| CS2910 - Optimization | Summer 2023 |
|---|---|

# Lecture 3 — Duality

*Lecturer: Yunwei Ren* *Scribed by Zhidan Li*

## Contents



## 1 Overview: Optimization under Constraints

In this lecture, we will show how to solve a family of optimization problems under constraints. We want to solve the following optimization problem:

$$\begin{aligned} \min\ & f(x) \\ s.t.\ & h_i(x) \le 0, \quad \forall i \in [m], \\ & \ell_j(x) = 0, \quad \forall j \in [n]. \end{aligned} \tag{1}$$

Note that $f$, $\{h_i\}_{i\in[m]}$ and $\{\ell_j\}_{j\in[n]}$ might not be necessarily convex, but we assume that they are all $C^1$ functions.

Let $D$ be the feasible region of (1). When $f$ is a convex function and $D$ is a convex set, the projected gradient descent works if $\Pi_D(x)$ can be computed efficiently. In this lecture, we will introduce the duality of the optimization problems to make (1) much easier to solve.

## 2 Lagrange Multiplier and Weak Duality

To solve (1), it makes sense to transform it into one without constraints. For the region $D \subseteq \mathbb{R}^d$, define the indicator function $\iota_D : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ as:

$$\iota_D(x) = \begin{cases} 0 & x \in D, \\ \infty & x \notin D. \end{cases}$$

Then, the optimization problem (1) is equivalent to the following problem without constraints:

$$\min_{x \in \mathbb{R}^d} f(x) + \iota_D(x). \tag{2}$$

Since affine functions are often easier to analyze, for $\iota_D$, we have the following form of it.

**Lemma 1.** *Under the definition above, it holds that*

$$\iota_D(x) = \sup_{u \in \mathbb{R}^m_{\geq 0}, v \in \mathbb{R}^n} \left\{ \sum_{i=1}^m u_i h_i(x) + \sum_{j=1}^n v_j \ell_j(x) \right\}.$$

*Proof.* When $x \in D$, it is clear that $\iota_D(x) \leq 0$. We choose $u = 0$, then it holds that $\iota_D(x)$ can be 0.

When $x \notin D$, there are two cases: there exists $h_i(x) > 0$; there exists $\ell_j(x) \neq 0$.

- **Case 1:** $\exists h_i(x) > 0$. Then we let $u_i \to \infty$ and other $u_{i'} = 0$, $v = 0$. It is clear that $\iota_D(x) \to \infty$.
- **Case 2:** $\exists \ell_j(x) \neq 0$. Then we choose $v_j \to \text{sign}(\ell_j)\infty$ and other $v_{j'} = 0$, $u = 0$. It is clear that $\iota_D(x) \to \infty$.

□

By Lemma 1, the optimization problem (2) is equivalent to the following optimization problem:

$$\min_{x \in \mathbb{R}^d} \sup_{u \in \mathbb{R}^m_{\geq 0}, v \in \mathbb{R}^n} \left( f(x) + \sum_{i=1}^m u_i h_i(x) + \sum_{j=1}^n v_j \ell_j(x) \right). \tag{3}$$

For the sake of simplicity, define the function $L : \mathbb{R}^d \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^n$ as

$$L(x, u, v) \triangleq f(x) + \sum_{i=1}^m u_i h_i(x) + \sum_{j=1}^n v_j \ell_j(x).$$

We call (3) the *primal problem* ((P), for short) and the function $L$ the *Lagrangian function*. Note that, the function $g(u, v) \triangleq \inf_{x \in \mathbb{R}^d} L(x, u, v)$ is a concave function on $(u, v)$, it is much easier to maximize $g(u, v)$ than the original optimization. In words, if we can solve the inner unconstrained optimization problem $\inf_{x \in \mathbb{R}^d} L(x, u, v)$, the outer optimization problem is relatively easy.

Consider the following optimization problem

$$\sup_{u \in \mathbb{R}^m_{\geq 0}, v \in \mathbb{R}^n} \min_{x \in \mathbb{R}^d} L(x, u, v). \tag{4}$$

We call it the *dual problem* ((D), for short) of (3). **Note that** (P) and (D) are **not** equivalent in general.

