Optimization Theory and Applications

Kun Zhu (zhukun@nuaa.edu.cn)

January 2, 2019

General Constrained Problems

- So far we have considered only problems with equality constraints: $\mathbf{h}(\mathbf{x}) = \mathbf{0}$
- We now consider problems that have inequality constraints: $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, where $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^p$
- As before, we give necessary conditions for problems with equality and inequality constraints

Special Case: Only Inequality Constraints

- Consider the problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}, \end{array}$$

  where $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), \ldots, g_p(\mathbf{x})]^T$

- As usual, we assume f, $\mathbf{g} \in \mathcal{C}^1$
- A point $\mathbf{x}$ is feasible if $g_1(\mathbf{x}) \leq 0, \ldots, g_p(\mathbf{x}) \leq 0$

Special Case: Only Inequality Constraints

- ***Definition***: We say that the jth constraint $g_j \leq 0$ is *active* at $\mathbf{x}^*$ if $g_j(\mathbf{x}^*) = 0$. It is *inactive* if $g_j(\mathbf{x}^*) < 0$
- Note that if a constraint is inactive at $\mathbf{x}^*$, then it is inactive at all points in some neighborhood of $\mathbf{x}^*$. Hence, locally around $\mathbf{x}^*$, the inactive constraints can be "ignored"
- Define $J(\mathbf{x}^*) = \{j : g_j(\mathbf{x}^*) = 0\}$, the set of indices of constraints that are active
- ***Definition***: A feasible point $\mathbf{x}^*$ is *regular* if the vectors $\nabla g_j(\mathbf{x}^*), j \in \mathcal{J}(\mathbf{x}^*)$, are linearly independent

Special Case: Only Inequality Constraints

- Let $\mathbf{x}^*$ be a local minimizer of the original problem (with inequality constraint) and regular
- Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & g_j(\boldsymbol{x}) = 0, \ j \in J(\boldsymbol{x}^*) \end{array}$$

- Note that $\mathbf{x}^*$ is also a local minimizer for the above problem
- Therefore, the Lagrange conditions hold at $\mathbf{x}^*$ for the above problem

Special Case: Only Inequality Constraints

- Hence, by the Lagrange Theorem, there exists μ_j^*, $j \in \mathcal{J}(\mathbf{x}^*)$, such that

$$Df(\mathbf{x}^*) + \sum_{j \in \mathcal{J}(\mathbf{x}^*)} \mu_j^* Dg_j(\mathbf{x}^*) = \mathbf{0}^T$$

- Let us define $\mu_j^* = 0$ for $j \notin \mathcal{J}(\mathbf{x}^*)$ (i.e., all inactive j)

- Then, we can write the above condition as

$$Df(\mathbf{x}^*) + \mu^{*T} D\mathbf{g}(\mathbf{x}^*) = \mathbf{0}^T$$

where $\mu^* = [\mu_1^*, \ldots, \mu_p^*]^T$

Special Case: Only Inequality Constraints

- Note that

$$\mu^{*T}\mathbf{g}(\mathbf{x}^*) = 0$$

  since for each j, either $g_i(\mathbf{x}^*) = 0$ (active j) or $\mu_j^* = 0$ (inactive j)

- In other words, for all $j \notin \mathcal{J}(\mathbf{x}^*)$ (inactive), we have $\mu_j^* = 0$
- It turns out that we can say more about μ^*: every component of it is ≥ 0
- To see this, we only need to concentrate on those $j \in \mathcal{J}(\mathbf{x}^*)$, since the other μ_j^* are 0

Special Case: Only Inequality Constraints

- We can illustrate the above fact using a picture

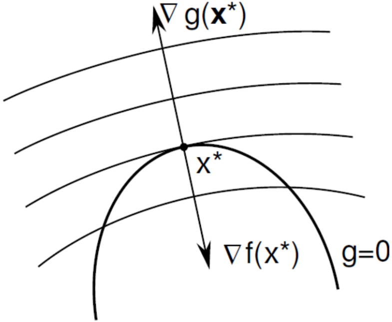
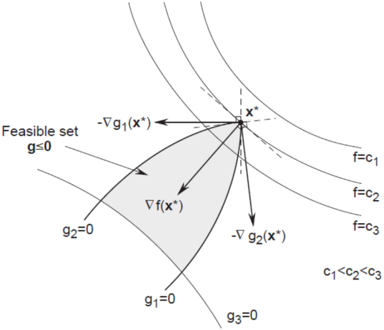


