CHAPTER 20

Implementation Issues for Interior-Point Methods

In this chapter, we discuss implementation issues that arise in connection with the path-following method.

The most important issue is the efficient solution of the systems of equations discussed in the previous chapter. As we saw, there are basically three choices, involving either the reduced KKT matrix,

$$B = \begin{bmatrix} -E^{-2} & A \\ A^T & D^{-2} \end{bmatrix}, \tag{20.1}$$

or one of the two matrices associated with the normal equations:

$$AD^2A^T + E^{-2} \tag{20.2}$$

or

$$A^TE^2A + D^{-2}. \tag{20.3}$$

(Here, $E^{-2} = Y^{-1}W$ and $D^{-2} = X^{-1}Z$.)

In the previous chapter, we explained that dense columns/rows are bad for the normal equations and that therefore one might be better off solving the system involving the reduced KKT matrix. But there is also a reason one might prefer to work with one of the systems of normal equations. The reason is that these matrices are positive definite. We shall show in the first section that there are important advantages in working with positive definite matrices. In the second section, we shall consider the reduced KKT matrix and see to what extent the nice properties possessed by positive definite matrices carry over to these matrices.

After finishing our investigations into numerical factorization, we shall take up a few other relevant tasks, such as how one extends the path-following algorithm to handle problems with bounds and ranges.

1. Factoring Positive Definite Matrices

As we saw in the proof of Theorem 19.1, the matrix (20.2) appearing in the primal normal equations is positive semidefinite (and so is (20.3), of course). In fact, it is even better—it's positive definite. A matrix B is *positive definite* if $\xi^T B\xi > 0$ for all vectors $\xi \neq 0$. In this section, we will show that, if we restrict our row/column reordering to symmetric reorderings, that is, reorderings where the rows and columns undergo the same permutation, then there is no danger of encountering a pivot

R.J. Vanderbei, *Linear Programming*, International Series in Operations Research & Management Science 196, DOI 10.1007/978-1-4614-7630-6_20,

element whose value is zero. Hence, the row/column permutation can be selected ahead of time based only on the aim of maintaining sparsity.

If we restrict ourselves to symmetric permutations, each pivot element is a diagonal element of the matrix. The following result shows that we can start by picking an arbitrary diagonal element as the first pivot element:

THEOREM 20.1. *If B is positive definite, then $b_{ii} > 0$ for all i.*

The proof follows trivially from the definition of positive definiteness:

$$b_{ii} = e_i^T B e_i > 0.$$

The next step is to show that after each stage of the elimination process, the remaining uneliminated matrix is positive definite. Let us illustrate by breaking out the first row/column of the matrix and looking at what the first step of the elimination process does. Breaking out the first row/column, we write

$$B = \begin{bmatrix} a & b^T \\ b & C \end{bmatrix}.$$

Here, a is the first diagonal element (a scalar), b is the column below a, and C is the matrix consisting of all of B except the first row/column. One step of elimination (as described in Chapter 8) transforms B into

$$\left[\begin{array}{c|c} a & b^T \\ \hline b & C - \frac{bb^T}{a} \end{array}\right].$$

The following theorem tells us that the uneliminated part is positive definite:

THEOREM 20.2. *If B is positive definite, then so is $C - bb^T/a$.*

PROOF. The fact that B is positive definite implies that

$$\begin{bmatrix} x & y^T \end{bmatrix} \begin{bmatrix} a & b^T \\ b & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2y^T bx + y^T Cy \tag{20.4}$$

is positive whenever the scalar x or the vector y is nonzero (or both). Fix a vector $y \neq 0$, and put $x = -\frac{1}{a} b^T y$. Using these choices in (20.4), we get

$$0 < \frac{1}{a} y^T bb^T y - 2\frac{1}{a} y^T bb^T y + y^T Cy = y^T \left(C - \frac{bb^T}{a} \right) y.$$

Since y was an arbitrary nonzero vector, it follows that $C - bb^T/a$ is positive definite. □

Hence, after one step of the elimination, the uneliminated part is positive definite. It follows by induction then that the uneliminated part is positive definite at every step of the elimination.

Here's an example:

$$B = \begin{bmatrix} 2 & -1 & & & -1 \\ -1 & 3 & -1 & -1 & \\ & -1 & 2 & -1 & \\ & -1 & -1 & 3 & -1 \\ -1 & & & -1 & 3 \end{bmatrix}.$$

At the end of the four steps of the elimination (without permutations), we end up with

$$\begin{bmatrix} 2 & -1 & & & -1 \\ -1 & \frac{5}{2} & -1 & -1 & -\frac{1}{2} \\ & -1 & \frac{8}{5} & -\frac{7}{5} & -\frac{1}{5} \\ & -1 & -\frac{7}{5} & \frac{11}{8} & -\frac{11}{8} \\ -1 & -\frac{1}{2} & -\frac{1}{5} & -\frac{11}{8} & 1 \end{bmatrix}.$$

