EE 227A: Convex Optimization and Applications **February 7, 2012**

Lecture 7: **Weak Duality**

Lecturer: Laurent El Ghaoui

7.1 Lagrange Dual problem

7.1.1 Primal problem

In this section, we consider a possibly *non-convex* optimization problem

$$p^* := \min_x \ f_0(x) \ : \ f_i(x) \leq 0, \quad i = 1, \cdots, m, \tag{7.1}$$

where the functions $f_0, f_1, \ldots, f_m$ We denote by $\mathcal{D}$ the domain of the problem (which is the intersection of the domains of all the functions involved), and by $\mathcal{X} \subseteq \mathcal{D}$ its feasible set.

We will refer to the above as the *primal* problem, and to the decision variable x in that problem, as the *primal variable.* One purpose of Lagrange duality is to find a lower bound on a minimization problem (or an upper bounds for a maximization problem). Later, we will use duality tools to derive optimality conditions for convex problems.

7.1.2 Dual problem

Lagrangian. To the problem we associate the Lagrangian $\mathcal{L} : \mathbf{R}^n \times \mathbf{R}^m \rightarrow \mathbf{R}$, with values

$$\mathcal{L}(x, \lambda) := f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$$

The variables $\lambda \in \mathbf{R}^m$, are called *Lagrange multipliers.*

We observe that, for every feasible $x \in \mathcal{X}$, and every $\lambda \geq 0$, $f_0(x)$ is bounded below by $\mathcal{L}(x, \lambda)$:

$$\forall \, x \in \mathcal{X}, \ \ \forall \, \lambda \in \mathbf{R}^m_+ \ : \ f_0(x) \geq \mathcal{L}(x, \lambda). \tag{7.2}$$

The Lagrangian can be used to express the primal problem (7.1) as an *unconstrained* one. Precisely:

$$p^* = \min_x \ \max_{\lambda \geq 0, \, \nu} \ \mathcal{L}(x, \lambda), \tag{7.3}$$

where we have used the fact that, for any vectors $f \in \mathbf{R}^m$, we have

$$\max_{\lambda \geq 0} \ \lambda^T f = \begin{cases} 0 & \text{if } f \leq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Lagrange dual function. We then define the *Lagrange dual function* (dual function for short) the function

$$g(\lambda) := \min_x \; \mathcal{L}(x, \lambda).$$

Note that, since g is the pointwise minimum of affine functions ($\mathcal{L}(x, \cdot)$ is affine for every x), it is concave. Note also that it may take the value $-\infty$.

From the bound (7.2), by minimizing over x in the right-hand side, we obtain

$$\forall \, x \in \mathcal{X}, \;\; \forall \, \lambda \geq 0 \; : \; f_0(x) \geq \min_{x'} \; \mathcal{L}(x', \lambda,) = g(\lambda),$$

which, after minimizing over x the left-hand side, leads to the lower bound

$$\forall \, \lambda \in \mathbf{R}^m_+, \;\; \nu \in \mathbf{R}^p \; : \; p^* \geq g(\lambda).$$

Lagrange dual problem. The best lower bound that we can obtain using the above bound is $p^* \geq d^*$, where

$$d^* = \max_{\lambda \geq 0, \, \nu} \; g(\lambda).$$

We refer to the above problem as the dual problem, and to the vector $\lambda \in \mathbf{R}^m$ as the dual variable. The dual problem involves the maximization of a concave function under convex (sign) constraints, so it is a convex problem. The dual problem always contains the implicit constraint $\lambda \in \mathbf{dom}\, g$.

We have obtained:

Theorem 1 (Weak duality). *For the general (possibly non-convex) problem (7.1),weak duality holds: $p^* \geq d^*$.*

Case with equality constraints. If equality constraints are present in the problem, we can represent them as two inequalities. It turns out that this leads to the same dual, as if we would directly use a single dual variable for each equality constraint, which is not restricted in sign. To see this, consider the problem

$$p^* := \min_x \; f_0(x) \; : \; \begin{array}{ll} f_i(x) \leq 0, & i = 1, \cdots, m, \\ h_i(x) = 0, & i = 1, \cdots, p. \end{array}$$

We write the problem as

$$p^* := \min_x \; f_0(x) \; : \; \begin{array}{l} f_i(x) \leq 0, \quad i = 1, \cdots, m, \\ h_i(x) \leq 0, \;\; -h_i(x) \leq 0, \quad i = 1, \cdots, p. \end{array}$$

Using a mulitplier $\nu_i^{\pm}$ for the constraint $\pm h_i(x) \leq 0$, we write the associated Lagrangian as

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

where $\nu := \nu^+ - \nu^-$ does not have any sign constraints.