Illustration of the KKT Theorem

- An example with multiple inequality constraints



- The constraint $g_3(\mathbf{x}) \leq 0$ is inactive
- By the KKT theorem, we have

$$\nabla f(\mathbf{x}^*) + \mu_1^* \nabla g_1(\mathbf{x}^*) + \mu_2^* \nabla g_2(\mathbf{x}^*) = \mathbf{0}$$

Illustration of the KKT Theorem

- We can see from the figure that $\nabla f(\mathbf{x}^*)$ must be a linear combination of the vectors $-\nabla g_1(\mathbf{x}^*)$ and $-\nabla g_2(\mathbf{x}^*)$ with positive coefficients

- This corresponds to

$$\nabla f(\mathbf{x}^*) = -\sum_{j \in \mathcal{J}(\mathbf{x}^*)} \mu_j^* \nabla g_j(\mathbf{x}^*)$$

where $\mu_j^* \geq 0$

Summary of The Special Case: KKT Theorem

- Consider the problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}, \end{array}$$

- ***Karush-Kuhn-Tucker (KKT) Theorem*** (for the special case): Suppose $\mathbf{x}^*$ is a local minimizer and is regular. Then, there exists $\mu^* \in \mathbb{R}^p$ such that
  - Dual feasibility
  $$\mu^* \geq 0$$
  - Optimality condition
  $$Df(\mathbf{x}^*) + \mu^{*T} D\mathbf{g}(\mathbf{x}^*) = \mathbf{0}^T$$
  - Complementary slackness condition
  $$\mu^{*T}\mathbf{g}(\mathbf{x}^*) = 0$$
  - Primal feasibility condition
  $$\mathbf{g}(\mathbf{x}^*) \leq 0$$

Summary of Special Case: Only Inequality Constraints

- The conditions are called ***KKT conditions*** (note that we usually include the constraints as part of the KKT conditions)
- The vector μ^* is called the *KKT multiplier vector*
- Note that for feasible $\mathbf{x}^*$ and μ^*

$$\mu^{*T}\mathbf{g}(\mathbf{x}^*) = 0 \iff \mu_i^* g_i(\mathbf{x}^*) = 0 \text{ for all } i = 1, \ldots, p$$

- Actually, there is a more general version of the theorem, where we have both equality and inequality constraints (see later)

Example of the Special Case

- Consider the problem

$$\begin{aligned} \text{minimize} \quad & x_1^2 + x_2^2 + x_1 x_2 - 3x_1 \\ \text{subject to} \quad & x_1, x_2 \geq 0, \end{aligned}$$

- $f(\mathbf{x}) = x_1^2 + x_2^2 + x_1 x_2 - 3x_1$
- $g_1(\mathbf{x}) = -x_1$, $g_2(\mathbf{x}) = -x_2$
- The KKT conditions for the problem are

1. $\boldsymbol{\mu} = [\mu_1, \mu_2]^T \geq 0$;
2. $Df(\boldsymbol{x}) - \boldsymbol{\mu}^T = \mathbf{0}^T$;
3. $\boldsymbol{\mu}^T \boldsymbol{x} = 0$.
4. $\boldsymbol{x} \geq \mathbf{0}$.

- Try it

Example of the Special Case

- We have

$$Df(\mathbf{x}) = [2x_1 + x_2 - 3, x_1 + 2x_2]$$

- This gives

$$\begin{aligned} 2x_1 + x_2 - \mu_1 &= 3 \\ x_1 + 2x_2 - \mu_2 &= 0 \\ \mu_1 x_1 + \mu_2 x_2 &= 0 \\ \mu_1, \mu_2, x_1, x_2 &\geq 0. \end{aligned}$$

- We now have four variables, three equations, and the inequality constraints on each variable
- Try to solve it (hint: consider points making the constraints active)