From this eliminated form, we extract the lower triangular matrix, the diagonal matrix, and the upper triangular matrix to write B as

$$B = \begin{bmatrix} 2 & & & & \\ -1 & \frac{5}{2} & & & \\ & -1 & \frac{8}{5} & & \\ & -1 & -\frac{7}{5} & \frac{11}{8} & \\ -1 & -\frac{1}{2} & -\frac{1}{5} & -\frac{11}{8} & 1 \end{bmatrix} \begin{bmatrix} 2 & & & & \\ & \frac{5}{2} & & & \\ & & \frac{8}{5} & & \\ & & & \frac{11}{8} & \\ & & & & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 & -1 & & & -1 \\ & \frac{5}{2} & -1 & -1 & -\frac{1}{2} \\ & & \frac{8}{5} & -\frac{7}{5} & -\frac{1}{5} \\ & & & \frac{11}{8} & -\frac{11}{8} \\ & & & & 1 \end{bmatrix}.$$

As we saw in Chapter 8, it is convenient to combine the lower triangular matrix with the diagonal matrix to get a new lower triangular matrix with ones on the diagonal. But the current lower triangular matrix is exactly the transpose of the upper triangular matrix. Hence, to preserve symmetry, we should combine the diagonal matrix with both the lower and the upper triangular matrices. Since it only appears once, we must multiply and divide by it (in the middle of the product). Doing this, we get

$$B = \begin{bmatrix} 1 & & & & \\ -\frac{1}{2} & 1 & & & \\ & -\frac{2}{5} & 1 & & \\ & -\frac{2}{5} & -\frac{7}{8} & 1 & \\ -\frac{1}{2} & -\frac{1}{5} & -\frac{1}{8} & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & & & & \\ & \frac{5}{2} & & & \\ & & \frac{8}{5} & & \\ & & & \frac{11}{8} & \\ & & & & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & & & -\frac{1}{2} \\ & 1 & -\frac{2}{5} & -\frac{2}{5} & -\frac{1}{5} \\ & & 1 & -\frac{7}{8} & -\frac{1}{8} \\ & & & 1 & -1 \\ & & & & 1 \end{bmatrix}.$$

The lower triangular matrix in this representation is usually denoted by L and the diagonal matrix by D (not to be confused with the D at the beginning of the chapter). Hence, this factorization can be summarized as

$$B = LDL^T$$

and is referred to as an *LDL^T-factorization.* Of course, once a factorization is found, it is easy to solve systems of equations using forward and backward substitution as discussed in Chapter 8.

1.1. Stability. We began our discussion of factoring positive definite matrices with the comment that a symmetric permutation can be chosen purely with the aim of preserving sparsity, since it is guaranteed that no pivot element will ever vanish. However, the situation is even better than that—we can show that whenever a pivot element is small, so is every other nonzero in the uneliminated part of the same row/column. Before saying why, we need to set down a few technical results.

THEOREM 20.3. *If $\bar{b}_{ii}$ denotes a diagonal element in the uneliminated submatrix at some stage of an elimination and b_{ii} denotes the original value of that diagonal element, then $0 < \bar{b}_{ii} \leq b_{ii}$.*

PROOF. The positivity of $\bar{b}_{ii}$ follows from the fact the uneliminated submatrix is positive definite. The fact that it is bounded above by b_{ii} follows from the fact that each step of the elimination can only decrease diagonal elements, which can be seen by looking at the first step of the elimination. Using the notation introduced just after Theorem 20.1,

$$c_{ii} - \frac{b_i^2}{a} \leq c_{ii}.$$

$\square$

THEOREM 20.4. *If B is symmetric and positive definite, then $|b_{ij}| < \sqrt{b_{ii}b_{jj}}$ for all $i \neq j$.*

PROOF. Fix $i \neq j$ and let $\xi = re_i + e_j$. That is, ξ is the vector that's all zero except for the ith and jth position, where it's r and 1, respectively. Then,

$$0 < \xi^T B \xi = b_{ii}r^2 + 2b_{ij}r + b_{jj},$$

for all $r \in \mathbb{R}$. This quadratic expression is positive for all values of r if and only if it is positive at its minimum, and it's easy to check that it is positive at that point if and only if $|b_{ij}| < \sqrt{b_{ii}b_{jj}}$. $\square$

These two theorems, together with the fact that the uneliminated submatrix is symmetric and positive definite, give us bounds on the off-diagonal elements. Indeed, consider the situation after a number of steps of the elimination. Using bars to denote matrix elements in the uneliminated submatrix and letting M denote an upper bound on the diagonal elements before the elimination process began (which, without loss of generality, could be taken as 1), we see that, if $\bar{b}_{jj} < \epsilon$, then

$$\bar{b}_{ij} < \sqrt{\epsilon M}. \tag{20.5}$$

This bound is exceedingly important and is special to positive definite matrices.

2. Quasidefinite Matrices

In this section, we shall study factorization techniques for the reduced KKT matrix (20.1). The reduced KKT matrix is an example of a quasidefinite matrix. A symmetric matrix is called *quasidefinite* if it can be written (perhaps after a symmetric permutation) as

$$B = \begin{bmatrix} -E & A \\ A^T & D \end{bmatrix},$$

where E and D are positive definite matrices. Quasidefinite matrices inherit some of the nice properties of positive definite matrices. In particular, one can perform an arbitrary symmetric permutation of the rows/columns and still be able to form a factorization of the permuted matrix.

