SIAM J. MATRIX ANAL. APPL.
Vol. 21, No. 3, pp. 1004–1025

STAIRCASE FAILURES EXPLAINED BY ORTHOGONAL VERSAL FORMS*

ALAN EDELMAN† AND YANYUAN MA‡

Abstract. Treating matrices as points in n^2-dimensional space, we apply geometry to study and explain algorithms for the numerical determination of the Jordan structure of a matrix. Traditional notions such as sensitivity of subspaces are replaced with angles between tangent spaces of manifolds in n^2-dimensional space. We show that the subspace sensitivity is associated with a small angle between complementary subspaces of a tangent space on a manifold in n^2-dimensional space. We further show that staircase algorithm failure is related to a small angle between what we call staircase invariant space and this tangent space. The matrix notions in n^2-dimensional space are generalized to pencils in $2mn$-dimensional space. We apply our theory to special examples studied by Boley, Demmel, and Kågström.

Key words. staircase algorithm, Jordan structure, Kronecker structure, versal deformation, SVD

AMS subject classification. 65F99

PII. S089547989833574X

1. Introduction.

The problem of accurately computing Jordan–Kronecker canonical structures of matrices and pencils has captured the attention of many specialists in numerical linear algebra. Standard algorithms for this process are denoted "staircase algorithms" because of the shape of the resulting matrices [22, p. 370]. Little is understood concerning how and why they fail, and in this paper, we study the geometry of matrices in n^2-dimensional space and pencils in $2mn$-dimensional space to explain these failures. This follows a geometrical program to complement and perhaps replace traditional numerical concepts associated with matrix subspaces that are usually viewed in n-dimensional space.

This paper targets readers who are already familiar with the staircase algorithm. We refer them to [22, p. 370] and [10] for excellent background material and list other literature in section 1.1 for those who wish to have a comprehensive understanding of the algorithm. On the mathematical side, it is also helpful if the reader has some knowledge of Arnold's theory of versal forms, though a dedicated reader should be able to read this paper without such knowledge, perhaps skipping section 3.2.

The most important contributions of this paper may be summarized as follows:

- A geometrical explanation of staircase algorithm failures is given.
- Three significant subspaces are identified that decompose matrix or pencil space: $\mathcal{T}_b$, $\mathcal{R}$, $\mathcal{S}$. The most important of these spaces is $\mathcal{S}$, which we choose to call the "staircase invariant space."
- The idea that the staircase algorithm computes an Arnold normal form that is numerically more appropriate than Arnold's "matrices depending on parameters" is discussed.

*Received by the editors March 16, 1998; accepted for publication (in revised form) by P. Van Dooren May 3, 1999; published electronically March 8, 2000.
http://www.siam.org/journals/simax/21-3/33574.html

†Department of Mathematics, Massachusetts Institute of Technology, Room 2-380, Cambridge, MA 02139-4307 (edelman@math.mit.edu, http://www-math.mit.edu/~edelman). This author was supported by NSF grants 9501278-DMS and 9404326-CCR.

‡Department of Mathematics, Massachusetts Institute of Technology, Room 2-333, Cambridge, MA 02139-4307 (yanyuan@math.mit.edu, http://www-math.mit.edu/~yanyuan). This author was supported by NSF grant 9501278-DMS.

- A first order perturbation theory for the staircase algorithm is given.
- The theory is illustrated using an example by Boley [3].

The paper is organized as follows: In section 1.1 we briefly review the literature on staircase algorithms. In section 1.2 we introduce concepts that we call *pure, greedy,* and *directed* staircases to emphasize subtle distinctions on how the algorithm might be used. Section 1.3 contains some important messages that result from the theory to follow.

Section 2 presents two similar-looking matrices with very different staircase behavior. Section 3 studies the relevant n^2-dimensional geometry of matrix space, while section 4 applies this theory to the staircase algorithm. The main result may be found in Theorem 6.

Sections 5, 6, and 7 mimic sections 2, 3, and 4 for matrix pencils. Section 8 applies the theory toward special cases introduced by Boley [3] and Demmel and Kågström [12].

1.1. Jordan–Kronecker algorithm history. The first staircase algorithm was given by Kublanovskaya for Jordan structure in 1966 [32], where a normalized QR factorization is used for rank determination and nullspace separation. Ruhe [35] first introduced the use of the SVD into the algorithm in 1970. The SVD idea was further developed by Golub and Wilkinson [23, section 10]. Kågström and Ruhe [28, 29] wrote the first library-quality software for the complete Jordan normal form reduction, with the capability of returning after different steps in the reduction. Recently, Chaitin-Chatelin and Frayssé [6] developed a nonstaircase "qualitative" approach.

The staircase algorithm for the Kronecker structure of pencils was given by Van Dooren [13, 14, 15] and Kågström and Ruhe [30]. Kublanovskaya [33] fully analyzed the AB algorithm; however, earlier work on the AB algorithm goes back to the 1970s. Kågström [26, 27] gave an RGDSVD/RGQZD algorithm and this provided a base for later work on software. Error bounds for this algorithm were given by Demmel and Kågström [8, 9]. Beelen and Van Dooren [2] gave an improved algorithm which requires $O(m^2n)$ operations for $m \times n$ pencils. Boley [3] studied the sensitivity of the algebraic structure. Error bounds are given by Demmel and Kågström [10, 11].

Staircase algorithms are used both theoretically and practically. Elmroth and Kågström [19] used the staircase algorithm to test the set of 2×3 pencils; hence to analyze the algorithm Demmel and Edelman [7] used the algorithm to calculate the dimension of matrices and pencils with a given form. Van Dooren [14], Emami-Naeini and Van Dooren [20], Kautsky, Nichols, and Van Dooren [31], Boley and Van Dooren [5], and Wicks and DeCarlo [36] considered systems and control applications. Software for control theory was provided by Demmel and Kågström [12].

A number of papers used geometry to understand Jordan–Kronecker structure problems. Fairgrieve [21] regularized by taking the most degenerate matrix in a neighborhood; Edelman, Elmroth, and Kågström [17, 18] studied versality and stratifications; and Boley [4] concentrates on stratifications.

1.2. The staircase algorithms. Staircase algorithms for the Jordan–Kronecker form work by making sequences of rank decisions in combination with eigenvalue computations. We coin the terms *pure*, *greedy*, and *directed* staircases to emphasize a few variations on how the algorithm might be used. Pseudocode for the Jordan versions appears near the end of this subsection. In combination with these three choices, one can choose an option of *zeroing*. These choices are explained below.

The three variations for purposes of discussion are considered in exact arithmetic. The *pure* version is the pure mathematician's algorithm: It gives precisely the Jordan structure of a given matrix. The *greedy* version (also useful for a pure mathematician!) attempts to find the most "interesting" Jordan structure near the given matrix. The *directed* staircase attempts to find a nearby matrix with a preconceived Jordan structure. Roughly speaking, the difference between pure, greedy, and directed is whether the Jordan structure is determined by the matrix, by a user-controlled neighborhood of the matrix, or directly by the user, respectively.

In the *pure* staircase algorithm, rank decisions are made using the singular value decomposition. An explicit distinction is made between zero singular values and nonzero singular values. This determines the exact Jordan form of the input matrix.

The *greedy* staircase algorithm attempts to find the most interesting Jordan structure near the given matrix. Here the word "interesting" (or "degenerate") is used in the same sense as it is with precious gems—the rarer, the more interesting. Algorithmically, as many singular values as possible are thresholded to zero with a user-defined threshold. The more singular values that are set to zero, the rarer in the sense of codimension (see [7, 17, 18]).

The *directed* staircase algorithm allows the user to decide in advance what Jordan structure is desired. The Jordan structure dictates which singular values are set to 0. Directed staircase is used in a few special circumstances. For example, it is used when separating the zero Jordan structure from the right singular structure (used in `GUPTRI` [10, 11]). Moreover, Elmroth and Kågström imposed structures by the staircase algorithm in their investigation of the set of 2×3 pencils [19]. Recently, Lippert and Edelman [34] use directed staircase to compute an initial guess for a Newton minimization approach to computing the nearest matrix with a given form in the Frobenius norm.

In the greedy and directed modes, if we explicitly *zero* the singular values, we end up computing a new matrix in staircase form that has the same Jordan structure as a matrix *near* the original one. If we do not explicitly zero the singular values, we end up computing a matrix that is orthogonally similar to the original one (in the absence of roundoff errors), which is *nearly* in staircase form. For example, in `GUPTRI` [11], the choice of whether to zero the singular values is made by the user with an input parameter named `zero` which may be true or false.

To summarize the many choices associated with a staircase algorithm, there are really five distinct algorithms worth considering: The pure algorithm stands on its own; otherwise, the two choices of combinatorial structure (greedy and directed) may be paired with the choice to zero or not. Thereby we have the five algorithms:

1. pure staircase,
2. greedy staircase with zeroing,
3. greedy staircase without zeroing,
4. directed staircase with zeroing, and
5. directed staircase without zeroing.

Notice that in the pure staircase, we do not specify zeroing or not zeroing, since both will give the same result vacuously.

Of course, algorithms run in finite precision. One further detail is that there is some freedom in the singular value calculations which leads to an ambiguity in the staircase form: In the case of unequal singular values, an order must be specified, and when singular values are equal, there is a choice of basis to be made. We will not specify any order for the SVD, except that all singular values considered to be zero appear first.

In the ith loop iteration, we use w_i to denote the number of singular values that are considered to be zero. For the directed algorithm, w_i are input; otherwise, w_i are computed. In pseudocode, we have the following staircase algorithms for computing the Jordan form corresponding to eigenvalue λ.

INPUT:
1) matrix A
2) specify pure, greedy, or direct mode
3) specify zeroing or not zeroing

OUTPUT:
1) matrix A that may or may not be in staircase form
2) Q (optional)

$i = 0, \quad Q = I$
$A_{tmp} = A - \lambda I$
while A_{tmp} not full rank
  $i = i + 1$
  Let $n' = \sum_{j=1}^{i-1} w_j$ and $n_{tmp} = n - n' = dim(A_{tmp})$
  Use the SVD to compute an n_{tmp} by n_{tmp} unitary matrix V whose leading w_i columns span the nullspace or an approximation
    Choice I: Pure: Use the SVD algorithm to compute w_i and the exact nullspace
    Choice II: Greedy: Use the SVD algorithm and threshold the small singular values with a user specified tolerance, thereby defining w_i. The corresponding singular vectors become the first w_i vectors of V.
    Choice III: Directed: Use the SVD algorithm, the w_i are defined from the input Jordan structure. The w_i singular vectors are the first w_i columns of V.
  $A = \text{diag}(I_{n'}, V^*) \cdot A \cdot \text{diag}(I_{n'}, V), \quad Q = Q \cdot \text{diag}(I_{n'}, V)$
  Let A_{tmp} be the lower right $n_{tmp} - w_i$ by $n_{tmp} - w_i$ corner of A
  $A_{tmp} = A_{tmp} - \lambda I$
endwhile
If zeroing, return A in the form $\lambda I +$ a block strictly upper triangular matrix.

