DSC 211 Introduction to Optimization Winter 2024 Instructor: Jun-Kun Wang
Scribe: Merlin Chang, Marialena Sfyraki February 8, 2024

Lecture 10: Kahn-Karush-Tucker Optimality Conditions

1 Duality Theory Cont'd

Recall that we are tying to find the minimizer of a function subject to constrains, where

$$\begin{aligned} &\inf_x f(x) \\ &\text{s.t. } f_j(x) \leq 0, \quad j = 1, \ldots, m. \\ &\text{s.t. } \textbf{affine } \; h_i(x) = 0, \quad i = 1, \ldots, p. \end{aligned} \tag{1}$$

We have the following definitions:

Definition 1. ***(Lagrangian)***

$$L(x, \lambda, \mu) := f(x) + \sum_{j=1}^{m} \lambda_j f_j(x) + \sum_{i=1}^{p} \mu_i h_i(x). \tag{2}$$

Definition 2. ***(Dual function)***

$$g(\lambda, \mu) := \inf_x L(x, \lambda, \mu) \tag{3}$$

Definition 3. ***(Dual Problem)***

$$\sup_{\lambda \geq 0, \mu} g(\lambda, \mu) \tag{4}$$

Definition 4. ***(Strong Duality)*** *Strong duality means that there's no duality gap between the primal and the dual, in other words,*

$$\sup_{\lambda \geq 0; \mu} g(\lambda, \mu) = \inf_{x \in C} f(x).$$

Definition 5. ***(Slater condition)*** *Slater's condition outlines a* ***sufficient condition*** *for strong duality to hold. Namely, if*

- *The primal problem is convex;*
- *The primal problem has a strictly feasible point, $\bar{x}$ such that all the inequality constraints defining C are strict at $\bar{x}$, i.e., $f_j(\bar{x}) < 0, \forall j \in [m]$, and $h_i(\bar{x}) = 0, \forall i \in [p]$.*

then, strong duality holds, which implies:

$$d^* = \sup_{\lambda \geq 0; \mu} g(\lambda, \mu) = \inf_{x \in C} f(x) = p^*.$$

Remark. For linear programs, namely in the forms of:

$$\min_x c^\top x$$
$$\text{s.t. } Ax \leq b$$

strong duality holds if either primal or dual is feasible, and it does not require strict feasibility. [1]

Theorem 1. *If $f, f_1, \ldots, f_m$ are convex functions and $h_i(\cdot)$ are affine, the Slater condition guarantees the strong duality. For proof of this theorem, please refer to chapter 5.4 of Algorithms for Convex Optimization by Vishnoi.*

2 Kahn-Karush-Tucker Condition

For problems in the form 1, the Kahn-Karush-Tucker conditions are the following:

Definition 6. ***(KKT conditions)*** *We say the primal variables $x_* \in \mathbb{R}^d$ and the dual variables $\lambda_* \in \mathbb{R}^m$, $\mu_* \in \mathbb{R}^p$ satisfy KKT conditions if*

- *(Primal feasibility)* $\forall j \in [m] : f_j(x_*) \leq 0$ *and* $\forall i \in [p] : h_i(x_*) = 0$.
- *(Dual feasibility)* $\lambda_* \geq 0$.
- *(Stationarity)* $\partial_x L(x_*, \lambda_*, \mu_*) = 0$.
- *(Complementary slackness)* $\forall j \in [m] : \lambda_j f_j(x_*) = 0$.

Remark. The complementary slackness condition has the following implication.

- $\lambda_i^* > 0 \implies f_i(x^*) = 0$.
- $f_i(x^*) < 0 \implies \lambda_i^* = 0$.

3 Strong Duality and KKT conditions

Now, let $x_* \in \mathbb{R}^d$ be the primal feasible points and let $\lambda_* \in \mathbb{R}^m$ and $\mu_* \in \mathbb{R}^p$ be the dual feasible points.