The idea is that, after each step of the elimination, the remaining uneliminated part of the matrix is still quasidefinite. To see why, let's break out the first row/column of the matrix and look at the first step of the elimination process. Breaking out the first row/column of B, we write

$$B = \begin{bmatrix} -a & -b^T & f^T \\ -b & -C & G \\ f & G^T & D \end{bmatrix},$$

where a is a scalar, b and f are vectors, and C, D, and G are matrices (of the appropriate dimensions). One step of the elimination process transforms B into

$$\left[\begin{array}{c|cc} -a & -b^T & f^T \\ \hline -b & -\left(C - \frac{bb^T}{a}\right) & G + \frac{bf^T}{a} \\ f & G^T + \frac{fb^T}{a} & D + \frac{ff^T}{a} \end{array}\right].$$

The uneliminated part is

$$\begin{bmatrix} -\left(C - \frac{bb^T}{a}\right) & G + \frac{bf^T}{a} \\ G^T + \frac{fb^T}{a} & D + \frac{ff^T}{a} \end{bmatrix}.$$

Clearly, the lower-left and upper-right blocks are transposes of each other. Also, the upper-left and lower-right blocks are symmetric, since C and D are. Therefore, the whole matrix is symmetric. Theorem 20.2 tells us that $C - bb^T/a$ is positive definite and $D + ff^T/a$ is positive definite, since the sum of a positive definite matrix and a positive semidefinite matrix is positive definite (see Exercise 20.2). Therefore, the uneliminated part is indeed quasidefinite.

Of course, had the first pivot element been selected from the submatrix D instead of E, perhaps the story would be different. But it is easy to check that it's the same. Hence, no matter which diagonal element is selected to be the first pivot element, the resulting uneliminated part is quasidefinite. Now, by induction it follows that every step of the elimination process involves choosing a pivot element from the diagonals of a quasidefinite matrix. Since these diagonals come from either a positive definite submatrix or the negative of such a matrix, it follows that they are always nonzero (but many of them will be negative). Therefore, just as for positive definite matrices, an arbitrary symmetric permutation of a quasidefinite matrix can be factored without any risk of encountering a zero pivot element.

Here's an example:

$$B = \begin{array}{cc} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} -1 & & & -2 & 1 \\ & -2 & & & 2 \\ & & -3 & 1 & \\ -2 & & 1 & 2 & \\ 1 & 2 & & & 1 \end{bmatrix} \end{array}. \tag{20.6}$$

(The blocks are easy to pick out, since the negative diagonals must be from $-E$, whereas the positive ones are from D.) Let's eliminate by picking the diagonals in the order $1, 5, 2, 4, 3$. No permutations are needed in preparation for the first step of the elimination. After this step, we have

$$\begin{array}{cc} & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} -1 & & & -2 & 1 \\ & -2 & & & 2 \\ & & -3 & 1 & \\ -2 & & 1 & 6 & -2 \\ 1 & 2 & & -2 & 2 \end{bmatrix} \end{array}.$$

Now, we move row/column 5 to the pivot position, slide the other rows/columns down/over, and eliminate to get

$$\begin{array}{cc} & \begin{array}{ccccc} 1 & 5 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 5 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} -1 & 1 & & & -2 \\ 1 & 2 & 2 & & -2 \\ & 2 & -4 & & 2 \\ & & & -3 & 1 \\ -2 & -2 & 2 & 1 & 4 \end{bmatrix} \end{array}.$$

Row/column 2 is in the correct position for the third step of the elimination, and therefore, without further ado, we do the next step in the elimination:

$$\begin{array}{cc} & \begin{array}{ccccc} 1 & 5 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 5 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} -1 & 1 & & & -2 \\ 1 & 2 & 2 & & -2 \\ & 2 & -4 & & 2 \\ & & & -3 & 1 \\ -2 & -2 & 2 & 1 & 5 \end{bmatrix} \end{array}.$$

Finally, we interchange rows/columns 3 and 4 and do the last elimination step to get

$$\begin{array}{cc} & \begin{array}{ccccc} 1 & 5 & 2 & 4 & 3 \end{array} \\ \begin{array}{c} 1 \\ 5 \\ 2 \\ 4 \\ 3 \end{array} & \begin{bmatrix} -1 & 1 & & -2 & \\ 1 & 2 & 2 & -2 & \\ & 2 & -4 & 2 & \\ -2 & -2 & 2 & 5 & 1 \\ & & & 1 & -\frac{16}{5} \end{bmatrix} \end{array}.$$

