ME 7247: Advanced Control Systems	**Fall 2022–23**

Lecture 18: Bounding and Duality

Tuesday, November 8, 2022

Lecturer: Laurent Lessard *Scribe: Milad Alipour Shahraki*

In this lecture, we continue our discussion on optimization problems and convex optimization problems in particular. So we talked last time about what the general optimization problem looks like. Now we will talk about the classification of optimization problems. After that we will look into the bounding and duality.

1 Classification of Optimization Problems

Recall a general optimization problem:

$$\begin{array}{llll} \underset{x}{\text{minimize}} & f(x) & & \text{(Objective)} \\ \text{s.t.} & g_i(x) \leq 0 & i = 1, \dots, m & \text{(Inequality Constraints)} \\ & h_j(x) = 0 & j = 1, \dots, p & \text{(Equality Constraints)} \end{array}$$

There are different kinds of optimization problems with different names and those names are based on the type of constraints. The classification based on the nature of the objective and constraints is summarized in the following table:

Objective	**Constraints**	**Name**
Linear	$g_i(x)$ linear, $h_j(x)$ linear	Linear Program (LP)
Quadratic	$g_i(x)$ linear, $h_j(x)$ linear	Quadratic Program (QP)
Quadratic	$g_i(x)$ quadratic, $h_j(x)$ linear	Quadratically Constrained QP (QCQP)
Convex	$g_i(x)$ Convex, $h_j(x)$ Linear	Convex Program (CP)

Table 1: Classification of optimization problems. In all problems above, *quadratic* means *convex quadratic*. So a function of the form $x^\mathsf{T} Q x + 2p^\mathsf{T} x + r$, where $Q \succeq 0$.

Note: These optimization problems are nested. For example, an LP is a special case of QP, which is a special case of QCQP, which is a special case of a convex program.

‘

1.1 Geometry of the Problem

Now we can analyze the geometry of the optimization problems.

- **Linear Program (LP):** The feasible set is a polytope for an LP. Therefore, for a linear objective function, the minimum or maximum is always going to occur at a vertex unless one of the faces happens to be perfectly aligned with f in which case all the points on that face will be optimal. (Fig. 1)

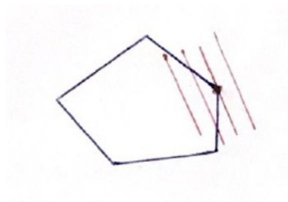

Figure 1: Linear Program (LP)

- **Quadratic Program (QP):** In this case, the feasible set forms a polyhedron. We have convex quadratic as our objective; therefore, the contours are going to be ellipsoids. Now if the minimum point of this convex quadratic objective lies inside the polyhedron, then this is the optimal point but if it lies outside the polyhedron, then the optimal point lies somewhere on the boundary. (Fig. 2)

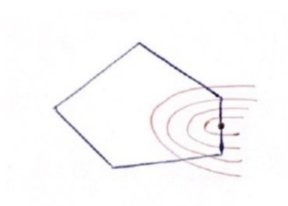

Figure 2: Quadratic Program (QP)

- **Quadratically Constrained Quadratic Program (QCQP):** In case of a QCQP, the feasible set will be an intersection of multiple quadratic (ellipsoids) and linear constraints. So the boundary at some sides may be curved. The optimal point then would lie either inside, or on the boundary of the feasible set. (Fig. 3)

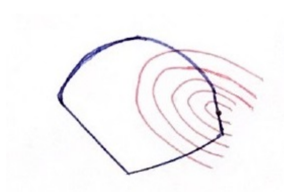

Figure 3: Quadratically Constrained Quadratic Program (QCQP)

1.2 Problem Transformation

It is possible to transform a QCQP or a general convex program into an equivalent problem with a linear objective, by writing it in *epigraph form.* Consider the general optimization problem:

$$\begin{aligned}\min_{x} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_j(x) = 0 \quad j = 1, \ldots, p\end{aligned}$$

Now we can take the convex problem, and convert it to another convex problem but with a linear objective by introducing the variable t:

$$\begin{aligned}\min_{x} \quad & t \\ \text{s.t.} \quad & f(x) \leq t \\ & g_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_j(x) = 0 \quad j = 1, \ldots, p\end{aligned}$$

This is now an optimization problem with a linear objective that has the exact solution as before. Therefore, we can always increase the dimension of the space by one and have the solution end up on the boundary of the feasible set.

