Duality Theory for Constrained Optimization

A tricky thing in duality theory is to decide what we call the domain or *ground set* $\mathcal{D}$ and what we call the constraints g_i's or h_j's. Based on whether constraints are explicitly stated or implicitly stated in the form of the ground set, the dual problem could be very different. Thus, many duals are possible for the given primal.

For the rest of the discussion $\mathcal{D}$ will mostly mean $\Re^n$

Formally: The Dual Theory for Constrained Optimization

Consider the general constrained minimization problem

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{D}} \quad & f(\mathbf{x}) \\
\text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, 2, \ldots, m \\
\text{subject to} \quad & h_j(\mathbf{x}) = 0,\ j = 1, 2, \ldots, n
\end{aligned}
\tag{71}
$$

- Consider forming the lagrange function by associating prices (called lagrange multipliers) λ_i and μ_j , with constraints involving g_i and h_j respectively.

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{n} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{n} \mu_j h_j(\mathbf{x}) = f(\mathbf{x}) + \lambda^T \mathbf{g}(\mathbf{x}) + \mu^T \mathbf{h}(\mathbf{x})$$

- At each feasible $\mathbf{x}$, for fixed $\lambda_i \geq 0\, \forall i \in \{1..m\}$,

Formally: The Dual Theory for Constrained Optimization

Consider the general constrained minimization problem

$$\begin{array}{ll} \min_{\mathbf{x}\in\mathcal{D}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0,\ i = 1, 2, \ldots, m \\ \text{subject to} & h_j(\mathbf{x}) = 0,\ j = 1, 2, \ldots, n \end{array} \tag{71}$$

- Consider forming the lagrange function by associating prices (called lagrange multipliers) λ_i and μ_j , with constraints involving g_i and h_j respectively.

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{n} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{n} \mu_j h_j(\mathbf{x}) = f(\mathbf{x}) + \lambda^T \mathbf{g}(\mathbf{x}) + \mu^T \mathbf{h}(\mathbf{x})$$

- At each feasible $\mathbf{x}$, for fixed $\lambda_i \geq 0\, \forall i \in \{1..m\}$,

$$f(\mathbf{x}) \geq \quad L(\mathbf{x}, \lambda, \mu) \quad \text{if } g_i(\mathbf{x}) \leq 0 \ \&\ h_j(\mathbf{x}) = 0 \tag{72}$$

Formally: The Dual Theory for Constrained Optimization

- For $\lambda_i \geq 0 \, \forall i \in \{1..m\}$ and μ_j, minimizing the right hand side of (72) over all feasible $\mathbf{x}$

$$f(\mathbf{x}) \geq \min_{\mathbf{x} \;\; s.t \;\; g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x}) = 0} L(\mathbf{x}, \lambda, \mu) \quad \triangleq L^*(\lambda, \mu) \tag{73}$$

- $L^*(\lambda, \mu)$ is a pointwise (w.r.t $\mathbf{x} \in g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x}) = 0$) minimum of linear functions ($L(\mathbf{x}, \lambda, \mu)$) and is therefore always a concave function of \lambdas and \mus

Formally: The Dual Theory for Constrained Optimization

- For $\lambda_i \geq 0 \, \forall i \in \{1..m\}$ and μ_j, minimizing the right hand side of (72) over all feasible $\mathbf{x}$

$$f(\mathbf{x}) \geq \min_{\mathbf{x} \ s.t \ g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x})=0} L(\mathbf{x}, \lambda, \mu) \triangleq L^*(\lambda, \mu) \quad (73)$$

- $L^*(\lambda, \mu)$ is a pointwise (w.r.t $\mathbf{x} \in g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x}) = 0$) minimum of linear functions ($L(\mathbf{x}, \lambda, \mu)$) and is therefore always a concave function.
- Since $f(\mathbf{x}) \geq L^*(\lambda, \mu)$ for all primal feasible $\mathbf{x}$ and dual feasible *i.e.,* $\lambda_i \geq 0$ and μ_j, , we can maximize the lower bound $L^*(\lambda, \mu)$ to give the following **Dual Problem**

Dual opt is always a convex optimization problem and always gives you a lower bound to the primal

$$\begin{aligned} &\max_{\lambda \in \Re^m, \mu \in \Re^p} && L^*(\lambda, \mu) \\ &\text{subject to} && \lambda \geq \mathbf{0} \end{aligned} \quad (74)$$

Theorem

(i) The dual function $L^*(\lambda, \mu)$ is always concave. (ii) If p^* is solution of (71) and d^* of (74) then $p^* \geq d^*$

Formally: The Dual Theory for Constrained Optimization (contd.)