While the staircase algorithm often works very well, it has been known to fail. We can say that the greedy algorithm fails if it does not detect a matrix with the least generic form [7] possible within a given tolerance. We say that the directed algorithm fails if the staircase form it produces is very far (orders of magnitude, in terms of the usual Frobenious norm of matrix space) from the staircase form of the nearest matrix with the intended structure. In this paper, we mainly concentrate on the greedy staircase algorithm and its failure, but the theory is applicable to both approaches. We emphasize that we are intentionally vague about how far is "far" as this may be application dependent, but we will consider several orders of magnitude to constitute this notion.

1.3. Geometry of staircase and Arnold forms.

Our geometrical approach is inspired by Arnold's theory of versality [1]. For readers already familiar with Arnold's theory, we point out that we have a new normal form that enjoys the same properties as Arnold's original form, but is more useful numerically. For numerical analysts, we point out that these ideas are important for understanding the staircase algorithm. Perhaps it is safe to say that numerical analysts have had an "Arnold normal form" for years, but we did not recognize it as such—the computer was doing it for us automatically.

TABLE 1.1

		Angles			Components	
A	Staircase fails	$\langle \mathcal{S}, \mathcal{T}_b \oplus \mathcal{R} \rangle$	$\langle \mathcal{T}_b, \mathcal{R} \rangle$	$\langle \mathcal{S}, \mathcal{R} \rangle$	$\mathcal{S}$	$\mathcal{R}$
No weak stair	no	large	large	$\pi/2$	small	small
Weak stair	no	large	small	$\pi/2$	small	large
Weak stair	yes	small	small	$\pi/2$	large	large

The strength of the normal form that we introduce in section 3 is that it provides a first order rounding theory of the staircase algorithm. We will show that instead of decomposing the perturbation space into the normal space and a tangent space at a matrix A, the algorithm chooses a so-called staircase invariant space to take the place of the normal space. When some directions in the staircase invariant space are very close to the tangent space, the algorithm can fail.

From the theory, we decompose the matrix space into three subspaces that we call $\mathcal{T}_b$, $\mathcal{R}$, and $\mathcal{S}$, the precise definitions of which are given in Definitions 1 and 3. Here, $\mathcal{T}_b$ and $\mathcal{R}$ are two subspaces of the tangent space and $\mathcal{S}$ is a certain complementary space of the tangent space in the matrix space. For the eager reader, we point out that angles between these spaces are related to the behavior of the staircase algorithm; note that $\mathcal{R}$ is always orthogonal to $\mathcal{S}$. (We use $\langle \cdot, \cdot \rangle$ to represent the angle between two spaces.) See Table 1.1.

Here, by a weak stair [16] we mean the near rank deficiency of any superdiagonal block of the strictly block upper triangular matrix A.

2. A staircase algorithm's failure to motivate the theory. Consider the two matrices

$$A_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & \delta \\ 0 & 0 & 0 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 0 & \delta & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

where δ =`1.5e-9` is approximately on the order of the square root of the double precision machine $\epsilon = 2^{-52}$, roughly, `2.2e-16`. Both of these matrices clearly have the Jordan structure $J_3(0)$, but the staircase algorithm on A_1 and A_2 can behave very differently.

To test this, we used the `GUPTRI` [11] algorithm. `GUPTRI`[1] requires an input matrix A and two tolerance parameters `EPSU` and `GAP`. We ran `GUPTRI` on $\tilde{A}_1 \equiv A_1 + \epsilon \mathtt{E}$ and $\tilde{A}_2 \equiv A_2 + \epsilon \mathtt{E}$, where

$$\mathtt{E} = \begin{pmatrix} .3 & .4 & .2 \\ .8 & .3 & .6 \\ .4 & .9 & .6 \end{pmatrix},$$

and $\epsilon =$ `2.2e-14` is roughly 100 times the double-precision machine ϵ. The singular values of each of the two matrices $\tilde{A}_1$ and $\tilde{A}_2$ are $\sigma_1 =$ `1.0000e00`, $\sigma_2 =$ `1.4901e-09`, and $\sigma_3 =$ `8.8816e-15`. We set `GAP` to be always ≥ 1 and let `EPSU` $= a/(\|\tilde{A}_i\| * \mathtt{GAP})$,

[1]GUPTRI [10, 11] is a "greedy" algorithm with a sophisticated thresholding procedure based on two input parameters EPSU and GAP ≥ 1. We threshold σ_{k-1} if $\sigma_{k-1} < GAP \times \max(\sigma_k, EPSU \times \|A\|)$ (defining $\sigma_{n+1} \equiv 0$). The first argument of the maximum σ_k ensures a large gap between thresholded and nonthresholded singular values. The second argument ensures that σ_{k-1} is small. Readers who look at the GUPTRI software should note that singular values are ordered from smallest to largest, contrary to modern convention.

TABLE 2.1

a	Computed Jordan structure for $\tilde{A}_1$	Computed Jordan structure for $\tilde{A}_2$
$a \geq \sigma_2$	$J_2(0) \oplus J_1(0) + O(10^{-9})$	$J_2(0) \oplus J_1(0) + O(10^{-9})$
$\gamma \leq a < \sigma_2$	$J_3(0) + O(10^{-6})$	$J_3(0) + O(10^{-14})$
$a < \gamma$	$J_1(0) \oplus J_1(\alpha) \oplus J_1(\beta) + O(10^{-14})$	$J_1(0) \oplus J_1(\alpha) \oplus J_1(\beta) + O(10^{-14})$

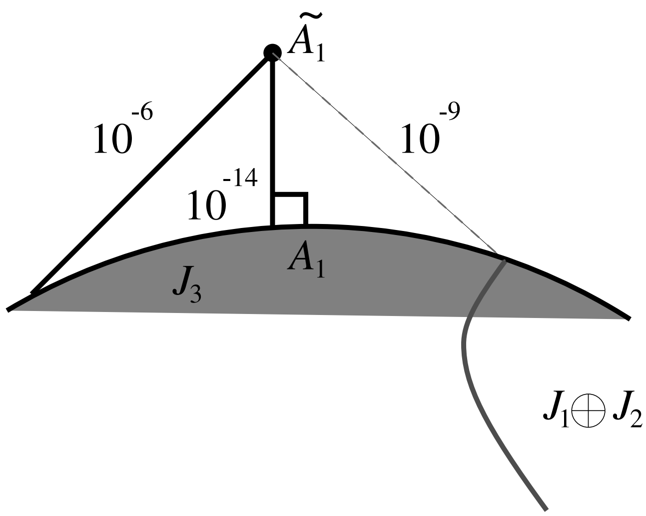

FIG. 2.1. *The staircase algorithm fails to find A_1 at distance 2.2e-14 from $\tilde{A}_1$ but does find a $J_3(0)$ or a $J_2(0) \oplus J_1(0)$ if given a much larger tolerance. (The latter is δ away from $\tilde{A}_1$.)*

where we vary the value of a. (The tolerance is effectively a.) Our observations are tabulated in Table 2.1.

Here, we use $J_k(\lambda)$ to represent a $k \times k$ Jordan block with eigenvalue λ. In the table, typically, $\alpha \neq \beta \neq 0$. Setting a small (smaller than $\gamma = 1.9985$`e-14` here, which is the smaller singular value in the second stage), the software returns two nonzero singular values in the first and second stages of the algorithm and one nonzero singular value in the third stage. Setting `EPSU` $\times$ `GAP` large (larger than σ_2 here), we zero two singular values in the first stage and one in the second stage, giving the structure $J_2(0) \oplus J_1(0)$ for both $\tilde{A}_1$ and $\tilde{A}_2$. (There is a matrix within $O(10^{-9})$ of A_1 and A_2 of the form $J_2(0) \oplus J_1(0)$.) The most interesting case is in between. For appropriate `EPSU` $\times$ `GAP` $\approx a$ (between γ and σ_2 here), we zero one singular value in each of the three stages, getting a $J_3(0)$ which is $O(10^{-14})$ away for A_2, while we can only get a $J_3(0)$ which is $O(10^{-6})$ away for A_1. In other words, the staircase algorithm fails for A_1 but not for A_2. As pictured in Figure 2.1, the A_1 example indicates that a matrix of the correct Jordan structure may be within the specified tolerance, but the staircase algorithm may fail to find it.

Consider the situation when A_1 and A_2 are transformed using a random orthogonal matrix Q. As a second experiment, we pick

$$Q \approx \begin{pmatrix} -.39878 & .20047 & -.89487 \\ -.84538 & -.45853 & .27400 \\ -.35540 & .86577 & .35233 \end{pmatrix}$$

TABLE 2.2

a	Computed Jordan structure for $\tilde{A}_1$	Computed Jordan structure for $\tilde{A}_2$
$a \geq \sigma_2$	$J_2(0) \oplus J_1(0) + O(10^{-5})$	$J_2(0) \oplus J_1(0) + O(10^{-6})$
$\gamma \leq a < \sigma_2$	$J_1(0) \oplus J_1(\alpha) \oplus J_1(\beta)$ ☹	$J_3(0) + O(10^{-6})$
$a < \gamma$	$J_1(0) \oplus J_1(\alpha) \oplus J_1(\beta) + O(10^{-14})$	$J_1(0) \oplus J_1(\alpha) \oplus J_1(\beta) + O(10^{-14})$

and take $\tilde{A}_1 = Q(A_1 + \epsilon E)Q^T$, $\tilde{A}_2 = Q(A_2 + \epsilon E)Q^T$. This will impose a perturbation of order ϵ. We ran `GUPTRI` on these two matrices; Table 2.2 shows the result.

In the table, $\gamma =$ `2.6980e-14`, all other values are the same as in the previous table.

In this case, `GUPTRI` is still able to detect a J_3 structure for $\tilde{A}_2$, although the one it finds is $O(10^{-6})$ away. But it fails to find any J_3 structure at all for $\tilde{A}_1$. The comparison of A_1 and A_2 in the two experiments indicates that the explanation is more subtle than the notion of a weak stair (a superdiagonal block that is almost column rank deficient) [16].

