Lagrange Duality

Daniel P. Palomar
Hong Kong University of Science and Technology (HKUST)

ELEC5470 - Convex Optimization
Fall 2017-18, HKUST, Hong Kong

Outline of Lecture

- Lagrangian
- Dual function
- Dual problem
- Weak and strong duality
- KKT conditions
- Summary

(Acknowledgement to Stephen Boyd for material for this lecture.)

Lagrangian

- Consider an optimization problem in standard form (not necessarily convex)

$$\begin{array}{lll} \underset{x}{\text{minimize}} & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0 & i = 1, \ldots, m \\ & h_i(x) = 0 & i = 1, \ldots, p \end{array}$$

  with variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, and optimal value $p^\star$.

- The *Lagrangian* is a function $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \rightarrow \mathbf{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$, defined as

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

  where λ_i is the Lagrange multiplier associated with $f_i(x) \leq 0$ and ν_i is the Lagrange multiplier associated with $h_i(x) = 0$.

Lagrange Dual Function

- The *Lagrange dual function* is defined as the infimum of the Lagrangian over x: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\
&= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)
\end{aligned}
$$

- Observe that:
  - the infimum is unconstrained (as opposed to the original constrained minimization problem)
  - g is concave regardless of original problem (infimum of affine functions)
  - g can be $-\infty$ for some λ, ν

Lower bound property: if $\lambda \geq 0$, then $g(\lambda, \nu) \leq p^\star$.

Proof. Suppose $\tilde{x}$ is feasible and $\lambda \geq 0$. Then,

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu).$$

Now choose minimizer of $f_0(\tilde{x})$ over all feasible $\tilde{x}$ to get $p^\star \geq g(\lambda, \nu)$. □

- We could try to find the best lower bound by maximizing $g(\lambda, \nu)$. This is in fact the dual problem.

Dual Problem

- The *Lagrange dual problem* is defined as

$$\begin{array}{ll} \underset{\lambda,\nu}{\text{maximize}} & g(\lambda,\nu) \\ \text{subject to} & \lambda \geq 0. \end{array}$$

- This problem finds the best lower bound on $p^\star$ obtained from the dual function.

- It is a convex optimization (maximization of a concave function and linear constraints).

- The optimal value is denoted $d^\star$.

- λ, ν are dual feasible if $\lambda \geq 0$ and $(\lambda, \nu) \in \operatorname{dom} g$ (the latter implicit constraints can be made explicit in problem formulation).

Example: Least-Norm Solution of Linear Equations

- Consider the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^T x \\ \text{subject to} & Ax = b. \end{array}$$

- The Lagrangian is

$$L(x, \nu) = x^T x + \nu^T (Ax - b).$$

- To find the dual function, we need to solve an unconstrained minimization of the Lagrangian. We set the gradient equal to zero

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \Longrightarrow x = -(1/2) A^T \nu$$

and we plug the solution in L to obtain g:

$$g(\nu) = L\left(-(1/2)\,A^T\nu, \nu\right) = -\frac{1}{4}\nu^T AA^T\nu - b^T\nu$$

- The function g is, as expected, a concave function of ν.
- From the lower bound property, we have

$$p^\star \geq -\frac{1}{4}\nu^T AA^T\nu - b^T\nu \text{ for all } \nu.$$

- The dual problem is the QP

$$\underset{\nu}{\text{maximize}} \quad -\tfrac{1}{4}\nu^T AA^T\nu - b^T\nu.$$

Example: Standard Form LP

- Consider the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^T x \\ \text{subject to} & Ax = b, \quad x \geq 0. \end{array}$$

- The Lagrangian is

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= \left(c + A^T \nu - \lambda\right)^T x - b^T \nu. \end{aligned}$$

- L is a linear function of x and it is unbounded if the term multiplying x is nonzero.

- Hence, the dual function is

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T\nu & c + A^T\nu - \lambda = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

- The function g is a concave function of (λ, ν) as it is linear on an affine domain.

