Convex Optimization: Reading Notes 4

GKxx

May 2, 2022

1 Lagrange duality

Definition 1.1 (Lagrangian). *For an optimization problem*

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leqslant 0, \quad i = 1, \cdots, m, \\ & h_i(x) = 0, \quad i = 1, \cdots, p, \end{aligned} \tag{1}$$

with variable $x \in \mathbb{R}^n$*, domain* $\mathcal{D} = \bigcap_{i=0}^m \operatorname{\mathbf{dom}} f_i \cap \bigcap_{i=1}^p h_i$ *and optimal value* p^**, we define its* Lagrangian *to be*

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

with domain

$$\operatorname{\mathbf{dom}} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p.$$

The variables λ_i *and* ν_i *are called the* Lagrangian multipliers *or* dual variables.

Definition 1.2 (Lagrange dual function). *For the optimization problem defined as in (1), the* Lagrange dual function *is*

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

Remark 1.3. *The Lagrange dual function is concave.*

Proposition 1.4 (Lower bound property). $g(\lambda, \nu) \leqslant p^*$ *for every* $\lambda \succeq 0$.

Proof. Suppose the optimal value p^* is achieved at $x = \tilde{x}$. For every $\lambda \succeq 0$, we have that

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leqslant L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x),$$

where $f_i(x) \leqslant 0$ and $h_j(x) = 0$ for every $i = 1, \cdots, m, j = 1, \cdots, p$, so

$$g(\lambda, \nu) \leqslant f_0(\tilde{x}).$$

$\square$

Definition 1.5 (Dual feasibility)**.** *A pair* (λ, ν) *is referred to as* dual feasible *if* $\lambda \succeq 0$ *and* $(\lambda, \nu) \in \textbf{dom}\, g$.

Remark 1.6 (Conjugate function)**.** *The conjugate function of* f *is*

$$f^*(y) = \sup_x \left(y^T x - f(x)\right).$$

Proposition 1.7. *For an optimization problem with linear constraints*

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & Ax \preceq b \\ & Cx = d, \end{aligned}$$

the Lagrangian is

$$L(x, \lambda, \nu) = f_0(x) + \lambda^T(Ax - b) + \nu^T(Cx - d),$$

and the Lagrange dual function is

$$\begin{aligned} g(\lambda, \nu) &= \inf_x \left(f_0(x) + \left(A^T\lambda + C^T\nu\right)^T x\right) - b^T\lambda - d^T\nu \\ &= f_0^*\left(-A^T\lambda - C^T\nu\right) - b^T\lambda - d^T\nu. \end{aligned}$$

Remark 1.8. *The Lagrange dual problem is convex.*

Definition 1.9 (Lagrange dual problem)**.** *For an optimization problem in the form of (1), the* Lagrange dual problem *is*

$$\begin{aligned} \max \quad & g(\lambda, \nu) \\ \text{s.t.} \quad & \lambda \succeq 0. \end{aligned}$$

The optimal value is denoted by d^*. *The difference* $p^* - d^*$ *is called the* duality gap.

Example 1.10 (Linear program)**.** *The Lagrange dual of a linear program*

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \preceq b \end{aligned} \tag{2}$$

is

$$\begin{aligned} \max \quad & -b^T\lambda \\ \text{s.t.} \quad & A^T\lambda + c = 0 \\ & \lambda \succeq 0. \end{aligned} \tag{3}$$

Proof. The Lagrangian of problem (2) is

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

so the Lagrange dual function is

$$g(\lambda) = \inf_x \left(c + A^T\lambda\right)^T x - b^T\lambda.$$

Note that unless $c+A^T\lambda = 0$, the infimum of a linear function $\left(c + A^T\lambda\right)^T x$ is $-\infty$. Therefore

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

Therefore, the dual problem is

$$\begin{aligned} \max \quad & -b^T\lambda \\ \text{s.t.} \quad & A^T\lambda + c = 0 \\ & \lambda \succeq 0. \end{aligned}$$

□

Theorem 1.11 (Weak duality). $d^* \leqslant p^*$. *It follows directly from the lower bound property.*

Definition 1.12 (Strong duality). *If* $d^* = p^*$ *for some problem, we say that the* strong duality *holds.*

Definition 1.13 (Slater's condition). *For a convex problem*

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leqslant 0, \quad i = 1, \cdots, m \\ & Ax = b \end{aligned} \tag{4}$$

with domain $\mathcal{D}$*, we say that the* Slater's condition *holds if there exists* $x \in \textbf{relint}\,\mathcal{D}$ *such that* $f_i(x) < 0$ *for every* $i = 1, \cdots, m$*. Such a point is also called* strictly feeasible*.*

Definition 1.14 (Refined Slater's condition). *The refined Slater's condition does not require affine inequalities to hold with strict inequality.*

Theorem 1.15. *Both the Slater's condition and its refined condition imply strong duality.*

Proof. We consider the convex problem of the form (4). Assume that the matrix A has full rank, otherwise we can eliminate some of the equality constraints that are expressible as linear combinations of other constraints. We further assume that the domain $\mathcal{D}$ has nonempty interior, which means $\textbf{relint}\,\mathcal{D} = \textbf{int}\,\mathcal{D}$. Suppose that Slater's condition holds.