From this final matrix, it is easy to extract the LDL^T-factorization of the permutation of B:

$$
B = \begin{array}{c} \\ 1 \\ 5 \\ 2 \\ 4 \\ 3 \end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 5 & 2 & 4 & 3 \end{array} \\
\begin{bmatrix}
-1 & 1 & & -2 & \\
1 & 1 & 2 & & \\
& 2 & -2 & & \\
-2 & & & 2 & 1 \\
& & & 1 & -3
\end{bmatrix}
\end{array}
$$

$$
= \begin{array}{c} 1 \\ 5 \\ 2 \\ 4 \\ 3 \end{array}
\begin{bmatrix}
1 & & & & \\
-1 & 1 & & & \\
& 1 & 1 & & \\
2 & -1 & -\frac{1}{2} & 1 & \\
& & & \frac{1}{5} & 1
\end{bmatrix}
\begin{bmatrix}
-1 & & & & \\
& 2 & & & \\
& & -4 & & \\
& & & 5 & \\
& & & & -\frac{16}{5}
\end{bmatrix}
\begin{array}{c}
\begin{array}{ccccc} 1 & 5 & 2 & 4 & 3 \end{array} \\
\begin{bmatrix}
1 & -1 & & 2 & \\
& 1 & 1 & -1 & \\
& & 1 & -\frac{1}{2} & \\
& & & 1 & \frac{1}{5} \\
& & & & 1
\end{bmatrix}
\end{array} .
$$

As always, the factorization is computed so that one can solve systems of equations. Given the factorization, all that is required is a forward substitution, a backward substitution, a scaling, and two permutations.

2.1. Instability. We have shown that an arbitrary symmetric permutation of the rows/columns of a quasidefinite matrix can be factored into LDL^T. That is, mathematically none of the pivot elements will be zero. But they can be small, and their smallness can cause troubles not encountered with positive definite matrices. To explain, let's look at an example. Consider the linear programming problem

$$
\begin{array}{lrcl}
\text{maximize} & x_1 + \ \ x_2 & & \\
\text{subject to} & x_1 + 2x_2 & \leq & 1 \\
& 2x_1 + \ \ x_2 & \leq & 1 \\
& x_1,\ x_2 & \geq & 0
\end{array}
$$

and its dual

$$
\begin{array}{lrcl}
\text{minimize} & y_1 + \ \ y_2 & & \\
\text{subject to} & y_1 + 2y_2 & \geq & 1 \\
& 2y_1 + \ \ y_2 & \geq & 1 \\
& y_1,\ y_2 & \geq & 0\,.
\end{array}
$$

Drawing pictures, it is easy to see that the optimal solution is

$$
\begin{aligned}
x_1^* = x_2^* = y_1^* = y_2^* &= \frac{1}{3} \\
z_1^* = z_2^* = w_1^* = w_2^* &= 0.
\end{aligned}
$$

Therefore, as the optimal solution is approached, the diagonal elements in $X^{-1}Z$ and $Y^{-1}W$ approach zero. Therefore, at late stages in the path-following method, one is faced with the problem of factoring the following reduced KKT matrix:

$$B = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} -\epsilon_1 & & 1 & 2 \\ & -\epsilon_2 & 2 & 1 \\ 1 & 2 & \delta_1 & \\ 2 & 1 & & \delta_2 \end{bmatrix} \end{array},$$

where ϵ_1, ϵ_2, δ_1, and δ_2 are small positive numbers. Consider what happens if we permute the rows/columns so that the original diagonal elements appear in the following order: $1, 3, 4, 2$. The permuted matrix is seen to be

$$B = \begin{array}{c} \\ 1 \\ 3 \\ 4 \\ 2 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 3 & 4 & 2 \end{array} \\ \begin{bmatrix} -\epsilon_1 & 1 & 2 & \\ 1 & \delta_1 & & 2 \\ 2 & & \delta_2 & 1 \\ & 2 & 1 & -\epsilon_2 \end{bmatrix} \end{array}.$$

After the first two steps of the elimination, we have

$$\left[\begin{array}{cc|cc} -\epsilon_1 & 1 & 2 & \\ 1 & \delta_1 + \frac{1}{\epsilon_1} & \frac{2}{\epsilon_1} & 2 \\ \hline 2 & \frac{2}{\epsilon_1} & \left(\delta_2 + \frac{4}{\epsilon_1}\right) - \frac{4/\epsilon_1^2}{\left(\delta_1 + \frac{1}{\epsilon_1}\right)} & 1 - \frac{4/\epsilon_1}{\left(\delta_1 + \frac{1}{\epsilon_1}\right)} \\ & 2 & 1 - \frac{4/\epsilon_1}{\left(\delta_1 + \frac{1}{\epsilon_1}\right)} & -\epsilon_2 - \frac{4}{\left(\delta_1 + \frac{1}{\epsilon_1}\right)} \end{array}\right]. \tag{20.7}$$

Using exact arithmetic, the uneliminated part simplifies to

$$\begin{bmatrix} \delta_2 + \frac{4\delta_1}{1+\epsilon_1\delta_1} & 1 - \frac{4}{1+\epsilon_1\delta_1} \\ 1 - \frac{4}{1+\epsilon_1\delta_1} & -\epsilon_2 - \frac{4\epsilon_1}{1+\epsilon_1\delta_1} \end{bmatrix}.$$