Note: In general, there are specialty solvers that are designed for just solving linear programs or just solving quadratic programs, and there are solvers that solve convex problems. The more specialized the solver is, the faster it will be. So if you have an LP, you should always use LP solvers, as a QP or general CP solver would work but would be much less efficient.

2 Bounding and Duality

In this section, we are going to talk about the bounding and duality.

2.1 Bounding

Oftentimes when we have an optimization problem, solving it is going to be in some sort of a approximate way, or in an iterative fashion, where at every step we get a little bit closer to the solution. In other words, finding the lower and upper bounds can help us know how close we are to the solution, especially when we cannot solve the optimization problem exactly, or it is difficult to solve. The nice thing about convex optimization problems is that there are really good ways for obtaining both upper and lower bounds. Recall a simplified optimization problem:

$$\begin{aligned}\min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in C\end{aligned} \tag{1}$$

where C denotes the feasible set that satisfies all the constraints.

Bounds using the objective. Suppose we can find can functions $\underline{f}(x)$ and $\bar{f}(x)$ such that $\underline{f}(x) \leq f(x) \leq \bar{f}(x)$ for all x, then we can bound the solution to the optimization problem using

$$\boxed{\begin{array}{ll} \min\limits_x & \bar{f}(x) \\ \text{s.t.} & x \in C \end{array}} \quad \geq \quad \boxed{\begin{array}{ll} \min\limits_x & f(x) \\ \text{s.t.} & x \in C \end{array}} \quad \geq \quad \boxed{\begin{array}{ll} \min\limits_x & \underline{f}(x) \\ \text{s.t.} & x \in C \end{array}}$$

This is useful when we have a problem that is not convex and the only source of non-convexity is the objective function. In this case, we could look for functions that bound it (and are easier to optimize), which will allow us to bound the solution of the difficult problem. The inequality stays the same if we replace the min by a max.

Bounds using the feasible set. Another way is to find a set that is larger than and a set that is smaller than our feasible set in (1) such as $\underline{S} \subseteq S \subseteq \bar{S}$. This leads to:

$$\boxed{\begin{array}{ll} \min\limits_x & f(x) \\ \text{s.t.} & x \in \underline{S} \end{array}} \quad \geq \quad \boxed{\begin{array}{ll} \min\limits_x & f(x) \\ \text{s.t.} & x \in S \end{array}} \quad \geq \quad \boxed{\begin{array}{ll} \min\limits_x & f(x) \\ \text{s.t.} & x \in \bar{S} \end{array}}$$

This is useful when we have a difficult constraint that we can replace by a set containing it or a set contained by it. If we replace the mins by maxes, the inequalities will be reversed.

Due to the epigraph form, it should come as no surprise that if there is a way to bound an optimization problem by bounding its objective function, there should be a corresponding way to bound the problem by bounding its feasible set. This will allow us to deal with non-convex constraints. We can outer bound them, or we can inner bound them by using nicer sets over which we can solve the optimization problem. (Fig. 4)

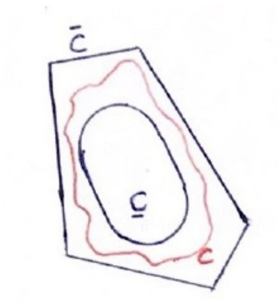

Figure 4: Convex lower and upper bounds of a non-convex set.

Note: If we want to find a set that is larger than S we can simply remove some of the constraints. Similarly, we can find a set that is smaller than S by adding constraints.

There is also a really simple way to find a bound which only works for upper bounding and that is just pick any feasible point! This leverages the fact that $f(x) \geq \min_x f(x)$ for all x. So we can always find an upper bound to a minimization problem by just picking any feasible point. Likewise, we can lower-bound any maximization problem by picking any feasible point. There is no analogous approach when it comes to lower-bounding. This brings us to a concept called *duality*.

2.2 Duality

The way we should think about duality is that it is a way of lower-bounding a minimization problem or upper-bounding a maximization problem. Consider the original optimization problem:

$$P = \boxed{\begin{array}{lll} \min\limits_x & f(x) & \\ \text{s.t.} & g_i(x) \leq 0 & i = 1, \ldots, m \\ & h_j(x) = 0 & j = 1, \ldots, p \end{array}} \tag{2}$$

where P stands for primal which is the opposite of dual. Usually, the original optimization problem is called the *primal* problem. Now we are going to derive something that called the *dual* optimization problem. It is an optimization problem but instead of being a minimization, it is a maximization. Solving the dual gives us a lower bound to the solution of the primal.

