Large Scale Optimization for Machine Learning:
Lecture 9

Lecturer: Meisam Razaviyayn Scribe: Ziyu He

Sep 21, 2017

1 Lagrangian and Dual Function

The previous lectures introduced a branch of methods to solve constrained optimization problems, namely, feasible direction methods. In this lecture, based on the theory of Lagrangian duality, we will discuss another way to handle constrained problems. First, let's consider the following generic formulation of optimization problem that we will work with throughout this lecture:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0, \quad \forall i = 1, 2, \ldots, m. \\ & x \in \mathcal{X} \end{aligned}$$

Where $\mathcal{X}$ is the set defined by the constraints that are easy to handle. We can define $\mathcal{X}$ based on our needs, e.g., $\mathcal{X} \triangleq \cap_{i=0}^{m} \mathbf{dom} f_i(\mathbf{x})$. Also note that for now we don't assume convexity of this problem.

1.1 Lagrangian Function

Definition 1 *The Lagrangian function $L(\mathbf{x}, \boldsymbol{\lambda}) : \mathbb{R}^{n+m} \to \mathbb{R}$ is defined as*

$$L(\mathbf{x}, \boldsymbol{\lambda}) \triangleq f_0(\mathbf{x}) + \lambda_1 f_1(\mathbf{x}) + \ldots + \lambda_m f_m(\mathbf{x})$$

where $\boldsymbol{\lambda}$ is called the Lagrange multiplier or dual variable.

The domain of $L(\mathbf{x}, \boldsymbol{\lambda})$ can be defined as $\mathbf{dom} L = \mathcal{X} \times \mathbb{R}^m$. We can intuitively interpret Lagrangian function as augmenting the objective function

$f_0(\mathbf{x})$ by taking the constraints $f_i(\mathbf{x}) \geq 0, i = 1...m$ into consideration while penalizing their potential violations with price $\boldsymbol{\lambda}$.

Example: Given the following optimization problem,

$$\begin{aligned} \min_{x} \quad & f_0(x) \\ \text{s.t.} \quad & 5x + 2 \leq 0 \\ & 3x^2 - 10 \leq 0 \end{aligned}$$

Then the Lagrangian function is,

$$L(x, \boldsymbol{\lambda}) = x^2 + \lambda_1(5x + 2) + \lambda_2(3x^2 - 10)$$

1.2 Dual Function

Definition 2 *(Lagrange) dual function* $g : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\}$

$$g(\boldsymbol{\lambda}) \triangleq \inf_{\mathbf{x} \in \mathcal{X}} L(\mathbf{x}, \boldsymbol{\lambda}) = \inf_{\mathbf{x} \in \mathcal{X}} \big(f_0(\mathbf{x}) + \lambda_1 f_1(\mathbf{x}) + \ldots + \lambda_m f_m(\mathbf{x})\big)$$

We can treat the dual function as an evaluation of price $\boldsymbol{\lambda}$ and have the following properties of dual function $g(\boldsymbol{\lambda})$.

Property 1. Dual function $g(\boldsymbol{\lambda})$ is always concave regardless of the primal problem's convexity (from now on we will denote the original problem as *primal*). It is obvious that $L(\mathbf{x}, \boldsymbol{\lambda})$ is an affine (hence concave) function of $\boldsymbol{\lambda}$, so that $g(\boldsymbol{\lambda})$ is the point-wise infimum of a family of concave functions of $\boldsymbol{\lambda}$ which means $g(\boldsymbol{\lambda})$ is concave.

Property 2. It is possible for dual function $g(\boldsymbol{\lambda})$ to take on the value $-\infty$ for some $\boldsymbol{\lambda}$, which typically happens when $L(\mathbf{x}, \boldsymbol{\lambda})$ is unbounded below in $\mathbf{x}$ for some $\boldsymbol{\lambda}$. We can show this with the following example:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^T \mathbf{x} \leq b_i, \quad \forall i = 1, 2, \ldots, m \end{aligned}$$

Then,

$$\begin{aligned}
L(\mathbf{x}, \boldsymbol{\lambda}) &= \mathbf{c}^T\mathbf{x} + \sum_i \lambda_i(\mathbf{a}_i^T\mathbf{x} - b_i) \\
g(\boldsymbol{\lambda}) &= \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) \\
&= \inf_{\mathbf{x}} \mathbf{c}^T\mathbf{x} + \sum_i \lambda_i(\mathbf{a}_i^T\mathbf{x} - b_i) \\
&= \inf_{\mathbf{x}} (\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda})^T\mathbf{x} - \mathbf{b}^T\boldsymbol{\lambda}
\end{aligned}$$

