Chapter 12
Dual Pivot Rule

In this book, pivot rules used in the dual simplex method are referred to as dual pivot rule.[1] Like in the primal simplex context, a dual pivot rule is crucial to algorithm's efficiency.

Assume that $\bar{x} = B^{-1}b \not\geq 0$. The dual Dantzig conventional rule selects a row index $p \in B$ such that $\bar{x}_p$ is the minimum among components of the basic solution. Thus, the dual objective value will increase the most possible for a unit stepsize. This rule is far from ideal, like its primal counterpart. From (4.24) and (4.25), it is known that the increment of the dual objective value will be

$$\hat{f} - \bar{f} = \bar{x}_{j_p}(\bar{z}_q / \bar{a}_{p\,q}).$$

According to the "most-improvement" criterion, p and q are determined such that preceding increment attains the largest possible. Unfortunately, the related computational effort is too high to be practicable, just as its primal counterpart.

This chapter will address some very promising dual pivot rules, which can be regarded as dual variants of those presented in the previous chapter.

For simplicity, assume again that the current basis matrix consists of the first m columns of A, i.e.,

$$B = \{1, \cdots, m\}, \quad N = \{m+1, \cdots, n\}.$$

[1] It is referred to as row rule. After a pivot row is selected, there is a very limited choice of a column pivot.

P.-Q. PAN, *Linear Programming Computation*, DOI 10.1007/978-3-642-40754-3_12,

12.1 Dual Steepest-Edge Rule

Consider the dual problem

$$\begin{aligned} \max \quad & g = b^{\mathrm{T}} y, \\ \text{s.t.} \quad & A^{\mathrm{T}} y \le c. \end{aligned}$$

Let $\bar{y}$ be the current dual basic feasible solution, satisfying

$$B^{\mathrm{T}} \bar{y} = c_B, \tag{12.1}$$

$$N^{\mathrm{T}} \bar{y} \le c_N. \tag{12.2}$$

Define

$$\begin{aligned} y(\beta) &= \bar{y} - \beta h^i, \\ h^i &= B^{-\mathrm{T}} e_i, \quad i = 1, \cdots, m. \end{aligned}$$

From the preceding two expressions and (12.1), for any $i \in \{1, \cdots, m\}$ and $\beta \ge 0$ it holds that

$$\begin{aligned} a_i^{\mathrm{T}} y(\beta) &= a_i^{\mathrm{T}} \bar{y} - \beta = c_i - \beta \le c_i, \\ a_k^{\mathrm{T}} y(\beta) &= a_k^{\mathrm{T}} \bar{y} = c_k, \quad k = 1, \cdots, m, k \ne i. \end{aligned}$$

It is known that $-h^i, i = 1, \cdots, m$ is a edge direction, emanating from the vertex $\bar{y}$ of the feasible region $\{y | A^{\mathrm{T}} y \le c\}$. The determination of row index i implies that the basic variable x_i leaves the basis, so constraint $a_i^{\mathrm{T}} y \le c_i$ may be satisfied as a strict inequality. Since

$$-b^{\mathrm{T}} h^i = -e_i^{\mathrm{T}} B^{-1} b = -\bar{x}_i,$$

when $\bar{x}_i < 0$, the edge direction $-h^i$ forms an acute angle with the dual objective gradient b, as is an uphill direction. Therefore, the objective value will never decrease if a row index $p \in \{1, \cdots, m\}$ is selected such that $\bar{x}_p < 0$; it strictly increase if dual nondegeneracy is assumed. In particular, the following rule selects the edge that forms the smallest angle with b (Forrest and Goldfarb 1992).

Rule 12.1.1 (Dual steepest-edge row rule) Select pivot row index

$$p \in \arg\min\{\bar{x}_i / \|h^i\| \mid i = 1, \cdots, m\}.$$

Like in the primal context, practicability of the preceding rule lies in computing

$$\|h^i\|^2, \quad i = 1, \cdots, m.$$

in a recurrence manner.