**Proposition 2** (max-min inequality/weak duality)**.** *Let $X, Y$ be two topological spaces. For $f : X \times Y \to \mathbb{R}$, it holds that*

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) \geq \sup_{y \in Y} \inf_{x \in X} f(x, y).$$

*When the equality holds, we say the primal problem has* ***strong duality****.*

*Proof.* By definition, for all $y_0 \in Y$, it holds that

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) \geq \inf_{x \in X} f(x, y_0).$$

Then it immediately holds that

$$\inf_{x \in X} \sup_{y \in Y} f(x, y) \geq \sup_{y \in Y} \inf_{x \in X} f(x, y).$$

□

Although the optimal value of the dual problem might not give us an optimal value of the primal problem, it still offers a certification of the optimal value.

**Corollary 3** (primal-dual solutions as certification)**.** *For $x_* \in D$, $u_* \in \mathbb{R}^m_{\geq 0}$, $v_* \in \mathbb{R}^n$, if $f(x_*) = g(u_*, v_*)$, then $x_*$ (respectively, $u_*, v_*$) is an optimal solution to (P) (respectively (D)).*

*Proof.* For all $x \in D$, by the weak duality (Proposition 2), it holds that

$$f(x) \geq L(x, u_*, v_*) \geq g(u_*, v_*) = f(x_*).$$

Thus we show $f(x_*)$ is the optimal value.

The optimality of $(u_*, v_*)$ comes directly from Proposition 2. □

# 3 Slater's Condition

(This section refers to Chapter 5, [BV04]). Now we focus on under which conditions, the strong duality holds. For the primal optimization problem (2), define the set $\mathcal{G}$ as

$$\mathcal{G} \triangleq \left\{ (r, s, t) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} : \exists x \in \mathbb{R}^d, (\forall i \in [m], j \in [n], h_i(x) = r_i, \ell_j(x) = s_j) \wedge (f(x) = t) \right\}.$$

Let $p_*$ be the optimal value of (3). By constraints, it holds that

$$p_* = \inf \{ t \mid (r, s, t) \in \mathcal{G} : r \leq 0, s = 0 \}.$$

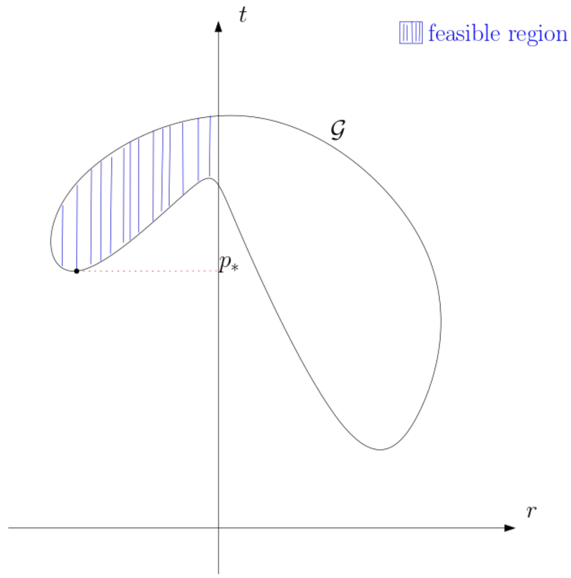


Figure 1: an easy case of $\mathcal{G}$, feasible region and the optimal value

Now we consider the Lagrangian function $L(x, u, v)$. On $G$, we write it as the following form

$$L(x, u, v) = t + u^\top r + v^\top s = (u, v, 1)^\top (r, s, t), r_i = h_i(x), s_j = \ell_j(x), \forall i \in [m], j \in [n].$$

Using this form of $L$, we claim

$$g(u, v) = \inf \left\{ (u, v, 1)^\top (r, s, t) \mid (r, s, t) \in \mathcal{G} \right\}.$$

Since $(u, v, 1)^\top (r, s, t)$ is a affine function on $(r, s, t)$, the infimum of $g(u, v)$ can be view as the intersection of some 'tangent hyperplane' with 'slope' perpendicular to $(u, v, 1)$ at some point $(r, s, t)$ on $t$-axis. The following figure illustrates an easy case of this geometric interpretation when $\mathcal{G}$ lies on $\mathbb{R}^2$.

