Lagrangian Dual Interpretations and Duality Applications

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/˜yyye

Chapters 11.7-8, 14.1-2

Recall Rules to Construct the Lagrangian Dual

$$\text{(GCO)} \qquad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & c_i\mathbf{x}) \quad (\le, =, \ge) \quad 0, \; i = 1, ..., m, \end{array}$$

- All multipliers are dual variables.
- Derive the LDC

$$\nabla f(\mathbf{x}) = \mathbf{y}^T \nabla \mathbf{c}(\mathbf{x})$$

If no $\mathbf{x}$ appeared in an equation, set it as an equality constraint for the dual; otherwise, express $\mathbf{x}$ in terms of $\mathbf{y}$ and replace $\mathbf{x}$ in the Lagrange function, which becomes the Dual objective. (This may be very difficult ...)

- Add the MSC as dual constraints.

The Lagrangian Dual of LP with Bound Constraints

Sometimes the dual can be constructed by simple reasoning: consider

$$\begin{array}{lll} (LP) & \text{minimize} & \mathbf{c}^T\mathbf{x} \\ & \text{subject to} & A\mathbf{x} = \mathbf{b},\ -\mathbf{e} \le \mathbf{x} \le \mathbf{e}\ (\|\mathbf{x}\|_\infty \le 1); \end{array}$$

Let the Lagrangian multipliers be $\mathbf{y}$ for equality constraints. Then the Lagrangian dual objective would be

$$\phi(\mathbf{y}) = \inf_{-\mathbf{e}\le\mathbf{x}\le\mathbf{e}} L(\mathbf{x},\mathbf{y}) = \inf_{-\mathbf{e}\le\mathbf{x}\le\mathbf{e}} \left[(\mathbf{c} - A^T\mathbf{y})^T\mathbf{x} + \mathbf{b}^T\mathbf{y}\right];$$

where if $(\mathbf{c} - A^T\mathbf{y})_j \le 0$, $x_j = 1$; and otherwise, $x_j = -1$.

Therefore, the Lagrangian dual is

$$\begin{array}{lll} (LDP) & \text{maximize} & \mathbf{b}^T\mathbf{y} - \|\mathbf{c} - A^T\mathbf{y}\|_1 \\ & \text{subject to} & \mathbf{y} \in R^m. \end{array}$$

The Lagrangian Dual of LP with the Log-Barrier I

For a fixed $\mu > 0$, consider the problem

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} - \mu \sum_{j=1}^n \log(x_j) \\ \text{s.t.} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

Again, the non-negativity constraints can be “ignored” if the feasible region has an ”interior”, that is, any minimizer must have $\mathbf{x}(\mu) > \mathbf{0}$. Thus, the Lagrangian function would be simply given by

$$L(\mathbf{x}, \mathbf{y}) = \mathbf{c}^T\mathbf{x} - \mu \sum_{j=1}^n \log(x_j) - \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) = (\mathbf{c} - A^T\mathbf{y})^T\mathbf{x} - \mu \sum_{j=1}^n \log(x_j) + \mathbf{b}^T\mathbf{y}.$$

Then, the Lagrangian dual objective (we implicitly need $\mathbf{x} > \mathbf{0}$ for the function to be defined)

$$\phi(\mathbf{y}) := \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{x}} \left[(\mathbf{c} - A^T\mathbf{y})^T\mathbf{x} - \mu \sum_{j=1}^n \log(x_j) + \mathbf{b}^T\mathbf{y} \right].$$

The Lagrangian Dual of LP with the Log-Barrier II

First, from the view point of the dual, the dual needs to choose $\mathbf{y}$ such that $\mathbf{c} - A^T\mathbf{y} > \mathbf{0}$, since otherwise the primal can choose $\mathbf{x} > \mathbf{0}$ to make $\phi(\mathbf{y})$ go to $-\infty$.