Proof by first principles:

Formal Proof for Part (i): Consider two values of the dual variables, *viz.*, $\lambda_1 \geq \mathbf{0}$ and $\lambda_2 \geq \mathbf{0}$ as well as μ_1 and μ_2 (with no constraints). Let $\lambda = \theta\lambda_1 + (1-\theta)\lambda_2$ and $\mu = \theta\mu_1 + (1-\theta)\mu_2$ for any $\theta \in [0,1]$. Then,

$$\begin{aligned}
L^*(\lambda,\mu) &= \min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x}) + \lambda^T g(\mathbf{x}) + \mu^T h(\mathbf{x}) \\
&= \min_{\mathbf{x}\in\mathcal{D}} \theta \left[f(\mathbf{x}) + \lambda_1^T g(\mathbf{x}) + \mu_1^T h(\mathbf{x})\right] + (1-\theta)\left[f(\mathbf{x}) + \lambda_2^T g(\mathbf{x} + \mu_2^T h(\mathbf{x}))\right] \\
&\geq \min_{\mathbf{x}\in\mathcal{D}} \theta \left[f(\mathbf{x}) + \lambda_1^T g(\mathbf{x}) + \mu_1^T h(\mathbf{x})\right] + \min_{\mathbf{x}\in\mathcal{D}} (1-\theta)\left[f(\mathbf{x}) + \lambda_2^T g(\mathbf{x}) + \mu_2^T h(\mathbf{x})\right] \\
&= \theta L^*(\lambda_1,\lambda_1) + (1-\theta) L^*(\lambda_2,\lambda_2)
\end{aligned}$$

min of sum >= sum of mins

This proves that $L^*(\lambda)$ is a concave function.

Formally: The Dual Theory for Constrained Optimization (contd.)

Formal Proof for Part (ii):
If $\widehat{\mathbf{x}}$ is a feasible solution to the primal problem (71) and $\widehat{\lambda}$ is a feasible solution to the dual problem (74), then

$$f(\widehat{\mathbf{x}}) \geq f(\widehat{\mathbf{x}}) + \widehat{\lambda}^T \mathbf{g}(\widehat{\mathbf{x}}) \geq \min_{\textit{Feasible } \widehat{\mathbf{x}} \in \mathcal{D}} f(\widehat{\mathbf{x}}) + \widehat{\lambda}^T \mathbf{g}(\widehat{\mathbf{x}}) = L^*(\widehat{\lambda})$$

That is,

$$f(\widehat{\mathbf{x}}) \geq L^*(\widehat{\lambda})$$

A direct consequence of this is that (since the min over x and max over \lambda are over disjoint sets of variables)

$$p^* = \min_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x}) \geq \max_{\lambda \geq \mathbf{0}} L^*(\lambda) = d^*$$

This proves the second part of the theorem.

The Dual Theory for Constrained Optimization: Examples and Graphical Interpretation

- The dual is concave (or the negative of the dual is convex) irrespective of the primal. Solving the dual is therefore always a convex programming problem.
- In some sense, the dual is better structured than the primal. However, the dual cannot be drastically simpler than the primal.
- For example, if the primal is not a Linear Program, the dual cannot be an LP.
- Similarly, the dual can be quadratic only if the primal is quadratic.
- We will look at two examples to give a flavour of how the duality theory works.

Useful in (a) sometimes simplified parametrization of dual problem
(b) algorithms that explicitly invoke dual (eg: primal - dual interior point)
(c) Characterizing convergence by estimating duality gap

Example Derivations of Dual

Example Derivations of the Dual: Linear Programs

$$\begin{aligned} \min_{\mathbf{x} \in \Re^n} \quad & \mathbf{c}^T \mathbf{x} \\ \text{subject to} \quad & -A\mathbf{x} + \mathbf{b} \leq \mathbf{0} \end{aligned}$$

The lagrangian for this problem is:

Example Derivations of the Dual: Linear Programs

$$\begin{array}{ll} \min\limits_{\mathbf{x}\in\Re^n} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & -A\mathbf{x}+\mathbf{b}\leq\mathbf{0} \end{array}$$

The lagrangian for this problem is:

$$L(\mathbf{x},\lambda)=\mathbf{c}^T\mathbf{x}+\lambda^T\mathbf{b}-\lambda^TA\mathbf{x}=\mathbf{b}^T\lambda+\mathbf{x}^T\left(\mathbf{c}-A^T\lambda\right)$$

The next step is to get L^*, which we obtain using the first derivative test:

Example Derivations of the Dual: Linear Programs

$$\begin{array}{ll} \min\limits_{\mathbf{x}\in\Re^n} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & -A\mathbf{x}+\mathbf{b}\leq\mathbf{0} \end{array}$$

The lagrangian for this problem is:

$$L(\mathbf{x},\lambda)=\mathbf{c}^T\mathbf{x}+\lambda^T\mathbf{b}-\lambda^TA\mathbf{x}=\mathbf{b}^T\lambda+\mathbf{x}^T\left(\mathbf{c}-A^T\lambda\right)$$

The next step is to get L^*, which we obtain using the first derivative test:

$$L^*(\lambda)=\min_{\mathbf{x}\in\Re^n}\mathbf{b}^T\lambda+\mathbf{x}^T\left(\mathbf{c}-A^T\lambda\right)=\begin{cases}\mathbf{b}^T\lambda & \text{if } A^T\lambda=\mathbf{c}\\ -\infty & \text{if } A^T\lambda\neq\mathbf{c}\end{cases}$$

Example Derivations of the Dual: Linear Programs (contd.)

