Lecture 4

Based on Chapter 2 of Wright's book

Optimality Conditions

Consider the following problem:

$$\min c^T x \text{ subject to } Ax = b,\, x \geq 0, \tag{2.1}$$

where

$$c, x \in \mathbb{R}^n, \qquad b \in \mathbb{R}^m, \qquad A \in \mathbb{R}^{m \times n}.$$

The dual of (2.1) is

$$\max b^T \lambda \text{ subject to } A^T \lambda + s = c,\, s \geq 0, \tag{2.2}$$

where $\lambda \in \mathbb{R}^m$ and $s \in \mathbb{R}^n$.

The optimality conditions are:

$$\begin{aligned} A^T \lambda + s &= c, && (2.4a) \\ Ax &= b, && (2.4b) \\ x_i s_i &= 0, \quad i = 1, 2, \ldots, n, && (2.4c) \\ (x, s) &\geq 0. && (2.4d) \end{aligned}$$

Strict Feasibility

Let $\Omega_P = \{x^* \,|\, x^* \text{ solves } (2.1)\}, \qquad \Omega_D = \{(\lambda^*, s^*) \,|\, (\lambda^*, s^*) \text{ solves } (2.2)\}.$

A vector x is said to be *strictly feasible* for the primal problem if

$$Ax = b, \qquad x > 0;$$

that is, the positivity condition $x \geq 0$ is replaced by a *strict* inequality. Similarly, the vector (λ, s) is strictly feasible for the dual problem if

$$A^T\lambda + s = c, \qquad s > 0.$$

Theorem 2.3 *Suppose that the primal and dual problems are feasible; that is, $\mathcal{F} \neq \emptyset$. Then if the dual problem has a strictly feasible point, the primal solution set Ω_P is nonempty and bounded. Similarly, if the primal problem has a strictly feasible point, the set*

$$\{s^* \,|\, (\lambda^*, s^*) \in \Omega_D \ \textit{for some } \lambda^* \in \mathbb{R}^m\}$$

is nonempty and bounded.

Proof. We prove the first statement and leave the second as an exercise.

Denote the strictly feasible dual point by $(\bar{\lambda}, \bar{s})$, and let $\hat{x}$ be any primal feasible point (not necessarily strict). Then as in (2.5) we have

$$0 \le \bar{s}^T \hat{x} = c^T \hat{x} - b^T \bar{\lambda}. \tag{2.10}$$

Now consider the set of points $\mathcal{T}$ defined by

$$\mathcal{T} = \{x \mid Ax = b, \quad x \ge 0, \quad c^T x \le c^T \hat{x}\}.$$

The set $\mathcal{T}$ is nonempty (since $\hat{x} \in \mathcal{T}$) and is obviously closed. For any $x \in \mathcal{T}$, we have from (2.5) and (2.10) that

$$\sum_{i=1}^{n} \bar{s}_i x_i = \bar{s}^T x = c^T x - b^T \bar{\lambda} \le c^T \hat{x} - b^T \bar{\lambda} = \bar{s}^T \hat{x}.$$

Since all terms in the summation on the left are nonnegative, we have

$$x_i \le \frac{1}{\bar{s}_i} \bar{s}^T \hat{x} \quad \Rightarrow \quad \|x\|_\infty \le \left(\max_{i=1,2,\ldots,n} \frac{1}{\bar{s}_i} \right) \bar{s}^T \hat{x}.$$

Since x was an arbitrary element of $\mathcal{T}$, we conclude from this inequality that $\mathcal{T}$ is bounded, as well as nonempty and closed. Hence, the function $c^T x$ must attain its minimum value on $\mathcal{T}$. That is, there exists a point x^* such that

$$x^* \in \mathcal{T}, \qquad c^T x^* \le c^T x \text{ for all } x \in \mathcal{T}.$$

Obviously, the set of points x^* with this property coincides with Ω_P. Hence, Ω_P is nonempty and bounded, since it is a subset of the bounded set $\mathcal{T}$. □

Strict Complementarity

The converse of Theorem 2.3 is also true: If the primal solution set Ω_P is nonempty and bounded, the dual problem has a strictly feasible point (similarly for Ω_D and the primal feasible set). The proof makes use of a "theorem of the alternative" (see, for example, Mangasarian [81, Chapter 2]), and we omit it here.