In this paper we present a geometrical theory that clearly predicts the difference between A_1 and A_2. The theory is based on how close certain directions that we will denote *staircase invariant directions* are to the tangent space of the manifold of matrices similar to the matrix with specified canonical form. It turns out that for A_1, these directions are nearly in the tangent space, but not for A_2. This is the crucial difference!

The tangent directions and the staircase invariant directions combine to form a "versal deformation" in the sense of Arnold [1], but one with more useful properties for our purposes.

3. Staircase invariant space and versal deformations.

3.1. The staircase invariant space and related subspaces. We consider block matrices as in Figure 3.1. Dividing a matrix A into blocks of row and column sizes $n_1, \ldots, n_k$, we obtain a *general block matrix*. A block matrix is *conforming to* A if it is also partitioned into blocks of size $n_1, \ldots, n_k$ in the same manner as A. If a general block matrix has nonzero entries only in the upper triangular blocks excluding the diagonal blocks, we call it a *block strictly upper triangular matrix*. If a general block matrix has nonzero entries only in the lower triangular blocks including the diagonal blocks, we call it a *block lower triangular matrix*. A matrix A is in *staircase form* if we can divide A into blocks of sizes $n_1 \geq n_2 \geq \cdots \geq n_k$ such that (s.t.) A is a strictly block upper triangular matrix and every superdiagonal block has full column rank. If a general block matrix has only nonzero entries on its diagonal blocks and each diagonal block is an orthogonal matrix, we call it a *block diagonal orthogonal matrix*. We call the matrix e^B a *block orthogonal matrix (conforming to* A*)* if B is a block antisymmetric matrix (conforming to A). (That is, B is antisymmetric with zero diagonal blocks. Here, we abuse the word "conforming" since e^B does not have a block structure.)

DEFINITION 1. *Suppose A is a matrix in staircase form. We call S a* staircase invariant matrix *of A if $S^TA = 0$ and S is block lower triangular. We call the space of matrices consisting of all such S the staircase invariant space of A, and denote it by $\mathcal{S}$.*

We remark that the columns of S will not be independent except possibly when $A = 0$; S can be the zero matrix as an extreme case. However the generic sparsity structure of S may be determined by the sizes of the blocks. For example, let A have

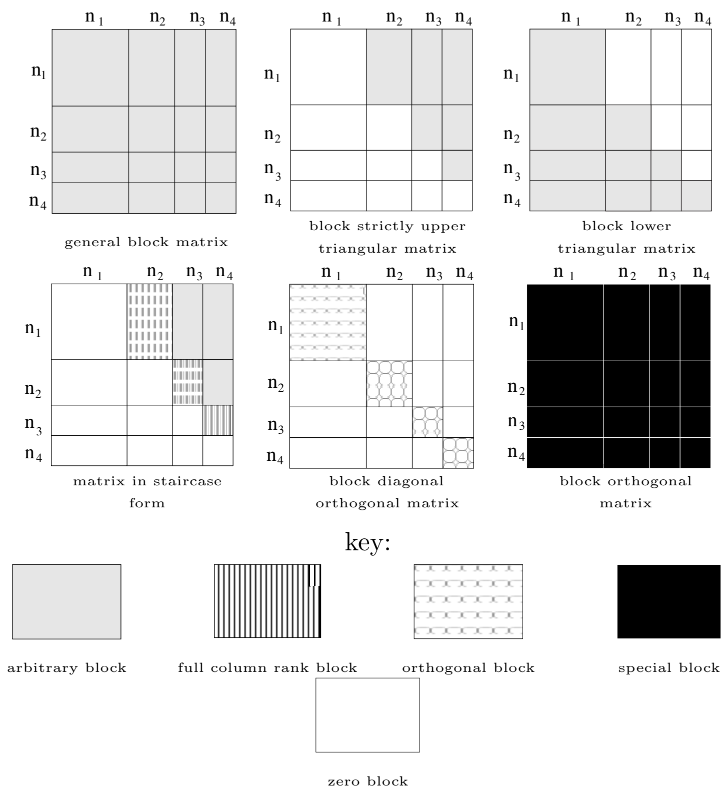

FIG. 3.1. *A schematic of the block matrices defined in the text.*

the staircase form

$$A = \left(\begin{array}{ccc|cc|cc|c}
0 & 0 & 0 & \times & \times & \times & \times & \times \\
0 & 0 & 0 & \times & \times & \times & \times & \times \\
0 & 0 & 0 & \times & \times & \times & \times & \times \\
 & & & 0 & 0 & \times & \times & \times \\
 & & & 0 & 0 & \times & \times & \times \\
 & & & & & 0 & 0 & \times \\
 & & & & & 0 & 0 & \times \\
 & & & & & & & 0
\end{array}\right).$$

Then

$$S = \left(\begin{array}{ccc|cc|cc|c}
\times & \times & \times & & & & & \\
\times & \times & \times & & & & & \\
\times & \times & \times & & & & & \\
\times & \times & \times & \circ & \circ & & & \\
\times & \times & \times & \circ & \circ & & & \\
\times & \times & \times & \times & \times & \times & \times & \\
\times & \times & \times & \times & \times & \times & \times & \\
\times & \times & \times & \times & \times & \times & \times & \times
\end{array}\right)$$

is a staircase invariant matrix of A if every column of S is a left eigenvector of A. Here, the $\circ$ notation indicates 0 entries in the block lower triangular part of S that are a consequence of the requirement that every column be a left eigenvector. This may be formulated as a general rule: If we find more than one block of size $n_i \times n_i$, then only those blocks on the lowest block row appear in the sparsity structure of S. For example, the $\circ$'s do not appear because they are above another block of size 2. As a special case, if A is strictly upper triangular, then S is 0 above the bottom row as is shown below. Readers familiar with Arnold's normal form will notice that if A is a given single Jordan block in normal form, then S contains the versal directions.

$$A = \begin{pmatrix} & \times & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \times & \times \\ & & & \times & \times & \times & \times \\ & & & & \times & \times & \times \\ & & & & & \times & \times \\ & & & & & & \times \\ & & & & & & \end{pmatrix}, \quad S = \begin{pmatrix} & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ \times & \times & \times & \times & \times & \times & \times \end{pmatrix}.$$

Definition 2. *Suppose A is a matrix. We call $\mathcal{O}(A) \equiv \{XAX^{-1} : X$ is a nonsingular matrix$\}$ the* orbit *of a matrix A. We call $\mathcal{T} \equiv \{AX - XA : X$ is any matrix$\}$ the* tangent space *of $\mathcal{O}(A)$ at A.*

Theorem 1. *Let A be an $n \times n$ matrix in staircase form; then the staircase invariant space $\mathcal{S}$ of A and the tangent space $\mathcal{T}$ form an oblique decomposition of $n \times n$ matrix space, i.e., $\mathbb{R}^{n^2} = \mathcal{S} \oplus \mathcal{T}$.*

Proof. Assume that $A_{i,j}$, the (i,j) block of A, is $n_i \times n_j$ for $i, j = 1, \dots, k$ and, of course, $A_{i,j} = 0$ for all $i \le j$.

There are n_1^2 degrees of freedom in the first block column of S because there are n_1 columns and each column may be chosen from the n_1-dimensional space of left eigenvectors of A. Indeed there are n_i^2 degrees of freedom in the ith block, because each of the n_i columns may be chosen from the n_i-dimensional space of left eigenvectors of the matrix obtained from A by deleting the first $i-1$ block rows and columns. The total number of degrees of freedom is $\sum_{i=1}^k n_i^2$, which combined with $\dim(\mathcal{T}) = n^2 - \sum_{i=1}^k n_i^2$ [7], gives the dimension of the whole space n^2.

If $S \in \mathcal{S}$ is also in $\mathcal{T}$, then S has the form $AX - XA$ for some matrix X. Our first step will be to show that X must have block upper triangular form after which we will conclude that $AX - XA$ is strictly block upper triangular. Since S is block lower triangular, it will then follow that if it is also in $\mathcal{T}$, it must be 0.

Let i be the first block column of X which does not have block upper triangular structure. Clearly the ith block column of XA is 0 below the diagonal block, so that the ith block column of $S = AX - XA$ contains vectors in the column space of A. However, every column of S is a left eigenvector of A from the definition and therefore is orthogonal to the column space of A. (Notice that we do not require these column vectors of S to be independent; the one Jordan block case is a good example.) Thus the ith block column of S is 0, and from the full column rank conditions on the superdiagonal blocks of A, we conclude that X is 0 below the block diagonal. ∎

Definition 3. *Suppose A is a matrix. We call $\mathcal{O}_b(A) \equiv \{Q^TAQ : Q = e^B, B$ is a block antisymmetric matrix conforming to $A\}$ the* block orthogonal orbit *of a matrix A. We call $\mathcal{T}_b \equiv \{AX - XA : X$ is a block antisymmetric matrix conforming to $A\}$ the* block tangent space *of the block orthogonal orbit $\mathcal{O}_b(A)$ at A. We call $\mathcal{R} \equiv \{$block strictly upper triangular matrix conforming to $A\}$ the* strictly upper block space *of A.*

Note that because of the complementary structure of the two matrices R and S, we can see that $\mathcal{S}$ is always orthogonal to $\mathcal{R}$.