Theorem 2 (Strong Duality and KKT conditions)**.** ***Strong duality***, *i.e.,*

$$f(x_*) = g(\lambda_*, \mu_*)$$

*implies that x_*, λ_*, μ_* satisfy the **KKT conditions**. Furthermore, if $f(\cdot), f_1(\cdot), \ldots, f_m(\cdot)$ are **convex** and $h_1(\cdot)$, $h_2(\cdot)$, $\ldots, h_p(\cdot)$ are affine, then the converse is also true: **KKT conditions implies the strong duality**.*

Remark. The above theorem has the following implications:

- Recall the dual value is always not greater than the primal value, that is

  $$\sup_{\lambda \geq 0; \mu} g(\lambda, \mu) \leq \inf_{x \in C} f(x).$$

  Therefore, when they have zero duality gap, that is

  $$g(\lambda_*, \mu_*) = f(x_*),$$

  then

  $$x_* \text{ is primal optimal ; } \lambda_*, \mu_* \text{ are dual optimal.}$$

- If $f(\cdot), f_1(\cdot), \ldots, f_m(\cdot)$ are **convex** and $h_1(\cdot)$, $h_2(\cdot)$, $\ldots, h_p(\cdot)$ are affine, then KKT conditions implies the strong duality. That is, if x_*, λ_*, μ_* satisfies KKT, then

  $$x_* \text{ is primal optimal ; } \lambda_*, \mu_* \text{ are dual optimal.}$$

4 Applications of KKT conditions

Example 1. Recall that the projection onto a set C is defined as $\text{Proj}_C(y) := \arg\min_{x \in C} \|y - x\|_2$. Now, consider that we want to find the projection onto the l_2-norm ball $C := \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$. This is equivalent to the following optimization problem

$$\begin{aligned} &\min_x \|x - y\|_2^2 \\ &\text{s.t. } \|x\|_2^2 \leq 1 \end{aligned}$$

We will show using the KKT conditions that the desired projection is equal to

$$\text{Proj}_C(y) = \frac{y}{\max\{1, \|y\|_2\}}.$$

Step 1. Getting the Lagrangian $L(x, \lambda)$ by introducting the Lagrangian Multiplier $\lambda \in \mathbb{R}^d_{\geq 0}$,

$$\begin{aligned} L(x, \lambda) &= f(x) + \lambda\left(\|x\|_2^2 - 1\right) \\ &= \|x - y\|_2^2 + \lambda\left(\|x\|_2^2 - 1\right). \end{aligned}$$

What is the "primal feasibility" of the KKT conditions in this case?

$$x : \|x\|_2^2 \leq 1$$

What is the "dual feasibility" of the KKT conditions in this case?

$$\lambda \geq 0$$

What is the "stationarity" of the KKT conditions in this case?

$$\begin{aligned} \nabla_x L(x, \lambda) = 0 &\Rightarrow 2(x - y) + 2\lambda x = 0 \\ &\Leftrightarrow y = (1 + \lambda)x \Leftrightarrow x = \frac{y}{(1 + \lambda)} \end{aligned} \tag{5}$$

What is the "complementary slackness" of the KKT conditions in this case?

$$\lambda\left(\|x\|_2^2 - 1\right) = 0 \tag{6}$$

As x must satisfy primal feasibility condition, we will now distinguish the following cases:

1. Case: $\|x\|_2^2 < 1$.
   By complementary slackness, we have $\lambda = 0$. In this case, by the stationarity condition(5), we also know that $y = x$. Intuitively, this implies that is y is already in the 2-norm ball, then we can simply output y as the solution to the projection problem.

2. Case: $\|x\|_2^2 = 1$.
   By stationarity, we know that

$$\begin{aligned} &\left\|\frac{y}{(1 + \lambda)}\right\|_2^2 = 1 \\ \Leftrightarrow\ &\frac{1}{(1 + \lambda)^2}\|y\|_2^2 = 1 \implies \|y\|_2 = (1 + \lambda) \\ \implies\ &x = \frac{y}{1 + \lambda} = \frac{y}{\|y\|_2} \end{aligned}$$

   The results make sense as if the output of the project is a point which satisfy $\|x\|_2^2 = 1$, then y's projection is the normalized vector y.