Now for any given $\mathbf{y}$ such that $\mathbf{c} - A^T\mathbf{y} > \mathbf{0}$, the $\inf$ problem has a unique finite close-form minimizer $\mathbf{x}$

$$x_j = \frac{\mu}{(\mathbf{c} - A^T\mathbf{y})_j}, \ \forall j = 1, ..., n.$$

Thus,

$$\phi(\mathbf{y}) = \mathbf{b}^T\mathbf{y} + \mu \sum_{j=1}^{n} \log(\mathbf{c} - A^T\mathbf{y})_j + n\mu(1 - \log(\mu)).$$

Therefore, the dual problem, for any fixed μ, can be written as

$$\max_{\mathbf{y}} \ \phi(\mathbf{y}) = n\mu(1 - \log(\mu)) + \max_{\mathbf{y}} \, [\mathbf{b}^T\mathbf{y} + \mu \sum_{j=1}^{n} \log(\mathbf{c} - A^T\mathbf{y})_j].$$

This is actually the LP dual with the Log-Barrier on dual inequality constraints $\mathbf{c} - A^T\mathbf{y} \geq \mathbf{0}$.

The Dual of SVM

$$\begin{aligned} \text{minimize}_{\mathbf{x},x_0,\beta} \quad & \beta + \mu\|\mathbf{x}\|^2 \\ \text{subject to} \quad & \mathbf{a}_i^T\mathbf{x} + x_0 + \beta \geq 1,\ \forall i,\ (\mathbf{y}_a \geq \mathbf{0}) \\ & -\mathbf{b}_j^T\mathbf{x} - x_0 + \beta \geq 1,\ \forall j,\ (\mathbf{y}_b \geq \mathbf{0}) \\ & \beta \geq 0.\ (\alpha \geq 0) \end{aligned}$$

$$L(\mathbf{x}, x_0, \beta, \mathbf{y}_a, \mathbf{y}_b, \alpha) = \beta + \mu\|\mathbf{x}\|^2 - \mathbf{y}_a^T(A^T\mathbf{x} + x_0\mathbf{e} + \beta\mathbf{e} - \mathbf{e}) - \mathbf{y}_b^T(-B^T\mathbf{x} - x_0\mathbf{e} + \beta\mathbf{e} - \mathbf{e}) - \alpha\beta.$$

$$\begin{aligned} \nabla_{\mathbf{x}} L(\cdot) &= 2\mu\mathbf{x} - A\mathbf{y}_a + B\mathbf{y}_b = \mathbf{0},\ (\text{replace } \mathbf{x}) \\ \nabla_{x_0} L(\cdot) &= -\mathbf{e}^T\mathbf{y}_a + \mathbf{e}^T\mathbf{y}_b = 0,\ (\text{dual constraint}) \\ \nabla_{\beta} L(\cdot) &= 1 - \mathbf{e}^T\mathbf{y}_a - \mathbf{e}^T\mathbf{y}_b - \alpha = 0.\ (\text{dual constraint}) \end{aligned}$$

Then the dual objective is

$$\frac{-1}{4\mu}\|A\mathbf{y}_a - B\mathbf{y}_b\|^2 + \mathbf{e}^T\mathbf{y}_a + \mathbf{e}^T\mathbf{y}_b.$$

The Lagrangian Dual of LP with the Fisher Market

$$\begin{array}{rl} \max & \sum_{i \in B} w_i \log(\mathbf{u}_i^T \mathbf{x}_i) \\ \text{s.t.} & \sum_{i \in B} \mathbf{x}_i = \mathbf{b}, \quad \forall j \in G \\ & x_{ij} \geq 0, \quad \forall i, j, \end{array}$$

The Lagrangian function would be simply given by

$$L(\mathbf{x}_i \geq \mathbf{0}, i \in B, \mathbf{y}) = \sum_{i \in B} w_i \log(\mathbf{u}_i^T \mathbf{x}_i) - \mathbf{y}^T (\sum_{i \in B} \mathbf{x}_i - \mathbf{b}) = \sum_{i \in B} (w_i \log(\mathbf{u}_i^T \mathbf{x}_i) - \mathbf{y}^T \mathbf{x}_i) + \mathbf{b}^T \mathbf{y}.$$

Then, the Lagrangian dual objective, for any given $\mathbf{y} > \mathbf{0}$, would be

$$\phi(\mathbf{y}) := \sup_{\mathbf{x}_i \geq \mathbf{0}, i \in B} L(\mathbf{x}_i, i \in B, \mathbf{y}) = \inf_{\mathbf{x}_i \geq \mathbf{0}, i \in B} \sum_{i \in B} (w_i \log(\mathbf{u}_i^T \mathbf{x}_i) - \mathbf{y}^T \mathbf{x}_i) + \mathbf{b}^T \mathbf{y}.$$

