Optimization Methods		Date:	*2 March 2018*
Instructor: *Sujit Prakash Gujar*		Scribes:	Ashutosh Mishra Arun H Garimella Nikhil Vemuri

Lecture 15: Non Linear Programming

1 Recap

In the previous lectures, we saw integer programming and we noticed that feasible points in integer programming questions are scattered along and around the boundaries. Hence, the search space is exponential to find the best optimal.

We saw the LP relaxed objective value due to total unimodular constraint of matrix A. We also saw that TSP is not solved in polynomial time when LP relaxed using ellipsoid, so we search for another separation oracle which is nothing but solving MINCUT problem. This separation oracle gives polynomial time solution for TSP.

We discussed about LP and also saw that for two player zero sum games, we showed Nash Equilibrium exists using LP duality. We also showed nash equilibrium can be computed in polynomial time using LP for this specific class of games which is two player-zero sum games.

2 Introduction

We can say that our problem is non linear if:

- If the constraints of the problem are not linear.
- If the objective function is not linear.

If either of the conditions exist, then our optimization problem becomes non linear.
We have several examples of such unconstrained optimization, some of them are listed as follows:

1) Fermat-Weber Problem: Suppose there are m facilities i.e $(f_1, f_2, f_3....f_m)$. The goal is to find a distribution center x whose distance from all the m facilities is minimized. The figure depicts the mathematical formulation. 1(a). Mathematically, the problem can be stated as:

$$\min_{x \in \mathcal{R}^n} \sum_{i=1}^{m} ||x - f_i|| \tag{1}$$

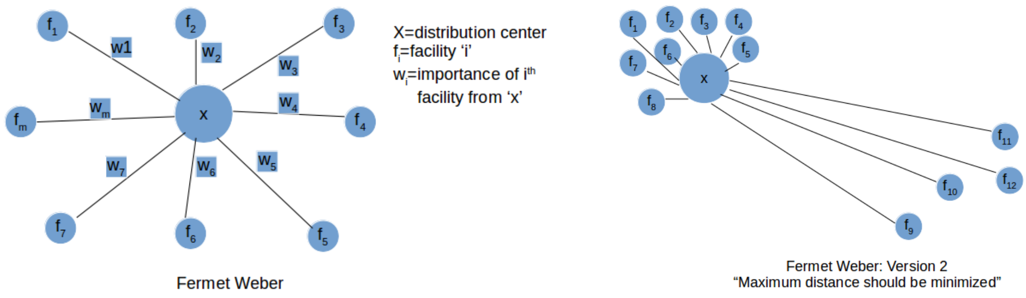
Some facilities can have higher importance than others, so the mathematical equation (1) can be reformulated as:

$$\min_{x \in \mathcal{R}^n} \sum_{i=1}^{m} w_i ||x - f_i|| \tag{2}$$

where, w_i is the weight assigned to each facility i based on some importance criteria.

NOTE: If the objective function has L1 norm in it's formulation, then objective function is still linear but if the objective function has L2 norm in it's formulation, the problem becomes unconstrained.

2) Fermat Weber Another Version: It may happen that the distribution of the facilities is uneven. So, the distribution center will be placed in the region where the facilities are more clustered together, but since our objective is to serve all of the facilities equally. Hence in such case, we minimize



(a) Fermat Weber (1) (b) Fermat Weber (2)

Figure 1: Fermat Weber cases

the maximum distance to any of the m facilities from distribution center x (distribution center) as depicted in 1(b).

$$\min_{x \in \mathcal{R}^n} f(x) = max(||x - f_1||, ||x - f_2||, ...||x - f_m|| \tag{3}$$

The same problem (3) can also be modelled as a constrained non linear program, with the following objective function:

$$\min_{x,\delta} \delta \tag{4}$$

$$\text{subject to} ||x - f_i|| \leq \delta \tag{5}$$

(4) is easy to solve for, hence we pose the problem like this.
3) Portfolio Management System: The problem statement states that returns obtained by investment in some security should be greater than a certain threshold. Additionally, we want to minimize the overall varaince of our distribution. Hence, our objective function becomes:

$$\text{minimize} \sum_{i=1}^{k} x_i^2 Var(p_i) \tag{6}$$

$$\text{subject to} \sum_{i=1}^{k} p_i x_i \geq T, \tag{7}$$

$$\sum_{i=1}^{k} x_i \leq B, \tag{8}$$

$$x_i \geq 0, \quad i = 1, \ldots, k. \tag{9}$$

where, p_i and x_i is the reward and amount of securities for i^{th} stock. T is the minimum expected reward and B is total amount which has to be invested.