Thus, inequality constraints in the original problem are associated with sign constraints on the corresponding multipliers, while the multipliers for the equality constraints are not explicitly constrained.

7.1.3 Minimax inequality

Weak duality can also be obtained as a consequence of the following *minimax inequality*, which is valid for *any* function ϕ of two vector variables x, y, and any subsets $\mathcal{X}$, $\mathcal{Y}$:

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \phi(x, y) \leq \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \phi(x, y). \tag{7.4}$$

To prove this, start from

$$\forall\, x, y \ : \ \min_{x' \in \mathcal{X}} \phi(x', y) \leq \max_{y' \in \mathcal{Y}} \phi(x, y').$$

and take the minimum over $x \in \mathcal{X}$ on the right-hand side, then the maximum over $y \in \mathcal{Y}$ on the left-hand side.

Weak duality is indeed a direct consequence of the above. To see this, start from the unconstrained formulation (7.3), and apply the above inequality with $\phi = \mathcal{L}$ the Lagrangian of the original problem, and $y = \lambda$ the vector of Lagrange multipliers.

Interpretation as a game. We can interpret the minimax inequality result in the context of a one-shot, zero-sum game. Assume that you have two players A and B, where A controls the decision variable x, while B controls y. We assume that both players have full knowledge of the other player's decision, once it is made. The player A seeks to minimize a payoff (to player B) $\mathcal{L}(x, y)$, while B seeks to maximize that payoff. The right-hand side in (7.4) is the optimal pay-off if the first player is required to play first. Obviously, the first player can do better by playing second, since then he or she knows the opponent's choice and can adapt to it.

7.2 Examples

7.2.1 Linear optimization problem

Inequality form. Consider the LP in standard inequality form

$$p^* = \min_x \ c^T x \ : \ Ax \leq b,$$

where $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, and the inequality in the constraint $Ax \leq b$ is interpreted component-wise.

The Lagrange function is

$$\mathcal{L}(x, \lambda) = c^T x + \lambda^T (Ax - b)$$

and the corresponding dual function is

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

The dual problem reads

$$d^* = \max_\lambda g(\lambda) = \max_\lambda -b^T \lambda \ : \ \lambda \geq 0, \ \ A^T \lambda + c = 0.$$

The dual problem is an LP in standard (sign-constrained) form, just as the primal problem was an LP in standard (inequality) form.

Weak duality implies that

$$c^T x + b^T \lambda \geq 0$$

for every x, λ such that $Ax \leq b$, $A^T \lambda = -c$. This property can be proven directly, by replacing c by $-A^T \lambda$ in the left-hand side of the above inequality, and exploiting $Ax \leq b$ and $\lambda \geq 0$.

Standard form. We can also consider an LP in standard form:

$$p^* = \min_x c^T x \ : \ Ax = b, \ \ x \geq 0.$$

The equality constraints are associated with a dual variable ν that is not constrained in the dual problem.

The Lagrange function is

$$\mathcal{L}(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (b - Ax)$$

and the corresponding dual function is

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

The dual problem reads

$$d^* = \max_{\lambda \geq 0, \, \nu} g(\lambda, \nu) = \max_\nu b^T \nu \ : \ c \geq A^T \nu.$$

This is an LP in inequality form.

7.2.2 Minimum Euclidean distance problem

Consider the problem of minimizing the Euclidean distance to a given affine space:

$$\min \frac{1}{2}\|x\|_2^2 \ : \ Ax = b, \tag{7.5}$$

where $A \in \mathbf{R}^{p \times n}$, $b \in \mathbf{R}^p$. We assume that A is full row rank, or equivalently, $AA^T \succ 0$. The Lagrangian is

$$\mathcal{L}(x, \nu) = \frac{1}{2}\|x\|_2^2 + \nu^T(Ax - b),$$

and the Lagrange dual function is

$$g(\nu) = \min_x \ \mathcal{L}(x, \nu) = \min_x \ \frac{1}{2}\|x\|_2^2 + \nu^T(Ax - b).$$

In this example, the dual function can be computed analytically, using the optimality condition $\nabla_x \mathcal{L}(x, \nu) = x + A^T\nu = 0$. We obtain $x = -A^T\nu$, and

$$g(\nu) = -\frac{1}{2}\nu^T AA^T \nu - b^T\nu.$$

The dual problem expresses as

$$d^* = \max_\nu \ g(\nu) = \max_\nu \ -\frac{1}{2}\nu^T AA^T\nu - b^T\nu.$$

The dual problem can also be solved analytically, since it is unconstrained (the domain of g is the entire space $\mathbf{R}^p$). We obtain $\nu^* = -(AA^T)^{-1}b$, and

$$d^* = \frac{1}{2}b^T(AA^T)^{-1}b.$$

We have thus obtained the bound $p^* \geq d^*$.