Example 2. Now, consider that we want to find the projection onto the l_1-norm ball $C := \{x \in \mathbb{R}^d : \|x\|_1 \leq 1\}$. This is equivalent to the following modified optimization problem, for simplicity in finding gradient,

$$\min_x \frac{1}{2}\|x - y\|_2^2$$
$$\text{s.t. } \|x\|_1 \leq 1$$

Remark. The inequality($\|x\|_1 < 1$) in the constraint is very important, without it the set C is not a convex set.
Remark. Unlike the projection onto $l2$-norm, projection onto $l1$-norm does not have a *closed form* solution, this will be clear once we find the solution.

We will proceed using KKT theorem to show that the solution to the above problem is the following:

$$x = y \text{ if } \|y\|_1 \leq 1;$$

otherwise,

$$x[i] = \text{sign}(y[i]) \left(|y[i]| - \lambda\right)_+, \forall i \in [d]$$

where λ is a number such that $\sum_{i=1}^d \left(|y[i]| - \lambda\right)_+ = 1$ and $(z)_+ := \max\{0, z\}$.

We begin with constructing the Lagrangian $L(x, \lambda)$ for $\lambda \in \mathbb{R}^d_{\geq 0}$. Where

$$L(x, \lambda) = \frac{1}{2}\|x - y\|_2^2 + \lambda(\|x\|_1 - 1).$$

What is the "primal feasibility" of the KKT conditions in this case?

$$x : \|x\|_1 \leq 1$$

What is the "dual feasibility" of the KKT conditions in this case?

$$\lambda \geq 0$$

What is the "stationarity" of the KKT conditions in this case? There are some added complications since the $l1$-norm isn't a differentiable function.
Recall the subgradient of the $l1$-norm,

$\|x\|_1 = \sum_{i=1} |x[i]|$. Subgradient of $\|x\|_1 = \begin{bmatrix} g_1 \in \partial(|x[1]|) \\ g_2 \in \partial(|x[2]|) \\ \vdots \\ g_d \in \partial(|x[d]|) \end{bmatrix}$ where we have if $|x[i]| > 0$, then $\partial x[i] = \nabla x[i] = \pm 1$. If $x[i] = 0$, then $\partial x[i] = [-1, 1]$.

We want to set the subgradient of the Lagrangian to 0, which is

$$\begin{aligned}&\partial_x L(x,\lambda) = (x-y) + \lambda g_x \ni \mathbf{0}\\ &\Leftrightarrow y[i] = x[i] + \lambda g_i, \text{ where } g_i \in \partial(|x[i]|)\end{aligned}$$

Since the subgradient of an absolute value is anything in the set $[-1, 1]$, we can generalize it into

$$\partial|x[i]| = \text{sign}(x[i])$$

Thus, we will arrive at the following,

$$\mathbf{0} \in \partial_x L(x,\lambda) \implies 0 \in x[i] - y[i] + \lambda\text{sign}(x[i]) \implies y[i] \in x[i] + \lambda\text{sign}(x[i]),$$

then if we express $x[i]$ in terms of $y[i]$, we get

$$x[i] = \text{sign}(y[i])\left(|y[i]| - \lambda\right)_+, \forall i \in [d] \tag{7}$$

$$= \begin{cases} y[i] - \lambda, & y[i] > \lambda \\ 0, & y[i] \in [-\lambda, \lambda] \\ y[i] + \lambda, & y[i] < -\lambda \end{cases}. \tag{8}$$

What is the "complementary slackness" of the KKT conditions in this case?

$$\lambda(\|x\|_1 - 1) = 0 \implies \lambda = 0 \text{ or } \|x\|_1 = 1$$

Now, using the KKT conditions, we can derive the final solution. Starting from the dual feasibility condition, we have the following cases:

- $\lambda = 0$
  From stationarity, we have $x[i] = \text{sign}(y[i])|y[i]| = y[i]$, which implies that $\|x\|_1 = \|y\|_1$. From the primal feasibility condition, we know that this is the case where y in within the $l1$-norm ball. $\|y\|_1 \leq 1$.