Here, the diagonal elements are small but not zero and the off-diagonal elements are close to -3. But on a computer that stores numbers with finite precision, the computation comes out quite differently when the ϵ_i's and the δ_i's are smaller than the square root of the machine precision (i.e., about 10^{-8} for double-precision floating-point numbers). In the elimination process, the parenthesized expressions in (20.7) are evaluated before the other operations. Hence, in finite precision arithmetic, these expressions produce

$$\delta_2 + \frac{4}{\epsilon_1} = \frac{4}{\epsilon_1} \qquad \text{and} \qquad \delta_1 + \frac{1}{\epsilon_1} = \frac{1}{\epsilon_1},$$

and so (20.7) becomes

$$\begin{array}{c} \\ 1 \\ 3 \\ 4 \\ 2 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 3 & 4 & 2 \end{array} \\ \left[\begin{array}{cc|cc} -\epsilon_1 & 1 & 2 & \\ 1 & \frac{1}{\epsilon_1} & \frac{2}{\epsilon_1} & 2 \\ \hline 2 & \frac{2}{\epsilon_1} & 0 & -3 \\ & 2 & -3 & -4\epsilon_1 \end{array}\right] \end{array},$$

which clearly presents a problem for the next step in the elimination process.

Now let's consider what happens if the elimination process is applied directly to B without permuting the rows/columns. Sparing the reader the tedious details, the end result of the elimination is the following matrix:

$$\begin{bmatrix} -\epsilon_1 & & 1 & 2 \\ & -\epsilon_2 & 2 & 1 \\ 1 & 2 & \delta_1 + \frac{1}{\epsilon_1} + \frac{4}{\epsilon_2} & \frac{2}{\epsilon_1} + \frac{2}{\epsilon_2} \\ 2 & 1 & \frac{2}{\epsilon_1} + \frac{2}{\epsilon_2} & \delta_2 + \frac{4}{\epsilon_1} + \frac{1}{\epsilon_2} - \frac{(\frac{2}{\epsilon_1}+\frac{2}{\epsilon_2})^2}{\delta_1+\frac{1}{\epsilon_1}+\frac{4}{\epsilon_2}} \end{bmatrix}. \tag{20.8}$$

As before, in finite precision arithmetic, certain small numbers get lost:

$$\delta_2 + \frac{4}{\epsilon_1} = \frac{4}{\epsilon_1} \qquad \text{and} \qquad \delta_1 + \frac{1}{\epsilon_1} = \frac{1}{\epsilon_1}.$$

Making these substitutions in (20.8), we see that the final matrix produced by the elimination process using finite precision arithmetic is

$$\begin{bmatrix} -\epsilon_1 & & 1 & 2 \\ & -\epsilon_2 & 2 & 1 \\ 1 & 2 & \frac{1}{\epsilon_1} + \frac{4}{\epsilon_2} & \frac{2}{\epsilon_1} + \frac{2}{\epsilon_2} \\ 2 & 1 & \frac{2}{\epsilon_1} + \frac{2}{\epsilon_2} & 0 \end{bmatrix}.$$

Just as before, the fact that small numbers got lost has resulted in a zero appearing on the diagonal where a small but nonzero (in this case positive) number belongs. However, the situation is fundamentally different this time. With the first ordering, a -3 remained to be eliminated under the zero diagonal element, whereas with the second ordering, this did not happen. Of course, it didn't happen in this particular example because the 0 appeared as the last pivot element, which has no elements below it to be eliminated. But that is not the general reason why the second ordering does not produce nonzeros under zero pivot elements. In general, a zero (which should be a small positive) can appear anywhere in the lower-right block (relative to the original quasidefinite partitioning). But once the elimination process gets to this block, the remaining uneliminated part of the matrix is positive definite. Hence, the estimate in (20.5) can be used to tell us that all the nonzeros below a zero diagonal are in fact small themselves. A zero appearing on the diagonal only presents a problem if there are nonzeros below it that need to be eliminated. If there are none, then the elimination can simply proceed to the next pivot element (see Exercise 20.1).

Let's summarize the situation. We showed in the last chapter that the possibility of dense rows/columns makes it unattractive to work strictly with the normal equations. Yet, although the quasidefinite reduced KKT system can be used, it is numerically less stable. A compromise solution seems to be suggested. One could take a structured approach to the reordering heuristic. In the structured approach, one decides first whether it seems better to begin pivoting with elements from the upper-left block or from the lower-right block. Once this decision is made, one should pivot out all the diagonal elements from this block before working on the other block, with the exception that pivots involving dense rows/columns be deferred to the end of the elimination process. If no dense columns are identified, this strategy mimics the normal equations approach. Indeed, after eliminating all the diagonal elements in the upper-left block, the remaining uneliminated lower-right block contains exactly the matrix for the system of dual normal equations. Similarly, had the initial choice been to pivot out all the diagonal elements from the lower-right block, then the remaining uneliminated upper-left block becomes the matrix for the system of primal normal equations.

With this structured approach, if no dense rows/columns are identified and deferred, then the elimination process is numerically stable. If, on the other hand, some dense rows/columns are deferred, then the factorization is less stable. But in practice, this approach seems to work well. Of course, one could be more careful and monitor the diagonal elements. If a diagonal element gets small (relative to the other uneliminated nonzeros in the same row/column), then one could flag it and then calculate a new ordering in which such pivot elements are deferred to the end of the elimination process.