The $\mathcal{B} \cup \mathcal{N}$ Partition and Strict Complementarity

For every solution (x^*, λ^*, s^*), we know from (2.4c) that

$$x_j^* = 0 \text{ and/or } s_j^* = 0 \text{ for all } j = 1, 2, \ldots, n.$$

We can define two index sets $\mathcal{B}$ and $\mathcal{N}$ as follows.

$$\mathcal{B} = \{j \in \{1, 2, \ldots, n\} \,|\, x_j^* \neq 0 \text{ for some } x^* \in \Omega_P\}, \quad (2.11a)$$

$$\mathcal{N} = \{j \in \{1, 2, \ldots, n\} \,|\, s_j^* \neq 0 \text{ for some } (\lambda^*, s^*) \in \Omega_D\}. \quad (2.11b)$$

- Obviously $\mathcal{B} \cap \mathcal{N} = \emptyset$.

Theorem 2.4 (Goldman–Tucker) $\mathcal{B} \cup \mathcal{N} = \{1, 2, \ldots, n\}$. *That is, there exist at least one primal solution $x^* \in \Omega_P$ and one dual solution $(\lambda^*, s^*) \in \Omega_D$ such that $x^* + s^* > 0$.*

Primal-dual solutions (x^*, λ^*, s^*) with the property $x^* + s^* > 0$ are known as *strictly complementary* solutions. Theorem 2.4 guarantees that at least one such solution exists. A given linear program may have multiple primal-dual solutions, some that are strictly complementary and others that are not, as the following trivial example illustrates. Consider

$$\min_{x \in \mathbb{R}^3} x_1 \quad \text{subject to } x_1 + x_2 + x_3 = 1, \; x \geq 0, \tag{2.12}$$

whose dual is

$$\max_{\lambda \in \mathbb{R},\, s \in \mathbb{R}^3} \lambda \quad \text{subject to} \quad \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \lambda + s = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad s \geq 0. \tag{2.13}$$

Primal-dual solutions for (2.12) and (2.13) are

$$x^* = (0, t, 1-t)^T, \quad \lambda^* = 0, \quad s^* = (1, 0, 0)^T \quad \text{for any } t \in [0, 1]. \tag{2.14}$$

For t in the *open* interval $(0, 1)$, it is clear that (x^*, λ^*, s^*) is a strictly complementary solution. If t takes on one of its extreme values 0 or 1, however, the primal-dual solution is no longer strictly complementary, since there is an index j for which x_j^* and s_j^* are both zero.

It is easy to show that there is a primal solution x^* for which $x_j^* > 0$ for *all* $j \in \mathcal{B}$; that is, $x_{\mathcal{B}}^* > 0$. By the definition (2.11a), there exist solutions

$\bar{x}^j \in \Omega_P$ for each $j \in \mathcal{B}$ with the property that $\bar{x}_j^j > 0$. The average of these solutions, defined by

$$x^* = \frac{1}{|\mathcal{B}|} \sum_{j \in \mathcal{B}} \bar{x}^j, \tag{2.15}$$

is feasible, has the same objective value as each $\bar{x}^j$, and hence is also a solution. It is easy to check that this x^* has the desired property $x_{\mathcal{B}}^* > 0$.

Similarly, we can show that there is a dual solution (λ^*, s^*) such that $s_{\mathcal{N}}^* > 0$.

However, many linear programming problems have *no* strictly feasible points—that is, $\mathcal{F}^o = \emptyset$—although they still may be feasible ($\mathcal{F} \neq \emptyset$) and still may have finite optimal solutions. Consider the simple problem

$$\min_{x \in \mathbb{R}^3} x_1 \ \text{ subject to } x_1 + x_3 = 0, \ x \geq 0,$$

and its dual

$$\max_{\lambda \in \mathbb{R}, s \in \mathbb{R}^3} 0 \ \text{ subject to } \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \lambda + \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad s \geq 0.$$

Any feasible primal-dual vector $(x, \lambda, s) \in \mathcal{F}$ has $x_1 = x_3 = s_2 = 0$, so $\mathcal{F}^o = \emptyset$. The optimal objective function value for this problem is 0, and the primal-dual solution set is defined by

$$x^* = \begin{bmatrix} 0 \\ x_2^* \\ 0 \end{bmatrix}, \qquad s^* = \begin{bmatrix} 1 - \lambda^* \\ 0 \\ -\lambda^* \end{bmatrix},$$

where x_2^* and λ^* are any numbers for which $x_2^* \geq 0$ and $\lambda^* \leq 0$.

Although many primal-dual algorithms require a strictly feasible starting point, they can be adapted to handle the case in which no such point exists.

Lemma 2.5 *Suppose that $\mathcal{F}^o \neq \emptyset$. Then for each $K \geq 0$, the set*

$$\{(x, s) \,|\, (x, \lambda, s) \in \mathcal{F} \text{ for some } \lambda, \text{ and } x^T s \leq K\}$$

is bounded.