Example of the Special Case

- To find a solution for $\mathbf{x}^*$, μ^*, we first try $x_1^* = 0$ and $x_2^* = 0$, and notice that it is impossible. Why?
- Then we try $x_1^* = 0$
  - By the first equation, we must have $x_2^* > 0$. Thus, $\mu_2^* = 0$
  - Solving the equations we obtain

    $$x_2^* = 0, \quad \mu_1^* = -3$$

    which is not valid

Example of the Special Case

- Next, we try $x_2^* = 0$, which then implies $\mu_1^* = 0$
  - Solving the equations, we obtain

    $$x_1^* = \frac{3}{2} \quad \mu_2^* = \frac{3}{2}$$

    which is a valid solution to the KKT conditions
- What about the case $x_1^* > 0$ and $x_2^* > 0$?
- Note that, to solve conditions that have inequalities, we have to try solutions that are at the boundary (active constraints)

Variants of KKT Conditions

- We can easily modify the KKT conditions to problems with maximization or inequality constraints of the form $\mathbf{g}(\mathbf{x}) \geq 0$
- In the case of maximization, either we change the sign f, or we can change the sign of μ^*
- Similarly, in the case of constraints of the form $\mathbf{g}(\mathbf{x}) \geq 0$, either we change the sign $\mathbf{g}$, or we can change the sign of μ^*

Variants of KKT Conditions

- Specifically, consider the problem

$$\begin{array}{ll} \text{maximize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}. \end{array}$$

- The KKT conditions for the above problem are

$$\begin{array}{l} 1.\ \boldsymbol{\mu}^* \leq \mathbf{0} \\ 2.\ Df(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*T} D\boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0}^T \\ 3.\ \boldsymbol{\mu}^{*T}\boldsymbol{g}(\boldsymbol{x}^*) = 0 \\ 4.\ \boldsymbol{g}(\boldsymbol{x}^*) \leq \mathbf{0} \end{array}$$

- The only difference is the sign of μ^*

Variants of KKT Conditions

- Similarly, for the problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0}, \end{array}$$

- The KKT conditions for the above problem are

1. $\boldsymbol{\mu}^* \leq \boldsymbol{0}$
2. $Df(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*T} D\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}^T$
3. $\boldsymbol{\mu}^{*T}\boldsymbol{g}(\boldsymbol{x}^*) = 0$
4. $\boldsymbol{g}(\boldsymbol{x}^*) \geq \boldsymbol{0}$

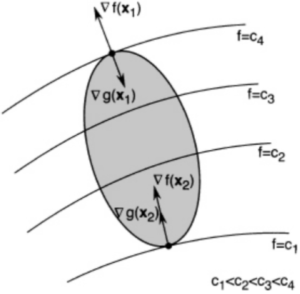
- Question: what if we have both maximization and $\mathbf{g}(\mathbf{x}) \geq 0$?

Variants of KKT Conditions

- If we have both maximization and $\mathbf{g}(\mathbf{x}) \geq 0$, then the KKT conditions are the same as the original (standard) case (except for the constraint).

Variants of KKT Conditions

- Example: In the following figure, the two points $\mathbf{x}_1$ and $\mathbf{x}_2$ are feasible points, that is $g(\mathbf{x}_1) \geq 0$ and $g(\mathbf{x}_2) \geq 0$, and they satisfy the KKT condition



x_1 is a maximizer and x_2 is a minimizer

Variants of KKT Conditions

- The point $\mathbf{x}_1$ is a maximizer. The KKT condition for this point is

  1. $\mu_1 \geq 0$.
  2. $\nabla f(x_1) + \mu_1 \nabla g(x_1) = \mathbf{0}$.
  3. $\mu_1 g(x_1) = 0$.
  4. $g(x_1) \geq 0$.

- The point $\mathbf{x}_2$ is a minimizer. The KKT condition for this point is

  1. $\mu_2 \leq 0$.
  2. $\nabla f(x_2) + \mu_2 \nabla g(x_2) = \mathbf{0}$.
  3. $\mu_2 g(x_2) = 0$.
  4. $g(x_2) \geq 0$.

Problems with Equality and Inequality Constraints

- Consider the optimization problem

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \\ & \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}, \end{aligned}$$

  where $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$, and $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^p$

- Our goal is to derive necessary conditions for the above general problem

Problems with Equality and Inequality Constraints

- Definition: a feasible point $\mathbf{x}^*$ is regular if the vectors

$$\nabla h_i(\mathbf{x}^*), i = 1, \ldots, m, \nabla g_j(\mathbf{x}^*), j \in \mathcal{J}(\mathbf{x}^*)$$

  are linearly independent
- By convention we consider every equality constraint $h_i = 0$ to be active
- Hence, regularity means the gradients of all active constraint functions are linearly independent

Problems with Equality and Inequality Constraints

- ***Theorem***: suppose $\mathbf{x}^*$ is a local minimizer and is regular. Then, there exists $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that

1. $\boldsymbol{\mu}^* \geq \mathbf{0}$
2. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*T} D\boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0}^T$
3. $\boldsymbol{\mu}^{*T}\boldsymbol{g}(\boldsymbol{x}^*) = 0$
4. $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$
5. $\boldsymbol{g}(\boldsymbol{x}^*) \leq \mathbf{0}$.