As a result, the optimal value of the dual problem can be written as the following form:

$$d_* = \sup_{u \in \mathbb{R}^m_{\geq 0}, v \in \mathbb{R}^n} \min_{x \in \mathbb{R}^d} L(x, u, v) = \sup_{u \in \mathbb{R}^m_{\geq 0}, v \in \mathbb{R}^n} \inf \left\{ (u, v, 1)^\top (r, s, t) \mid (r, s, t) \in \mathcal{G} \right\}.$$

Intuitively and geometrically speaking, we might say, $d^*$ is the 'highest' intersection on $t$-axis among all tangent supporting hyperplanes which are tangent to $\mathcal{G}$ at the point in the feasible region. View Figure 3 as an instance on $\mathbb{R}^2$.

On the other hand, to make the set contain the point $(0, 0, p_*)$, we introduce the following set $\mathcal{A}$ which can be viewed as an epigraph of $\mathcal{G}$:

$$\mathcal{A} \triangleq \left\{ (r, s, t) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} : \exists x \in \mathbb{R}^d \text{ s.t } h_i(x) \leq r_i, \ell_j(x) \leq s_j, \forall i \in [m], j \in [m] \right\}$$

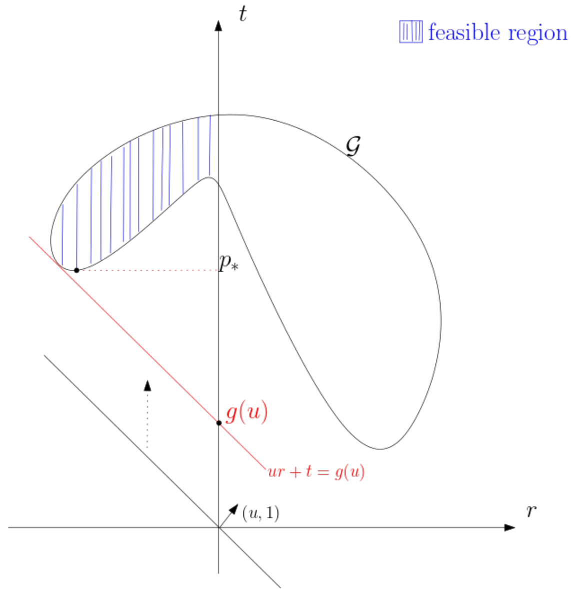


Figure 2: the geometric interpretation of $g(u, v)$. The red line is a tangent line with slope $-u$, and let $r = 0$ we get the point $(0, g(u))$ on this line.

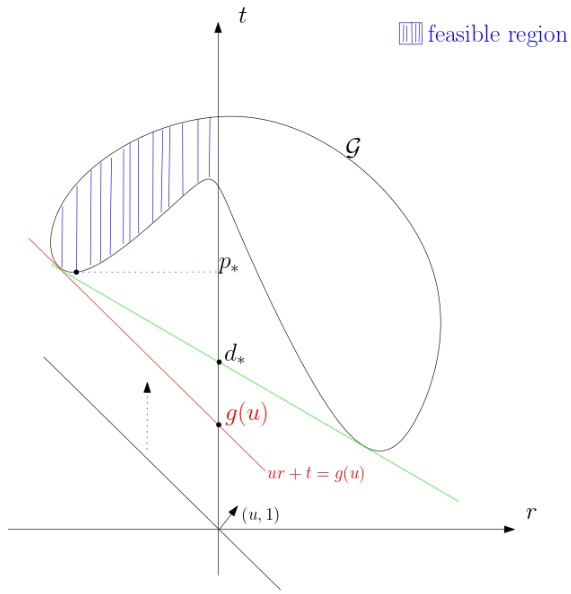


Figure 3: the green line is the tangent line achieves the optimal $d$. Note that it must be a tangent line at the point in the feasible region.