4) Other examples: SVM, Deep Learning problems also have non linear optimization objective function which is solved using various other techniques depending on the kind of the problem.

3 Non Linear Optimization for Unconstrained Problem

This lecture focused on dealing with unconstrained non linear optimization problem like:

$$\min_{x \in \mathbb{R}^N} f(x) \tag{10}$$

where, f: $\mathbb{R} \to \mathbb{R}$

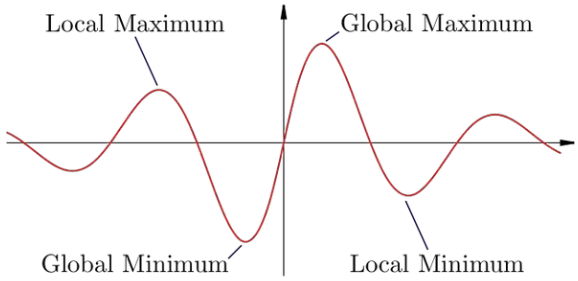


Fig. 2: Local/Global Maxima/Minima

1) Local Minimum: A vector $x^* \in \mathbb{R}^n$ is an unconstrained local minimum of f if $\exists\ \epsilon > 0$ such that $f(x^*) \leq f(x)\ \forall x \in ||x - x^*|| < \epsilon$.

2) Strict Local Minimum: A vector $x^* \in \mathbb{R}^n$ is an unconstrained strictly local minimum of f if $\exists\ \epsilon > 0$ such that $f(x^*)<f(x)\ \forall x \in ||x - x^*|| < \epsilon$ and $\forall x \neq x*$.

3) Global Minimum: A vector $x^* \in \mathbb{R}^n$ is an unconstrained global minimum of f if $f(x^*) \leq f(x)\ \forall x \in \mathbb{R}^n$.

4) Local Maximum: A vector $x^* \in \mathbb{R}^n$ is an unconstrained local maximum of f if $\exists\ \epsilon > 0$ such that $f(x^*) \geq f(x)\ \forall x \in ||x - x^*|| < \epsilon$.

5) Strict Local Maximum: A vector $x^* \in \mathbb{R}^n$ is an unconstrained strictly local maximum of f if $\exists\ \epsilon > 0$ such that $f(x^*)>f(x)\ \forall x \in ||x - x^*|| < \epsilon$ and $\forall x \neq x*$.

6) Global Maximum: A vector $x^* \in \mathbb{R}^n$ is an unconstrained global maximum of f if $f(x^*) \geq f(x)\ \forall x \in \mathbb{R}^n$.

We can observe different different local maxima and minima in the Fig. (2), but one global maximia and minima. Global maximia or minima can be local maxima minima as well but not vice-a-versa.

4 Optimiality Conditions

4.1 First order necessary condition

We consider a continuous and a differentiable function $f(x)$ such that $f : \mathbb{R}^n \rightarrow \mathbb{R}$. At any point, say x_0 and it's immediate neighbor x, we define the variation from x_0 to x as the first order variation stated as:

$$f(x) \approx f(x_0) + \delta^T \bigtriangledown f(x_0) \tag{11}$$

where,

$$\bigtriangledown f(x_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathrm{x}_0) \\ \frac{\partial f}{\partial x_2}(\mathrm{x}_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathrm{x}_0) \end{bmatrix} \tag{12}$$

Now if x^* is locally minima, then

$$\delta^T \bigtriangledown f(x_0) >= 0 \tag{13}$$

But if we chose, $\delta = (\epsilon, 0, 0, 0, 0, ..0)$, where ϵ is assumed to be small then,

$$\epsilon \frac{\partial f}{\partial x_i}(x^*) \geq 0 \forall \quad i = 1, \dots, n \tag{14}$$

ϵ could also take negative value also since there is no bound on it, therefore,

$$-\epsilon \frac{\partial f}{\partial x_i}(x^*) \geq 0 \forall \quad i = 1, \dots, n \tag{15}$$