The function L^* can be thought of as the extended value extension of the same function restricted to the domain $\left\{\lambda | A^T\lambda = \mathbf{c}\right\}$. Therefore, the dual problem can be formulated as:

Example Derivations of the Dual: Linear Programs (contd.)

The function L^* can be thought of as the extended value extension of the same function restricted to the domain $\left\{\lambda | A^T\lambda = \mathbf{c}\right\}$. Therefore, the dual problem can be formulated as:

$$\begin{aligned} \max_{\lambda \in \Re^m} \quad & \mathbf{b}^T\lambda \\ \text{subject to} \quad & A^T\lambda = \mathbf{c} \\ & \lambda \geq \mathbf{0} \end{aligned} \qquad (75)$$

This is the dual of the standard LP. What if the original LP was the following?

You can refer to the generalization of LPs to conic linear programs (with constraint that Ax + b belongs to cone) and their conic dual linear programs discussed at length in previous (pre-midsem part of) offerings of this course at https://www.cse.iitb.ac.in/~cs709/calendar2015.html

$$\begin{aligned} \min_{\mathbf{x} \in \Re^n} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & -A\mathbf{x} + \mathbf{b} \leq \mathbf{0} \quad \mathbf{x} \geq \mathbf{0} \end{aligned} \qquad (***)$$

Now we have a variety of options based on what constraints are introduced into the ground set (or domain) and what are explicitly treated as constraints. Some working out will convince us that treating $\mathbf{x} \in \Re^n$ as the constraint and the explicit constraints as part of the ground set is a very bad idea. One dual for this problem can be derived similarly as (75).

Example Derivations of the Dual: Variant of LP (contd.)

Let us look at a modified version of the problem in (76).

$$\begin{aligned} \min_{\mathbf{x} \in \Re^n} \quad & \mathbf{c}^T\mathbf{x} - \sum_{i=1}^{n} \ln x_i \\ \text{subject to} \quad & -A\mathbf{x} + \mathbf{b} = \mathbf{0} \\ & \mathbf{x} > \mathbf{0} \end{aligned}$$

(possibly underdetermined)

We first formulate the lagrangian for this problem.

We can choose to have x > 0 as part of the domain since ln(xi) already kind of discourages xi from getting close to 0 [Spirit of Barrier methods]

Example Derivations of the Dual: Variant of LP (contd.)

Let us look at a modified version of the problem in (76).

$$\begin{aligned} \min_{\mathbf{x}\in\Re^n} \quad & \mathbf{c}^T\mathbf{x} - \textstyle\sum_{i=1}^n \ln x_i \\ \text{subject to} \quad & -A\mathbf{x} + \mathbf{b} = \mathbf{0} \\ & \mathbf{x} > \mathbf{0} \end{aligned}$$

We first formulate the lagrangian for this problem.

$$L(\mathbf{x},\lambda) = \mathbf{c}^T\mathbf{x} - \sum_{i=1}^n \ln x_i + \lambda^T\mathbf{b} - \lambda^T A\mathbf{x} = \mathbf{b}^T\lambda + \mathbf{x}^T\left(\mathbf{c} - A^T\lambda\right) - \sum_{i=1}^n \ln x_i$$

The domain (or ground set) for this problem is $\mathbf{x} > \mathbf{0}$, which is open.

Example Derivations of the Dual: Variant of LP

- The expression for L^* can be obtained using the first derivative test, while keeping in mind that L can be made arbitrarily small (tending to $-\infty$) unless $(\mathbf{c} - A^T\lambda) > \mathbf{0}$.
- This is because, even if one component of $\mathbf{c} - A^T\lambda$ is less than or equal to zero, the value of L can be made arbitrarily small by decreasing the value of the corresponding component of $\mathbf{x}$ in the $\sum_{i=1}^{n} \ln x_i$ part.
- Further, the sum $\mathbf{b}^T\lambda + \mathbf{x}^T\left(\mathbf{c} - A^T\lambda\right) - \sum_{i=1}^{n} \ln x_i$ can be separated out into the individual components of λ_i, and this can be exploited while determining the critical point of L.

$$L^*(\lambda) = \min_{\mathbf{x}>\mathbf{0}} L(\mathbf{x}, \lambda) = \begin{cases} \mathbf{b}^T\lambda + n - \sum_{i=1}^{n} \ln \frac{1}{(\mathbf{c}-A^T\lambda)_i} & \text{if } (\mathbf{c} - A^T\lambda) > \mathbf{0} \\ -\infty & \text{otherwise} \end{cases}$$

Example Derivations of the Dual: Variant of LP (contd.)