The Lagrangian Dual of LP with the Fisher Market II

For each $i \in B$, the sup-solution is

$$x_{ij^*} = \frac{w_i}{y_{j^*}} > 0,\ j^* = \arg\min_j \frac{y_j}{u_{ij}},\ x_{ij} = 0\ \forall j \neq j^*.$$

Thus,

$$\phi(\mathbf{y}) = \mathbf{b}^T\mathbf{y} - \sum_{i \in B} w_i \log\left(\min_j[\frac{y_j}{u_{ij}}]\right) + \sum_{i \in B} w_i(\log(w_i) - 1).$$

The gradient and Hessian of ϕ

Let $\mathbf{x}(\mathbf{y})$ be a minimizer. Then

$$\phi(\mathbf{y}) = f(\mathbf{x}(\mathbf{y})) - \mathbf{y}^T\mathbf{h}(\mathbf{x}(\mathbf{y}))$$

Thus,

$$\begin{aligned}\nabla\phi(\mathbf{y}) &= \nabla f(\mathbf{x}(\mathbf{y}))^T\nabla\mathbf{x}(\mathbf{y}) - \mathbf{y}^T\nabla\mathbf{h}(\mathbf{x}(\mathbf{y}))\nabla\mathbf{x}(\mathbf{y}) - \mathbf{h}(\mathbf{x}(\mathbf{y})) \\ &= (\nabla f(\mathbf{x}(\mathbf{y}))^T - \mathbf{y}^T\nabla\mathbf{h}(\mathbf{x}(\mathbf{y})))\nabla\mathbf{x}(\mathbf{y}) - \mathbf{h}(\mathbf{x}(\mathbf{y})) \\ &= -\mathbf{h}(\mathbf{x}(\mathbf{y})).\end{aligned}$$

Similarly, we can derive

$$\nabla^2\phi(\mathbf{y}) = -\nabla\mathbf{h}(\mathbf{x}(\mathbf{y}))\left(\nabla^2_{\mathbf{x}}L(\mathbf{x}(\mathbf{y}),\mathbf{y})\right)^{-1}\nabla\mathbf{h}(\mathbf{x}(\mathbf{y}))^T,$$

where $\nabla^2_{\mathbf{x}}L(\mathbf{x}(\mathbf{y}),\mathbf{y})$ is the Hessian of the Lagrangian function that is assumed to be positive definite at any (local) minimizer.

The Toy Example

$$\begin{array}{ll}\text{minimize} & (x_1-1)^2+(x_2-1)^2\\ \text{subject to} & x_1+2x_2-1=0,\quad 2x_1+x_2-1=0.\end{array}$$

$$L(\mathbf{x},\mathbf{y})=(x_1-1)^2+(x_2-1)^2-y_1(x_1+2x_2-1)-y_2(2x_1+x_2-1).$$

$$x_1=0.5y_1+y_2+1,\quad x_2=y_1+0.5y_2+1.$$

$$\phi(\mathbf{y})=-1.25y_1^2-1.25y_2^2-2y_1y_2-2y_1-2y_2.$$

$$\nabla\phi(\mathbf{y})=\begin{pmatrix}2.5y_1+2y_2+2\\ 2y_1+2.5y_21+2\end{pmatrix},$$

$$\nabla^2\phi(\mathbf{y})=-\begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}^{-1}\begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}^T=-\begin{pmatrix}2.5 & 2\\ 2 & 2.5\end{pmatrix}$$

The Fisher Example again

$$\begin{array}{ll} \text{minimize} & -5\log(2x_1+x_2)-8\log(3x_3+x_4) \\ \text{subject to} & x_1+x_3=1, \quad x_2+x_4=1, \quad \mathbf{x}\geq \mathbf{0}. \end{array}$$

$$L(\mathbf{x}(\geq \mathbf{0}),\mathbf{y}) = -5\log(2x_1+x_2)-8\log(3x_3+x_4)-y_1(x_1+x_3-1)-y_2(x_2+x_4-1).$$

Start from $\mathbf{y}^0 > \mathbf{0}$, at the kth step, compute $\mathbf{x}^{k+1}$ from