Assume that x_p is selected to leave and x_q to enter the basis. Then the inverse of the basis matrix is updated by (3.28), i.e.,

$$\hat{B}^{-1} = B^{-1} - \frac{(\bar{a}_q - e_p)e_p^{\mathrm{T}}B^{-1}}{\bar{a}_{pq}},$$

where $\bar{a}_q = B^{-1}a_q$. Premultiplying the preceding expression by e_i^{T} and transposing it leads to the recurrence for edge directions, i.e.,

$$\tilde{h}^p = (1/\bar{a}_{pq})h^p, \tag{12.3}$$

$$\tilde{h}^{\mathrm{i}} = h^{\mathrm{i}} - (\bar{a}_{iq}/\bar{a}_{pq})h^{\mathrm{p}}, \quad i = 1, \cdots, m, i \neq p. \tag{12.4}$$

from which the recurrence formulas for squares of norms of edge directions follow:

$$\|\tilde{h}^{\mathrm{p}}\|^2 = (1/\bar{a}_{pq})^2\|h^{\mathrm{p}}\|^2, \tag{12.5}$$

$$\|\tilde{h}^{\mathrm{i}}\|^2 = \|h^{\mathrm{i}}\|^2 - 2(\bar{a}_{iq}/\bar{a}_{pq})u_i + (\bar{a}_{iq}/\bar{a}_{pq})^2\|h^{\mathrm{p}}\|^2, \quad i = 1, \cdots, m, i \neq p. \tag{12.6}$$

where

$$B^{\mathrm{T}}h^{\mathrm{p}} = e_p, \quad Bu = h^{\mathrm{p}}. \tag{12.7}$$

Note that $\|h^p\|^2 = (h^p)^{\mathrm{T}}h^p$ can be directly calculated, because h^{p} and $\bar{a}_q$ are obtained otherwise independently.

Computations, involved in Rule 12.1.1, is usually cheaper than that in the (primal) steepest-edge rule. They both solve an additional system (12.6), hence $Bu = h^{\mathrm{p}}$. It is noted that formula (11.9) needs to solve $B^{\mathrm{T}}v = \bar{a}_q$. Differing from (11.9), which is expensive when pivot row $e_p^{\mathrm{T}}B^{-1}N$ is dense, there is no any inner product involved in the middle term of (12.6). In practice, it is often the case of $n - m \gg m$, in which initially computing squares of norms of edge directions in the primal rule is cumbersome, compared with in the dual rule. As for taking advantage of sparsity, handling $B^{-1}a_j$ is also inferior to $B^{-\mathrm{T}}e_i$.

Now let us turn to the dual problem of form

$$\begin{array}{ll} \max & g = b^{\mathrm{T}}y, \\ \text{s.t.} & A^{\mathrm{T}}y + z = c, \quad z \geq 0. \end{array} \tag{12.8}$$

A new pivot rule can be derived if the steepest-edge direction is considered in (y, z)-space.

It is clear that the dual basic feasible solution

$$(\bar{y}, \bar{z}_N, \bar{z}_B) = (B^{-\mathrm{T}}c_B, c_N - N^{\mathrm{T}}B^{-\mathrm{T}}c_B, 0)$$

is actually the unique solution to the $(m+n)\times(m+n)$ system

$$\begin{pmatrix} B^{\mathrm{T}} & 0 & I \\ N^{\mathrm{T}} & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} y \\ z_N \\ z_B \end{pmatrix} = \begin{pmatrix} c_B \\ c_N \\ 0 \end{pmatrix}$$

which is, geometrically, a vertex of the polyhedron

$$\{(y,z) \in \mathcal{R}^m \times \mathcal{R}^n \mid A^{\mathrm{T}} y + z = c, z \geq 0\}.$$

The inverse of the coefficient matrix of the system is

$$\begin{pmatrix} B^{-\mathrm{T}} & 0 & -B^{-\mathrm{T}} \\ -N^{\mathrm{T}} B^{-\mathrm{T}} & I & N^{\mathrm{T}} B^{-\mathrm{T}} \\ 0 & 0 & I \end{pmatrix}$$

It is easy to verify that the last m columns of the preceding are just edge directions, emanating from the vertex, i.e.,

$$h^i = \begin{pmatrix} -B^{-\mathrm{T}} \\ N^{\mathrm{T}} B^{-\mathrm{T}} \\ I \end{pmatrix} e_i, \quad i = 1, \cdots, m. \tag{12.9}$$

Recurrence formulas of these edge directions are of the form (12.3) and (12.4), and squares of norms of them are of form (12.5) and (12.6), though (12.7) should be replaced by

$$B^{\mathrm{T}} h = e_p, \quad Bu = h + \sum_{j=m+1}^{n} \sigma_j a_j, \tag{12.10}$$

where σ_j is the jth component of the pivot row vector, i.e.,

$$\sigma = A^{\mathrm{T}} h. \tag{12.11}$$

As σ is computed while pricing, the following quantity

$$\|h^p\|^2 = h^{\mathrm{T}} h + \sigma^{\mathrm{T}} \sigma. \tag{12.12}$$

can be directly computed.