Finally, the dual will be

$$
\begin{array}{ll}
\max\limits_{\lambda \in \Re^m} & \mathbf{b}^T\lambda + n + \sum_{i=1}^{n} \ln \frac{1}{\left(\mathbf{c} - A^T\lambda\right)_i} \\
\text{subject to} & -A^T\lambda + \mathbf{c} > \mathbf{0}
\end{array}
$$

Geometry of Duality

It turns out that all the intuitions we need are in two dimensions, which makes it fairly convenient to understand the idea.

Geometry of Duality

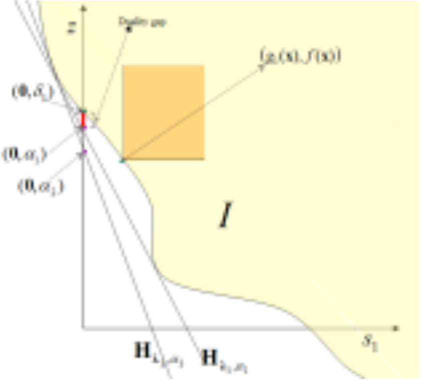

Figure 16: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- We will study the geometry of the dual in the *column space* $\Re^{m+1}$. Define $\mathcal{I} \subseteq \Re^{m+1}$ as

$$\mathcal{I} = \{(\mathbf{s}, z)\}\ \text{s.t} \begin{cases} \mathbf{s} \in \Re^m & \exists \mathbf{x} \in \mathcal{D}\ s.t\ g_i(\mathbf{x}) \leq s_i\ \forall i \\ z \in \Re & f(\mathbf{x}) \leq z \end{cases}$$

  - Recap: Any (linear) equality constraint $h(\mathbf{x}) = 0$ can be expressed using

    two inequality constraints
    h(x) <=0
    -h(x) <= 0

Geometry of Duality

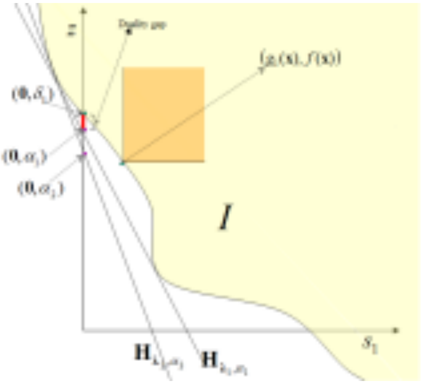

Figure 16: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- We will study the geometry of the dual in the *column space* $\Re^{m+1}$. Define $\mathcal{I} \subseteq \Re^{m+1}$ as

$$\mathcal{I} = \{(\mathbf{s}, z)\} \text{ s.t } \begin{cases} \mathbf{s} \in \Re^m & \exists \mathbf{x} \in \mathcal{D} \text{ } s.t \text{ } g_i(\mathbf{x}) \leq s_i \ \forall i \\ z \in \Re & f(\mathbf{x}) \leq z \end{cases}$$

  - Recap: Any (linear) equality constraint $h(\mathbf{x}) = 0$ can be expressed using two (convex) inequality constraints, *viz.*, $h(\mathbf{x}) \leq 0$ and $-h(\mathbf{x}) \leq 0$.

- Figure 16 illustrates in $\Re^2$ for $n = 1$, with s_1 along the $x-$axis and z along the $y-$axis.
- For $\mathbf{x} \in \mathcal{D}$, identify all points (s_1, z) for $s_1 \geq g_1(\mathbf{x})$ and $z \geq f(\mathbf{x})$.
  - These are points that lie to the right and above the point $(g_1(\mathbf{x}), f(\mathbf{x}))$.

Geometry of Duality: What is the Primal?

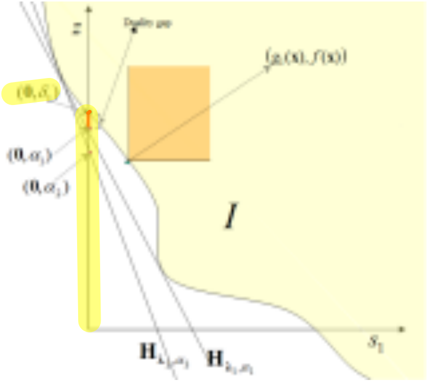

Figure 17: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Feasible region for the primal problem (67) is the region in $\mathcal{I}$ with $\mathbf{s} \leq \mathbf{0}$.
- Since all points above and to the right of a point in $\mathcal{I}$ also belong to $\mathcal{I}$, the solution to the primal problem corresponds to the point in $\mathcal{I}$ with $\mathbf{s} = \mathbf{0}$ and least possible value of z.
- In Figure 17, the solution to the primal corresponds to $(\mathbf{0}, \delta_1)$.

Geometry of Duality: What is the Primal?

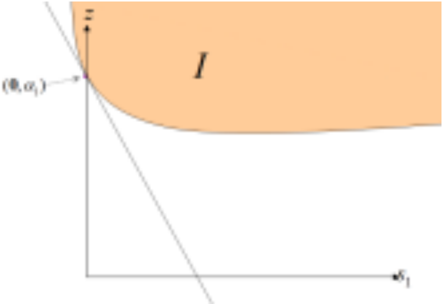

Figure 18: Example of the convex set $\mathcal{I}$ and hyperplane $\mathcal{H}_{\lambda,\alpha}$ for a single constrained well-behaved convex program.