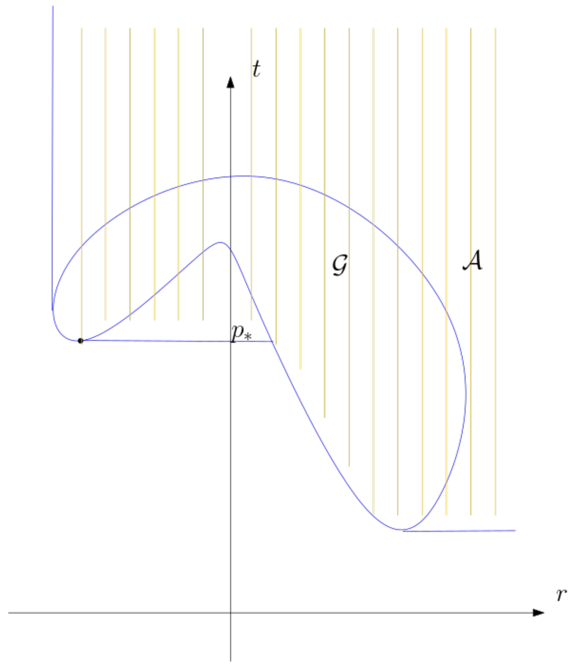


Figure 4: the epigraph

Now we introduce the Slater's condition.

**Definition 4** (Slater's condition)**.** *For a optimization problem* (3)*, we say it satisfies the Slater's condition if:*

- $f$ *is convex.*
- $h_i$ *is convex for all* $i \in [m]$*.*
- $\ell_j$ *is affine for all* $j \in [n]$*.*
- *There exists a point* $\hat{x} \in \operatorname{relint} D$*, i.e.,* $h_i(\hat{x}) < 0$ *for all* $i \in [m]$ *and* $\ell_j(\hat{x}) = 0$ *for all* $j \in [n]$*.*

**Lemma 5.** *If the optimization problem satisfies Slater's condition, then it has the strong duality.*

To prove Lemma 5, we firstly show the convexity of $\mathcal{A}$.

**Lemma 6.** *Under the Slater's condition,* $\mathcal{A}$ *is a convex set on* $\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}$*.*

The proof of Lemma 5 directly comes from the definition. We leave it as a mental training. Note that, even under the Slater's condition, $\mathcal{G}$ can be a non-convex set (let $f(x) = \frac{1}{2}x^2, h(x) = x$. It is trivial that $\mathcal{G}$ is not convex.)

Now we are ready to prove Lemma 5.

*Proof of Lemma 5.* For convenience, without loss of generality we assume that $p_*$ is finite. Since

$\ell_j$ is affine, we can write the optimization problem as the following form:

$$\begin{aligned} \min \ & f(x) \\ s.t. \ & h_i(x) \leq 0, \qquad \forall i \in [m], \\ & Ax - b = 0, \quad A \in \mathbb{R}^{n \times d}. \end{aligned}$$

We also assume $\operatorname{rank}(A) = n$. Consider the following set

$$\mathcal{B} \triangleq \{(0, 0, t) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} : t < p_*\}.$$

Directly from our construction, $\mathcal{B}$ is convex and $\mathcal{A} \cap \mathcal{B} = \varnothing$.

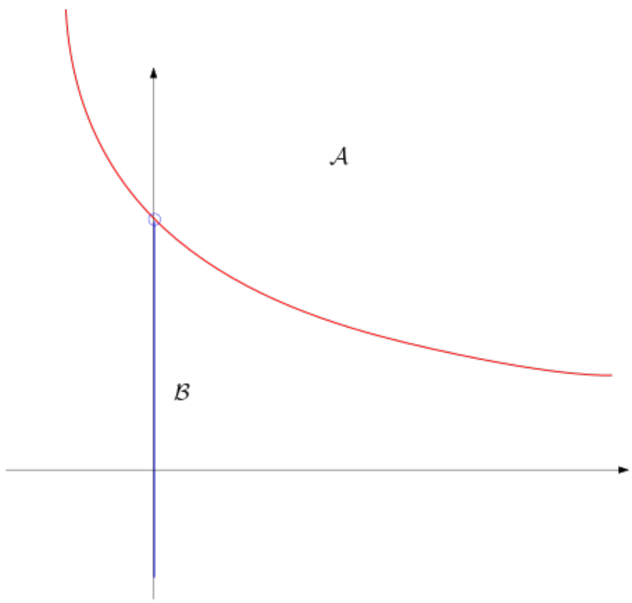


Figure 5: an example of $\mathcal{A}$ and $\mathcal{B}$

Since $\mathcal{A}, \mathcal{B}$ are convex and $\mathcal{A} \cap \mathcal{B} = \varnothing$, by the separating hyperplane theorem, there exists an affine function determined by $(\tilde{u}, \tilde{v}, \tilde{\mu}) \neq 0$ and a value $\gamma$ satisfying

- $(r, s, t) \in \mathcal{A} \implies \tilde{u}^\top r + \tilde{v}^\top s + \tilde{\mu} t \geq \gamma$.
- $(r, s, t) \in \mathcal{B} \implies \tilde{u}^\top r + \tilde{v}^\top s + \tilde{\mu} t \leq \gamma$.