3. Problems in General Form

In this section, we describe how to adapt the path-following algorithm to solving problems presented in the following general form:

$$\begin{array}{lrcccl} \text{maximize} & & & c^T x & & \\ \text{subject to} & a & \leq & Ax & \leq & b \\ & l & \leq & x & \leq & u. \end{array} \tag{20.9}$$

As in Chapter 9, some of the data elements are allowed to take on infinite values. However, let us consider first the case where all the components of a, b, l, and u are finite. Infinities require special treatment, which shall be discussed shortly.

Following the derivation of the path-following method that we introduced in Chapter 18, the first step is to introduce slack variables as appropriate to replace all inequality constraints with simple nonnegativity constraints. Hence, we rewrite the primal problem (20.9) as follows:

$$\begin{array}{lrcrcr} \text{maximize} & c^T x & & & & \\ \text{subject to} & Ax & + & f & = & b \\ & -Ax & + & p & = & -a \\ & x & + & t & = & u \\ & -x & + & g & = & -l \\ & f,\ p, & t, & g & \geq & 0. \end{array}$$

In Chapter 9, we showed that the dual problem is given by

$$\begin{array}{lrcl} \text{minimize} & b^T v - a^T q + u^T s - l^T h & & \\ \text{subject to} & A^T(v-q) - (h-s) & = & c \\ & v,\ q,\ s,\ h & \geq & 0, \end{array}$$

and the corresponding complementarity conditions are given by

$$\begin{aligned} f_i v_i &= 0 \qquad & i &= 1, 2, \ldots, m, \\ p_i q_i &= 0 \qquad & i &= 1, 2, \ldots, m, \\ t_j s_j &= 0 \qquad & j &= 1, 2, \ldots, n, \\ g_j h_j &= 0 \qquad & j &= 1, 2, \ldots, n. \end{aligned}$$

The next step in the derivation is to introduce the primal–dual central path, which we parametrize as usual by a positive real parameter μ. Indeed, for each $\mu > 0$, we define the associated central-path point in primal–dual space as the unique point that simultaneously satisfies the conditions of primal feasibility, dual feasibility, and μ-complementarity. Ignoring nonnegativity (which is enforced separately), these conditions are

$$\begin{aligned} Ax + f &= b \\ f + p &= b - a \\ x + t &= u \\ -x + g &= -l \\ \\ A^T y + s - h &= c \\ y + q - v &= 0 \\ \\ FVe &= \mu e \\ PQe &= \mu e \\ TSe &= \mu e \\ GHe &= \mu e. \end{aligned}$$

Note that we have replaced the primal feasibility condition, $-Ax + p = -a$, with the equivalent condition that $f + p = b - a$, and we have introduced into the dual problem new variables y defined by $y = v - q$. The reason for these changes is to put the system of equations into a form in which A and A^T appear as little as possible (so that solving the system of equations for the step direction will be as efficient as possible).

The last four equations are the μ-complementarity conditions. As usual, each upper case letter that appears on the left in these equations denotes the diagonal matrix having the components of the corresponding lower-case vector on its diagonal. The system is a nonlinear system of $5n + 5m$ equations in $5n + 5m$ unknowns. It has a unique solution in the strict interior of the following subset of primal–dual space:

$$\{(x, f, p, t, g, y, v, q, s, h) : f, p, t, g, v, q, s, h \geq 0\}. \tag{20.10}$$

This fact can be seen by noting that these equations are the first-order optimality conditions for an associated strictly convex barrier problem.

As μ tends to zero, the central path converges to the optimal solution to both the primal and dual problems. The path-following algorithm is defined as an iterative process that starts from a point in the strict interior of (20.10), estimates at each iteration a value of μ representing a point on the central path that is in some sense closer to the optimal solution than the current point, and then attempts to step toward this central-path point, making sure that the new point remains in the strict interior of the set given in (20.10).

Suppose for the moment that we have already decided on the target value for μ. Let $(x, \ldots, h)$ denote the current point in the strict interior of (20.10), and let $(x + \Delta x, \ldots, h + \Delta h)$ denote the point on the central path corresponding to the target value of μ. The defining equations for the point on the central path can be written as

$$
\begin{aligned}
A\Delta x + \Delta f &= b - Ax - f &&=: \rho \\
\Delta f + \Delta p &= b - a - f - p &&=: \alpha \\
\Delta x + \Delta t &= u - x - t &&=: \tau \\
-\Delta x + \Delta g &= -l + x - g &&=: \nu \\[1ex]
A^T\Delta y + \Delta s - \Delta h &= c - A^T y - s + h &&=: \sigma \\
\Delta y + \Delta q - \Delta v &= -y - q + v &&=: \beta \\[1ex]
FV^{-1}\Delta v + \Delta f &= \mu V^{-1}e - f - V^{-1}\Delta V \Delta f &&=: \gamma_f \\
QP^{-1}\Delta p + \Delta q &= \mu P^{-1}e - q - P^{-1}\Delta P \Delta q &&=: \gamma_q \\
ST^{-1}\Delta t + \Delta s &= \mu T^{-1}e - s - T^{-1}\Delta T \Delta s &&=: \gamma_s \\
HG^{-1}\Delta g + \Delta h &= \mu G^{-1}e - h - G^{-1}\Delta G \Delta h &&=: \gamma_h,
\end{aligned}
$$

where we have introduced notations $\rho, \ldots, \gamma_h$ as shorthands for the right-hand side expressions. This is almost a linear system for the direction vectors $(\Delta x, \ldots, \Delta h)$. The only nonlinearities appear on the right-hand sides of the complementarity equations (i.e., in $\gamma_f, \ldots, \gamma_h$). As we saw before, Newton's method suggests that we simply drop these nonlinear terms to obtain a linear system for the "delta" variables.