Proof. Let $(\bar{x}, \bar{\lambda}, \bar{s})$ be any vector in $\mathcal{F}^o$ and (x, λ, s) be any point in $\mathcal{F}$ with $x^T s \leq K$. Since $A\bar{x} = b$ and $Ax = b$, we have $A(\bar{x} - x) = 0$. Similarly, $A^T(\bar{\lambda} - \lambda) + (\bar{s} - s) = 0$. These two equations imply that

$$(\bar{x} - x)^T(\bar{s} - s) = -(\bar{x} - x)^T A^T(\bar{\lambda} - \lambda) = 0.$$

Rearranging this system and using $x^T s \leq K$, we find that

$$\bar{x}^T s + \bar{s}^T x \leq K + \bar{x}^T \bar{s}. \tag{2.17}$$

Since $(\bar{x}, \bar{s}) > 0$, the quantity ξ defined by

$$\xi = \min_{i=1,2,\ldots,n} \min(\bar{x}_i, \bar{s}_i)$$

is positive. Substituting in (2.17), we obtain

$$\xi e^T(x + s) \leq K + \bar{x}^T \bar{s},$$

which implies that

$$0 \leq x_i \leq \frac{1}{\xi}(K + \bar{x}^T \bar{s}), \qquad 0 \leq s_i \leq \frac{1}{\xi}(K + \bar{x}^T \bar{s}), \qquad i = 1, 2, \ldots, n,$$

Lemma 2.6 (Farkas's lemma) *For each matrix* $G \in \mathbb{R}^{p \times n}$ *and each vector* $g \in \mathbb{R}^n$, *either*

I. $Gd \geq 0$, $g^T d < 0$, *has a solution* $d \in \mathbb{R}^n$,

II. $G^T \pi = g$, $\pi \geq 0$, *has a solution* $\pi \in \mathbb{R}^p$,

but never both.

Corollary 2.7 *For each pair of matrices* $G \in \mathbb{R}^{p \times n}$ *and* $H \in \mathbb{R}^{q \times n}$ *and each vector* $g \in \mathbb{R}^n$, *either*

I. $Gd \geq 0$, $Hd = 0$, $g^T d < 0$, *has a solution* $d \in \mathbb{R}^n$,

II. $G^T \pi + H^T \eta = g$, $\pi \geq 0$, *has a solution* $\pi \in \mathbb{R}^p$, $\eta \in \mathbb{R}^q$,

but never both.

Goldman Tucker Teorem

Proof. (Theorem 2.4) Let $\mathcal{J}$ be the set of indices in $\{1, 2, \ldots, n\}$ that do not belong to either $\mathcal{B}$ or $\mathcal{N}$. We prove the result by showing that $\mathcal{J}$ is empty.

We have already proved that $\mathcal{B} \cap \mathcal{N} = \emptyset$; hence, $\mathcal{B} \cup \mathcal{N} \cup \mathcal{J}$ is a partition of $\{1, 2, \ldots, n\}$. Let $A_{\mathcal{B}}$ and $A_{\mathcal{J}}$ denote the submatrices of columns of A that correspond to $\mathcal{B}$ and $\mathcal{J}$, respectively.

Select any index $i \in \mathcal{J}$. We show that i must also belong to either $\mathcal{N}$ or $\mathcal{B}$, depending on whether there exists a vector w satisfying the following properties:

$$\begin{aligned} A_{\cdot i}^T w &< 0, \\ -A_{\cdot j}^T w &\geq 0 \qquad \text{for } j \in \mathcal{J} \backslash \{i\}, \\ A_{\mathcal{B}}^T w &= 0, \end{aligned} \tag{2.22}$$

where $A_{\cdot i}$ denotes the ith column of A. Suppose first that w satisfying (2.22) exists. Let (x^*, λ^*, s^*) be a primal-dual solution for which $s_{\mathcal{N}}^* > 0$, and define the vector $(\bar{\lambda}, \bar{s})$ as

$$\bar{\lambda} = \lambda^* + \epsilon w, \qquad \bar{s} = c - A^T\bar{\lambda} = s^* - \epsilon A^T w,$$

choosing $\epsilon > 0$ small enough that

$$\begin{aligned}
\bar{s}_i &= s_i^* - \epsilon A_{\cdot i}^T w > 0, \\
\bar{s}_j &= s_j^* - \epsilon A_{\cdot j}^T w \geq 0, \qquad j \in \mathcal{J}\backslash\{i\}, \\
\bar{s}_{\mathcal{B}} &= s_{\mathcal{B}}^* = 0, \\
\bar{s}_{\mathcal{N}} &= s_{\mathcal{N}}^* - \epsilon A_{\mathcal{N}}^T w > 0.
\end{aligned}$$

It follows from these relations that $(\bar{\lambda}, \bar{s})$ is feasible for the dual problem. In fact, it is also optimal, since any primal solution vector x^* must have $x_{\mathcal{N}}^* = 0$ and therefore $\bar{s}^T x^* = 0$. Therefore, by the definition (2.11b), we must have $i \in \mathcal{N}$.