When we have $\mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda} \neq \mathbf{0}$ we can always choose $\mathbf{x} \in \mathcal{X}$ (i.e. $\mathbb{R}^n$) to make $L(\mathbf{x}, \boldsymbol{\lambda})$ arbitrarily small, thus

$$g(\boldsymbol{\lambda}) = \begin{cases} -\infty & \text{if } \mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda} \neq \mathbf{0} \\ -\mathbf{b}^T\boldsymbol{\lambda} & \text{if } \mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda} = \mathbf{0} \end{cases}$$

1.3 Lower Bound Property of Dual Function

Here we will introduce an additional important property of dual function which shows that $g(\boldsymbol{\lambda})$ provides us a lower bound of the primal optimal value under some condition.

Theorem 1 $g(\boldsymbol{\lambda}) \leq f_0(\mathbf{x}),\ \forall \boldsymbol{\lambda} \succeq \mathbf{0},\ \forall \mathbf{x}$ *primal feasible*

Proof Assume we have an arbitrary primal feasible $\tilde{\mathbf{x}}$ i..e, $f_i(\mathbf{x}) \leq \mathbf{0}, i = 1...m$ and we have $\boldsymbol{\lambda} \succeq \mathbf{0}$, then we have:

$$\begin{aligned}
f_0(\tilde{\mathbf{x}}) &\geq f_0(\tilde{\mathbf{x}}) + \sum_i \boldsymbol{\lambda} f_i(\tilde{\mathbf{x}}) \\
&\geq \inf_{\mathbf{z} \in \mathcal{X}} (f_0(\mathbf{z}) + \sum_i \boldsymbol{\lambda} f_i(\mathbf{z})) \\
&= g(\boldsymbol{\lambda})
\end{aligned}$$

Note that we get the second inequality from the fact that feasible set of the primal problem is a subset of $\mathcal{X}$, since $\tilde{\mathbf{x}}$ is arbitrary we can conclude our proof.

In other words, as long as $\boldsymbol{\lambda} \succeq \mathbf{0}$, we have $g(\boldsymbol{\lambda})$ as lower bound of objective value $f_0(\mathbf{x})$ for any feasible $\mathbf{x}$ including the primal optimal solution $\mathbf{x}^*$:

Corollary 1 $g(\boldsymbol{\lambda}) \leq p^*,\ \forall \lambda \succeq \mathbf{0},$ *where* $p^* \triangleq f_0(\mathbf{x}^*)$ *and* $\mathbf{x}^*$ *primal optimal.*

2 Lagrange Dual Problem

2.1 Lagrange Dual Problem

Inspired by the fact that for each $\boldsymbol{\lambda} \succeq \mathbf{0}$ we have $g(\boldsymbol{\lambda})$ as a lower bound of primal optimal value p^*, we construct the (Lagrangian) dual problem

$$\begin{aligned} d^* \triangleq \max_{\boldsymbol{\lambda}} \quad & g(\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \succeq \mathbf{0} \end{aligned}$$

to find the "best" lower bound of p^* for the primal problem

$$\begin{aligned} p^* \triangleq \min_{\mathbf{x}} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0, \quad \forall i = 1, \ldots, m \\ & \mathbf{x} \in \mathcal{X} \end{aligned}$$

and have the following properties:

Property 1 The dual problem is always convex regardless of the convexity of primal problem (since we always have $g(\boldsymbol{\lambda})$ being concave and our dual problem maximizes $g(\boldsymbol{\lambda})$ over $\boldsymbol{\lambda} \succeq \mathbf{0}$).

Property 2 Apparently we have $d^* \leq p^*$ (from the lower bound property of $g(\boldsymbol{\lambda})$). We denote this property as *weak duality* and we define $p^* - d^*$ as *duality gap*. Note that weak duality will hold for any problem.

2.2 Strong Duality and Slater's Condition

We say *strong duality* holds when the *duality gap* $\triangleq p^* - d^* = 0 \Leftrightarrow p^* = d^*$

Unfortunately, *strong duality* does not hold in general. But for a convex primal problem we can conclude strong duality when some special conditions hold (we will denote such conditions as *constraint qualifications*). In other words, *constraint qualifications* + *primal convexity* $\Rightarrow$ *strong duality*.