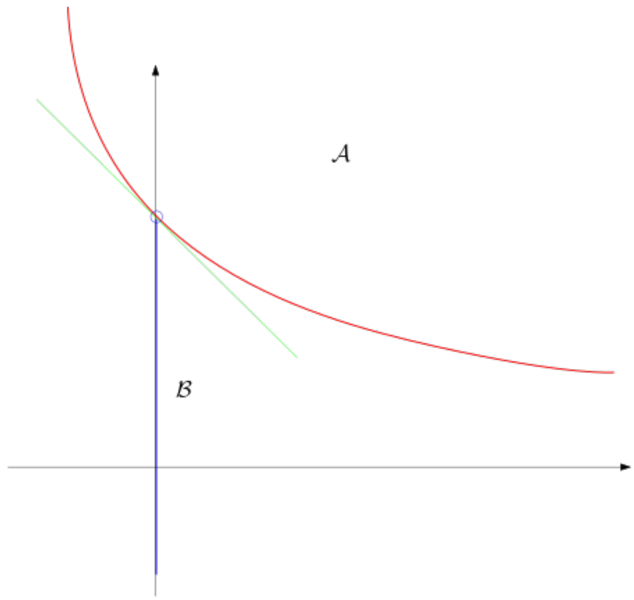


Figure 6: the green line separates the two convex sets

From our construction of $\mathcal{A}$, it must hold $\tilde{u} \geq 0$ and $\tilde{\mu} \geq 0$ (otherwise, $\tilde{u}^\top r + \tilde{\mu} t$ cannot be bounded). Also, since $t < p_*$ for every $(0, 0, t) \in \mathcal{B}$, it holds that $\tilde{\mu} p_* \leq \gamma$. Thus for every $x \in D$,

$$\sum_{i=1}^{m} \tilde{u}_i h_i(x) + \tilde{v}^\top (Ax - b) + \tilde{\mu} f(x) \geq \gamma \geq \tilde{\mu} p_*.$$

We separate the remaining part of our proof into two cases: $\tilde{\mu} > 0$; $\tilde{\mu} = 0$.

- **Case 1:** $\tilde{\mu} > 0$. Consider the pair $(\tilde{u}/\tilde{\mu}, \tilde{v}/\tilde{\mu})$. For all $x \in D$,

$$L(x, \tilde{u}/\tilde{\mu}, \tilde{v}/\tilde{\mu}) = f(x) + \frac{\sum_{i=1}^{m} \tilde{u}_i h_i(x) + \tilde{v}^\top (Ax - b)}{\tilde{\mu}} \geq p_*.$$

  Then combining it with the weak duality, we establish the strong duality.

- **Case 2:** $\tilde{\mu} = 0$. The separating hyperplane theorem implies

$$\sum_{i=1}^{m} \tilde{u}_i h_i(x) + \tilde{v}^\top (Ax - b) \geq \gamma \geq 0.$$

  Consider the point $\hat{x} \in \text{relint} D$. Since $A\hat{x} - b = 0$, it holds that

$$\sum_{i=1}^{m} \tilde{u}_i h_i(x) \geq 0.$$

  Since $h_i(\hat{x}) < 0$, it must hold that $\tilde{u} = 0$. Then for all $x \in D$,

$$\tilde{v}^\top (Ax - b) \geq 0.$$

  Since $(\tilde{u}, \tilde{v}, \tilde{\mu}) \neq 0$, we obtain $\tilde{v} \neq 0$. By our assumption, $\hat{x}$ is in the relative interior of $D$ and $\tilde{v}(A\hat{x} - b) = 0$, there must be some $x \in D$ near $\hat{x}$ such that $\tilde{v}^\top (Ax - b) < 0$ unless $\tilde{v}A = 0$. This leads to a contradiction to the assumption that $\text{rank}(A) = n$.