Clearly, the main computational burden is to solve the system shown above. It is important to note that this is a large, sparse, indefinite, linear system. It is also symmetric if one negates certain rows and rearranges rows and columns appropriately:

$$
\left[\begin{array}{ccccc|ccccc}
-FV^{-1} & & & & & & -I & & & \\
 & & & & & I & & & & I \\
 & & & & & -I & & & I & \\
 & & & & & & I & I & & \\
 & & & & & A & I & & & \\
\hline
 & I & -I & & A^T & & & & & \\
-I & & & I & I & & & & & \\
 & & & I & & & & QP^{-1} & & \\
 & & I & & & & & & HG^{-1} & \\
 & I & & & & & & & & ST^{-1}
\end{array}\right]
\left[\begin{array}{c}
\Delta v \\ \Delta s \\ \Delta h \\ \Delta q \\ \Delta y \\ \hline \Delta x \\ \Delta f \\ \Delta p \\ \Delta g \\ \Delta t
\end{array}\right]
=
\left[\begin{array}{c}
-\gamma_f \\ \tau \\ \nu \\ \alpha \\ \rho \\ \hline \sigma \\ \beta \\ \gamma_q \\ \gamma_h \\ \gamma_s
\end{array}\right].
$$

Because this system is symmetric, we look for symmetric pivots to solve it. That is, we choose our pivots only from the diagonal elements. It turns out that we can eliminate Δv, Δp, Δg, and Δt using the nonzero diagonals $-FV^{-1}$, QP^{-1},

HG^{-1}, and ST^{-1}, respectively, in any order without causing any nondiagonal fill-in. Indeed, the equations for Δv, Δp, Δg, and Δt are

$$\begin{aligned} \Delta v &= VF^{-1}(\gamma_f - \Delta f) \\ \Delta p &= PQ^{-1}(\gamma_q - \Delta q) \\ \Delta g &= GH^{-1}(\gamma_h - \Delta h) \\ \Delta t &= TS^{-1}(\gamma_s - \Delta s), \end{aligned} \tag{20.11}$$

and after elimination from the system, we get

$$\left[\begin{array}{cccc|cc} -TS^{-1} & & & & I & \\ & -GH^{-1} & & & -I & \\ & & -PQ^{-1} & & & I \\ & & & & A & I \\ \hline I & -I & & A^T & & \\ & & I & I & & VF^{-1} \end{array}\right] \left[\begin{array}{c} \Delta s \\ \Delta h \\ \Delta q \\ \Delta y \\ \hline \Delta x \\ \Delta f \end{array}\right] = \left[\begin{array}{c} \hat{\tau} \\ \hat{\nu} \\ \hat{\alpha} \\ \rho \\ \hline \sigma \\ \hat{\beta} \end{array}\right],$$

where again we have introduced abbreviated notations for the components of the right-hand side:

$$\begin{aligned} \hat{\tau} &= \tau - TS^{-1}\gamma_s \\ \hat{\nu} &= \nu - GH^{-1}\gamma_h \\ \hat{\alpha} &= \alpha - PQ^{-1}\gamma_q \\ \hat{\beta} &= \beta + VF^{-1}\gamma_f. \end{aligned}$$

Next we use the pivot elements $-TS^{-1}$, $-GH^{-1}$, and $-PQ^{-1}$ to solve for Δs, Δh, and Δq, respectively:

$$\begin{aligned} \Delta s &= -ST^{-1}(\hat{\tau} - \Delta x) \\ \Delta h &= -HG^{-1}(\hat{\nu} + \Delta x) \\ \Delta q &= -QP^{-1}(\hat{\alpha} - \Delta f). \end{aligned} \tag{20.12}$$