In this book, the pivot rule, based on (12.5)–(12.7), is referred to as *dual steepest-edge pivot rule I*, and that based on (12.5), (12.6) and (12.10)–(12.12) referred to as *dual steepest-edge pivot rule II*. Associated numerical results will be given in Table 12.1 in the next section, in a comparison with according approximate rules there.

12.2 Approximate Dual Steepest-Edge Rule

Harris (1973) derived a dual variant of the Devex rule, as may be regarded as an approximation of the dual steepest-edge rule II.

She constructed a "reference framework", a set of m indices of components of z. Assume that subvector $\hat{h}^{\mathrm{i}}$ consists of components, located on the reference framework, of edge direction h^{i}. Weights $s_i, i = 1, \cdots, m$ are endowed to $\bar{x}_i$ to approximate $\|\hat{h}^{\mathrm{i}}\|$.

Rule 12.2.1 (Dual Devex row rule) Select pivot row index

$$p \in \arg\min\{\bar{x}_i / s_i | i = 1, \ldots, m\},$$

where weights s_i, $i = 1, \ldots, m$ are determined as follows, so that the rule can be regarded as an approximation of the steepest-edge rule under the reference framework.

At the beginning, index set B is taken as the reference framework, and s_i are set to 1 for all $i = 1, \ldots, m$ (see (12.9)); So, the dual Devex rule is just the same as the dual Dantzig rule in this case. Subsequently, s_i are updated iteration by iteration. Assume that σ is defined by (12.11), and components, located on the reference framework, form subvector $\hat{\sigma}$. As (12.3) and (12.4) are still valid when vectors in them are replaced by subvectors, consisting of components located on the reference framework, we have the following updates:

$$\begin{aligned} \bar{s}_p &= \max\{1, \|\hat{\sigma}\| / |\bar{a}_{pq}|\}, \\ \bar{s}_i &= \max\{s_i, |\bar{a}_{iq} / \bar{a}_{pq}| \|\hat{\sigma}\|\}, \quad i = 1, \cdots, m, i \neq p, \end{aligned}$$

the last of which comes from using the larger norm of vectors $\hat{h}^{\mathrm{i}}$ and $-(\bar{a}_{iq}/\bar{a}_{pq})h^{\mathrm{p}}$ instead of the norm of their sum (see (12.9)). In each iteration, $\bar{a}_q = B^{-1}a_q$ is computed separately. As σ is available, $\|\hat{\sigma}\|$ can be computed directly.

Similar to the (primal) Devex rule, when errors caused by repeatedly using of the updating formulas accumulate too high, it is necessary to determine a new reference framework and set all weights to 1 again. As weight $\bar{t}_q$ is calculated directly, it is convenient to monitor errors. It should be restarted when the calculated value differs from the updated value by a relatively large margin, e.g., when the former exceeds some times of the latter (Harris used double).

When errors, caused by repeatedly using of the updating formulas, accumulate too much, it is necessary to determine a new reference framework and set all weights to 1 again, just like the approximate steepest-edge rules. It is also possible to monitor errors: the process should be restarted when the calculated value of $\|\hat{\sigma}\|$ differs from the updated value s_p by a relatively large margin.

Table 12.1 CPU time and ratio

Dual pivot rule	Dantzig	Devex	Steepest-edge II	Dynamic steepest-edge	Projective steepest-edge	Steepest-edge I
Hours	177.78	67.43	12.72	10.41	7.36	6.36
Time ratio	27.95	10.60	2.00	1.64	1.16	1.00

So-called *dual projective steepest-edge rule* results from using $\|\hat{h}^{\mathrm{i}}\|$ rather than $\bar{s}_i$. This rule also yields from modifying the dual steepest-edge rule under the reference framework. In fact, replacing h^{i} in (12.5) and (12.6) by $\hat{h}^{\mathrm{i}}$ leads to recurrence formulas of $\|\hat{h}^{\mathrm{i}}\|^2$, whereas u in (12.6) can be obtained by solving system

$$Bu = \sum_{j=m+1}^{n} \hat{\sigma}_j a_j .$$

A further variant of the dual steepest-edge rule results from expending the reference framework whenever it is reset, that is, by adding the pivot column index to the current reference framework if it is not in it already. Accordingly, the recurrence formulas are modified slightly. Such a variant is called *dual dynamic steepest-edge rule*.