Combining the two cases above, we prove the strong duality. □

### 3.1 An example of solving optimization via duality

Now we show an example of solving optimization problems via duality. Consider the problem finding the projection to the polytope.

$$\begin{aligned} &\min_{x \in \mathbb{R}^d} \frac{1}{2}\|x - x_0\|^2 \\ &s.t. \ \langle w_i, x\rangle \le b_i, \forall i \in [m]. \end{aligned} \tag{5}$$

We compute its Lagrangian function

$$L(x, u) = \frac{1}{2}\|x - x_0\|^2 + \sum_{i=1}^{m} u_i \left(\langle w_i, x\rangle - b_i\right)$$

and let $g(u) = \inf_{x \in \mathbb{R}^d} L(x, u)$. By Slater's condition, to solve the primal optimization problem, it suffices to solve its dual problem

$$\max_{u \in \mathbb{R}^m_{\ge 0}} g(u).$$

Observe that, $L(x, u)$ is a strongly convex function. Then for $u \ge 0$, the minimizer of $L(x, u)$ is attained at the point $x(u) = x_0 - \sum_{i=1}^m u_i w_i$.

Then it holds that

$$g(u) = \langle u, \mathbf{b}\rangle - \frac{1}{2} u^\top W u$$

where

$$\begin{aligned} \mathbf{b} &\triangleq \{\langle w_i, x_0\rangle - b_i\}_{i \in [m]} \in \mathbb{R}^m \\ W &\triangleq (\langle w_i, w_j\rangle)_{i,j \in [m]} \in \mathbb{R}^{m \times m}. \end{aligned}$$

To (approximately) maximize $g(u)$ we can apply the projected gradient descent.

## 4 Karush-Kuhn-Tucker Conditions

For unconstrained optimization problem

$$\min_{x \in \mathbb{R}^d} f(x),$$

the necessary condition for $x_*$ to be a optimal point is $\nabla f(x_*) = 0$. How about the constraint optimization problem?

**Definition 7** (KKT conditions)**.** *We say a point $(x, u, v) \in \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^n$ satisfies **Karush-Kuhn-Tucker conditions** (KKT conditions for short) if the following hold:*

(a) ***Stationarity:***

$$\nabla_x L(x, u, v) = \nabla f(x) + \sum_{i=1}^{m} u_i \nabla h_i(x) + \sum_{j=1}^{n} v_j \nabla \ell_j(x) = 0.$$

(b) ***Complementary slackness:*** $u_i h_i(x) = 0, \forall i \in [n]$.

(c) ***Primal feasibility:*** $x \in D$, *i.e.*,

$$h_i(x) \le 0, \forall i \in [m],$$
$$\ell_j(x) = 0, \forall j \in [n].$$

(d) ***Dual feasibility:*** $u \ge 0$.

*Remark* 1. To see when KKT conditions hold, refer the wikipedia page KKT conditions.

## 4.1 Interpretation for KKT conditions

Before proving the necessity and the sufficiency of KKT conditions, we firstly give an interpretation of KKT conditions.

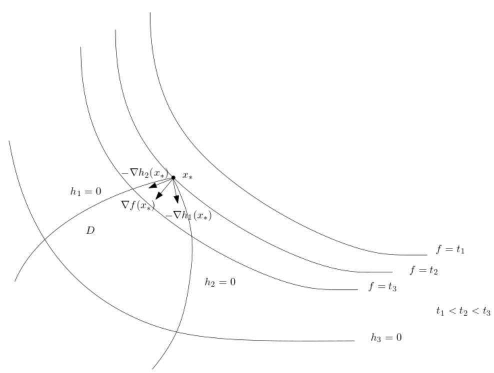


Figure 7: the geometric interpretation of KKT conditions

We interpret the primal problem as moving a particle in the space $\mathbb{R}^d$, subject to the three kinds of force fields:

- $f$ is a potential field. In this field, we need to minimize $f(x_*)$. Then the force of the field is $-\nabla f$.
- $h_i$ are one-sided constraint surfaces. The particle can move inside $h_i \le 0$. However, once $h_i(x) = 0$, the particle will be pushed inward.
- $\ell_j$ are two-sided constraint surfaces. That is, the particle is only allowed to move in these surfaces.