- The difference between the above KKT conditions and the previous one (with no equality constraints) is that we need to incorporate the Lagrange multiplier vector λ^*

Problems with Equality and Inequality Constraints

- The idea behind the proof of the general KKT theorem is the same as what we have seen for the special case with no equality constraints
- Basically, the proof involves applying the Lagrange theorem to the associated problem with only equality constraints involving active constraints at $\mathbf{x}^*$

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0} \\ & g_j(\boldsymbol{x}) = 0, \ j \in J(\boldsymbol{x}^*), \end{aligned}$$

and, as before, we have $\mu^* \geq 0$ and $\mu^{*T}\mathbf{g}(\mathbf{x}^*) = 0$

Example: Savings in Bank

- Bank interest paid monthly at rate $r > 0$
- We wish to deposit some money into the bank every month for n months, such that the total is D dollars
- Goal: maximize the total amount of money accumulated at the end of n months
- Let x_i be amount deposited in beginning of ith month
- Optimization problem:

$$\begin{array}{ll} \text{maximize} & (1+r)^n x_1 + (1+r)^{n-1} x_2 + \cdots + (1+r) x_n \\ \text{subject to} & x_1 + \cdots + x_n = D \\ & x_1, \ldots, x_n \geq 0 \end{array}$$

Example: Savings in Bank

- Write

$$
\begin{aligned}
f(\boldsymbol{x}) &= -\left((1+r)^n x_1 + (1+r)^{n-1} x_2 + \cdots + (1+r) x_n\right) \\
h(\boldsymbol{x}) &= x_1 + \cdots + x_n - D \\
\boldsymbol{g}(\boldsymbol{x}) &= -[x_1, \ldots, x_n]^T = -\boldsymbol{x}.
\end{aligned}
$$

- We have

$$
\begin{aligned}
Df(\boldsymbol{x}) &= -[(1+r)^n, (1+r)^{n-1}, \ldots, (1+r)] \\
Dh(\boldsymbol{x}) &= [1, 1, \ldots, 1] \\
D\boldsymbol{g}(\boldsymbol{x}) &= -\boldsymbol{I}_n.
\end{aligned}
$$

Example: Savings in Bank

- The KKT conditions are

$$\begin{aligned}
\mu_1, \ldots, \mu_n &\geq 0 \\
-(1+r)^{n-i+1} + \lambda - \mu_i &= 0,\ i = 1, \ldots, n \\
\mu_1 x_1 + \cdots + \mu_n x_n &= 0 \\
x_1 + \cdots + x_n &= D \\
x_1, \ldots, x_n &\geq 0.
\end{aligned}$$

- Suppose that $x_1^* > 0$. Then $\mu_1^* = 0$, and so we have

$$\begin{aligned}
\lambda^* &= (1+r)^n, \\
\mu_i^* &= (1+r)^n - (1+r)^{n-i+1} > 0,\ i = 2, \ldots, n, \\
x_1^* &= D,\ x_i^* = 0,\ i = 2, \ldots, n.
\end{aligned}$$

Example: Savings in Bank

- The previous solution corresponds to depositing D dollars in the first month
- Are there any other solutions?
- Suppose $x_i^* > 0, i \neq 1$ (hence $\mu_i^* = 0$). We then conclude that

$$\begin{aligned} \lambda^* &= (1+r)^{n-i+1}, \\ \mu_{i-1}^* &= (1+r)^{n-i+1} - (1+r)^{n-i+2} < 0, \end{aligned}$$

  which is clearly not valid
- Hence, there are no other solutions

Example: Power Control for Capacity Optimization

- We consider the optimization problem

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1, \end{array}$$

  where $\alpha_i > 0$

- This problem arises in information theory, in allocating power to a set of n communication channels