7.2.3 A non-convex boolean problem

For a given matrix $W = W^T \succ 0$, we consider the problem

$$p^* = \max_x \ x^T W x \ : \ x_i^2 \leq 1, i = 1, \ldots, n.$$

Lagrange relaxation. In this maximization problem, Lagrange duality will provide an upper bound on the problem. This is called a "relaxation", as we go above the true maximum, as if we'd relax (ignore) constraints.

Lagrange dual. The Lagrangian writes

$$\mathcal{L}(x,\lambda) = x^T W x + \sum_{i=1}^n \lambda_i (1 - x_i^2) = \mathbf{Tr}\, D_\lambda + x^T (W - D_\lambda) x.$$

where $D_\lambda := \mathbf{diag}(\lambda)$.

To find the dual function, we need to maximize the Lagrangian with respect to the primal variable x. We express this problem as

$$g(\lambda) = \max_x \; \mathcal{L}(x,\lambda) = \min_t \; t \;:\; \forall\, x, \;\; t \geq \mathbf{Tr}\, D_\lambda + x^T (W - D_\lambda) x.$$

The last inequality holds if and only if

$$\begin{pmatrix} D_\lambda - W & 0 \\ 0 & t - \mathbf{Tr}\, D_\lambda \end{pmatrix} \succeq 0.$$

Hence the dual function is the optimal value of an SDP in one variable:

$$g(\lambda) = \min_t \; t \;:\; \begin{pmatrix} D_\lambda - W & 0 \\ 0 & t - \mathbf{Tr}\, D_\lambda \end{pmatrix} \succeq 0.$$

We can solve this problem explicitly:

$$g(\lambda) = \begin{cases} \mathbf{Tr}\, D_\lambda & \text{if } D_\lambda \succeq W \\ -\infty & \text{otherwise.} \end{cases}$$

The dual problem involves minimizing (that is, getting the best upper bound) the dual function over the variable $\lambda \geq 0$:

$$d^* = \min_\lambda \; \lambda^T \mathbf{1} \;:\; \mathbf{diag}(\lambda) \succeq W.$$

The above is an SDP, in variable λ. Note that $\lambda > 0$ is automatically enforced by the PSD constraint.

Geometric interpretation. The Lagrange relaxation of the primal problem can be interpreted geometrically, as follows. For $t > 0$, $\lambda > 0$, consider the ellipsoids

$$\mathcal{E}_t = \left\{ x \;:\; x^T W x \leq t \right\}, \;\; \mathcal{E}_\lambda = \left\{ x \;:\; x^T D_\lambda x \leq \mathbf{Tr}\, D_\lambda \right\}.$$

The primal problem amounts to find the smallest $t \geq 0$ for which the ellipsoid $\mathcal{E}_t$ contains the ball $\mathcal{B}_\infty := \{x \;:\; \|x\|_\infty \leq 1\}$. Note that for every $\lambda > 0$, $\mathcal{E}_\lambda$ contains the ball $\mathcal{B}_\infty$. To find an upper bound on the problem, we can find the smallest t for which there exist $\lambda > 0$ such that $\mathcal{E}_t \supseteq \mathcal{E}_\lambda$. The latter condition is precisely $\mathbf{diag}(\lambda) \succeq W$, $t \geq \mathbf{Tr}\, D_\lambda$.

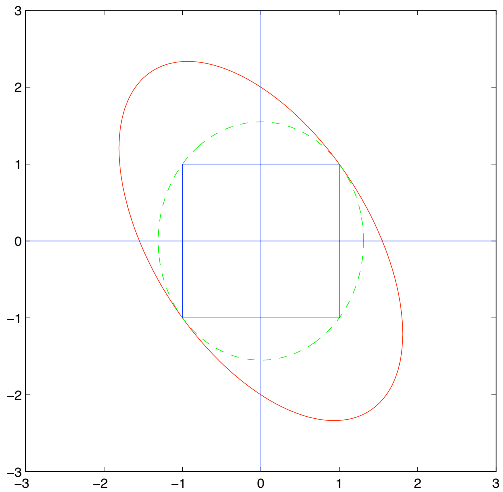

Figure 7.1. Geometric interpretation of dual problem in the boolean quadratic problem. In 2D the relaxation turns out to be exact.