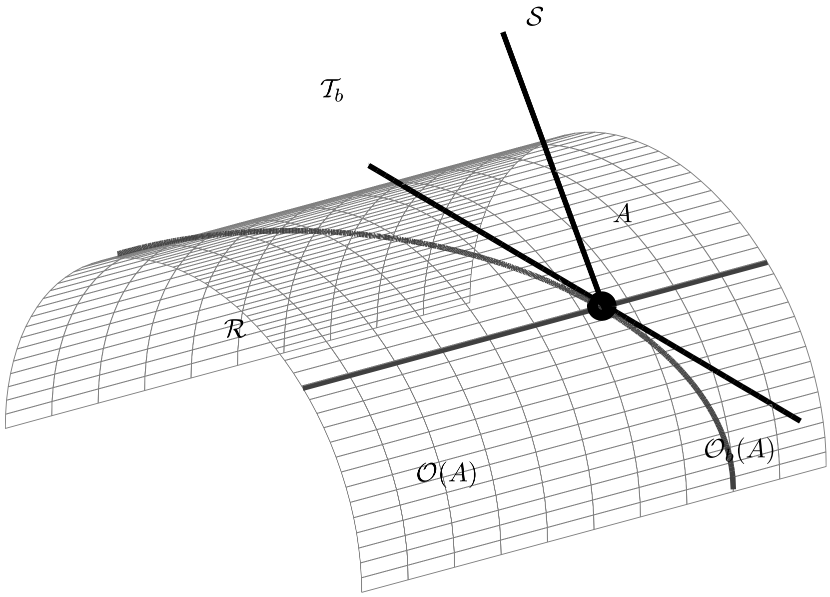

FIG. 3.2. *A diagram of the orbits and related spaces. The similarity orbit at A is indicated by a surface $\mathcal{O}(A)$; the block orthogonal orbit is indicated by a curve $\mathcal{O}_b(A)$ on the surface; the tangent space of $\mathcal{O}_b(A)$, $\mathcal{T}_b$, is indicated by a line, $\mathcal{R}$, which lies on $\mathcal{O}(A)$ is pictured as a line too; and the staircase invariant space $\mathcal{S}$ is represented by a line pointing away from the plane.*

THEOREM 2. *Let A be an $n \times n$ matrix in staircase form; then the tangent space $\mathcal{T}$ of the orbit $\mathcal{O}(A)$ can be split into the block tangent space $\mathcal{T}_b$ of the orbit $\mathcal{O}_b(A)$ and the strictly upper block space $\mathcal{R}$, i.e., $\mathcal{T} = \mathcal{T}_b \oplus \mathcal{R}$.*

Proof. We know that the tangent space $\mathcal{T}$ of the orbit at A has dimension $n^2 - \sum_{i=1}^k n_i^2$. If we decompose X into a block upper triangular matrix and a block antisymmetric matrix, we can decompose every $AX - XA$ into a block strictly upper triangular matrix and a matrix in $\mathcal{T}_b$. Since $\mathcal{T} = \mathcal{T}_b + \mathcal{R}$, each of $\mathcal{T}_b$ and $\mathcal{R}$ has dimension $\leq 1/2(n^2 - \sum_{i=1}^k n_i^2)$, they must both be exactly of dimension $1/2(n^2 - \sum_{i=1}^k n_i^2)$. Thus we know that they actually form a decomposition of $\mathcal{T}$, and the strictly upper block space $\mathcal{R}$ can also be represented as $\mathcal{R} \equiv \{AX - XA : X$ is block upper triangular matrix conforming to $A\}$. ◻

COROLLARY 1. $\mathbb{R}^{n^2} = \mathcal{T}_b \oplus \mathcal{R} \oplus \mathcal{S}$. *See Figure* 3.2.

In Definition 3, we really do not need the whole set $\{e^B : B$ is block antisymmetric$\} \equiv \{e^B\}$, we merely need a small neighborhood around $B = 0$. Readers may well wish to skip ahead to section 4, but for those interested in mathematical technicalities we review a few simple concepts. Suppose that we have partitioned $n = n_1 + \cdots + n_k$. An orthogonal decomposition of n-dimensional space into k mutually orthogonal subspaces of dimensions $n_1, n_2, \ldots, n_k$ is a point on the *flag manifold*. (When $k = 2$, this is the *Grassmann manifold*.) Equivalently, a point on the flag manifold is specified by a filtration, i.e., a nested sequence of subspaces V_i of dimension $n_1 + \cdots + n_i$ $(i = 1, \ldots, k)$:

$$0 \subset V_1 \subset \cdots \subset V_k = \mathbb{C}^n.$$

The corresponding decomposition can be written as

$$\mathbb{C}^n = V_k = V_1 \oplus V_2 \backslash V_1 \oplus \cdots \oplus V_k \backslash V_{k-1}.$$

This may be expressed concretely. If from a unitary matrix U, we define only V_i for $i = 1, \ldots, k$ as the span of the first $n_1 + n_2 + \cdots + n_i$ columns, then we have $V_1 \subset \cdots \subset V_k$, i.e., a point on the flag manifold. Of course, many unitary matrices U will correspond to the same flag manifold point. In an open neighborhood of $\{e^B\}$, near the point $e^0 = I$, the map between $\{e^B\}$ and an open subset of the flag manifold is a one-to-one homeomorphism. The former set is referred to as a local cross section [25, Lemma 4.1, p. 123] in Lie algebra. No two unitary matrices in a local cross section would have the same sequence of subspaces $V_i, i = 1, \ldots, k$.

3.2. Staircase as a versal deformation. Next, we are going to build up the theory of our versal form. Following Arnold [1], a *deformation* of a matrix A is a matrix $A(\lambda)$ with entries that are power series in the complex variables λ_i, where $\lambda = (\lambda_1, \ldots, \lambda_k)^T \in \mathbb{C}^k$, convergent in a neighborhood of $\lambda = 0$, with $A(0) = A$.

Good introductions to versal deformations may be found in [1, section 2.4] and [17]. The key property of a versal deformation is that it has enough parameters so that no matter how the matrix is perturbed, it may be made equivalent by analytic transformations to the versal deformation with some choice of parameters. The advantage of this concept for a numerical analyst is that we might make a rounding error in any direction and yet still think of this as a perturbation to a standard canonical form.

Let $N \subset M$ be a smooth submanifold of a manifold M. We consider a smooth mapping $A : \Lambda \to M$ of another manifold Λ into M and let λ be a point in Λ such that $A(\lambda) \in N$. The mapping A is called *transversal* to N at λ if the tangent space to M at $A(\lambda)$ is the direct sum

$$TM_{A(\lambda)} = A_* T\Lambda_\lambda \oplus TN_{A(\lambda)}.$$

Here, $TM_{A(\lambda)}$ is the tangent space of M at $A(\lambda)$, $TN_{A(\lambda)}$ is the tangent space of N at $A(\lambda)$, $T\Lambda_\lambda$ is the tangent space of Λ at λ, and A_* is the mapping from $T\Lambda_\lambda$ to $TM_{A(\lambda)}$ induced by A. (It is the Jacobian.)

THEOREM 3. *Suppose A is in staircase form. Fix $S_i \in \mathcal{S}$, $i = 1, \ldots, k$, s.t. span$\{S_i\} = \mathcal{S}$ and $k \geq \dim(\mathcal{S})$. It follows that*

$$A(\lambda) \equiv A + \sum_i \lambda_i S_i \tag{3.1}$$

is a versal deformation of every particular $A(\lambda)$ for λ small enough. $A(\lambda)$ is mini*versal at $\lambda = 0$ if $\{S_i\}$ is a basis of $\mathcal{S}$.*

Proof. Theorem 1 tells us the mapping $A(\lambda)$ is transversal to the orbit at A. From the equivalence of transversality and versality [1], we know that $A(\lambda)$ is a versal deformation of A. Since the dimension of the staircase invariant space $\mathcal{S}$ is the codimension of the orbit, $A(\lambda)$ given by (3.1) is a miniversal deformation if the S_i are a basis for $\mathcal{S}$ (i.e., $k = \dim(\mathcal{S})$). Moreover, $A(\lambda)$ is a versal deformation of every matrix in a neighborhood of A; in other words, the space $\mathcal{S}$ is transversal to the orbit of every $A(\lambda)$. Take a set of matrices X_i s.t. the $X_iA - AX_i$ form a basis of the tangent space $\mathcal{T}$ of the orbit at A. We know $\mathcal{T} \oplus \mathcal{S} = \mathbb{R}^{n^2}$ (here $\oplus$ implies $\mathcal{T} \cap \mathcal{S} = 0$), so there is a fixed minimum angle θ between $\mathcal{T}$ and $\mathcal{S}$. For small enough λ, we can guarantee that the $X_iA(\lambda) - A(\lambda)X_i$ are still linearly independent of each other and span a subspace of the tangent space at $A(\lambda)$ that is at least, say, $\theta/2$ away from $\mathcal{S}$. This means that the tangent space at $A(\lambda)$ is transversal to $\mathcal{S}$. □

Arnold's theory concentrates on general similarity transformations. As we have seen above, the staircase invariant directions are a perfect versal deformation. This idea can be refined to consider similarity transformations that are block orthogonal. Everything is the same as above, except that we add the block strictly upper triangular matrices R to compensate for the restriction to block orthogonal matrices. We now spell this out in detail, as follows.

DEFINITION 4. *If the matrix $C(\lambda)$ is block orthogonal for every λ, then we refer to the deformation as a* block orthogonal deformation*.*

We say that two deformations $A(\lambda)$ and $B(\lambda)$ are block orthogonally equivalent *if there exists a block orthogonal deformation $C(\lambda)$ of the identity matrix such that $A(\lambda) = C(\lambda)B(\lambda)C(\lambda)^{-1}$.*

We say that a deformation $A(\lambda)$ is block orthogonally versal *if any other deformation $B(\mu)$ is block orthogonally equivalent to the deformation $A(\phi(\mu))$. Here, ϕ is a mapping analytic at 0 with $\phi(0) = 0$.*

THEOREM 4. *A deformation $A(\lambda)$ of A is block orthogonally versal iff the mapping $A(\lambda)$ is transversal to the block orthogonal orbit of A at $\lambda = 0$.*

Proof. The proof follows Arnold [1, sections 2.3 and 2.4] except that we use the block orthogonal version of the relevant notions, and we remember that the tangents to the block orthogonal group are the commutators of A with the block antisymmetric matrices. □

Since we know that $\mathcal{T}$ can be decomposed into $\mathcal{T}_b \oplus \mathcal{R}$, we get the following theorem.

THEOREM 5. *Suppose a matrix A is in staircase form. Fix $S_i \in \mathcal{S}, i = 1, \ldots, k$, s.t. span$\{S_i\} = \mathcal{S}$ and $k \geq \dim(\mathcal{S})$. Fix $R_j \in \mathcal{R}, j = 1, \ldots, l$, s.t. span$\{R_j\} = \mathcal{R}$ and $l \geq \dim(\mathcal{R})$. It follows that*

$$A(\lambda) \equiv A + \sum_i \lambda_i S_i + \sum_j \lambda_j R_j$$

is a block orthogonally versal deformation of every particular $A(\lambda)$ for λ small enough. $A(\lambda)$ is block orthogonally mini*versal at A if $\{S_i\}$, $\{R_j\}$ are bases of $\mathcal{S}$ and $\mathcal{R}$.*

It is not hard to see that the theory we set up for matrices with all eigenvalues 0 can be generalized to a matrix A with different eigenvalues. The staircase form is a block upper triangular matrix, each of its diagonal blocks of the form $\lambda_i I + A_i$, with A_i in staircase form defined at the beginning of this chapter, and superdiagonal blocks arbitrary matrices. Its staircase invariant space is spanned by the block diagonal matrices, each diagonal block being in the staircase invariant space of the corresponding diagonal block A_i. $\mathcal{R}$ space is spanned by the block strictly upper triangular matrices s.t. every diagonal block is in the $\mathcal{R}$ space of the corresponding A_i. $\mathcal{T}_b$ is defined exactly the same as in the one eigenvalue case. All our theorems are still valid. When we give the definitions or apply the theorems, we do not really use the values of the eigenvalues. All that is important is how many different eigenvalues A has. In other words, we are working with bundle instead of orbit.