- Straightforward to prove that if $f(\mathbf{x})$ and each of the constraints $g_i(\mathbf{x})$, $1 \leq i \leq n$ are convex functions, then $\mathcal{I}$ must be a convex set.

Prove this simply by definition of I

Geometry of Duality: What is the Primal?

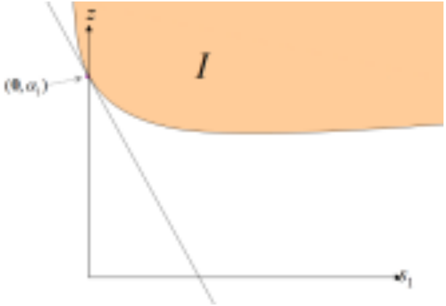

Figure 18: Example of the convex set $\mathcal{I}$ and hyperplane $\mathcal{H}_{\lambda,\alpha}$ for a single constrained well-behaved convex program.

H/W

- Straightforward to prove that if $f(\mathbf{x})$ and each of the constraints $g_i(\mathbf{x}),\ 1 \leq i \leq n$ are convex functions, then $\mathcal{I}$ must be a convex set.
- **Recap:** $\mathcal{I} \subseteq \Re^{m+1}$ as

$$\mathcal{I} = \{(\mathbf{s}, z)\} \text{ s.t } \begin{cases} \mathbf{s} \in \Re^m & \exists \mathbf{x} \in \mathcal{D} \text{ s.t } g_i(\mathbf{x}) \leq s_i \ \forall i \\ z \in \Re & f(\mathbf{x}) \leq z \end{cases}$$

The point corresponding to the convex combination of s1 and s2 will be the convex combinations of corresponding x1 and x2

Geometry of Duality: What is the Dual?

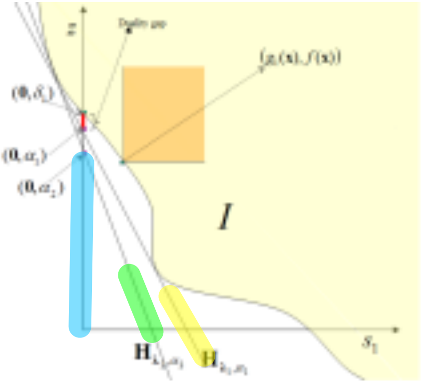

Figure 19: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Define a hyerplane $\mathcal{H}_{\lambda,\alpha}$, parametrized by $\lambda \in \Re^m$ and $\alpha \in \Re$ as

$$\mathcal{H}_{\lambda,\alpha} = \left\{ (\mathbf{s}, z) \middle| \lambda^T.\mathbf{s} + z = \alpha \right\}$$

- Consider all $\mathcal{H}_{\lambda,\alpha}$ that lie below $\mathcal{I}$. For example, in the Figure 19, both hyperplanes $\mathcal{H}_{\lambda_1,\alpha_1}$ and $\mathcal{H}_{\lambda_2,\alpha_2}$ lie below the set $\mathcal{I}$.
- Of all $\mathcal{H}_{\lambda,\alpha}$ that lie below $\mathcal{I}$, consider the hyperplane whose intersection with the line $\mathbf{s} = \mathbf{0}$, corresponds to as high a value of z as possible.
- This hyperplane must be a

supporting hyperplane to I

Geometry of Duality: What is the Dual?

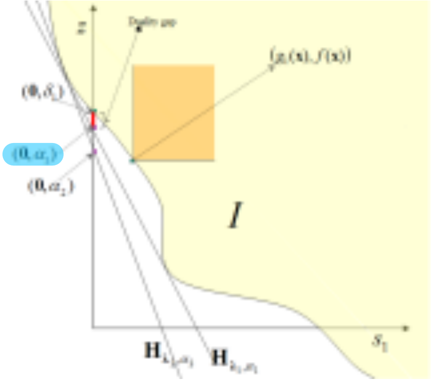

Figure 19: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Define a hyerplane $\mathcal{H}_{\lambda,\alpha}$, parametrized by $\lambda \in \Re^m$ and $\alpha \in \Re$ as

$$\mathcal{H}_{\lambda,\alpha} = \left\{ (\mathbf{s}, z) \,\middle|\, \lambda^T.\mathbf{s} + z = \alpha \right\}$$

- Consider all $\mathcal{H}_{\lambda,\alpha}$ that lie below $\mathcal{I}$. For example, in the Figure 19, both hyperplanes $\mathcal{H}_{\lambda_1,\alpha_1}$ and $\mathcal{H}_{\lambda_2,\alpha_2}$ lie below the set $\mathcal{I}$.
- Of all $\mathcal{H}_{\lambda,\alpha}$ that lie below $\mathcal{I}$, consider the hyperplane whose intersection with the line $\mathbf{s} = \mathbf{0}$, corresponds to as high a value of z as possible.
- This hyperplane must be a supporting hyperplane. $\mathcal{H}_{\lambda_1,\alpha_1}$ happens to be such a supporting hyperplane. Its point of intersection $(\mathbf{0}, \alpha_1)$ precisely corresponds to the solution to the dual problem.