There exists several constraint qualifications for strong duality (a summary can be found in [1]), a simple example is *Slater's condition*:

Definition 3 *Slater's condition:* $\exists \mathbf{x} \in \operatorname{relint}(\mathcal{X})$ *with* $f_i(\mathbf{x}) < 0, i = 1, \ldots, m$

where $\operatorname{relint}(\mathcal{X})$ denotes the relative interior of set $\mathcal{X}$:

Definition 4 *The interior of a set* $\mathcal{A}$*:* $\operatorname{int}(\mathcal{A}) = \{\mathbf{x} \in \mathcal{A} | \exists \varepsilon > 0, N_\varepsilon(\mathbf{x}) \subseteq \mathcal{A}\}$

Definition 5 *The relative interior of a set* $\mathcal{A}$*:* $\operatorname{relint}(\mathcal{A}) = \{\mathbf{x} \in \mathcal{A} | \exists \varepsilon > 0, N_\varepsilon(\mathbf{x}) \cap \operatorname{Aff}(\mathcal{A}) \subseteq \mathcal{A}\}$*,* $\operatorname{Aff}(\mathcal{A})$ *being the affine hull of* $\mathcal{A}$*.*

In Slater's condition, the affine inequalities do not need to hold with strict inequality. That is to say if we denote the affine inequalities as $f_{A_{(j)}}(\mathbf{x}) \leq 0, j = 1...k$, we can relax Slater's condition as $\exists \mathbf{x} \in \operatorname{relint}(\mathcal{X})$ *s.t.* $f_{A_{(j)}}(\mathbf{x}) \leq 0,\ j = 1...k$ *and* $f_i(\mathbf{x}) < 0,\ \forall i \in [m] \backslash \{A_{(j)} | j = 1...k\}$. In other words, we only need to check non-affine inequality constraints.

Proof of strong duality under Slater's condition and primal convexity can be found in 5.3.2. of [2].

<u>Example of a Slater point:</u>

$$\begin{aligned} \min_x \quad & f_0(x) \\ \text{s.t.} \quad & x^2 \leq 1 \\ & 5x + 1 \leq 2 \end{aligned}$$

Note that since second constraint is affine, we only need to check the first condition. Since $\mathcal{X} \triangleq \mathbb{R}$, $\exists x$ s.t. $x^2 < 1$. Hence Slater's condition holds and we have strong duality for this problem.

2.3 Linear and Quadratic Examples

(1) A linear primal optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \preceq \mathbf{b} \end{aligned}$$

We have already shown that the dual function should be:

$$g(\boldsymbol{\lambda}) = \begin{cases} -\infty & \text{if } \mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda} \neq \mathbf{0} \\ -\mathbf{b}^T\boldsymbol{\lambda} & \text{if } \mathbf{c} + \mathbf{A}^T\boldsymbol{\lambda} = \mathbf{0} \end{cases}$$

so that we have dual problem:

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \quad & g(\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \succeq \mathbf{0} \end{aligned}$$

We can see that this problem is equivalent to the following problem

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \quad & -\mathbf{b}^T\boldsymbol{\lambda} \\ \text{s.t.} \quad & \mathbf{A}^T\boldsymbol{\lambda} + \mathbf{c} = \mathbf{0} \\ & \boldsymbol{\lambda} \succeq \mathbf{0} \end{aligned}$$

A detailed discussion of this equivalence can be found in 5.2.1 of [2].

(2) Consider the quadratic linear optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}^T\mathbf{Q}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{aligned}$$

Then we have Lagrangian function:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

and dual function as follow:

$$g(\boldsymbol{\lambda}) = \inf_{\mathbf{x}} \mathbf{x}^T\mathbf{Q}\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

If $\mathbf{Q} \succ \mathbf{0}$ we have the primal convexity and we have the minimizer of Lagrange function as:

$$\begin{aligned} & \nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = 2\mathbf{Q}\mathbf{x}^* + \mathbf{A}^T\boldsymbol{\lambda} = 0 \Longleftrightarrow \mathbf{x}^* = -\frac{1}{2}\mathbf{Q}^{-1}\mathbf{A}^T\boldsymbol{\lambda} \\ & \Longrightarrow g(\boldsymbol{\lambda}) = -\frac{1}{4}\boldsymbol{\lambda}^T\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\mathbf{b} \end{aligned}$$

Thus we have dual problem:

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \quad & -\frac{1}{4}\boldsymbol{\lambda}^T\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\mathbf{b} \\ \text{s.t.} \quad & \boldsymbol{\lambda} \succeq \mathbf{0} \end{aligned}$$

(3) Another quadratic optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2}||\mathbf{x}||^2 \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{aligned}$$

We have the duality as:

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \quad & -\frac{1}{2}\boldsymbol{\lambda}^T\mathbf{A}\mathbf{A}^T\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\mathbf{b} \\ \text{s.t.} \quad & \boldsymbol{\lambda} \succeq \mathbf{0} \end{aligned}$$

Duality In Algorithms

Many Algorithms produce at iterate r:

- Primal Feasible point $\mathbf{x}^r$
- Dual Feasible point $\boldsymbol{\lambda}^r$

Since the optimal objective value is bounded above and below by the $f_0(\mathbf{x}^r)$ and $g(\boldsymbol{\lambda}^r)$, we can use the difference between these two points as a stopping criteria. Note when strong duality holds $f_0(\mathbf{x}^r) - g(\boldsymbol{\lambda}^r) \to 0$ (if our algorithm can converge for both primal and dual problem).

Thus although the optimal value is not known, we can estimate the accuracy of the algorithm at every iteration r. For instance, we can define $\mathbf{x}$ being $\epsilon-$suboptimal for primal problem and $\boldsymbol{\lambda}$ being $\epsilon-$suboptimal for dual problem if we have $f_0(\mathbf{x}) - g(\boldsymbol{\lambda}) = \epsilon$ (see 5.5.1 of [2]).

3 Optimality Conditions

From previous examples, we can see that dual problems are guaranteed to be convex regardless of primal convexity, and sometimes their feasible sets are simpler than the primal feasible sets. In other words, we might come across with specific cases when the dual problems are easier to solve than the primal problems (e.g., if we apply gradient projection method to the third example in 2.3). So one might want to know if we can somehow connect the primal optimal solution $\mathbf{x}^*$ and dual optimal solution $\boldsymbol{\lambda}^*$ together especially when strong duality holds. As a matter of fact, strong duality can provide us such information. To be more accurate, we will see that based on strong duality we can define both necessary and sufficient condition for a pair of $(\mathbf{x}, \boldsymbol{\lambda})$ being primal/dual optimal.

First let's have a look at what strong duality can tell us (assume that $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ primal/dual optimal):

$$\begin{aligned} f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*) \\ &= \inf_{\mathbf{x}\in\mathcal{X}} \big(f_0(\mathbf{x}) + \textstyle\sum_i \lambda_i^* f_i(\mathbf{x})\big) \\ &\leq f_0(\mathbf{x}^*) + \textstyle\sum_i \lambda_i^* f_i(\mathbf{x}^*) \\ &\leq f_0(\mathbf{x}^*) \end{aligned}$$

The third and fourth inequalities are obtained from $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ being primal/dual feasible. Note that now we have $f_0(\mathbf{x}^*) \geq f_0(\mathbf{x}^*)$ which means we should have:

$$\begin{cases} \inf_{\mathbf{x}\in\mathcal{X}} \big(f_0(\mathbf{x}) + \sum_i \lambda_i^* f_i(\mathbf{x})\big) = f_0(\mathbf{x}^*) + \sum_i \lambda_i^* f_i(\mathbf{x}^*) \\ \sum_i \lambda_i^* f_i(\mathbf{x}^*) = 0 \end{cases}$$

The first equality means that $\mathbf{x}^*$ is the minimizer of function $L(\mathbf{x}, \boldsymbol{\lambda}^*)$ when $\mathbf{x} \in \mathcal{X}$. If we have $\mathcal{X} \triangleq \mathbb{R}^n$, then according to the first-order optimality condition we should have:

$$\nabla f_0(\mathbf{x}^*) + \textstyle\sum_i \lambda_i^* \nabla f_i(\mathbf{x}^*) = 0$$

which will be useful later. Technically, if we have $\mathcal{X} \neq \mathbb{R}^n$ we should apply the first-order optimality condition under constraints and get the following conclusion instead:

$$\left(\nabla f_0(\mathbf{x}^*) + \sum_i \lambda_i^* \nabla f_i(\mathbf{x}^*)\right)^T (\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in \mathcal{X}$$

But we will stick to the first conclusion. Since $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ primal/dual feasible, we have $\lambda_i^* f_i(\mathbf{x}^*) \leq 0, \forall i$. Thus for the second equality $\sum_i \lambda_i^* f_i(\mathbf{x}^*) = 0$ to hold, we must have: $\lambda_i^* f_i(\mathbf{x}^*) = 0, \forall i$. We define this property as *complementary slackness.*

Definition 6 *Complementary slackness: if we have* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *being primal/dual optimal, then we have* $\lambda_i^* f_i(\mathbf{x}^*) = 0, \forall i$.