Suppose, alternatively, that no vector w satisfies (2.22). Applying Corollary 2.7, we deduce that the following system must have a solution:

$$-\sum_{j \in \mathcal{J}\setminus\{i\}} \pi_j A_{\cdot j} + A_{\mathcal{B}}\eta = A_{\cdot i}, \qquad \pi_j \geq 0 \text{ for all } j \in \mathcal{J}\setminus\{i\}. \tag{2.23}$$

By defining a vector $v \in \mathbb{R}^{|\mathcal{J}|}$ as

$$v_i = 1, \qquad v_j = \pi_j \quad \text{for all } j \in \mathcal{J}\setminus\{i\},$$

we can rewrite (2.23) as

$$A_{\mathcal{J}}v = A_{\mathcal{B}}\eta, \qquad v \geq 0, \qquad v_i > 0. \tag{2.24}$$

Now let x^* be a primal solution for which $x^*_{\mathcal{B}} > 0$, and define $\bar{x}$ by

$$\bar{x}_{\mathcal{B}} = x^*_{\mathcal{B}} - \epsilon\eta, \qquad \bar{x}_{\mathcal{J}} = \epsilon v, \qquad \bar{x}_{\mathcal{N}} = 0.$$

Substituting from (2.24), we have that $A\bar{x} = b$, and for sufficiently small $\epsilon > 0$, we also have that $\bar{x} \geq 0$. So $\bar{x}$ is feasible and, in fact, optimal because of $\bar{x}_{\mathcal{N}} = 0$. Since $v_i = 1$, we also have that $\bar{x}_i = \epsilon > 0$; hence, $i \in \mathcal{B}$ by (2.11a).

We have shown that any index $i \in \mathcal{J}$ also belongs to either $\mathcal{B}$ or $\mathcal{N}$. Therefore, by the definition of $\mathcal{J}$, we have $\mathcal{J} = \emptyset$; therefore, the proof is complete. □

A General Complexity Theorem for Path-Following Methods

Theorem 3.2 shows that if the reduction in μ at each iteration depends on the dimension n in a certain way, and if the initial duality measure is not too large, the algorithm has polynomial complexity.

Theorem 3.2 *Let $\epsilon \in (0,1)$ be given. Suppose that our algorithm for solving (2.4) generates a sequence of iterates that satisfies*

$$\mu_{k+1} \leq \left(1 - \frac{\delta}{n^\omega}\right) \mu_k, \qquad k = 0, 1, 2, \ldots, \tag{3.10}$$

for some positive constants δ and ω. Suppose too that the starting point (x^0, λ^0, s^0) satisfies

$$\mu_0 \leq 1/\epsilon^\kappa \tag{3.11}$$

for some positive constant κ. Then there exists an index K with

$$K = O\left(n^{\omega}|\log \epsilon|\right)$$

such that

$$\mu_k \leq \epsilon \qquad \textit{for all } k \geq K.$$

Proof. By taking logarithms of both sides in (3.10), we obtain

$$\log \mu_{k+1} \leq \log\left(1 - \frac{\delta}{n^{\omega}}\right) + \log \mu_k.$$

By repeatedly applying this formula and using (3.11), we have

$$\log \mu_k \leq k \log\left(1 - \frac{\delta}{n^{\omega}}\right) + \log \mu_0 \leq k \log\left(1 - \frac{\delta}{n^{\omega}}\right) + \kappa \log \frac{1}{\epsilon}.$$

The estimate for the log function

$$\log(1+\beta) \le \beta \qquad \text{for all } \beta > -1$$

(see Lemma 4.1) implies that

$$\log \mu_k \le k\left(-\frac{\delta}{n^\omega}\right) + \kappa \log \frac{1}{\epsilon}.$$

Therefore, the convergence criterion $\mu_k \le \epsilon$ is satisfied if we have

$$k\left(-\frac{\delta}{n^\omega}\right) + \kappa \log \frac{1}{\epsilon} \le \log \epsilon.$$

This inequality holds for all k that satisfy

$$k \ge K = (1+\kappa)\frac{n^\omega}{\delta} \log \frac{1}{\epsilon},$$

so the proof is complete. □