- From the lower bound property, we have

$$p^\star \geq -b^T\nu \quad \text{if } c + A^T\nu \geq 0.$$

- The dual problem is the LP

$$\begin{array}{ll} \underset{\nu}{\text{maximize}} & -b^T\nu \\ \text{subject to} & c + A^T\nu \geq 0. \end{array}$$

Example: Two-Way Partitioning

- Consider the problem

$$
\begin{array}{ll}
\underset{x}{\text{minimize}} & x^T W x \\
\text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n.
\end{array}
$$

- It is a nonconvex problem (quadratic equality constraints). The feasible set contains 2^n discrete points.
- The Lagrangian is

$$
\begin{aligned}
L(x, \nu) &= x^T W x + \sum_{i=1}^{n} \nu_i \left(x_i^2 - 1 \right) \\
&= x^T \left(W + \text{diag}(\nu) \right) x - 1^T \nu.
\end{aligned}
$$

- L is a quadratic function of x and it is unbounded if the matrix $W + \text{diag}(\nu)$ has a negative eigenvalue.

- Hence, the dual function is

$$g(\nu) = \inf_x L(x, \nu) = \begin{cases} -1^T\nu & W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

- From the lower bound property, we have

$$p^\star \geq -1^T\nu \quad \text{if } W + \text{diag}(\nu) \succeq 0.$$

- As an example, if we choose $\nu = -\lambda_{\min}(W)\,1$, we get the bound

$$p^\star \geq n\lambda_{\min}(W).$$

- The dual problem is the SDP

$$\begin{array}{ll} \underset{\nu}{\text{maximize}} & -1^T\nu \\ \text{subject to} & W + \text{diag}(\nu) \succeq 0. \end{array}$$

Weak and Strong Duality

- From the lower bound property, we know that $g(\lambda, \nu) \leq p^\star$ for feasible (λ, ν). In particular, for a (λ, ν) that solves the dual problem.

- Hence, *weak duality* always holds (even for nonconvex problems):

$$d^\star \leq p^\star.$$

- The difference $p^\star - d^\star$ is called *duality gap*.

- Solving the dual problem may be used to find nontrivial lower bounds for difficult problems.

- Even more interesting is when equality is achieved in weak duality. This is called *strong duality*:

$$d^\star = p^\star.$$

- Strong duality means that the duality gap is zero.

- Strong duality:
  - is very desirable (we can solve a difficult problem by solving the dual)
  - does not hold in general
  - usually holds for convex problems
  - conditions that guarantee strong duality in convex problems are called *constraint qualifications*.

Slater's Constraint Qualification

- Slater's constraint qualification is a very simple condition that is satisfied in most cases and ensures strong duality for convex problems.

- Strong duality holds for a convex problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0 \qquad i = 1, \ldots, m \\ & Ax = b \end{array}$$

if it is strictly feasible, i.e.,

$$\exists x \in \operatorname{int} \mathcal{D} : \quad f_i(x) < 0 \quad i = 1, \ldots, m, \quad Ax = b.$$

- It can be relaxed by using $\mathbf{relint}\,\mathcal{D}$ (interior relative to affine hull) instead of $\mathbf{int}\,\mathcal{D}$; linear inequalities do not need to hold with strict inequality, ...

- There exist many other types of constraint qualifications.

Example: Inequality Form LP

- Consider the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^T x \\ \text{subject to} & Ax \le b. \end{array}$$

- The dual problem is

$$\begin{array}{ll} \underset{\lambda}{\text{maximize}} & -b^T \lambda \\ \text{subject to} & A^T \lambda + c = 0, \quad \lambda \ge 0. \end{array}$$

- From Slater's condition: $p^\star = d^\star$ if $A\tilde{x} < b$ for some $\tilde{x}$.

- In this case, in fact, $p^\star = d^\star$ except when primal and dual are infeasible.