Then, stationarity means the sum of forces $\nabla f(x_*)$ must be balanced by a linear sum of the forces $\nabla h_i(x_*)$ and $\nabla \ell_j(x_*)$ (intuitively the point must be stationary).

Complementary slackness means, if $h_i(x_*) < 0$, then force $\nabla h_i(x_*)$ must be zero, since when the particle is not on the boundary, the one-sided force cannot be active.

Primal feasibility means the particle must be in the feasible region, and dual feasibility means all forces $\nabla g_i(x_*)$ must be one-sided.

### 4.2 Necessity and sufficiency of KKT conditions

Now we show the necessity and the sufficiency of KKT conditions.

**Lemma 8** (neccesity of KKT conditions)**.** *Suppose the function*

$$x(u,v) \triangleq \operatorname*{argmin}_{x \in \mathbb{R}^d} L(x,u,v)$$

*is $C^1$ function. Then a point $(x_*, u_*, v_*)$ is a solution to the primal problem in the sense that*

$$L(x_*, u_*, v_*) = \sup_{u \in \mathbb{R}^m_{\geq 0}, v \in \mathbb{R}^n} \inf_{x \in \mathbb{R}^d} L(x,u,v)$$

*only if KKT conditions hold.*

*Proof.* We verify KKT conditions by definition.

- **Dual feasibility:** It is trivial since $u \in \mathbb{R}^m_{\geq 0}$.
- **Stationarity:** The stationarity comes from the fact that

$$x_* \in \operatorname*{argmin}_{x \in \mathbb{R}^d} L(x, u_*, v_*).$$

- **Primal feasibility:** Assume $x_* \notin D$. There are two cases: $\exists i \in [m], h_i(x_*) > 0$; $\exists j \in [n], \ell_j(x_*)$. We assume $\exists i \in [m], h_i(x_*) > 0$ and the other case is similar (left as another mental exercise). By the chain rule,

$$\partial_{u_i} L(x(u,v),u,v) = \langle \nabla_x L(x(u,v),u,v), \nabla_{u_i} x(u,v) \rangle + h_i(x(u,v)).$$

  Then, for $\delta > 0$,

$$\begin{aligned}
&L(x(u_* + \delta \mathbf{e}_i, v_*), u_* + \delta \mathbf{e}_i, v_*) \\
&= L(x(u_*, v_*), u_*, v_*) + \langle \nabla_x L(x(u_*, v_*), u_*, v_*), \nabla_{u_i} x(u_*, v_*) \rangle \, \delta \\
&+ h_i(x(u_*, v_*))\delta + o(\delta) \\
&= L(x_*, u_*, v_*) + h_i(x_*)\delta + o(\delta) && \text{(stationarity)} \\
&> L(x_*, u_*, v_*) && \text{(when } \delta \text{ is small.)}
\end{aligned}$$

  which leads to a contradiction to the optimality of $(x_*, u_*, v_*)$

- **Complementary slackness:** Assume that $h_i(x_*) < 0$ and $(u_*)_i > 0$ for some $i \in [m]$. Similarly to the calculation above we show

$$L(x(u_* - \delta \mathbf{e}_i, v_*), u_* - \delta \mathbf{e}_i, v_*) = L(x_*, u_*, v_*) - h_i(x_*)\delta + o(\delta) > L(x_*, u_*, v_*)$$

  thus also leading to a contradiction.

□

On the other hand, we establish that it suffices to show KKT conditions of a point to show its optimality.

**Lemma 9** (sufficiency of KKT conditions)**.** *Suppose that Slater's condition holds. Then a point* $(x_*, u_*, v_*)$ *is a solution to the primal if KKT conditions hold.*

*Proof.* By the stationarity and the convexity of the function $x \mapsto L(x, u_*, v_*)$, we know $x_*$ is a global minimizer of $L(x, u_*, v_*)$. Then

$$g(u_*, v_*) = L(x_*, u_*, v_*) = f(x_*)$$

where the last equality holds by the primal feasibility and complementary slackness. Then by Corollary 3, we show the optimality. □

## References

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.