbb{R}$. *Suppose each* $b_i \overset{i.i.d.}{\sim} N(\mu_i, \sigma^2)$, *that is*

$$\begin{aligned} P(b_i|\mu_i, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(b_i - \mu_i)^2}{2\sigma^2}\} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{b_i^2}{2\sigma^2}\} \exp\{-\frac{\frac{1}{2}\mu_i^2 - b_i\mu_i}{\sigma^2}\}. \end{aligned}$$

It is convention to choose the parameters that multiply b_i *as the linear function of the variables* $\mathbf{a}_i$ *with the parametric coefficient* $\mathbf{x}$. *Here we make the assumption that*

$$\theta_i = \mu_i = \langle \mathbf{a}_i, \mathbf{x} \rangle.$$

We wish to examine how we find a good $\mathbf{x}$ *to make this work. Our strategy for this is to maximize the* likelihood *of all observations* $\{b_i\}$ *as a function of* $\mathbf{x}$, *i.e.*

$$\max_{\mathbf{x}} \prod_i \exp\{-\frac{1}{\sigma^2}(\frac{1}{2}\mu_i^2 - b_i\mu_i)\} \Rightarrow \max_{\mathbf{x}} \log \prod_i \exp\{-\frac{1}{2\sigma^2}(\langle \mathbf{a}_i, x\rangle^2 - b_i\langle \mathbf{a}_i, \mathbf{x}\rangle)\}.$$

To maximize this expression, we take the negative log of the expression, i.e. we want to minimize

$$\min_{\mathbf{x}} \ \frac{1}{\sigma^2} \sum_{i=1}^{n} (\frac{1}{2} \langle \mathbf{a}_i, \mathbf{x} \rangle^2 - b_i \langle \mathbf{a}_i, \mathbf{x} \rangle).$$

To write it more compactly, we denote,

$$A = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

And we have,

$$\sum_{i=1}^{m} \langle \mathbf{a}_i, x \rangle^2 = \|A\mathbf{x}\|^2, \qquad \sum_{i=1}^{m} b_i \langle \mathbf{a}_i, \mathbf{x} \rangle = \langle \mathbf{b}, A\mathbf{x} \rangle,$$

we get the minimization problem

$$\arg\min_{x} \ \frac{1}{2} \|A\mathbf{x}\|^2 - \langle b, A\mathbf{x} \rangle = \arg\min_{x} \ \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|^2$$

which is a linear least-squares regression problem.

References