The first thing we will do is replace $f(x)$ by a smaller function found by adding non-positive terms that come from the constraints.

$$P = \boxed{\begin{array}{lll} \min\limits_x & f(x) & \\ \text{s.t.} & g_i(x) \leq 0 & i = 1, \ldots, m \\ & h_j(x) = 0 & j = 1, \ldots, p \end{array}} \underset{\text{pick } \lambda_i \geq 0}{\geq} \boxed{\begin{array}{lll} \min\limits_x & f(x) + \sum\limits_{i=1}^m \lambda_i g_i(x) + \sum\limits_{j=1}^p \mu_j h_j(x) & \\ \text{s.t.} & g_i(x) \leq 0 \quad i = 1, \ldots, m & \\ & h_j(x) = 0 \quad j = 1, \ldots, p & \end{array}} \tag{3}$$

Here if we pick $\lambda_i \geq 0$, then because $g_i(x) \leq 0$, each of the terms of the first sum is less than or equal to zero. Also, since $h_j(x) = 0$, then each of the terms of the second sum is zero. Note that the λ_i and μ_j are not optimization variables. They are *parameters*. Each different choice of λ_i and μ_j gives a different lower bound. Further lower-bound by removing all the constraints:

$$P \geq \boxed{\begin{array}{lll} \min\limits_x & f(x) + \sum\limits_{i=1}^m \lambda_i g_i(x) + \sum\limits_{j=1}^p \mu_j h_j(x) & \\ \text{s.t.} & g_i(x) \leq 0 \quad i = 1, \ldots, m & \\ & h_j(x) = 0 \quad j = 1, \ldots, p & \end{array}} \geq \underbrace{\boxed{\min\limits_x \; f(x) + \sum\limits_{i=1}^m \lambda_i g_i(x) + \sum\limits_{j=1}^p \mu_j h_j(x)}}_{F(\lambda,\mu)} \tag{4}$$

Each λ, μ with $\lambda \geq 0$ leads to a different lower bound $F(\lambda, \mu)$. For certain choices, we might have $F(\lambda, \mu) = -\infty$, which is corresponds to the case where (4) is unbounded. This isn't a problem, it's just not a particularly useful lower bound! Lower bounds derived in (3) and (4) are functions of the λ_i and μ_j. We call $F(\lambda, \mu)$ the *dual function*. The *dual problem* consists of finding the best possible lower bound achievable using this approach. In other words, maximize the dual function.

$$p^\star = \underbrace{\boxed{\begin{array}{lll} \min\limits_x & f(x) & \\ \text{s.t.} & g_i(x) \leq 0 & i = 1, \ldots, m \\ & h_j(x) = 0 & j = 1, \ldots, p \end{array}}}_{\text{Primal Problem}} \geq \underbrace{\boxed{\begin{array}{ll} \max\limits_{\lambda,\mu} & F(\lambda, \mu) \\ \text{s.t.} & \lambda \geq 0 \end{array}}}_{\text{Dual Problem}} = d^\star \tag{5}$$

The original problem (2) is usually called the *Primal Problem* (with optimal objective $p^\star$), and the problem of maximizing the dual function is called the *Dual Problem* (with optimal objective $d^\star$). Considering (5), we can write

$$f(x) \text{ for any primal-feasible } x \quad \geq \quad p^\star \quad \geq \quad d^\star \quad \geq \quad F(\lambda, \mu) \text{ for any dual-feasible } (\lambda, \mu) \tag{6}$$

Therefore, we can simultaneously search for x's that give us smaller and smaller upper bounds, and search for (λ, μ)'s that give us larger and larger lower bounds. Eq. (6) is called *weak duality*, and weak duality always holds. If we have $p^\star = d^\star$, we call it *strong duality.* Obviously strong duality is a desirable property. With strong duality, we have a great way to bound the problem. We can solve the primal and dual at the same time, and have a way to estimate how close we are to the optimal solution without knowing its exact value.

It would be nice to have a way of knowing that strong duality holds without having to solve the problem. This requires some sort of property of the optimization problem that is easy to verify. Such properties are called *constraint qualifications.* Here are two useful ones.