Example: Convex QP

- Consider the problem (assume $P \succeq 0$)

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^T P x \\ \text{subject to} & Ax \leq b. \end{array}$$

- The dual problem is

$$\begin{array}{ll} \underset{\lambda}{\text{maximize}} & -(1/4)\,\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \geq 0. \end{array}$$

- From Slater's condition: $p^\star = d^\star$ if $A\tilde{x} < b$ for some $\tilde{x}$.

- In this case, in fact, $p^\star = d^\star$ always.

Example: Nonconvex QP

- Consider the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leq 1 \end{array}$$

which is nonconvex in general as $A \nsucceq 0$.

- The dual problem is

$$\begin{array}{ll} \underset{\lambda}{\text{maximize}} & -b^T (A + \lambda I)^{\#} b - \lambda \\ \text{subject to} & A + \lambda I \succeq 0 \\ & b \in \mathcal{R}(A + \lambda I) \end{array}$$

which can be rewritten as

$$\begin{array}{ll} \underset{t,\lambda}{\text{maximize}} & -t-\lambda \\ \text{subject to} & \begin{bmatrix} A+\lambda I & b \\ b^T & t \end{bmatrix} \succeq 0. \end{array}$$

- In this case, strong duality holds even though the original problem is nonconvex (not trivial).

Complementary Slackness

- Assume strong duality holds, $x^\star$ is primal optimal and $(\lambda^\star, \nu^\star)$ is dual optimal. Then,

$$\begin{aligned} f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x) \right) \\ &\leq f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star h_i(x^\star) \\ &\leq f_0(x^\star) \end{aligned}$$

- Hence, the two inequalities must hold with equality. Implications:
  - $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
  - $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \ldots, m$; this is called *complementary slackness*:

$$\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0.$$

Karush-Kuhn-Tucker (KKT) Conditions

KKT conditions (for differentiable f_i, h_i):

1. primal feasibility: $f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p$

2. dual feasibility: $\lambda \geq 0$

3. complementary slackness: $\lambda_i^{\star} f_i(x^{\star}) = 0$ for $i = 1, \ldots, m$

4. zero gradient of Lagrangian with respect to x:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

- We already known that if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions.

- What about the opposite statement?

- If x, λ, ν satisfy the KKT conditions for a convex problem, then they are optimal.

Proof. From complementary slackness, $f_0(x) = L(x, \lambda, \nu)$ and, from 4th KKT condition and convexity, $g(\lambda, \nu) = L(x, \lambda, \nu)$. Hence, $f_0(x) = g(\lambda, \nu)$. □

Theorem. *If a problem is convex and Slater's condition is satisfied, then x is optimal if and only if there exists λ, ν that satisfy the KKT conditions.*

Example: Waterfilling Solution

- Consider the maximization of the mutual information in a MIMO channel under Gaussian noise:

$$\begin{array}{ll} \underset{\mathbf{Q}}{\text{maximize}} & \log\det\left(\mathbf{R}_n + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right) \\ \text{subject to} & \text{Tr}\left(\mathbf{Q}\right) \leq P \\ & \mathbf{Q} \succeq \mathbf{0}. \end{array}$$

- This problem is convex: the logdet function is concave, the trace constraint is just a linear constraint, and the positive semidefiniteness constraint is an LMI.

- Hence, we can use a general-purpose method such as an interior-point method to solve it in polynomial time.

- However, this problem admits a closed-form solution as can be derived from the KKT conditions.

- The Lagrangian is

$$L(\mathbf{Q}; \mu, \mathbf{\Psi}) = -\log\det\left(\mathbf{R}_n + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right) + \mu\left(\mathsf{Tr}\left(\mathbf{Q}\right) - P\right) - \mathsf{Tr}\left(\mathbf{\Psi}\mathbf{Q}\right).$$

- The gradient of the Lagrangian is

$$\nabla_{\mathbf{Q}} L = -\mathbf{H}^\dagger\left(\mathbf{R}_n + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\right)^{-1}\mathbf{H} + \mu\mathbf{I} - \mathbf{\Psi}.$$