These forms are normal forms that have the same property as the Arnold's normal form: They are continuous under perturbation. The reason that we introduce block orthogonal notation is that the staircase algorithm is a realization to first order of the block orthogonally versal deformation, as we will see in the next section.

4. Application to matrix staircase forms. We are ready to understand the staircase algorithm described in section 1.2. We concentrate on matrices with all eigenvalues 0, since otherwise the staircase algorithm will separate other structures and continue recursively.

We use the notation stair(A) to denote the output A of the staircase algorithm as described in section 1.2. Now suppose that we have a matrix A which is in staircase form. To zeroth order, any instance of the staircase algorithm replaces A with $\hat{A} = Q_0^T A Q_0$, where Q_0 is block diagonal orthogonal. Of course this does not change the staircase structure of A; the Q_0 represents the arbitrary rotations within the subspaces and can depend on how the software is written and the subtlety of roundoff errors when many singular values are 0. Next, suppose that we perturb A by ϵE. According to Corollary 1, we can decompose the perturbation matrix uniquely as $E = S + R + T_b$ with $S \in \mathcal{S}$, $R \in \mathcal{R}$, and $T_b \in \mathcal{T}_b$. Theorem 6 states that, in addition to some block diagonal matrix Q_0, the staircase algorithm will apply a block orthogonal similarity transformation $Q_1 = I + \epsilon X + o(\epsilon)$ to $A + \epsilon E$ to kill the perturbation in T_b.

THEOREM 6. *Suppose that A is a matrix in staircase form and E is any perturbation matrix. The staircase algorithm (without zeroing) on $A + \epsilon E$ will produce an orthogonal matrix Q (depending on ϵ) and the output matrix* stair$(A + \epsilon E) =$ $Q^T(A + \epsilon E)Q = \hat{A} + \epsilon(\hat{S} + \hat{R}) + o(\epsilon)$, *where $\hat{A}$ has the same staircase structure as A, $\hat{S}$ is a staircase invariant matrix of $\hat{A}$, and $\hat{R}$ is a block strictly upper triangular matrix. If singular values are zeroed out, then the algorithm further kills $\hat{S}$ and outputs $\hat{A} + \epsilon\hat{R} + o(\epsilon)$.*

Proof. After the first stage of the staircase algorithm, the first block column is orthogonal to the other columns, and this property is preserved through the completion of the algorithm. Generally, after the ith iteration, the ith block column below (including) the diagonal block is orthogonal to all other columns to its right, and this property is preserved all through. So when the algorithm terminates, we will have a matrix whose columns below (including) the diagonal block are orthogonal to all the columns to the right; in other words, it is a matrix in staircase form plus a staircase invariant matrix.

We can always write the similarity transformation matrix as $Q = Q_0(I + \epsilon X + \mathrm{o}(\epsilon))$, where Q_0 is a block diagonal orthogonal matrix and X is a block antisymmetric matrix that does not depend on ϵ because of the local cross section property that we mentioned at the beginning of section 3. Notice that Q_0 is not a constant matrix decided by A; it depends on ϵE to its first order. We should have written $(Q_0)_0 + \epsilon(Q_0)_1 + \mathrm{o}(\epsilon)$ instead of Q_0 . However, we do not expand Q_0 since as long as it is a block diagonal orthogonal transformation, it does not change the staircase structure of the matrix. Hence, we get

$$
\begin{aligned}
\mathrm{stair}(A + \epsilon E) &= \mathrm{stair}(A + \epsilon S + \epsilon R + \epsilon T_b) \\
&= (I + \epsilon X^T + o(\epsilon))Q_0^T(A + \epsilon S + \epsilon R + \epsilon T_b)Q_0(I + \epsilon X + o(\epsilon)) \\
&= (I + \epsilon X^T + o(\epsilon))(\hat{A} + \epsilon\hat{S} + \epsilon\hat{R} + \epsilon\hat{T}_b)(I + \epsilon X + o(\epsilon)) \qquad (4.1) \\
&= \hat{A} + \epsilon(\hat{S} + \hat{R} + \hat{T}_b + \hat{A}X - X\hat{A}) + o(\epsilon) \\
&= \hat{A} + \epsilon(\hat{S} + \hat{R}) + o(\epsilon).
\end{aligned}
$$

Here, $\hat{A}, \hat{S}, \hat{R}$, and $\hat{T}_b$ are, respectively, $Q_0^T A Q_0, Q_0^T S Q_0, Q_0^T R Q_0$, and $Q_0^T T_b Q_0$. It is easy to check that $\hat{S}, \hat{R}, \hat{T}_b$ is still in the $\mathcal{S}, \mathcal{R}, \mathcal{T}_b$ space of $\hat{A}$. X is a block antisymmetric matrix satisfying $\hat{T}_b = X\hat{A} - \hat{A}X$. We know that X is uniquely determined because the dimensions of $\hat{T}_b$ and the block antisymmetric matrix space are the same.

The reason that $\hat{T}_b = X\hat{A} - \hat{A}X$, and hence the last equality in (4.1), holds is because the algorithm forces the output form as described in the first paragraph of this proof: $\hat{A} + \epsilon\hat{R}$ is in staircase form and $\epsilon\hat{S}$ is a staircase invariant matrix. Since $(\mathcal{S} \oplus \mathcal{R}) \cap \mathcal{T}_b$ is the zero matrix, the T_b term must vanish. ∎

To understand more clearly what this observation tells us, let us check some simple situations. If the matrix A is only perturbed in the direction $\mathcal{S}$ or $\mathcal{R}$, then the similarity transformation will simply be a block diagonal orthogonal matrix Q_0. If we ignore this transformation, which does not change any structure, we can think of the output to be unchanged from the input. This is the reason we call $\mathcal{S}$ the staircase invariant space. The reason we did not include R in the staircase invariant space is that $A + \epsilon R$ is still within $O_b(A)$. If the matrix A is only perturbed along the block tangent direction $\mathcal{T}_b$, then the staircase algorithm will kill the perturbation and do a block diagonal orthogonal similarity transformation.

Although the staircase algorithm decides this Q_0 step by step all through the algorithm (due to SVD rank decisions), we can actually think of the Q_0 as decided at the first step. We can even ignore this Q_0 because the only reason it comes up is that the SVD we use follows a specific method of sorting singular values when they are different and choosing the basis of the singular vector space when the same singular values appear.

We know that every matrix A can be reduced to a staircase form under an orthogonal transformation. In other words, we can always think of any general matrix M as P^TAP, where A is in staircase form. Thus, in general, the staircase algorithm always introduces an orthogonal transformation and returns a matrix in staircase form and a first order perturbation in its staircase invariant direction, i.e., $\text{stair}(M + \epsilon E) = \text{stair}(P^TAP + \epsilon E) = \text{stair}(A + \epsilon PEP^T)$.

It is now obvious that if a staircase form matrix A has its $\mathcal{S}$ and $\mathcal{T}$ almost normal to each other, then the staircase algorithm will behave very well. On the other hand, if $\mathcal{S}$ is very close to $\mathcal{T}$, then it will fail. To emphasize this, we write it as a conclusion.

CONCLUSION 1. *The angle between the staircase invariant space $\mathcal{S}$ and the tangent space $\mathcal{T}$ decides the behavior of the staircase algorithm. The smaller the angle, the worse the algorithm behaves.*

In the one Jordan block case, we have an if-and-only-if condition for $\mathcal{S}$ to be near $\mathcal{T}$.

THEOREM 7. *Let A be an $n \times n$ matrix in staircase form and suppose that all of its block sizes are 1×1; then $\mathcal{S}(A)$ is close to $\mathcal{T}(A)$ iff the following two conditions hold:*

(1) *(Row condition) there exists a nonzero row in A s.t. every entry on this row is $o(1)$.*
(2) *(Chain condition) there exists a chain of length $n-k$ with the chain value $O(1)$, where k is the lowest row satisfying (1).*

Here we call $A_{i_1,i_2}, A_{i_2,i_3}, \ldots, A_{i_t,i_{t+1}}$ a chain of length t and the product $A_{i_1,i_2}A_{i_2,i_3}\cdots A_{i_t,i_{t+1}}$ the chain value.

Proof sketch. Notice that $\mathcal{S}$ being close to $\mathcal{T}$ is equivalent to $\mathcal{S}$ being almost perpendicular to $\mathcal{N}$, the normal space of A. In this case, $\mathcal{N}$ is spanned by $\{I, A^T, A^{T2}, \ldots, A^{T(n-1)}\}$ and $\mathcal{S}$ consists of matrices with nonzero entries only in the last row. Considering the angle between any two matrices from the two spaces, it is straightforward to show that for $\mathcal{S}$ to be almost perpendicular to $\mathcal{N}$ is equivalent to the following:

(1) There exists a k s.t. the (n, k) entry of each of the matrices $I, A^T, \ldots, A^{T(n-1)}$ is $o(1)$ or 0.
(2) If the entry is $o(1)$, then it must have some other O(1) entry in the same matrix. Assume k is the largest choice if there are different k's. By a combi-

natorial argument, we can show that these two conditions are equivalent to the row and chain conditions, respectively, in our theorem. ∎

Remark 1. Note that saying that there exists an O(1) entry in a matrix is equivalent to saying that there exists a singular value of the matrix of O(1). So, the chain condition is the same as saying that the singular values of A^{n-k} are not all O(ϵ) or smaller.

Generally, we do not have an if-and-only-if condition for $\mathcal{S}$ to be close to $\mathcal{T}$. We only have a necessary condition, that is, only if at least one of the superdiagonal blocks of the original unperturbed matrix has a singular value almost zero, i.e., it has a weak stair, will $\mathcal{S}$ be close to $\mathcal{T}$. Actually, it is not hard to show that the angle between $\mathcal{T}_b$ and $\mathcal{R}$ is at most in the same order as the smallest singular value of the weak stair. So, when the perturbation matrix E is decomposed into $R + S + T_b$, R and T_b are typically very large, but whether S is large depends on whether $\mathcal{S}$ is close to $\mathcal{T}$.