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\geq\mathbf{0}} \ L(\mathbf{x}(\geq \mathbf{0}),\mathbf{y}^k),$$

then let

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \frac{1}{\beta}(A\mathbf{x}^{k+1}-\mathbf{b}).$$

Farkas Lemma for Nonlinear Constraints I

Consider the convex constrained system:

$$\text{(CCS)} \qquad \begin{array}{lrcl} \min & \mathbf{0}^T\mathbf{x} & & \\ \text{s.t.} & c_i(\mathbf{x}) & \geq & 0,\ i = 1, ..., m, \end{array}$$

where $c_i(.)$ are concave functions and the Lagrangian Function is given by

$$L(\mathbf{x}, \mathbf{y}) = -\mathbf{y}^T\mathbf{c}(\mathbf{x}) = -\sum_{i=1}^{m} y_i c_i(\mathbf{x}),\ \mathbf{y} \geq \mathbf{0}.$$

Again, let

$$\phi(\mathbf{y}) := \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}).$$

Theorem 1 *If there exists $\mathbf{y} \geq \mathbf{0}$ such that $\phi(\mathbf{y}) > 0$, then (CSS) is infeasible.*

The proof is directly from the dual objective function $\phi(\mathbf{y})$ is a homogeneous function and the dual has its objective value unbounded from above.

Farkas Lemma for Nonlinear Constraints II

Consider the system, for a parameter $b \geq 0$,

$$-x_1^2 - (x_2 - 1)^2 + b \geq 0, \quad (y_1 \geq 0)$$
$$-x_1^2 - (x_2 + 1)^2 + b \geq 0, \quad (y_2 \geq 0)$$

$$L(\mathbf{x}, \mathbf{y}) = y_1(x_1^2 + (x_2 - 1)^2 - b) + y_2(x_1^2 + (x_2 + 1)^2 - b).$$

Then, if $y_1 + y_2 \neq 0$,

$$\phi(\mathbf{y}) = \frac{4y_1y_2 - b(y_1 + y_2)^2}{y_1 + y_2}, \quad (y_1, y_2) \geq 0$$

When $b \geq 1$, $\phi(\mathbf{y}) \leq 0$; and, otherwise, one can choose $y_1 = y_2 = y > 0$ such that

$$\phi(\mathbf{y}) = 2(1 - b)y > 0$$

which implies that the original constrained system is infeasible.

The Augmented Lagrangian Function

For equality constraints $\{\mathbf{x} : \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$, in both theory and practice, we can consider an augmented Lagrangian function (ALF)

$$L_a(\mathbf{x}, \mathbf{y}, \mathbf{s}) = f(\mathbf{x}) - \mathbf{y}^T\mathbf{h}(\mathbf{x}) - \mathbf{s}^T\mathbf{c}(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{h}(\mathbf{x})\|^2$$

for some positive parameter ρ, which corresponds to an equivalent problem of (**??**):

$$f^* := \min \quad f(\mathbf{x}) + \frac{\beta}{2}\|\mathbf{h}(\mathbf{x})\|^2 \quad \text{s.t.} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}.$$

Note that, although at feasibility the additional square term in objective is redundant, it helps to improve strict convexity of the Lagrangian function.

For the Fisher example:

$$\begin{array}{rl} & L_a(\mathbf{x}(\geq \mathbf{0}), \mathbf{y}) \\ = & -5\log(2x_1 + x_2) - 8\log(3x_3 + x_4) - y_1(x_1 + x_3 - 1) - y_2(x_2 + x_4 - 1) \\ & +\frac{\beta}{2}((x_1 + x_3 - 1)^2 + (x_2 + x_4 - 1)^2). \end{array}$$

The Augmented Lagrangian Dual

Now the dual function:

$$\phi_a(\mathbf{y}) = \min_{\mathbf{x} \in X} L_a(\mathbf{x}, \mathbf{y}); \tag{1}$$

and the dual problem

$$(f^* \geq) \phi_a^* := \max \quad \phi_a(\mathbf{y}). \tag{2}$$

Note that the dual function approximately satisfies $\frac{1}{\beta}$-Lipschitz condition (see Chapter 14 of L&Y).

For the convex optimization case, say $\mathbf{h}(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$, we have

$$\nabla^2 L_a(\mathbf{x}, \mathbf{y}) = \nabla^2 f(\mathbf{x}) + \beta(A^T A).$$