- $\lambda > 0$
  By complementary slackness, we know that $\|x\|_1 = 1$, which implies

  $$\sum_{i=1}^{n} |\text{sign}(y[i])\left(|y[i]| - \lambda\right)_+| = \sum_{i=1}^{n} |(|y[i]| - \lambda)_+ = 1$$

  From this, we will make note that λ is the root for solving the above linear piece-wise equations.

To express the expression $x[i] = \text{sign}(y[i]) \left(|y[i]| - \lambda\right)_+$ in a more explicit format, we can consider the expression case by case, which is

$$x[i] = \text{sign}(y[i]) \left(|y[i]| - \lambda\right)_+ = \begin{cases} y[i] - \lambda, & y[i] > \lambda \\ 0, & y[i] \in [-\lambda, \lambda] \\ y[i] + \lambda, & y[i] < -\lambda \end{cases} \tag{9}$$

where λ is the root for $\sum_{i=1}^{n} (|y[i]| - \lambda)_+ = 1$.
Thus, the solution for projection to $l1$-norm ball is y if $\|y\|_1 \leq 1$, otherwise it can be constructed by component following equation (7).
Remark. The solution is not a closed form solution as it partially depend on λ. For solving piece-wise linear equations, please follow the guide by Ang[2].

5 Proof for Theorem 2

Proof. **The proof consists of two parts:**

- The direction of strong duality implies KKT can be shown by:

$$f(x^*) = g(\lambda^*, \mu^*) = \inf_x \left(f(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \mu_i^* h_i(x) \right) \tag{10}$$

$$\leq f(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \mu_i^* h_i(x^*) \tag{11}$$

$$\leq f(x^*) \tag{12}$$

  where (11) by the definition of the infimum and (12) is by the primal feasibility and dual feasibility. Now it is evident that (11) and (12) are actually equalities. That is, we have

$$f(x^*) = g(\lambda^*, \mu^*) = \inf_x \left(f(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \mu_i^* h_i(x) \right) \tag{13}$$

$$= f(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \mu_i^* h_i(x^*) \tag{14}$$

$$= f(x^*), \tag{15}$$

  where equality (14) implies the stationarity, $\partial_x L(x_*, \lambda_*, \mu_*) = 0$, and equality (15) implies the complementary slackness (since we also have the primal and dual feasibility above).

- Now we prove the other direction: if $f(\cdot), f_1(\cdot), \ldots, f_m(\cdot)$ are **convex** and $h_1(\cdot)$, $h_2(\cdot)$, $\ldots, h_p(\cdot)$ are affine, then KKT conditions imply strong duality.

  By definition, the dual function with optimal λ, μ is

$$
\begin{aligned}
& g(\lambda_*, \mu_*) \\
& = \inf_x L(x, \lambda_*, \mu_*), \text{ by definition.} \\
& = L(x_*, \lambda_*, \mu_*), \text{by convexity of } L \text{ and stationarlity condition.} \\
& = f(x_*) + \sum_{j=1}^m \lambda_j^* f_j(x_*) + \sum_{i=1}^p \mu_i^* h_i(x_*), \text{ by definition of Lagrangian.} \\
& = f(x_*) + \sum_{j=1}^m \lambda_j^* f_j(x_*), \text{ by primal feasibility condition.} \\
& = f(x_*), \text{ by complementary slackness condition}
\end{aligned}
$$

Thus completes the proof. □

Bibliographic notes

Chapter 5 in [3] provides a thorough treatment of Lagrangian, duality, conjugate functions, and KKT optimality conditions. For the proof of strong duality under Slater's condition, see also Chapter 5.4 of [4].

References

[1] Laurent El Ghaoui. Hyper-Textbook: Optimization Models and Applications. EECS UC Berkeley, 2021 `https://inst.eecs.berkeley.edu/~ee127/sp21/livebook/l_sdual_slater.html`

[2] Andersen Ang Solving piece-wise linear equation `https://angms.science/doc/misc/solvingPiecewiseLinear.pdf`

[3] Boyd, Stephen, and Vandenberghe, Lieven. Convex optimization. Cambridge University Press, 2021

[4] Nisheeth K. Vishnoi. Algorithms for Convex Optimization. Cambridge University Press, 2021