Notice that (4.1) is valid for sufficiently small ϵ. What range of ϵ is considered sufficiently small? Clearly, ϵ has to be smaller than the smallest singular value δ of the weak stairs. Moreover, the algorithm requires the perturbations along both $\mathcal{T}$ and $\mathcal{S}$ to be smaller than δ. Assuming the angle between $\mathcal{T}$ and $\mathcal{S}$ is θ, then generally, when θ is large, we would expect an ϵ smaller than δ to be sufficiently small. However, when θ is close to zero, for a random perturbation, we would expect an ϵ in the order of δ/θ to be sufficiently small. Here, again, we can see that the angle between $\mathcal{S}$ and $\mathcal{T}$ decides the range of effective ϵ. For small θ, when ϵ is not sufficiently small, we observed some discontinuity in the zeroth order term in (4.1) caused by the ordering of singular values during certain stages of the algorithm. Thus, instead of the identity matrix, we get a permutation matrix in the zeroth order term.

The theory explains why the staircase algorithm behaves so differently on the two matrices A_1 and A_2 in section 2. Using Theorem 7, we can see that A_1 is a staircase failure ($k = 2$), while A_2 is not ($k = 1$). By a direct calculation, we find that the tangent space and the staircase invariant space of A_1 is very close ($\sin(\langle \mathcal{S}, \mathcal{T} \rangle) = \delta/\sqrt{1+\delta^2}$), while this is not the situation for A_2 ($\sin(\langle \mathcal{S}, \mathcal{T} \rangle) = 1/\sqrt{3}$). When transforming to get $\tilde{A}_1$ and $\tilde{A}_2$ with Q, which is an approximate orthogonal matrix up to the order of square root of machine precision ϵ_m, another error in the order of $\sqrt{\epsilon_m}$ (10^{-7}) is introduced, it is comparable with δ in our experiment, so the staircase algorithm actually runs on a shifted version $A_1 + \delta E_1$ and $A_2 + \delta E_1$. That is why we see R as large as an $O(10^{-6})$ added to J_3 in the second table for $\tilde{A}_2$ (see Table 2.2). We might as well call A_2 a staircase failure in this situation, but A_1 suffers a much worse failure under the same situation, in that the staircase algorithm fails to detect a J_3 structure at all. This is because the tangent space and the staircase invariant space are so close that the S and T component are very large, hence (4.1) no longer applies.

5. A staircase algorithm failure to motivate the theory for pencils. The pencil analog to the staircase failure in section 2 is

$$(A_1, B_1) = \left(\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} \delta & 0 & 0 & 0 \\ 0 & \delta & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \right),$$

where δ = `1.5e-8`. This is a pencil with the structure $L_1 \oplus J_2(0)$. After we add a random perturbation of size `1e-14` to this pencil, `GUPTRI` fails to return back the original pencil no matter which `EPSU` we choose. Instead, it returns back a more generic $L_2 \oplus J_1(0)$ pencil $O(\epsilon)$ away.

On the other hand, for another pencil with the same $L_1 \oplus J_2(0)$ structure,

$$(A_2, B_2) = \left(\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \delta & 0 \end{bmatrix} \right),$$

`GUPTRI` returns an $L_1 \oplus J_2(0)$ pencil $O(\epsilon)$ away.

At this point, readers may correctly expect that the reason behind this is again the angle between two certain spaces, as in the matrix case.

6. Matrix pencils.

Parallel to the matrix case, we can set up a similar theory for the pencil case. For simplicity, we concentrate on the case when a pencil only has L-blocks and $J(0)$-blocks. Pencils containing L^T-blocks and nonzero (including ∞) eigenvalue blocks can always be reduced to the previous case by transposing and exchanging the two matrices of the pencil and/or shifting.

6.1. The staircase invariant space and related subspaces for pencils.

A pencil (A, B) is in *staircase form* if we can divide both A and B into block rows of sizes $r_1, \ldots, r_k$ and block columns of sizes $s_1, \ldots, s_{k+1}$, s.t. A is strictly block upper triangular with every superdiagonal block having full column rank and B is block upper triangular with every diagonal block having full row rank and the rows orthogonal to each other. Here we allow s_{k+1} to be zero. A pencil is called *conforming to* (A, B) if it has the same block structure as (A, B). A square matrix is called *row (column) conforming to* (A, B) if it has diagonal block sizes the same as the row (column) sizes of (A, B).

DEFINITION 5. *Suppose (A, B) is a pencil in staircase form and B_d is the block diagonal part of B. We call (S_A, S_B) a* staircase invariant pencil of (A, B) *if $S_A^T A = 0$, $S_B B_d^T = 0$, and (S_A, S_B) has complementary structure to (A, B). We call the space consisting of all such (S_A, S_B) the* staircase invariant space of (A, B) *and denote it by $\mathcal{S}$.*

For example, let (A, B) have the staircase form

$$(A, B) = \left(\left[\begin{array}{ccc|cc|cc|c} 0 & 0 & 0 & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times & \times & \times \\ & & & 0 & 0 & \times & \times & \times \\ & & & 0 & 0 & \times & \times & \times \\ & & & & & 0 & 0 & \times \\ & & & & & 0 & 0 & \times \\ & & & & & & & 0 \end{array} \right], \left[\begin{array}{ccc|cc|cc|c} \times & \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times & \times & \times \\ & & & \times & \times & \times & \times & \times \\ & & & \times & \times & \times & \times & \times \\ & & & & & \times & \times & \times \\ & & & & & \times & \times & \times \\ & & & & & & & \times \end{array} \right] \right);$$

then

$$(S_A, S_B) = \left(\left[\begin{array}{ccc|cc|cc|c} \circ & \circ & \circ & & & & & \\ \circ & \circ & \circ & & & & & \\ \circ & \circ & \circ & \circ & \circ & & & \\ \circ & \circ & \circ & \circ & \circ & & & \\ \times & \times & \times & \times & \times & \times & \times & \\ \times & \times & \times & \times & \times & \times & \times & \\ \times & \times & \times & \times & \times & \times & \times & \times \end{array} \right], \left[\begin{array}{ccc|cc|cc|c} \times & \times & \times & & & & & \\ \times & \times & \times & & & & & \\ \times & \times & \times & \circ & \circ & & & \\ \times & \times & \times & \circ & \circ & & & \\ \times & \times & \times & \circ & \circ & \circ & \circ & \end{array} \right] \right)$$

is a staircase invariant pencil of (A, B) if every column of S_A is in the left null space of A and every row of S_B is in the right null space of B. Notice that the sparsity structure of S_A and S_B is at most complementary to that of A and B, respectively, but S_A and S_B are often less sparse because of the requirement on the nullspace. To be precise, if we find more than one diagonal block with the same size, then among the blocks of this size, only the blocks on the lowest block row appear in the sparsity structure of S_A. If any of the diagonal blocks of B is a square block, then S_B has all zero entries throughout the corresponding block column.

As special cases, if A is a strictly upper triangular square matrix and B is an upper triangular square matrix with diagonal entries nonzero, then S_A has only nonzero entries in the bottom row and S_B is simply a zero matrix. If A is a strictly upper triangular $n \times (n+1)$ matrix and B is an upper triangular $n \times (n+1)$ matrix with diagonal entries nonzero, then (S_A, S_B) is the zero pencil.

DEFINITION 6. *Suppose (A,B) is a pencil. We call $\mathcal{O}(A,B) \equiv \{X(A,B)Y :$ X, Y are nonsingular square matrices $\}$ the* orbit *of a pencil (A,B). We call $\mathcal{T} \equiv \{X(A,B) - (A,B)Y : X, Y$ are any square matrices $\}$ the* tangent space *of $\mathcal{O}(A,B)$ at (A,B).*

THEOREM 8. *Let (A,B) be an $m \times n$ pencil in staircase form; then the staircase invariant space $\mathcal{S}$ of (A,B) and the tangent space $\mathcal{T}$ form an oblique decomposition of $m \times n$ pencil space, i.e., $\mathbb{R}^{2mn} = \mathcal{S} + \mathcal{T}$.*

Proof. The proof of the theorem is similar to that of Theorem 1; first we prove the dimension of $\mathcal{S}(A,B)$ is the same as the codimension of $\mathcal{T}(A,B)$, then we prove $\mathcal{S} \cap \mathcal{T} = \{0\}$ by induction. The reader may fill in the details. ∎

DEFINITION 7. *Suppose (A,B) is a pencil. We call $\mathcal{O}_b(A,B) \equiv \{P(A,B)Q : P = e^X, X$ is a block antisymmetric matrix row conforming to $(A,B), Q = e^Y, Y$ is a block antisymmetric matrix column conforming to $(A,B)\}$ the* block orthogonal orbit *of a pencil (A,B). We call $\mathcal{T}_b \equiv \{X(A,B) - (A,B)Y : X$ is a block antisymmetric matrix row conforming to (A,B), Y is a block antisymmetric matrix column conforming to $(A,B)\}$ the* block tangent space *of the block orthogonal orbit $\mathcal{O}_b(A,B)$ at (A,B). We call $\mathcal{R} \equiv \{U(A,B) - (A,B)V : U$ is a block upper triangular matrix row conforming to (A,B), V is a block upper triangular matrix column conforming to $(A,B)\}$ the* block upper pencil space *of (A,B).*

THEOREM 9. *Let (A,B) be an $m \times n$ pencil in staircase form; then the tangent space $\mathcal{T}$ of the orbit $\mathcal{O}(A,B)$ can be split into the block tangent space $\mathcal{T}_b$ of the orbit $\mathcal{O}_b(A,B)$ and the block upper pencil space $\mathcal{R}$, i.e., $\mathcal{T} = \mathcal{T}_b \oplus \mathcal{R}$.*

Proof. This can be proved by a very similar argument concerning the dimensions as for matrix, in which the dimension of $\mathcal{R}$ is $2\sum_{i<j} r_i s_j + \sum r_i s_i$, the dimension of $\mathcal{T}_b$ is $\sum_{i<j} r_i r_j + \sum_{i<j} s_i s_j$, the codimension of the orbit $\mathcal{O}(A,B)$ (or $\mathcal{T}$) is $\sum s_i r_i - \sum_{j>i} s_i s_j + 2\sum_{j>i} s_i r_j - \sum_{j>i} r_i r_j$ [7]. ∎

COROLLARY 2. $\mathbb{R}^{2mn} = \mathcal{T}_b \oplus \mathcal{R} \oplus \mathcal{S}$.