- The KKT conditions are

$$\begin{aligned} \mathsf{Tr}\left(\mathbf{Q}\right) \leq P, \quad \mathbf{Q} &\succeq \mathbf{0} \\ \mu \geq 0, \quad \mathbf{\Psi} &\succeq \mathbf{0} \\ \mathbf{H}^{\dagger}\left(\mathbf{R}_n + \mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}\right)^{-1}\mathbf{H} + \mathbf{\Psi} &= \mu\mathbf{I} \\ \mu\left(\mathsf{Tr}\left(\mathbf{Q}\right) - P\right) = 0, \quad \mathbf{\Psi}\mathbf{Q} &= \mathbf{0}. \end{aligned}$$

- Can we find a $\mathbf{Q}$ that satisfies the KKT conditions (together with some dual variables)?

- First, let's simplify the KKT conditions by defining the so-called *whitened channel*: $\widetilde{\mathbf{H}} = \mathbf{R}_n^{-1/2}\mathbf{H}$.

- Then, the third KKT condition becomes:

$$\widetilde{\mathbf{H}}^\dagger \left(\mathbf{I} + \widetilde{\mathbf{H}}\mathbf{Q}\widetilde{\mathbf{H}}^\dagger\right)^{-1} \widetilde{\mathbf{H}} + \mathbf{\Psi} = \mu \mathbf{I}.$$

- To simplify even further, let's write the SVD of the channel matrix as $\widetilde{\mathbf{H}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$ (denote the eigenvalues σ_i), obtaining:

$$\mathbf{\Sigma}^\dagger \left(\mathbf{I} + \mathbf{\Sigma}\widetilde{\mathbf{Q}}\mathbf{\Sigma}^\dagger\right)^{-1} \mathbf{\Sigma} + \widetilde{\mathbf{\Psi}} = \mu \mathbf{I}.$$

where $\widetilde{\mathbf{Q}} = \mathbf{V}^\dagger\mathbf{Q}\mathbf{V}$ and $\widetilde{\mathbf{\Psi}} = \mathbf{V}^\dagger\mathbf{\Psi}\mathbf{V}$.

- The KKT conditions are:

$$\begin{aligned} \mathsf{Tr}(\widetilde{\mathbf{Q}}) \leq P, \quad \widetilde{\mathbf{Q}} &\succeq \mathbf{0} \\ \mu \geq 0, \quad \widetilde{\mathbf{\Psi}} &\succeq \mathbf{0} \\ \mathbf{\Sigma}^{\dagger}\left(\mathbf{I}+\mathbf{\Sigma}\widetilde{\mathbf{Q}}\mathbf{\Sigma}^{\dagger}\right)^{-1}\mathbf{\Sigma}+\widetilde{\mathbf{\Psi}} &= \mu\mathbf{I} \\ \mu\left(\mathsf{Tr}(\widetilde{\mathbf{Q}})-P\right)=0, \quad \widetilde{\mathbf{\Psi}}\widetilde{\mathbf{Q}} &= \mathbf{0}. \end{aligned}$$

- At this point, we can make a guess: perhaps the optimal $\widetilde{\mathbf{Q}}$ and $\widetilde{\mathbf{\Psi}}$ are diagonal? Let's try ...

- Define $\widetilde{\mathbf{Q}} = \mathsf{diag}\,(\mathbf{p})$ ($\mathbf{p}$ is the power allocation) and $\widetilde{\boldsymbol{\Psi}} = \mathsf{diag}\,(\boldsymbol{\psi})$.

- The KKT conditions become:

$$
\begin{aligned}
\sum_i p_i \leq P, \quad p_i &\geq 0 \\
\mu \geq 0, \quad \psi_i &\geq 0 \\
\frac{\sigma_i^2}{1+\sigma_i^2 p_i} + \psi_i &= \mu \\
\mu\left(\sum_i p_i - P\right) = 0 \quad , \psi_i p_i &= 0.
\end{aligned}
$$

- Let's now look into detail at the KKT conditions.