1. If the problem is a linear program (LP), then it always satisfies strong duality.
2. If an optimization problem is convex and strictly feasible, then it satisfies strong duality. This is known as *Slater's condition.*

Note: Strict feasibility means that there exists at least one feasible x for which all the inequalities are strict. That is, there exists x such that $g_i(x) < 0$ for all i and $h_j(x) = 0$ for all j.

We can slightly relax Slater's condition by requiring that strict feasibility only needs to hold for the g_i's that are non-linear.

The dual optimization problem is always a convex optimization problem even if $g_i(x)$ and $h_j(x)$ are not convex. In other words, the $F(\lambda, \mu)$ is concave and therefore, the dual problem (maximizing a concave function) is a convex optimization problem. To understand this, ignore μ for now and consider just λ. If we have an optimization problem such as:

$$\min \left\{ c_1^\top \lambda, c_2^\top \lambda, \ldots, c_n^\top \lambda \right\} = \min_i \; c_i^\top \lambda \tag{7}$$

we end up with Fig. 5, which is a concave function. The same holds true if we have infinitely many c_i, which we write as $c(x)$ where x is a real number that indexes each c. Now replace (7) by $\min_x c(x)^\top \lambda$. The result is again a concave function. Similarly, $F(\lambda, \mu)$ is a concave function of (λ, μ), which means that the dual problem is to maximize a concave function; i.e. minimizing a convex function!

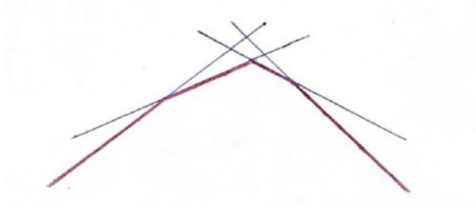

Figure 5: Illustration of the minimum of a collection of linear functions

2.2.1 Dual of a Linear Program

Consider a linear program of the form:

$$P = \boxed{\begin{array}{ll} \min_x & c^\top x \\ \text{s.t.} & Ax \le b \\ & Hx = g \end{array}} \tag{8}$$

We can find a lower bound for (8) by picking $\lambda \ge 0$ and writing

$$P = \boxed{\begin{array}{ll} \min_x & c^\top x \\ \text{s.t.} & Ax \le b \\ & Hx = g \end{array}} \quad \ge \quad \boxed{\min_x \quad c^\top x + \lambda^\top (Ax - b) + \mu^\top (Hx - g)}$$

which we can rearrange the terms and write

$$P \quad \ge \quad \boxed{\min_x \quad c^\top x + \lambda^\top (Ax - b) + \mu^\top (Hx - g)} \quad = \quad \min_x \quad (c^\top + \lambda^\top A + \mu^\top H)x - (\lambda^\top b + \mu^\top g)$$

$$= \min_x \quad \big(\underbrace{c + A^\top \lambda + H^\top \mu}_{q}\big)^\top x - (\lambda^\top b + \mu^\top g)$$

where $A \in \mathbb{R}^{m \times n}$ and $H \in \mathbb{R}^{p \times n}$. The only way that this minimization is going to be finite is if $q = 0$, so we have:

$$F(\lambda, \mu) = \begin{cases} -(\lambda^\top b + \mu^\top g), & \text{if } c + A^\top \lambda + H^\top \mu = 0 \\ -\infty, & \text{otherwise.} \end{cases}$$

So the dual problem is

$$D = \boxed{\begin{array}{ll} \max_{\lambda,\mu} & -\lambda^\top b - \mu^\top g \\ \text{s.t.} & c + A^\top \lambda + H^\top \mu = 0 \\ & \lambda \ge 0 \end{array}}$$

This is still a linear program. We can also write it as

$$D = \boxed{\begin{array}{ll} \max_{\lambda,\mu} & \begin{bmatrix} -b \\ -g \end{bmatrix}^\top \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \\ \text{s.t.} & \begin{bmatrix} -1 & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \le 0 \\ & \begin{bmatrix} A^\top & H^\top \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} + c = 0. \end{array}}$$

Therefore, we can conclude that the dual of a linear program is a linear program.

Note: Although we can solve the problem by either solving the primal problem or the dual problem, it is important to recognize that these optimization problems are over different spaces. The primal is an optimization over n variables ($x \in \mathbb{R}^n$) and $m+p$ constraints, whereas the dual is an optimization over $m+p$ variables ($(\lambda, \mu) \in \mathbb{R}^{m+p}$) and $m+n$ constraints. So the number of variables in the dual problem is the number of constraints in the primal problem.