We cite numerical results reported by Forrest and Goldfarb (1992) below. The hardware and software environments as well as 20 test problems were the same as those described for the primal case, presented in Sect. 11.3. Codes based on six dual rules: Dantzig, Devex, steepest-edge II, dynamic steepest-edge, project steepest-edge and steepest-edge I were tested. Table 12.1 gives total CPU times, required for solving all the problems, in the first row, and time ratios of the first five rules to the dual projective steepest-edge rule in the second row.

It is seen from the preceding table that the dual Dantzig rule is the slowest whereas the dual steepest-edge rule I is the fastest, with time ratio as high as 27.95, though the tests are favorable to the last four rules because they take advantages of the structures of computer IBM RISC system/6000 in pricing and solving systems.

In terms of reported numerical results, the dual rules appear to be inferior to their primal counterparts slightly (see Tables 11.1 and 12.1). However, Forrest and Goldfarb indicate that these data are unfair to the dual rules, because their primal codes were optimized while the dual ones were not; the numbers of hours required by the last four codes in Table 12.1 should be reduced by 10 % at least, though decrements of the dual Dantzig and Devex are not as so much. Therefore, the dual rules are actually efficient, compared to their primal counterparts.

The steepest-edge rule and Devex approximates as well as their dual variants are superior over the standard rules with large margins, and are widely used in many commercial codes, such as CPLEX.

12.3 Dual Largest-Distance Rule

This section derive a dual variant of the largest-distance rule (Sect. 11.4). The basic idea is to determine a leaving variable by finding an inequality constraint, mostly violated by the current solution in the reduced space.

Let $B = \{1, \cdots, m\}$, $N = A \backslash B$ be the current basis and nonbasis of the standard problem (1.8). Assume that the associated simplex tableau is dual but not primal feasible, i.e., $\bar{b} = \bar{x}_B \not\geq 0$.

For some $i \in B$, the inequality $\bar{x}_i < 0$ implies that the solution, $\bar{x}_N = 0$, in the reduced space violates constraint

$$\bar{b}_i - (w^i)^{\mathrm{T}} x_N \geq 0,$$

where $\bar{b} = B^{-1} b$ and

$$w^i = N^{\mathrm{T}} B^{-\mathrm{T}} e_i.$$

The signed distance from point $\bar{x}_N = 0$ to the boundary $\bar{b}_i - (w^i)^{\mathrm{T}} x_N = 0$ is $\bar{x}_i / \|w^i\|$, where any norm is allowed in principle though the Euclidean norm might be preferable.

The dual Dantzig conventional rule is

$$p \in \arg\min\{\bar{x}_i | i = 1, \cdots, m\},$$

which does not necessarily correspond to the mostly violated inequality constraint, since the according distance would be very small when $\|w^p\|$ is large, as the point $\bar{x}_N = 0$ may not far away from the boundary actually. So, it should not be expected that this rule performs satisfactorily.

If one determines a leaving variable that corresponds to the mostly violated inequality constraint, the following rule follows.

Rule 12.3.1 (Dual largest-distance row rule) Select a pivot row index

$$p \in \arg\min\{\bar{x}_i / \|w^i\| \quad | \, i = 1, \cdots, m\}.$$

Involving $\|w^i\|$, the preceding rule is cumbersome, compared with the (primal) largest-distance rule. Like in the primal context, however, $\|w^i\|^2$, $i = 1, \cdots, m$ can be computed recursively, so that the rule would be still practicable.

Consider the set of n-dimensional vectors

$$\sigma^i = A^{\mathrm{T}} B^{-\mathrm{T}} e_i, \quad i = 1, \cdots, m.$$

Note that w^i is an $(n - m)$-subvector of σ^i for any $i = 1, \ldots, m$. Assume that x_p leaves and x_q enters the basis, then it is not difficult to derive the following recurrence relation:

$$\tilde{\sigma}^p = (1/\bar{a}_{pq})\sigma^p,$$
$$\tilde{\sigma}^i = \sigma^i - (\bar{a}_{iq}/\bar{a}_{pq})\sigma^p, \quad i = 1, \cdots, m, i \neq p.$$

where pivot column $\bar{a}_q = B^{-1}a_q$ is available. Further, recurrence formulas of squares of their norms can be obtained, i.e.,