Define

$$\mathcal{A} = \{(u, v, t) \mid \exists x \in \mathcal{D} \text{ s.t. } f(x) \preceq u, Ax - b = v, f_0(x) \leqslant t\},$$

and

$$\mathcal{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid s < p^*\}.$$

It is easy to see by definition that $\mathcal{A}$ and $\mathcal{B}$ are convex. Pick $(u, v, t) \in \mathcal{A} \cap \mathcal{B}$, so that $\exists x \in \mathcal{D}$ s.t. $f_0(x) \leqslant t < p^*$, which is impossible. This shows that $\mathcal{A}$ and $\mathcal{B}$ are disjoint. By the separating hyperplane theorem, there exist $(\tilde{\lambda}, \tilde{v}, \mu) \neq 0$ and α such that

$$\forall (u, v, t) \in \mathcal{A}, \quad \tilde{\lambda}^T u + \tilde{v}^T v + \mu t \geqslant \alpha,$$

and

$$\forall (u, v, t) \in \mathcal{B}, \quad \tilde{\lambda}^T u + \tilde{v}^T v + \mu t = \mu t \leqslant \alpha.$$

Note that for every $(u, v, t) \in \mathcal{A}$, any (u', v, t') satisfying $u' \succ u, t' > t$ is also in $\mathcal{A}$. Therefore, $\tilde{\lambda} \succeq 0$ and $\mu \geqslant 0$ must hold, otherwise $\tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t$ would not be bounded below by α.

Since $\mu \geqslant 0$ and for every $s < p^*$ we have $\mu s \leqslant \alpha$, it follows that $\mu p^* \leqslant \alpha$, so we obtain

$$\mu p^* \leqslant \alpha \leqslant \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t$$

for every $(u, v, t) \in \mathcal{A}$. This gives that for every $x \in \mathcal{D}$,

$$\mu p^* \leqslant \alpha \leqslant \sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) + \mu f_0(x).$$

When $\mu > 0$, the above leads to

$$L\left(x, \frac{\tilde{\lambda}}{\mu}, \frac{\tilde{\nu}}{\mu}\right) \geqslant \frac{\alpha}{\mu} \geqslant p^*$$

for any $x \in \mathcal{D}$. Therefore

$$g\left(\frac{\tilde{\lambda}}{\mu}, \frac{\tilde{\nu}}{\mu}\right) = \inf_x L\left(x, \frac{\tilde{\lambda}}{\mu}, \frac{\tilde{\nu}}{\mu}\right) \geqslant p^*,$$

but the weak duality theorem says that

$$g(\lambda, \nu) \leqslant p^*, \quad \forall \lambda \succeq 0, \nu.$$

Hence $g\left(\frac{\tilde{\lambda}}{\mu}, \frac{\tilde{\nu}}{\mu}\right) = p^*$, so the strong duality holds.

If $\mu = 0$, the Slater's condition gives that for some $\tilde{x} \in \textbf{int}\,\mathcal{D}$,

$$f_i(\tilde{x}) < 0, i = 1, \cdots, m, \quad \text{and} \quad A\tilde{x} - b = 0.$$

Therefore

$$0 \leqslant \alpha \leqslant \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}),$$

but $\tilde{\lambda} \succeq 0$ and $f_i(\tilde{x}) < 0$ holds for every i. This implies $\tilde{\lambda} = 0$, so $\tilde{\nu}^T (Ax - b) \geqslant 0 \, \forall x \in \mathcal{D}$. From $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ and $\tilde{\lambda} = 0, \mu = 0$ we obtain $\tilde{\nu} \neq 0$. Since $\tilde{x} \in \textbf{int}\,\mathcal{D}$ and $\tilde{n}u^T (A\tilde{x} - b) = 0$, there must be some $x \in D$ with $\tilde{\nu}^T (Ax - b) < 0$ unless $\tilde{\nu}^T A = 0$, while this contradicts the assumption that A has full rank. □

Theorem 1.16. *Strong duality also holds for problems with a quadratic objective and one quadratic inequality constraint, provided that Slater's condition holds.*

2 Optimality conditions

Theorem 2.1 (Complementary slackness). *Suppose that the strong duality holds. If x^* is primal optimal and (λ^*, ν^*) is dual optimal, then*

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \cdots, m.$$

Proof. We have that

$$\begin{aligned} f(x^*) = g(\lambda^*, \nu^*) &= \inf_x L(x, \lambda^*, \nu^*) \\ &\leqslant L(x^*, \lambda^*, \nu^*) = f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \leqslant f_0(x^*). \end{aligned}$$

Since $h_i(x^*) = 0, i = 1, \cdots, p$, it follows that

$$\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0,$$

while $\lambda_i^* \geqslant 0, f_i(x^*) \leqslant 0$. Therefore

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \cdots, m.$$

□

Definition 2.2 (Karush-Kuhn-Tucker conditions). *Suppose $f_0, f_1, \cdots, f_m, h_1, \cdots, h_p$ are differentiable. For x, λ, ν, the* KKT conditions *are*