2.3 Complementary Slackness

Suppose we have strong duality, and if we evaluate the objective at the optimal x, i.e. $x^\star$, then

$$\begin{aligned}
f(x^\star) &= F(\lambda^\star, \mu^\star) \\
&= \min_x \quad f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_i h_i(x) \qquad \text{(by definition)} \\
&\leq f(x^\star) + \underbrace{\sum_{i=1}^{m} \lambda_i^\star g_i(x^\star)}_{\leq 0} + \underbrace{\sum_{j=1}^{p} \mu_i^\star h_i(x^\star)}_{=0} \quad \text{(upper bounding using the primal optimal } x^\star) \\
&\leq f(x^\star) \quad \text{(upper bounding by removing the above sum terms)}
\end{aligned}$$

We started with $f(x^\star)$, and we ended with $f(x^\star)$. This means that each of the stages of the above inequalities are equality hold. In particular, each term of the sums are equal to zero,

$$\boxed{\lambda_i^\star g_i(x^\star) = 0} \qquad \text{for } i = 1, \dots, m. \tag{9}$$

Since λ_i is the dual variable that corresponds to the constraint $g_i(x) \leq 0$, Eq. (9) tells us that if the constraint g_i is *slack* at the optimal point (inequality is strict), the corresponding dual variable will be *binding* (equal to zero). Likewise, if we solve the dual problem and find that a certain dual variable is slack ($\lambda_i > 0$), then the corresponding inequality constraint must be binding (equal to zero). In summary,

1. If a primal constraint is slack, then the corresponding dual variable is binding. In other words, if $g_i(x^\star) > 0$, then $\lambda_i^\star = 0$.
2. If a dual variable is slack, then the corresponding primal constraint is binding. In other words, if $\lambda_i^\star > 0$ is slack, then $g_i(x^\star) = 0$.

The property (9) is called *complementary slackness*. Note that the converses of these statements does not hold. For example, if we find out that $g_i(x^\star) = 0$, we cannot conclude anything about $\lambda_i^\star$. Now, we will talk about the necessary conditions for optimality.

2.4 KKT Conditions

The Karush–Kuhn–Tucker (KKT) conditions are necessary conditions for optimality for a constrained optimization problem.

Consider an optimization problem of the form

$$P \quad = \quad \begin{cases} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_j(x) = 0 \quad j = 1, \ldots, p \end{cases}$$

Suppose f, g_i, and h_j are differentiable functions, strong duality holds, and $x^\star$ solves P. Define the following function, called the *Lagrangian*:

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{n} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x).$$

There exists λ and μ such that the following four conditions hold:

1. Stationarity: $\nabla_x L(x^\star, \lambda, \mu) = 0$.
2. Primal Feasibility: $g_i(x^\star) \leq 0$ and $h_j(x^\star) = 0$ for all i and j.
3. Dual Feasibility: $\lambda_i \geq 0$ for all i.
4. Complementary Slackness: $\lambda_i g_i(x^\star) = 0$ for all i.

The KKT conditions are *necessary conditions for optimality.* So when a point is optimal, the KKT conditions hold. The converse is *not true.* So if the KKT conditions hold for a particular point, this does not mean the point is necessarily optimal. Given a general optimization problem we can write down the KKT conditions and try to find points that satisfy them, then we can use that as a list of candidates to check which one will be the optimal solution.

Unconstrained optimization. In the special case of an unconstrained optimization problem, all x are feasible. The KKT conditions reduce to only the stationarity condition, which is the familiar $\nabla_x f(x) = 0$. Again, this is a necessary condition for optimality but it is not sufficient. If a point is optimal, the derivative is zero. But if the derivative is zero, the point need not be optimal (it could be a minimum, a maximum, or a saddle point).

Only equality constraints. If the optimization problem only has equality constraints, then we do not have any λ dual variables. The KKT conditions reduce to only stationarity condition, which is $\nabla_x L(x^\star, \mu) = 0$, and the primal feasibility condition, which is $h_j(x^\star) = 0$. We can also write primal feasiblity as $\nabla_\mu L(x^\star, \mu) = 0$. In this scenario, the μ_j are called *Lagrange multipliers.*