KKT Optimality Conditions

With the conclusions from our previous discussion, now we can define a first-order necessary optimality condition for primal/dual feasible $(\mathbf{x}, \boldsymbol{\lambda})$ to be primal/dual optimal. Such condition is named after Karush-Kuhn-Tucker (KKT for short):

Theorem 2 *Assume* f_i*'s are differentiable and* $\mathcal{X} = \mathbb{R}^n$ *and* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *optimal with zero duality gap. Then,*

1. $f_i(\mathbf{x}^*) \leq 0 \quad \forall i$ $\Leftarrow$ *Primal Feasibility*
2. $\boldsymbol{\lambda}^* \succeq 0 \quad \forall i$ $\Leftarrow$ *Dual Feasibility*
3. $\lambda_i^* f_i(\mathbf{x}^*) = 0 \quad \forall i$ $\Leftarrow$ *Complementary Slackness*
4. $\nabla f_0(\mathbf{x}^*) + \sum_i \lambda_i^* \nabla f_i(\mathbf{x}^*) = 0$ $\Leftarrow \mathbf{x}^*$ *minimizer for* $L(\mathbf{x}, \boldsymbol{\lambda}^*)$

The first two conditions follow feasibility of $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$. The third condition (complementary slackness) and the forth condition come from strong duality. Note that KKT conditions can be applied to all optimization problems. If our problem is convex, then we have KKT conditions as sufficient optimality condition:

Theorem 3 *Assume* f_i*'s are differentiable and convex where* $\mathcal{X} = \mathbb{R}^n$. *If we* *have primal/dual feasible* $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}})$ *satisfying the following conditions (KKT):*

1. $f_i(\tilde{\mathbf{x}}) \leq 0 \quad \forall i \qquad\qquad \Leftarrow$ *Primal Feasibility*
2. $\tilde{\boldsymbol{\lambda}} \succeq 0 \quad \forall i \qquad\qquad \Leftarrow$ *Dual Feasibility*
3. $\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0 \quad \forall i \qquad \Leftarrow$ *Complementary Slackness*
4. $\nabla f_0(\tilde{\mathbf{x}}) + \sum_i \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) = 0 \qquad \Leftarrow \tilde{\mathbf{x}}$ *minimizer for* $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}})$

then $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}})$ *primal/dual optimal with zero duality gap.*

Proof Since we have f_i convex $\forall i$ and $\tilde{\lambda} \geq 0 \forall i \;\Rightarrow L(\mathbf{x}, \tilde{\boldsymbol{\lambda}})$ convex in $\mathbf{x}$ (first order optimality condition is sufficient). Thus with the forth condition we have $f_0(\tilde{\mathbf{x}}) + \sum_i \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = \inf_{\mathbf{x} \in \mathcal{X}} f_0(\mathbf{x}) + \sum_i \tilde{\lambda}_i f_i(\mathbf{x}) = g(\tilde{\boldsymbol{\lambda}})$. According to the third condition, this means

$$g(\tilde{\boldsymbol{\lambda}}) = f_0(\tilde{\mathbf{x}}).$$

Moreoever, the weak duality always holds, i.e., $g(\tilde{\boldsymbol{\lambda}}) \leq f_0(\mathbf{x}), \; \forall$ feasible $\mathbf{x}$. Thus the above inequality implies

$$f_0(\tilde{\mathbf{x}}) \leq f_0(\mathbf{x}), \; \forall \text{ feasible } \mathbf{x}.$$

In other words, $(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}})$ primal/dual optimal.

Besides providing us the connection between dual optimality and primal optimality, KKT conditions also play an important role in problem solving. In some special cases, it is possible to solve the primal and dual problem by solving KKT conditions analytically. More generally, many algorithms for convex optimization can be interpreted as solving the KKT conditions [2].

Note that we have mentioned that sometimes the dual problem can be easier to solve than the primal problem. With the insights brought by KKT if we have solved dual optimal $\boldsymbol{\lambda}^*$ under strong duality, we can solve the primal problem by solving $\min_{\mathbf{x} \in \mathcal{X}} L(\mathbf{x}, \boldsymbol{\lambda}^*)$ since primal optimal $\mathbf{x}^*$ is the minimizer of this problem (note that this problem will be unconstrained if we have $\mathcal{X} \triangleq \mathbb{R}^n$). Few examples can be found in 5.5.5 of [2].

References

[1] E. Rodrigo, K Elizabeth, and R. Ademir. "Constraint Qualification for Nonlinear Programming" (Technical report), Federal University of Parana.

[2] S. Boyd, and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.