6.2. Staircase as a versal deformation for pencils. The theory of versal forms for pencils [17] is similar to that for matrices. A *deformation* of a pencil (A,B) is a pencil $(A,B)(\lambda)$ with entries power series in the real variables λ_i. We say that two deformations $(A,B)(\lambda)$ and $(C,D)(\lambda)$ are *equivalent* if there exist two deformations $P(\lambda)$ and $Q(\lambda)$ of identity matrices such that $(A,B)(\lambda) = P(\lambda)(C,D)(\lambda)Q(\lambda)$.

THEOREM 10. *Suppose (A,B) is in staircase form. Fix $S_i \in \mathcal{S}$, $i = 1, \ldots, k$, s.t. span$\{S_i\} = \mathcal{S}$ and $k \geq \dim(\mathcal{S})$. It follows that*

$$(A,B)(\lambda) \equiv (A,B) + \sum_i \lambda_i S_i \tag{6.1}$$

is a versal deformation of every particular $(A,B)(\lambda)$ for λ small enough. $(A,B)(\lambda)$ is miniversal *at $\lambda = 0$ if $\{S_i\}$ is a basis of $\mathcal{S}$.*

DEFINITION 8. *We say two deformations $(A,B)(\lambda)$ and $(C,D)(\lambda)$ are* block orthogonally equivalent *if there exist two block orthogonal deformations $P(\lambda)$ and $Q(\lambda)$ of the identity matrix such that $(A,B)(\lambda) = P(\lambda)(C,D)(\lambda)Q(\lambda)$. Here, $P(\lambda)$ and $Q(\lambda)$ are exponentials of matrices which are conforming to (A,B) in row and column, respectively.*

We say that a deformation $(A,B)(\lambda)$ is block orthogonally versal *if any other deformation $(C,D)(\mu)$ is block orthogonally equivalent to the deformation $(A,B)(\phi(\mu))$. Here, ϕ is a mapping holomorphic at 0 with $\phi(0) = 0$.*

THEOREM 11. *A deformation $(A,B)(\lambda)$ of (A,B) is block orthogonally versal iff the mapping $(A,B)(\lambda)$ is transversal to the block orthogonal orbit of (A,B) at $\lambda = 0$.*

This is the corresponding result to Theorem 4.

Since we know that $\mathcal{T}$ can be decomposed into $\mathcal{T}_b \oplus \mathcal{R}$, we get the following theorem.

THEOREM 12. *Suppose a pencil (A,B) is in staircase form. Fix $S_i \in \mathcal{S}, i = 1,\ldots,k$, s.t. span$\{S_i\} = \mathcal{S}$ and $k \geq \dim(\mathcal{S})$. Fix $R_j \in \mathcal{R}, j = 1,\ldots,l$, s.t. span$\{R_j\} = \mathcal{R}$ and $l \geq \dim(\mathcal{R})$. It follows that*

$$(A,B)(\lambda) \equiv (A,B) + \sum_i \lambda_i S_i + \sum_j \lambda_j R_j$$

is a block orthogonally versal deformation of every particular $(A,B)(\lambda)$ for λ small enough. $(A,B)(\lambda)$ is block orthogonally mini*versal at (A,B) if $\{S_i\}$, $\{R_j\}$ are bases of $\mathcal{S}$ and $\mathcal{R}$.*

Notice that as in the matrix case, we can also extend our definitions and theorems to the general form containing L^T-blocks and nonzero eigenvalue blocks, and again, we will not specify what eigenvalues they are and hence get into the bundle case. We want to point out only one particular example here. If (A,B) is in the staircase form of $L_n + J_1(\cdot)$, then A will be a strictly upper triangular matrix with nonzero entries on the super diagonal and B will be a triangular matrix with nonzero entries on the diagonal except the $(n+1,n+1)$ entry. S_A will be the zero matrix and S_B will be a matrix with the only nonzero entry on its $(n+1,n+1)$ entry.

7. Application to pencil staircase forms. We concentrate on $L \oplus J(0)$ structures only, since otherwise the staircase algorithm will separate all other structures and continue similarly after a shift and/or transpose on that part only. As in the matrix case, the staircase algorithm basically decomposes the perturbation pencil into three spaces $\mathcal{T}_b$, $\mathcal{R}$, and $\mathcal{S}$ and kills the perturbation in $\mathcal{T}_b$.

THEOREM 13. *Suppose that (A,B) is a pencil in staircase form and E is any perturbation pencil. The staircase algorithm (without zeroing) on $(A,B)+\epsilon E$ will produce two orthogonal matrices P and Q (depending on ϵ) and the output pencil* $\mathrm{stair}((A,B)+\epsilon E) = P^T((A,B)+\epsilon E)Q = (\hat{A},\hat{B}) + \epsilon(\hat{S}+\hat{R}) + o(\epsilon)$*, where $(\hat{A},\hat{B})$ has the sane staircase structure as (A,B), $\hat{S}$ is a staircase invariant pencil of $(\hat{A},\hat{B})$, and $\hat{R}$ is in the block upper pencil space $\mathcal{R}$. If singular values are zeroed out, then the algorithm further kills $\hat{S}$ and output $(\hat{A},\hat{B}) + \epsilon\hat{R} + o(\epsilon)$.*

We use a formula to explain the statement more clearly:

$$\begin{aligned}
&(I+\epsilon X + o(\epsilon))P_1((A,B)+\epsilon S + \epsilon R + \epsilon T_b)Q_1(I - \epsilon Y + o(\epsilon)) \\
&\quad = (I+\epsilon X + o(\epsilon))((\hat{A},\hat{B}) + \epsilon\hat{S} + \epsilon\hat{R} + \epsilon\hat{T}_b)(I - \epsilon Y + o(\epsilon)) \\
&\quad = (\hat{A},\hat{B}) + \epsilon(\hat{S} + \hat{R} + \hat{T}_b + X(\hat{A},\hat{B}) - (\hat{A},\hat{B})Y) + o(\epsilon) \\
&\quad = (\hat{A},\hat{B}) + \epsilon(\hat{S}+\hat{R}) + o(\epsilon).
\end{aligned} \tag{7.1}$$

Similarly, we can see that when a pencil has its $\mathcal{T}$ and $\mathcal{S}$ almost normal to each other, the staircase algorithm will behave well. On the other hand, if $\mathcal{S}$ is very close to $\mathcal{T}$, then it will behave badly. This is exactly the situation in the two pencil examples in section 5. Although the two pencils are both ill-conditioned, a direct calculation shows that the first pencil has its staircase invariant space very close to the tangent space (the angle $\langle \mathcal{S}, \mathcal{T} \rangle = \delta/\sqrt{\delta^2+2}$) while the second one does not (the angle $\langle \mathcal{S}, \mathcal{T} \rangle = 1/\sqrt{2+\delta^2}$).

The if-and-only-if condition for $\mathcal{S}$ to be close to $\mathcal{T}$ is more difficult than in the matrix case. One necessary condition is that one super diagonal block of A almost is not full column rank or one diagonal block of B almost is not full row rank. This is usually referred to as *weak coupling*.

8. Examples: The geometry of the Boley pencil and others. Boley [3, Example 2, p. 639] presents an example of a 7×8 pencil (A, B) that is controllable (has generic Kronecker structure) yet it is known that an uncontrollable system (nongeneric Kronecker structure) is nearby at a distance `6e-4`. What makes the example interesting is that the staircase algorithm fails to find this nearby uncontrollable system while other methods succeed (for example, [24]). Our theory provides a geometrical understanding of why this famous example leads to staircase failure: The staircase invariant space is very close to the tangent space.

The pencil that we refer to is $(A, B(\epsilon))$, where

$$A = \begin{pmatrix} \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \text{ and } B(\epsilon) = \begin{pmatrix} 1 & -1 & -1 & -1 & -1 & -1 & -1 & 7 \\ \cdot & 1 & -1 & -1 & -1 & -1 & -1 & 6 \\ \cdot & \cdot & 1 & -1 & -1 & -1 & -1 & 5 \\ \cdot & \cdot & \cdot & 1 & -1 & -1 & -1 & 4 \\ \cdot & \cdot & \cdot & \cdot & 1 & -1 & -1 & 3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & -1 & 2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \epsilon & 1 \end{pmatrix}.$$

(The dots refer to zeros, and in the original Boley example $\epsilon = 1$.)

When $\epsilon = 1$, the staircase algorithm predicts a distance of 1, and is therefore off by nearly four orders of magnitude. To understand the failure, our theory works best for smaller values of ϵ, but it is still clear that even for $\epsilon = 1$, there will continue to be difficulties.

It is useful to express the pencil $(A, B(\epsilon))$ as $P_0 + \epsilon E$, where $P_0 = (A, B(0))$ and S is zero except for a 1 in the (7,7) entry of its B part. P_0 is in the bundle of pencils whose Kronecker form is $L_6 + J_1(\cdot)$ and the perturbation E is exactly in the unique staircase invariant direction (hence the notation "S") as we pointed out at the end of section 6.

The relevant quantity is then the angle between the staircase invariant space and the pencil space. An easy calculation reveals that the angle is very small: $\theta_S = 0.0028$ radians. In order to get a feeling for what range of ϵ first order theory applies, we calculated the exact distance $d(\epsilon) \equiv d(P(\epsilon), \text{bundle})$ using the nonlinear eigenvalue template software [34]. To first order, $d(\epsilon) = \theta_S \cdot \epsilon$. Figure 8.1 plots the distances first for $\epsilon \in [0, 2]$ and then a close-up for $\epsilon = [0, 0.02]$.

Our observation based on this data suggests that first order theory is good to two decimal places for $\epsilon \leq 10^{-4}$ and one place for $\epsilon \leq 10^{-2}$. To understand the geometry of staircase algorithmic failure, one decimal place or even merely an order of magnitude is quite sufficient.

In summary, we see clearly that the staircase invariant direction is at a small

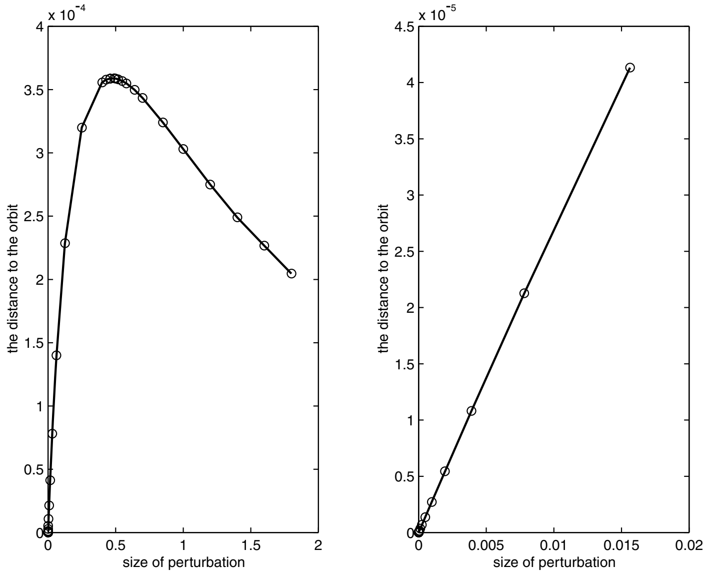

FIG. 8.1. *The picture to explain the change of the distance of the pencils $P_0 + \epsilon E$ to the bundle of $L_6 + J(\cdot)$ as ϵ changes. The second subplot is part of the first one at the points near $\epsilon = 0$.*

angle to the tangent space, and therefore the staircase algorithm will have difficulty finding the nearest pencil on the bundle or predicting the distance. This difficulty is quantified by the angle θ_S.