7.3 More on non-convex quadratic optimization

The Boolean problem examined previously is part of a general class of non-convex quadratic problems of the form

$$p^* := \max_x \ q_0(x) \ : \ q_i(x) \leq 0, \ \ i = 1, \ldots, m, \tag{7.6}$$

where $x \in \mathbf{R}^n$ is the decision variable, and q_i's are quadratic functions, of the form

$$q_i(x) := x^T Q_i x + 2 q_i^T x + p_i, \ \ i = 1, \ldots, m,$$

Lagrange relaxation. The idea is that if, for a given m-vector $\lambda \geq 0$, and scalar t, we have

$$\forall \, x \ : \ q_0(x) \leq \sum_{i=1}^m \lambda_i q_i(x) + t,$$

then for every x that is feasible for (7.6), the sum in the above is non-positive. Hence, $q_0(x) \geq t$, so that t is an upper bound on our problem. The condition above is easy to check, as it involves a single quadratic function: indeed, it is equivalent to the LMI in (t, λ):

$$\begin{pmatrix} Q_0 & q_0 \\ q_0^T & r_0 \end{pmatrix} \preceq \sum_{i=1}^m \lambda_i \begin{pmatrix} Q_i & q_i \\ q_i^T & r_i \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & t \end{pmatrix}. \tag{7.7}$$

Hence, the best upper bound that we can achieve using this approach is the SDP

$$\min_{t, \lambda} \ t \ : \ (7.7), \ \ \lambda \geq 0.$$

The $\mathcal{S}$-lemma. This mysterious name corresponds to a special case of non-convex quadratic optimization, where there is only a *single* constraint. (Refer to appendix B of [BV] for more details.) The problem bears the form

$$\max_x \; q_0(x) \;:\; q_1(x) \leq 0,$$

where both q_0, q_1 are arbitrary quadratic functions. The $\mathcal{S}$-lemma states that if there exist a point $x \in \mathbf{R}^n$ such that $q_1(x) < 0$, then the Lagrange relaxation is exact. The latter has the form of an SDP:

$$\min_{t,\lambda} \; t \;:\; \begin{pmatrix} Q_0 & q_0 \\ q_0^T & r_0 \end{pmatrix} \preceq \lambda \begin{pmatrix} Q_1 & q_1 \\ q_1^T & r_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & t \end{pmatrix}, \;\; \lambda \geq 0. \tag{7.8}$$

This shows in particular that the apparently non-convex problem of finding the direction of maximal variance for a given covariance matrix Σ is actually convex. Lagrange relaxation for the problem

$$\max_{x \,:\, \|x\|_2 = 1} \; x^T \Sigma x$$

yields the dual problem (check this!)

$$\min_t \; t \;:\; tI \succeq \Sigma.$$

From the $\mathcal{S}$-lemma, the bound is exact. The $\mathcal{S}$-lemma has many other applications.

Exercises

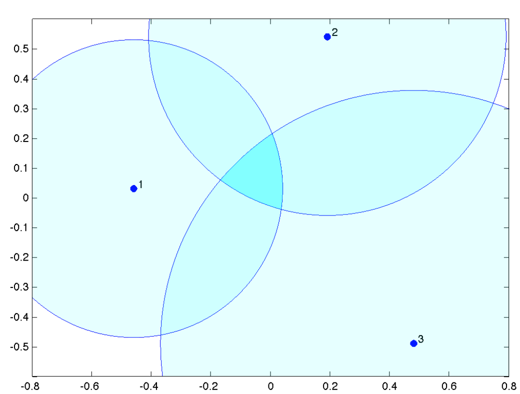

Figure 7.2. Localization problem with three range measurements in two dimensions.

1. *Anchor localization.* We are given anchor positions $x_i \in \mathbf{R}^3$, and associated distances from these anchor points to an unknown object, R_i, $i = 1, \ldots, m$. The problem is to estimate a position of the object, and associated measure of uncertainty around the estimated point. Geometrically, the measurements imply that the object is located at the intersection of the m spheres of centers x_i and radiuses R_i, $i = 1, \ldots, m$. The main problem is to provide one point in the intersection located at some kind of "center", and also a measure of the size of the intersection.