- First of all, observe that $\mu > 0$, otherwise we would have $\frac{\sigma_i^2}{1+\sigma_i^2 p_i} + \psi_i = 0$ which cannot be satisfied.

- Let's distinguish two cases in the power allocation:
  - if $p_i > 0$, then $\psi_i = 0 \Longrightarrow \frac{\sigma_i^2}{1+\sigma_i^2 p_i} = \mu \Longrightarrow p_i = \mu^{-1} - 1/\sigma_i^2$ (also note that $\mu = \frac{\sigma_i^2}{1+\sigma_i^2 p_i} < \sigma_i^2$)
  - if $p_i = 0$, then $\sigma_i^2 + \psi_i = \mu$ (note that $\mu = \sigma_i^2 + \psi_i \geq \sigma_i^2$.

- Equivalently,
  - if $\sigma_i^2 > \mu$, then $p_i = \mu^{-1} - 1/\sigma_i^2$
  - if $\sigma_i^2 \leq \mu$, then $p_i = 0$.

- More compactly, we can write the well-known *waterfilling solution*:

$$p_i = \left(\mu^{-1} - 1/\sigma_i^2\right)^+$$

where μ^{-1} is called water-level and is chosen to satisfy $\sum_i p_i = P$ (so that all the KKT conditions are satisfied).

- Therefore, the optimal solution is given by

$$\mathbf{Q}^\star = \mathbf{V}\text{diag}\left(\mathbf{p}\right)\mathbf{V}^\dagger$$

where

  – the optimal transmit directions are matched to the channel matrix
  – the optimal power allocation is the waterfilling.

Perturbation and Sensitivity Analysis

- Recall the original (unperturbed) optimization problem and its dual:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0 \quad \forall i \\ & h_i(x) = 0 \quad \forall i \end{array} \qquad \begin{array}{ll} \underset{\lambda,\nu}{\text{maximize}} & g(\lambda,\nu) \\ \text{subject to} & \lambda \geq 0 \end{array}$$

- Define the perturbed problem and dual as

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_i(x) \leq u_i \quad \forall i \\ & h_i(x) = v_i \quad \forall i \end{array} \qquad \begin{array}{ll} \underset{\lambda,\nu}{\text{maximize}} & g(\lambda,\nu) - u^T\lambda - v^T\nu \\ \text{subject to} & \lambda \geq 0 \end{array}$$

- x is primal variable and u, v are parameters
- Define $p^\star(u,v)$ as the optimal value as a function of u, v.

- **Global sensitivity**: Suppose strong duality holds for unperturbed problem and $\lambda^\star, \nu^\star$ are dual optimal for unperturbed problem. Then, from weak duality:

$$\begin{aligned} p^\star(u,v) &\geq g(\lambda^\star,\nu^\star) - u^T\lambda^\star - v^T\nu^\star \\ &= p^\star(0,0) - u^T\lambda^\star - v^T\nu^\star \end{aligned}$$

- Interpretation:
  - if $\lambda_i^\star$ large: $p^\star$ increases a lot if we tighten constraint i ($u_i < 0$)
  - if $\lambda_i^\star$ small: $p^\star$ does not decrease much if we loosen constraint i ($u_i > 0$)
  - if $\nu_i^\star$ large and positive: $p^\star$ increases a lot if we take $v_i < 0$
  - if $\nu_i^\star$ large and negative: $p^\star$ increases a lot if we take $v_i > 0$
  - etc.