1. *Primal feasible:* $f_i(x) \leqslant 0, i = 1, \cdots, m$ *and* $h_i(x) = 0, i = 1, \cdots, p$.
2. *Dual feasible:* $\lambda \succeq 0$.
3. *Complementary slackness:* $\lambda_i f_i(x) = 0, i = 1, \cdots, m$.
4. $\nabla_x L(x, \lambda, \nu) = 0$, *i.e.*

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

Theorem 2.3. *When strong duality holds, any primal optimal $\tilde{x}$ and dual optimal $\left(\tilde{\lambda}, \tilde{\nu}\right)$ must satisfy the KKT conditions.*

Theorem 2.4. *For convex problems, if $\left(\tilde{x}, \tilde{\lambda}, \tilde{\nu}\right)$ satisfy the KKT conditions, then $\tilde{x}$ is primal optimal and $\left(\tilde{\lambda}, \tilde{\nu}\right)$ is dual optimal, with zero duality gap.*

Proof. Suppose $\left(\tilde{x}, \tilde{\lambda}, \tilde{\nu}\right)$ satisfy the KKT conditions. The dual feasibility implies that $L\left(x, \tilde{\lambda}, \tilde{\nu}\right)$ is convex in x. Since $\nabla_x L\left(x, \tilde{\lambda}, \tilde{\nu}\right) = 0$, it follows that $\tilde{x}$ minimizes $L\left(x, \tilde{\lambda}, \tilde{\nu}\right)$. Therefore

$$g\left(\tilde{\lambda}, \tilde{\nu}\right) = \inf_x L\left(x, \tilde{\lambda}, \tilde{\nu}\right) = L\left(\tilde{x}, \tilde{\lambda}, \tilde{\nu}\right) = f_0\left(\tilde{x}\right) + \sum_{i=1}^{m} \lambda_i f_i\left(\tilde{x}\right).$$

From the complementary slackness we obtain $g\left(\tilde{\lambda}, \tilde{\nu}\right) = f_0\left(\tilde{x}\right)$, therefore strong duality holds, and $\left(\tilde{x}, \tilde{\lambda}, \tilde{\nu}\right)$ is optimal. □

Corollary 2.5. *If Slater's condition holds for a convex problem, then $\tilde{x}$ is optimal if and only if $\exists \left(\tilde{\lambda}, \tilde{\nu}\right)$* s.t. *$\left(\tilde{x}, \tilde{\lambda}, \tilde{\nu}\right)$ satisfy the KKT conditions.*

Definition 2.6 (Active set)**.** *The* active set *at a feasible point x is*

$$\mathcal{A}(x) = \{i \mid f_i(x) = 0, 1 \leqslant i \leqslant m\}.$$

Definition 2.7 (LICQ)**.** *The* linear independence constraint qualification *is said to hold at x if the set of active constraint gradients*

$$\{\nabla f_i(x), \nabla h_j(x) \mid i \in \mathcal{A}(x), j = 1, \cdots, p\}$$

is linearly independent at x.

Theorem 2.8. *Suppose x^* is a local optimal at which the LICQ holds. Then there exists Lagrange multipliers (λ^*, ν^*) such that (x^*, λ^*, ν^*) satisfies the KKT conditions.*

Definition 2.9 (MFCQ)**.** *The* Mangasarian-Fromovitz constraint qualification *is said to hold at x if the set of equality constrait gradients*

$$\{\nabla h_i(x) \mid i = 1, \cdots, p\}$$

is linearly independnet, and there exists d such that

$$\nabla h_i(x)^T d = 0 \, \forall i = 1, \cdots, p, \quad \text{and} \quad \nabla f_i(x)^T d < 0 \, \forall i \in \mathcal{A}(x).$$

Theorem 2.10. *LICQ implies MFCQ.*

Definition 2.11 (Strong Slater's condition)**.** *The* strong Slater's condition *is satisfied if the set of equality constraint gradients*

$$\{\nabla h_i(x) \mid i = 1, \cdots, p\}$$

is linearly independent and there exists a feasible point strictly satisfying all inequality constraints, i.e. $\exists x \in \mathcal{D}$ such that

$$f_i(x) < 0, i = 1, \cdots, m \quad \text{and} \quad h_i(x) = 0, i = 1, \cdots, p.$$

Theorem 2.12. *The Slater's condition (1.13) implies the existence of a nonempty, closed, convex set Λ_* such that for all $(\lambda^*, \nu^*) \in \Lambda_*$, (x, λ^*, ν^*) satisfies the KKT conditions. The strong Slater's condition (2.11) implies the existence of such a Λ_* that is bounded.*

For more common and useful CQs and their relationship with KKT conditions, see [1].

References

[1] G. Wachsmuth, "On licq and the uniqueness of lagrange multipliers," *Operations Research Letters*, vol. 41, no. 1, pp. 78–80, 2013.