If both (14) and (15) are satisfied then,

$$\bigtriangledown f(x^*) = 0 \tag{16}$$

Proposition: First order necessary condition
Let x^* be unconstrained local minimum of f: $\mathbb{R}^n \to \mathbb{R}$ which is continuous and differentiable, then

$$\bigtriangledown f(x^*) = 0 \tag{17}$$

Proof:

- Let us consider $d \in \mathbb{R}^n$ and $\alpha \geq 0$
- Let us define $g(\alpha) = f(x^* + \alpha * d) - f(x^*)$. Here, $g : \mathbb{R} \to \mathbb{R}$
- We define $\lim_{\alpha \to 0} \frac{g(f(x^*+\alpha d) - f(x^*))}{\alpha} \geq 0$
- Now, $\lim_{\alpha \to 0} \frac{g(\alpha)}{\alpha} = g'(\alpha)$. Also, $g'(\alpha) \geq 0$ because $f(x^* + \alpha d) - f(x^*) \geq 0$
- Hence, $g'(0) = d^T \bigtriangledown f(x^* + \alpha d) \geq 0$ if limit $\alpha = 0$
- Since there is no restriction on d, we can equivalently write as: $g'(0) = -d^T \bigtriangledown f(x^* + \alpha d) \geq 0$
- Therefore, $\bigtriangledown f(x^*) = 0$

But only knowing the value of gradient at x^* is not sufficient, to conclude that x^* is the minima we are looking for. Hence we go for second order condition.

4.2 Second order necessary condition

If the function f is twice differentiable, and if x^* is local minima, then with the help of Taylor series, we can write it as:

$$\bigtriangledown f(x^*)^T \Delta x + \frac{(\Delta x)^T \bigtriangledown^2 f(x^*) \Delta x}{2} \geq 0 \tag{18}$$

where, $\bigtriangledown^2 f(x^*)$ is a Hessian matrix. If we prove that Hessian is positive semi-definite then x^* will be local minimum. In other words,

$$\det \bigtriangledown^2 f(x^*) \geq 0 \tag{19}$$

Preposition: Second Order Necessary Condition

- Assume that $g(\alpha) = f(x^* + \alpha d) - f(x^*)$ as defined in the above preposition.
- Expanding $g(\alpha)$ using Taylor series, we get:

$$f(x^* + \alpha d) - f(x^*) = \alpha \bigtriangledown f(x^*)^T d + \frac{\alpha^2}{2} d^T \bigtriangledown^2 f(x^*) + o(\alpha^2) \tag{20}$$

- From above proposition, we know that $\lim_{\alpha \to 0} \frac{g(\alpha)}{\alpha} = g'(\alpha)$. Also, $g'(\alpha) \geq 0$ because $f(x^* + \alpha d) - f(x^*) \geq 0$

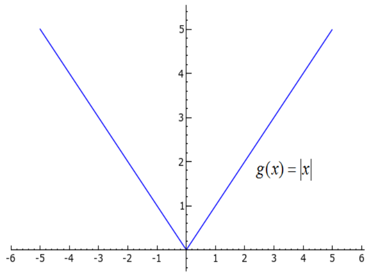


Fig. 3: mod x function

- Hence, we get

$$\frac{\alpha \bigtriangledown f(x^*)^T + \frac{\alpha^2}{2} d^T \bigtriangledown^2 f(x^*) + o(\alpha^2)}{2} \geq 0 \tag{21}$$

- As we know that $\bigtriangledown f(x^*) = 0$ and $\lim_{\alpha \to 0} \frac{o(\alpha^2)}{\alpha^2} = 0$
- Hence, $\frac{\alpha^2}{2} d^T \bigtriangledown^2 f(x^*) \geq 0$ and therefore, $\bigtriangledown^2 f(x^*)$ should be PSD(positive semi-definite) to hold the above statement.

5 Conclusion

So, we saw that if a function is first order differentiable, then $\bigtriangledown f(x^*) = 0$, but that does not mean that if gradient at $x^* = 0$, then first order condition is satisfied. Consider, $|x|$ as function as depicted in Fig (3). We find, global minimum to be at $x = 0$ but since it is not differentiable, first order condition is not satisfied.