$$\|\tilde{\sigma}^p\|^2 = (1/\bar{a}_{pq})^2\|\sigma^p\|^2,$$
$$\|\tilde{\sigma}^i\|^2 = \|\sigma^i\|^2 - 2(\bar{a}_{iq}/\bar{a}_{pq}){\sigma^i}^{\mathrm{T}}\sigma^p + (\bar{a}_{iq}/\bar{a}_{pq})^2\|\sigma^p\|^2, \quad i = 1, \cdots, m, i \neq p,$$

combining which, ${\sigma^i}^{\mathrm{T}}\sigma^p = e_i^{\mathrm{T}}B^{-1}N\sigma^p$ and $\|\sigma^i\|^2 = \|w^i\|^2 + 1$ leads to the required formulas

$$\|\tilde{w}^p\|^2 = (1/\bar{a}_{pq})^2(\|(w^p\|^2 + 1) - 1, \tag{12.13}$$
$$\|\tilde{w}^i\|^2 = \|w^i\|^2 - 2(\bar{a}_{iq}/\bar{a}_{pq})u_i + (\bar{a}_{iq}/\bar{a}_{pq})^2(\|w^p\|^2 + 1), \quad i = 1, \cdots, m, i \neq p, \tag{12.14}$$

where

$$Bu = \sum_{j=m+1}^{n} \sigma_j a_j,$$

and σ_j is the jth component of σ^p (i.e., the $(j-m)$th component of w^p). Note that components of w^{p} are available entries of the pivot row, and hence $\|w^p\|^2 = (w^p)^{\mathrm{T}}w^p$ can be calculated directly.

12.4 Dual Nested Rule

This section derives a dual variant of the nested rule (Sect. 11.5). Regarding negative components of the primal basic solution as elimination target, it focuses on the most stubborn ones among them.

Let $\epsilon > 0$ be primal feasibility tolerance. At the beginning of an iteration, a set I_1 of row indices is given the priority for associated basic indies to leave the basis. A row index in I_1, associated with the smallest component of $\bar{x}_B$, is determined. If the component is less than $-\epsilon$, then the corresponding basic index is selected to leave the basis, and the I_1 for the next iteration is born from the current I_1 by including all its row indices, associated with $\bar{x}_B$'s components less than $-\epsilon$. In the other case, do the same as before with $I_2 = B \backslash I_1$; if there is still no basic components less than $-\epsilon$, then optimality is attained.

The following is a nested variant of Dantzig conventional dual rule.

Rule 12.4.1 (Dual nested row rule: Dantzig) Given primal tolerance $\epsilon > 0$. Set $I_1 = B$ and $I_2 = \emptyset$.

1. Go to step 4 if $I_1 = \{i \in I_1 \mid \bar{x}_i < -\epsilon\} \neq \emptyset$.
2. Go to step 4 if $I_1 = \{i \in I_2 \mid \bar{x}_i < -\epsilon\} \neq \emptyset$.
3. Stop (optimality achieved).
4. Select a pivot row index $p \in \arg\min \{\bar{x}_i \mid i \in I_1\}$.
5. Update: $I_1 = I_1 \backslash p,\ I_2 = B \backslash I_1$.

In the first iteration or iterations in which I_2 is touched, the preceding rule actually proceeds the same as the standard dual row rule. After such an iteration, a series of iterations with nested pivoting follow, as might be called a "circle", where each I_1 is a proper subset of its predecessor. In the kth iteration of a circle, basic components associated with I_1 are less than $-\epsilon$ all the time. It is therefore reasonable to select such a stubborn one to leave the basis.

It is possible to turn any full dual rule to a nested version. The dual nested steepest-edge rule can be obtained from Rule 12.4.1 by using

$$p \in \arg\min \{\bar{x}_i / \|h^i\| \mid i \in I_1\}$$

in step 4 of it instead, the dual nested Devex rule obtained by using

$$p \in \arg\min\{\bar{x}_i / s_i \mid i \in I_1\}$$

and the dual nested largest-distance rule by using

$$p \in \arg\min\{\bar{x}_i / \|w^i\| \mid i \in I_1\}$$

(see relevant sections for notations).

In view of the performance of their primal counterparts, it might be expected that the dual nested rules and the dual nested largest-distance rules perform satisfactorily. The dual nested rules are easy to implement, fortunately. However, there is no any computational experience at present.