Geometry of Duality: The Dual - A bit more formally

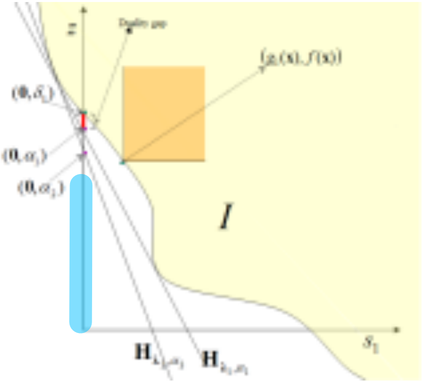

Figure 20: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Define two half-spaces corresponding to $\mathcal{H}_{\lambda,\alpha}$ as

$$\mathcal{H}^{+}_{\lambda,\alpha} = \left\{(\mathbf{s}, z) \,\middle|\, \lambda^T.\mathbf{s} + z \geq \alpha\right\}$$

$$\mathcal{H}^{-}_{\lambda,\alpha} = \left\{(\mathbf{s}, z) \,\middle|\, \lambda^T.\mathbf{s} + z \leq \alpha\right\}$$

- Define another set $\mathcal{L}$ as

$$\mathcal{L} = \left\{(\mathbf{s}, z) \,|\mathbf{s} = \mathbf{0}\right\}$$

Note that $\mathcal{L}$ is essentially the z axis

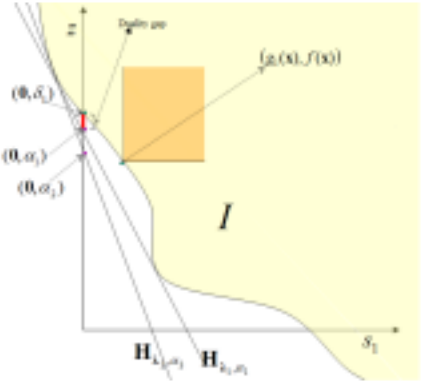

Figure 20: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Define two half-spaces corresponding to $\mathcal{H}_{\lambda,\alpha}$ as

$$\mathcal{H}^{+}_{\lambda,\alpha} = \left\{ (\mathbf{s}, z) \middle| \lambda^T.\mathbf{s} + z \geq \alpha \right\}$$

$$\mathcal{H}^{-}_{\lambda,\alpha} = \left\{ (\mathbf{s}, z) \middle| \lambda^T.\mathbf{s} + z \leq \alpha \right\}$$

- Define another set $\mathcal{L}$ as

$$\mathcal{L} = \left\{ (\mathbf{s}, z) \, | \mathbf{s} = \mathbf{0} \right\}$$

Note that $\mathcal{L}$ is essentially the z or function axis.

- The intersection of $\mathcal{H}_{\lambda,\alpha}$ with $\mathcal{L}$ is \alpha

Geometry of Duality: The Dual - A bit more formally

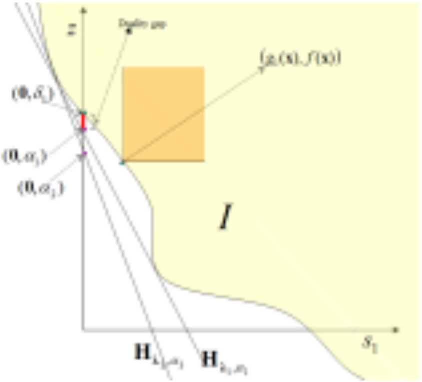

Figure 20: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Define two half-spaces corresponding to $\mathcal{H}_{\lambda,\alpha}$ as

$$\mathcal{H}^{+}_{\lambda,\alpha} = \left\{ (\mathbf{s}, z) \,\middle|\, \lambda^T.\mathbf{s} + z \geq \alpha \right\}$$

$$\mathcal{H}^{-}_{\lambda,\alpha} = \left\{ (\mathbf{s}, z) \,\middle|\, \lambda^T.\mathbf{s} + z \leq \alpha \right\}$$

- Define another set $\mathcal{L}$ as

$$\mathcal{L} = \left\{ (\mathbf{s}, z) \,|\, \mathbf{s} = \mathbf{0} \right\}$$

  Note that $\mathcal{L}$ is essentially the z or function axis.