   In this problem, we seek to compute an outer spherical approximation to the intersection, that is, a sphere of center x_0 and radius R_0, of minimal volume, such that it contains the intersection.

   (a) First show how to find a point in the intersection, or determine it is empty, via SOCP. To simplify, and without loss of generality, we assume from now on that 0 is inside the intersection. This means that the vector z with components $z_i := R_i^2 - x_i^T x_i$, $i = 1, \ldots, m$, is non-negative componentwise.

   (b) A first approach, which works well only in moderate dimensions (2D or 3D), simply entails gridding the boundary of the intersection. In 2D, we can parametrize the boundary explicitly, as a curve, using an angular parameter. For each angular direction $\theta \in [0, 2\pi]$, we can easily find the point that is located on the boundary of the intersection, in the direction given by θ: we simply maximize t such that the point $(t\cos\theta, t\sin\theta)$ is inside everyone of the spheres. (There is an explicit formula for the maximal value.)

   One we have computed N points on the boundary, $x^{(k)}$, $k = 1, \ldots, N$, we simply solve the SOCP

$$\min_{x_0, R_0} R_0 \;:\; R_0 \geq \|x_0 - x^{(k)}\|_2, \quad k = 1, \ldots, N.$$

   Compare the results with a uniform gridding of 13 and 63 points. Use the data

$$X = (x_1, x_2, x_3) = \begin{pmatrix} -0.46 & 0.19 & 0.48 \\ 0.03 & 0.54 & -0.49 \end{pmatrix},$$
$$R^T = (R_1, R_2, R_3) = \begin{pmatrix} 0.5 & 0.6 & 0.85 \end{pmatrix}.$$

   (c) Show that the optimal (minimum-volume) spherical approximation can be obtained by solving

$$R_0^2 = \min_{x_0} \max_{x} \left\{ \|x - x_0\|_2^2 \;:\; \|x - x_i\|_2^2 \leq R_i^2, \;\; i = 1, \ldots, m \right\}.$$

   (d) Using Lagrange relaxation, show that an upper bound on the optimal radius can be obtained as

$$\min_{x_0, y} F(x_0, y), \;\text{ with } F(x_0, y) := \sum_{i=1}^m y_i R_i^2 + \max_x \left(\|x - x_0\|_2^2 - \sum_{i=1}^m y_i \|x - x_i\|_2^2 \right).$$

(e) Show that

$$F(x_0, y) = \begin{cases} x_0^T x_0 + y^T z + \frac{\|Xy - x_0\|_2^2}{\sum_{i=1}^m y_i - 1} & \text{if } \sum_{i=1}^m y_i > 1, \\ x_0^T x_0 + y^T z & \text{if } \sum_{i=1}^m y_i = 1, \quad x_0 = Xy, \\ +\infty & \text{otherwise.} \end{cases}$$

(f) Solve the problem via CVX, and compare your approximation with the gridding approach.

2. *Reachable sets for discrete-time dynamical systems.* Consider the discrete-time linear system

$$x(t+1) = Ax(t) + Bp(t), \quad t = 0, 1, 2, \ldots$$

where $A \in \mathbf{R}^{n \times n}$ and $B \in \mathbf{R}^{n \times n_p}$. We assume that the initial condition $x(0)$ is zero, while the signal p is considered to be noise, and is only known to be norm-bounded, precisely $\|p(t)\|_2 \leq 1$ for ever $t \geq 0$. The goal of *reachability analysis* is to come up with bounds on the state at a certain time T.

We seek a minimum-volume sphere $\mathcal{S}$ that is guaranteed to contain $x(T)$, irrespective of the values of the perturbation signal $p(t)$ within its bounds. By applying the recursion, we can express $x(T)$ as a linear combination of $p(0), \ldots, p(T-1)$:

$$x(T) = \sum_{t=0}^{T-1} A^t B p(t) = Lp,$$

where $p = (p(0), \ldots, p(T-1))$, and $L := [L(0), \ldots, L(T-1)]$, with $L(t) := A^t B$.

(a) Show that a sufficient condition for $\mathcal{S}$ to contain the state vector at time T is

$$\exists\, \lambda \geq 0 \;:\; \forall\, p = (p(0), \ldots, p(T-1)), \;\; p^T L^T L p \leq R_0^2 + \sum_{t=0}^{T-1} \lambda(t)(p(t)^T p(t) - 1).$$

(b) Show how to compute the best approximation based on the condition above, via SDP.