- The variable x_i represents the transmit power allocated to the ith channel

- $\log(\alpha_i + x_i)$ gives the capacity or communication rate of the channel

- So the problem is to allocate a total power of one to the channels, in order to maximize the total communication rate

Example: Power Control for Capacity Optimization

- Introducing Lagrange multipliers $\lambda^* \in \mathbb{R}^n$ for the inequality constraints $\mathbf{x}^* \geq 0$, and a multiplier $v^* \in \mathbb{R}$ for the equality constraint $\mathbf{1}^T\mathbf{x} = 1$
- We obtain the KKT conditions
  - Optimality condition
    $$-\frac{1}{(\alpha_i + x_i^*)} - \lambda_i^* + v^* = 0, \quad i = 1, \ldots, n$$
  - Primal feasibility conditions
    $$\mathbf{x}^* \geq 0$$
    $$\mathbf{1}^T\mathbf{x}^* = 1$$
  - Complementary slackness conditions
    $$\lambda_i^* x_i^* = 0, \; i = 1, \ldots, n$$
    $$\lambda^* \geq 0$$

Example: Power Control for Capacity Optimization

- We start by noting that λ^* acts as a slack variable, the KKT conditions can be restated as

$$\mathbf{x}^* \geq 0$$

$$\mathbf{1}^T\mathbf{x}^* = 1$$

$$x_i^*(v^* - \tfrac{1}{\alpha_i + x_i^*}) = 0, \quad i = 1, \ldots, n$$

$$v^* \geq \tfrac{1}{\alpha_i + x_i^*}, \quad i = 1, \ldots, n$$

Example: Power Control for Capacity Optimization

- If $v^* < \frac{1}{\alpha_i}$, the last condition can only hold if $\mathbf{x}^* > 0$, which by the third condition implies that

$$v^* = \frac{1}{\alpha_i + x_i^*}$$

- Solving for x_i^*, we obtain that

$$x_i^* = \frac{1}{v^*} - \alpha_i$$

Example: Power Control for Capacity Optimization

- If $\nu^* \geq \frac{1}{\alpha_i}$, then $x_i^* > 0$ is impossible, because it would imply

  $$\nu^* \geq \frac{1}{\alpha_i} > \frac{1}{\alpha_i + x_i^*}$$

  which violates the complementary slackness condition

- Therefore, $x_i^* = 0$ if $\nu^* \geq \frac{1}{\alpha_i}$. Thus we have

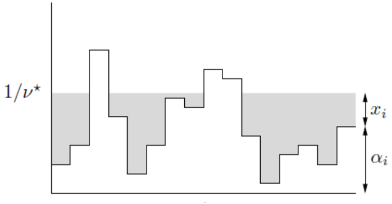
$$x_i^\star = \begin{cases} 1/\nu^\star - \alpha_i & \nu^\star < 1/\alpha_i \\ 0 & \nu^\star \geq 1/\alpha_i, \end{cases}$$

- More simply, $x_i^* = \max\{0, 1/\nu^* - \alpha_i\} = [1/\nu^* - \alpha_i]^+$

- This is the famous water-filling power allocation

Example: Power Control for Capacity Optimization

- We think of α_i as the ground level above patch i, then flood the region with water to a depth $1/v$
- The total amount of water used is then $\sum_{i=1}^{n} \max\{0, 1/v^* - \alpha_i\}$
- We then increase the flood level until we have used a total amount of water equal to one
- The depth of water above patch i is the optimal value x_i



Example: Power Control for Capacity Optimization

- Now the problem becomes finding the water level
- Substituting this expression for x_i^* into the condition $\mathbf{1}^T x^* = 1$, we obtain

$$\sum_{i=1}^{n} \max\{0, 1/\nu^\star - \alpha_i\} = 1.$$

Example: Power Control for Capacity Optimization

- A sample algorithm for obtaining the water level is described as follows

Algorithm 1 Algorithm for obtaining the water level

1. Initially, for each player k, sets the indicators of all channels to be 1 which assumes all subchannels to be active. Then calculates the water level.
2. For all active subchannels, if there exists $1/\mu_k - \mathrm{IN}_k(n) < 0$, set the indicator of the channel with smallest $1/\mu_k - \mathrm{IN}_k(n)$ to be 0 (i.e., inactive).
3. Calculates the new water level.
4. The steps from 2 to 3 are repeated until $1/\mu_k - \mathrm{IN}_k(n) \geq 0$ for all active subchannels.