- The intersection of $\mathcal{H}_{\lambda,\alpha}$ with $\mathcal{L}$ is the point $(\mathbf{0}, \alpha)$. That is, $(\mathbf{0}, \alpha) = \mathcal{L} \bigcap \mathcal{H}_{\lambda,\alpha}$
- **Dual:** Manipulate λ and α so that $\mathcal{I}$ lies in the half-space $\mathcal{H}^{+}_{\lambda,\alpha}$ **as tightly as possible.**

Geometry of Duality: The Dual - A bit more formally

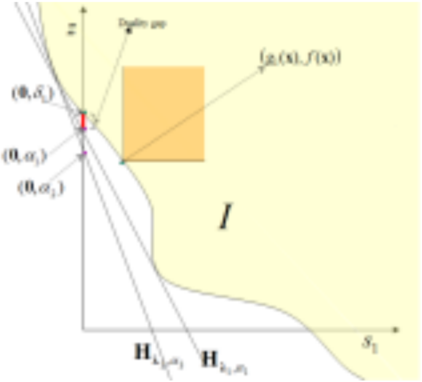

Figure 21: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Mathematically, we are interested in the problem of **maximizing the height of the point of intersection of $\mathcal{L}$ with $\mathcal{H}_{\lambda,\alpha}$ above the $\mathbf{s}$ plane, while ensuring that $\mathcal{I}$ remains a subset of $\mathcal{H}^{+}_{\lambda,\alpha}$.**

$$\begin{array}{ll} \max & \alpha \\ \text{subject to} & \mathcal{H}^{+}_{\lambda,\alpha} \supseteq \mathcal{I} \end{array}$$

By definitions of $\mathcal{I}$, $\mathcal{H}^{+}_{\lambda,\alpha}$ and the subset relation, this problem is equivalent to

Geometry of Duality: The Dual - A bit more formally

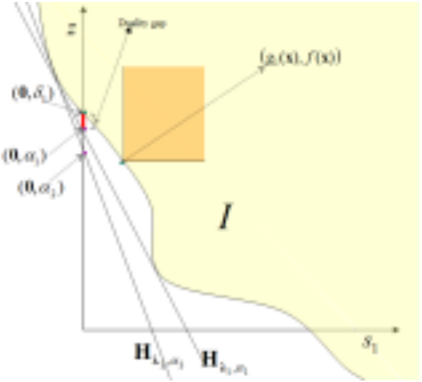

Figure 21: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Mathematically, we are interested in the problem of **maximizing the height of the point of intersection of $\mathcal{L}$ with $\mathcal{H}_{\lambda,\alpha}$ above the $\mathbf{s}$ plane, while ensuring that $\mathcal{I}$ remains a subset of $\mathcal{H}^+_{\lambda,\alpha}$.**

$$\begin{array}{ll} \max & \alpha \\ \text{subject to} & \mathcal{H}^+_{\lambda,\alpha} \supseteq \mathcal{I} \end{array}$$

By definitions of $\mathcal{I}$, $\mathcal{H}^+_{\lambda,\alpha}$ and the subset relation, this problem is equivalent to

$$\begin{array}{ll} \max & \alpha \\ \text{subject to} & \lambda^T.\mathbf{s} + z \geq \alpha \ \forall (\mathbf{s}, z) \in \mathcal{I} \end{array}$$

Geometry of Duality: The Dual - A bit more formally

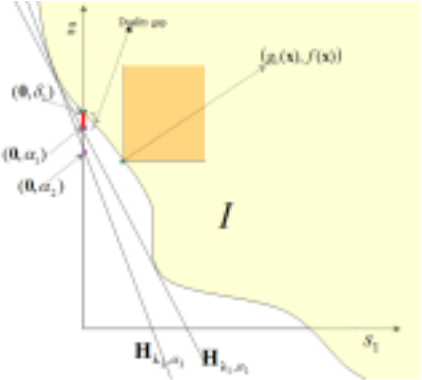

Figure 22: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Note: **If $(\mathrm{s}, z) \in \mathcal{I}$, then $(\mathrm{s}', z) \in \mathcal{I}$ for all $\mathrm{s}' \geq \mathrm{s}$ (as illustrated in Figure 22)**. Thus, we cannot afford to have any component of λ negative; if any of the λ_i's were negative,

then the upper half space constraint will be violated by cranking up s (while still remaining in I)

Geometry of Duality: The Dual - A bit more formally

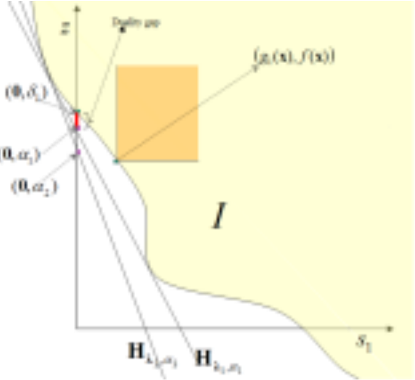

Figure 22: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Note: **If $(\mathbf{s}, z) \in \mathcal{I}$, then $(\mathbf{s}', z) \in \mathcal{I}$ for all $\mathbf{s}' \geq \mathbf{s}$ (as illustrated in Figure 22)**. Thus, we cannot afford to have any component of λ negative; if any of the λ_i's were negative, we could cranck up s_i arbitrarily to violate the inequality $\lambda^T.\mathbf{s} + z \geq \alpha$.
- Consequently, we can add the constraint $\lambda \geq \mathbf{0}$ to the forgoing problem without changing the solution.