After eliminating these variables, the system simplifies to

$$\left[\begin{array}{c|cc} & A & I \\ \hline A^T & D & \\ I & & E \end{array}\right] \left[\begin{array}{c} \Delta y \\ \hline \Delta x \\ \Delta f \end{array}\right] = \left[\begin{array}{c} \rho \\ \hline \sigma + ST^{-1}\hat{\tau} - HG^{-1}\hat{\nu} \\ \hat{\beta} + QP^{-1}\hat{\alpha}. \end{array}\right],$$

where

$$D = ST^{-1} + HG^{-1}$$

and

$$E = VF^{-1} + QP^{-1}.$$

Finally, we use the pivot element E to solve for Δf,

$$\Delta f = E^{-1}(\hat{\beta} + QP^{-1}\hat{\alpha} - \Delta y), \tag{20.13}$$

which brings us to the reduced KKT equations:

$$\left[\begin{array}{c|c} -E^{-1} & A \\ \hline A^T & D \end{array}\right]\left[\begin{array}{c} \Delta y \\ \hline \Delta x \end{array}\right] = \left[\begin{array}{c} \rho - E^{-1}(\hat{\beta} + QP^{-1}\hat{\alpha}) \\ \hline \sigma + ST^{-1}\hat{\tau} - HG^{-1}\hat{\nu} \end{array}\right]. \tag{20.14}$$

initialize $(x, f, p, t, g, y, v, q, s, h)$ such that $f, p, t, g, v, q, s, h > 0$
while (not optimal) {

$$\rho = b - Ax - w$$
$$\sigma = c - A^T y + z$$
$$\gamma = f^T v + p^T q + t^T s + g^T h$$
$$\mu = \delta \frac{\gamma}{n+m}$$
$$\gamma_f = \mu V^{-1} e - f$$
$$\gamma_q = \mu P^{-1} e - q$$
$$\gamma_s = \mu T^{-1} e - s$$
$$\gamma_h = \mu G^{-1} e - h$$
$$\hat{\tau} = u - x - t - TS^{-1}\gamma_s$$
$$\hat{\nu} = -l + x - g - GH^{-1}\gamma_h$$
$$\hat{\alpha} = b - a - f - p - PQ^{-1}\gamma_q$$
$$\hat{\beta} = -y - q + v + VF^{-1}\gamma_f$$
$$D = ST^{-1} + HG^{-1}$$
$$E = VF^{-1} + QP^{-1}$$

solve: $\begin{bmatrix} -E^{-1} & A \\ A^T & D \end{bmatrix}\begin{bmatrix} \Delta y \\ \Delta x \end{bmatrix} = \begin{bmatrix} \rho - E^{-1}(\hat{\beta} + QP^{-1}\hat{\alpha}) \\ \sigma + ST^{-1}\hat{\tau} - HG^{-1}\hat{\nu} \end{bmatrix}$

compute: Δf using (20.13), Δs, Δh, Δq using (20.12),
and Δv, Δp, Δg, Δt using (20.11)

$$\theta = r \left(\max_{ij} \left\{ -\frac{\Delta f_j}{f_j}, -\frac{\Delta p_i}{p_i}, -\frac{\Delta t_i}{t_i}, -\frac{\Delta g_j}{g_j}, -\frac{\Delta v_j}{v_j}, -\frac{\Delta q_i}{q_i}, -\frac{\Delta s_i}{s_i}, -\frac{\Delta h_j}{h_j} \right\} \right)^{-1} \wedge 1$$

$$\begin{array}{llll} x \leftarrow x + \theta\Delta x, & y \leftarrow y + \theta\Delta y, & f \leftarrow f + \theta\Delta f, & v \leftarrow v + \theta\Delta v \\ p \leftarrow p + \theta\Delta p, & q \leftarrow q + \theta\Delta q, & t \leftarrow t + \theta\Delta t, & s \leftarrow s + \theta\Delta s \\ g \leftarrow g + \theta\Delta g, & h \leftarrow h + \theta\Delta h & & \end{array}$$

}

FIGURE 20.1. The path-following method—general form.

Up to this point, none of the eliminations have produced any off-diagonal fill-in. Also, the matrix for system given in (20.14) is a symmetric quasidefinite matrix. Hence, the techniques given in Section 19.2 for solving such systems can be used. The algorithm is summarized in Figure 20.1.

Exercises

20.1 The matrix

$$B = \begin{bmatrix} 2 & & -2 & & \\ & 1 & & -1 & \\ -2 & & 2 & & \\ & -1 & & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}$$

is not positive definite but is positive semidefinite. Find a factorization $B = LDL^T$, where L is lower triangular with ones on the diagonal and D is a diagonal matrix with nonnegative diagonal elements. If such a factorization exists for every symmetric positive semidefinite matrix, explain why. If not, give a counterexample.

20.2 Show that the sum of a positive definite matrix and a positive semidefinite matrix is positive definite.

20.3 Permute the rows/columns of the matrix B given in (20.6) so that the diagonal elements from B appear in the order $2, 3, 4, 5, 1$. Compute an LDL^T-factorization of this matrix.

20.4 Show that, if B is symmetric and positive semidefinite, then $|b_{ij}| \leq \sqrt{b_{ii}b_{jj}}$ for all i, j.

Notes

Most implementations of interior-point methods assume the problem to be formulated with equality constraints. In this formulation, Lustig et al. (1994) give a good overview of the performance of interior-point algorithms compared with the simplex method.

The suggestion that it is better to solve equations in the KKT form instead of normal form was offered independently by a number of researchers (Gill et al. 1992; Turner 1991; Fourer and Mehrotra 1991; Vanderbei and Carpenter 1993).

The advantages of the primal–dual symmetric formulation were first reported in Vanderbei (1994). The basic properties of quasidefinite matrices were first given in Vanderbei (1995).