- **Local sensitivity**: Suppose strong duality holds for unperturbed problem, $\lambda^\star, \nu^\star$ are dual optimal for unperturbed problem, and $p^\star(u, v)$ is differentiable at $(0, 0)$. Then,

$$\frac{\partial p^\star(0,0)}{\partial u_i} = -\lambda_i^\star, \qquad \frac{\partial p^\star(0,0)}{\partial v_i} = -\nu_i^\star$$

Proof. (for $\lambda_i^\star$)From the global sensitivity result, we have

$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{\epsilon\downarrow 0} \frac{p^\star(te_i, 0) - p^\star(0,0)}{t} \geq \lim_{\epsilon\downarrow 0} \frac{-t\lambda_i^\star}{t} = -\lambda_i^\star$$

$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{\epsilon\uparrow 0} \frac{p^\star(te_i, 0) - p^\star(0,0)}{t} \leq \lim_{\epsilon\uparrow 0} \frac{-t\lambda_i^\star}{t} = -\lambda_i^\star.$$

Hence, the equality. □

Duality and Problem Reformulations

- Equivalent formulations of a problem can lead to very different duals.

- Reformulating the primal problem can be useful when the dual is difficult to derive or uninteresting.

- Common tricks:

  - introduce new variables and equality constraints
  - make explicit constraints implicit or vice-versa
  - transform objective or constraint functions (e.g., replace $f_0(x)$ by $\phi(f_0(x))$ with ϕ convex and increasing).

Example: Introducing New Variables

- Consider the problem

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_2 \,.$$

- We can rewrite it as

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & \|y\|_2 \\ \text{subject to} & y = Ax - b. \end{array}$$

- We can then derive the dual problem:

$$\begin{array}{ll} \underset{\nu}{\text{maximize}} & b^T\nu \\ \text{subject to} & A^T\nu = 0, \quad \|\nu\|_2 \le 1. \end{array}$$

Example: Implicit Constraints

- Consider the following LP with box constrains:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \leq x \leq \mathbf{1} \end{array}$$

- The dual problem is

$$\begin{array}{ll} \underset{\nu, \lambda_1, \lambda_2}{\text{maximize}} & -b^T \nu - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2 \\ \text{subject to} & c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \geq 0, \quad \lambda_2 \geq 0, \end{array}$$

which does not give much insight.

- If, instead, we rewrite the primal problem as

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \le x \le \mathbf{1} \\ \infty & \text{otherwise} \end{cases} \\ \text{subject to} & Ax = b \end{array}$$

then the dual becomes way more insightful:

$$\underset{\nu}{\text{maximize}} \quad -b^T \nu - \left\| A^T \nu + c \right\|_1$$

Duality for Problems with Generalized Inequalities

- The Lagrange duality can be naturally extended to generalized inequalities of the form

$$f_i(x) \preceq_{K_i} 0$$

where $\preceq_{K_i}$ is a generalized inequality on $\mathbf{R}^{k_i}$ with respect to the cone K_i.

- The corresponding dual variable has to satisfy

$$\lambda_i \succeq_{K_i^*} 0$$

where K_i^* is the dual cone of K_i.

Semidefinite Programming (SDP)

- Consider the following SDP ($F_i, G \in \mathbf{R}^{k\times k}$):

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n \preceq G. \end{array}$$

- The Lagrange multiplier is a matrix $\Psi \in \mathbf{R}^{k\times k}$ and the Lagrangian

$$L(x, \Psi) = c^T x + \mathsf{Tr}\left(\Psi\left(x_1 F_1 + \cdots + x_n F_n - G\right)\right)$$

- The dual problem is

$$\begin{array}{ll} \underset{\Psi}{\text{maximize}} & -\mathsf{Tr}\left(\Psi G\right) \\ \text{subject to} & \mathsf{Tr}\left(\Psi F_i\right) + c_i = 0, \; i = 1, \ldots, n \\ & \Psi \succeq 0. \end{array}$$

Summary

- We have introduced the Lagrange duality theory: Lagrangian, dual function, and dual problem.

- We have developed the optimality conditions for convex problems: the KKT conditions.

- We have illustrated the used of the KKT conditions to find the closed-form solution to a problem.

- We have overviewed some additional concepts such as duals of reformulations of problems, sensitivity analysis, generalized inequalities, and SDP.

References

Chapter 5 of

- Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

http://www.stanford.edu/~boyd/cvxbook/