$$
\begin{array}{ll}
\text{max} & \alpha \\
\text{subject to} & \lambda^T.\mathbf{s} + z \geq \alpha \ \forall(\mathbf{s}, z) \in \mathcal{I} \\
& \lambda \geq \mathbf{0}
\end{array}
$$

- Expect every point on $\partial\mathcal{I}$ to be of the form $(g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_m(\mathbf{x}), f(\mathbf{x}))$ for some $\mathbf{x} \in \mathcal{D}$. **Therefore ......**

Geometry of Duality: The Dual - A bit more formally

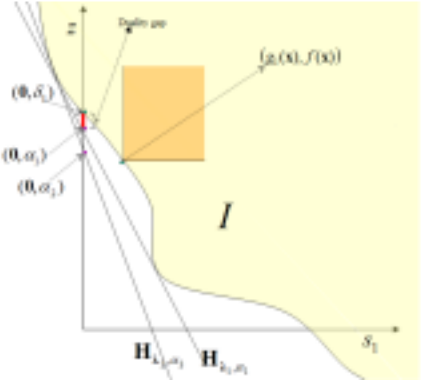

Figure 23: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Foregoing problem is equivalent to

$$\begin{array}{ll} \max & \alpha \\ \text{subject to} & \lambda^T.\mathbf{g}(\mathbf{x}) + f(\mathbf{x}) \geq \alpha \; \forall \mathbf{x} \in \mathcal{D} \\ & \lambda \geq \mathbf{0} \end{array}$$

- Recall that $L(\mathbf{x}, \lambda) = \lambda^T.\mathbf{g}(\mathbf{x}) + f(\mathbf{x})$. The **geometric** problem is therefore the same as

$$\begin{array}{ll} \max & \alpha \\ \text{subject to} & L(\mathbf{x}, \lambda) \geq \alpha \; \forall \mathbf{x} \in \mathcal{D} \\ & \lambda \geq \mathbf{0} \end{array}$$

Geometry of Duality: The Dual - A bit more formally

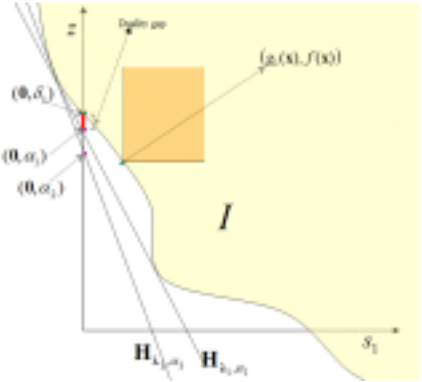

Figure 24: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Since, $L^*(\lambda) = \min_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x}, \lambda)$, we can deal with the equivalent problem

$$\begin{array}{ll} \max & \alpha \\ \text{subject to} & L^*(\lambda) \geq \alpha \\ & \lambda \geq \mathbf{0} \end{array}$$

- The **geometric** problem can be restated as

Geometry of Duality: The Dual - A bit more formally

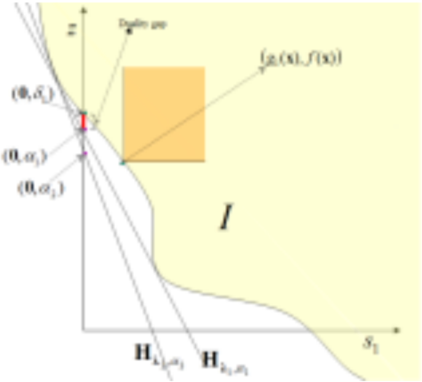

Figure 24: Example of the set $\mathcal{I}$ and hyperplanes $\mathcal{H}_{\lambda,\alpha}$ for a single constraint (*i.e.*, for $n = 1$).

- Since, $L^*(\lambda) = \min_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x}, \lambda)$, we can deal with the equivalent problem

$$\begin{array}{ll} \max & \alpha \\ \text{subject to} & L^*(\lambda) \geq \alpha \\ & \lambda \geq \mathbf{0} \end{array}$$

- The **geometric** problem can be restated as

$$\begin{array}{ll} \max & L^*(\lambda) \\ \text{subject to} & \lambda \geq \mathbf{0} \end{array}$$

This is precisely **the dual problem**. We thus get a geometric interpretation of the dual.

Geometry of Duality: Duality Gap and Convexity

- With reference to Figure 16, if the set $\mathcal{I}$ is not convex, there could be a **gap** between the $z-$intercept $(\mathbf{0}, \alpha_1)$ of the best supporting hyperplane $\mathcal{H}_{\lambda_1, \alpha_1}$ and the closest point $(\mathbf{0}, \delta_1)$ of $\mathcal{I}$ on the $z-$axis (solution to the primal).
- For non-convex $\mathcal{I}$, we can never prove in zero duality gap in general.
- Homework (Quiz 1, Problem 1): Write dual for constrained problem $\min_x f(x) = 5x^2 + 6x^3 - x^4$ on the closed interval $[-2, 10]$. Does it have a duality gap?