Since the Boley example is for $\epsilon = 1$, we computed the distance well past $\epsilon = 1$. The breakdown of first order theory is attributed to the curving of the bundle towards S. A three-dimensional schematic is portrayed in Figure 8.2.

The relevant picture for control theory is a planar intersection of the above picture. In control theory, we set the special requirement that the A matrix has the form $[0\ \ I]$. Pencils on the intersection of this hyperplane and the bundle are termed *uncontrollable*.

We analytically calculated the angle θ_c between S and the tangent space for the "uncontrollable surfaces." We found that $\theta_c = 0.0040$. Using the nonlinear eigenvalue template software [34], we numerically computed the true distance from $P_0 + \epsilon E$ to the uncontrollable surfaces and calculated the ratio of this distance to ϵ. We found that for $\epsilon < 8e - 4$, the ratio agrees with $\theta_c = 0.0040$ very well.

We did a similar analysis on the three pencils C_1, C_2, C_3 given by Demmel and Kågström [12]. We found that the *sin* values of the angles between $\mathcal{S}$ and $\mathcal{T}$ are, respectively, `2.4325e-02, 3.4198e-02`, and `8.8139e-03` and the *sin* values between $\mathcal{T}_b$ and $\mathcal{R}$ are, respectively, `1.7957e-02`, `7.3751e-03`, and `3.3320e-06`. This explains why we saw the staircase algorithm behave progressively worse on them. Especially, it explains why, when a perturbation about 10^{-3} is added to these pencils, C_3 behaves dramatically worse than C_1 and C_2. The component in $\mathcal{S}$ is almost of the same order as the entries of the original pencil.

So we conclude that the reason the staircase algorithm does not work well on this example is because $P_0 = (A, B(0))$ is actually a staircase failure, in that its tangent space, is very close to its staircase invariant space, and also the perturbation is so large that even if we know the angle in advance we cannot estimate the distance well.

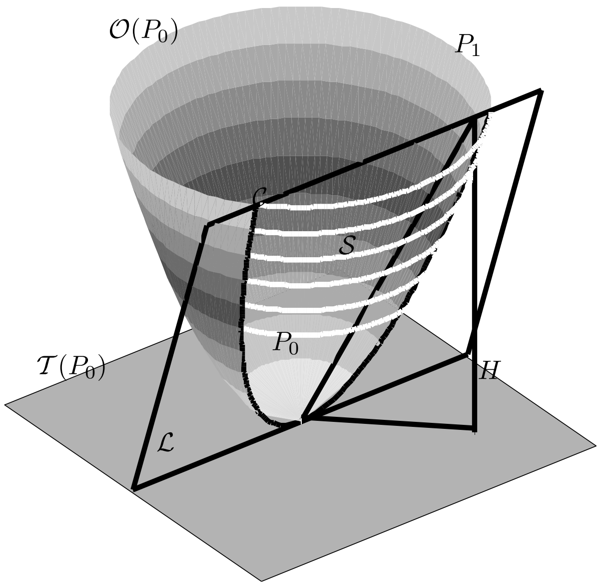

FIG. 8.2. *The staircase algorithm on the Boley example. The surface represents the orbit $\mathcal{O}(P_0)$. Its tangent space at the pencil P_0, $\mathcal{T}(P_0)$, is represented by the plane on the bottom. P_1 lies on the staircase invariant space $\mathcal{S}$ inside the "bowl." The hyperplane of uncontrollable pencils is represented by the plane cutting through the surface along the curve $\mathcal{C}$. It intersects $\mathcal{T}(P_0)$ along $\mathcal{L}$. The angle between $\mathcal{L}$ and $\mathcal{S}$ is θ_c. The angle between $\mathcal{S}$ and $\mathcal{T}(P_0)$, θ_S, is represented by the angle $\angle HP_0P_1$.*

Acknowledgments. The authors thank Bo Kågström and Erik Elmroth for their helpful discussion and their `conlab` software for easy interactive numerical testing. The staircase invariant directions were originally discovered for single Jordan blocks with Erik Elmroth while he was visiting MIT during the fall of 1996.

REFERENCES

[1] V. Arnold, *On matrices depending on parameters*, Russian Math. Surveys, 26 (1971), pp. 29–43.

[2] T. Beelen and P. V. Dooren, *An improved algorithm for the computation of Kronecker's canonical form of a singular pencil*, Linear Algebra Appl., 105 (1988), pp. 9–65.

[3] D. Boley, *Estimating the sensitivity of the algebraic structure of pencils with simple eigenvalue estimates*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 632–643.

[4] D. Boley, *The algebraic structure of pencils and block Toeplitz matrices*, Linear Algebra Appl., 279 (1998), pp. 255–279.

[5] D. Boley and P. V. Dooren, *Placing zeroes and the Kronecker canonical form*, Circuits Systems Signal Process., 13 (1994), pp. 783–802.

[6] F. Chaitin-Chatelin and V. Frayssé, *Lectures on Finite Precision Computations*, SIAM, Philadelphia, 1996.

[7] J. Demmel and A. Edelman, *The dimension of matrices (matrix pencils) with given Jordan (Kronecker) canonical forms*, Linear Algebra Appl., 230 (1995), pp. 61–87.

[8] J. Demmel and B. Kågström, *Stably computing the Kronecker structure and reducing subspace of singular pencils $A - \lambda B$ for uncertain data*, in Large Scale Eigenvalue Problems, J. Cullum and R. Willoughby, eds., North-Holland Math. Stud. 127, 1986, pp. 283–323.

[9] J. Demmel and B. Kågström, *Computing stable eigendecompositions of matrix pencils*, Linear Algebra Appl., 88/89 (1987), pp. 139–186.

[10] J. Demmel and B. Kågström, *The generalized Schur decomposition of an arbitrary pencil*

$A - \lambda B$*: Robust software with error bounds and applications.* I. *Theory and algorithms*, ACM Trans. Math. Software, 19 (1993), pp. 160–174.

[11] J. Demmel and B. Kågström, *The generalized Schur decomposition of an arbitrary pencil* $A-\lambda B$*: Robust software with error bounds and applications.* II. *Software and applications*, ACM Trans. Math. Software, 19 (1993), pp. 175–201.

[12] J. Demmel and B. Kågström, *Accurate solutions of ill-posed problems in control theory*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 126–145.

[13] P. V. Dooren, *The computation of Kronecker's canonical form of a singular pencil*, Linear Algebra Appl., 27 (1979), pp. 103–140.

[14] P. V. Dooren, *The generalized eigenstructure problem in linear system theory*, IEEE Trans. Automat. Control, 26 (1981), pp. 111–129.

[15] P. V. Dooren, *Reducing subspaces: Definitions, properties and algorithms*, in Matrix Pencils, B. Kågström and A. Ruhe, eds., Lecture Notes in Math. 973, Springer-Verlag, Berlin, 1983, pp. 58–73.

[16] P. V. Dooren, *private communication*, 1996.

[17] A. Edelman, E. Elmroth, and B. Kågström, *A geometric approach to perturbation theory of matrices and matrix pencils.* I. *Versal deformations*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 653–692.

[18] A. Edelman, E. Elmroth, and B. Kågström, *A geometric approach to perturbation theory of matrices and matrix pencils.* II. *A Stratification-enhanced staircase algorithm*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 667–699.

[19] E. Elmroth and B. Kågström, *The set of* 2*-by-*3 *matrix pencils—Kronecker structures and their transitions under perturbations*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 1–34.

[20] A. Emami-Naeini and P. V. Dooren, *Computation of zeros of linear multivariable systems*, Automatica, 18 (1982), pp. 415–430.

[21] T. F. Fairgrieve, *The Application of Singularity Theory to the Computation of Jordan Canonical Form*, Master's thesis, Univ. of Toronto, Toronto, ON, Canada, 1986.

[22] G. Golub and C. V. Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, London, 1996.

[23] G. H. Golub and J. H. Wilkinson, *Ill-conditioned eigensystems and the computation of the Jordan canonical form*, SIAM Rev., 18 (1976), pp. 578–619.

[24] M. Gu, *New methods for estimating the distance to uncontrollability*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 989–1003.

[25] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, New York, San Francisco, London, 1978.

[26] B. Kågström, *The generalized singular value decomposition and the general* $A - \lambda B$ *problem*, BIT, 24 (1984), pp. 568–583.

[27] B. Kågström, *RGSVD—An algorithm for computing the Kronecker structure and reducing subspaces of singular* $A-\lambda B$ *pencils*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 185–211.

[28] B. Kågström and A. Ruhe, *ALGORITHM* 560: *JNF, an algorithm for numerical computation of the Jordan normal form of a complex matrix [F*2*]*, ACM Trans. Math. Software, 6 (1980), pp. 437–443.

[29] B. Kågström and A. Ruhe, *An algorithm for numerical computation of the Jordan normal form of a complex matrix*, ACM Trans. Math. Software, 6 (1980), pp. 398–419.

[30] B. Kågström and A. Ruhe, *Matrix Pencils*, Lecture Notes in Math. 973, Springer-Verlag, New York, 1982.

[31] J. Kautsky, N. K. Nichols, and P. V. Dooren, *Robust pole assignment in linear state feedback*, Institute J. Control, 41 (1985), pp. 1129–1155.

[32] V. Kublanovskaya, *On a method of solving the complete eigenvalue problem of a degenerate matrix*, USSR Comput. Math. Phys., 6 (1966), pp. 1–14.

[33] V. Kublanovskaya, *AB-algorithm and its modifications for the spectral problem of linear pencils of matrices*, Numer. Math., 43 (1984), pp. 329–342.

[34] R. Lippert and A. Edelman, *Nonlinear eigenvalue problems*, in Templates for Eigenvalue Problems, Z. Bai, ed., to appear.

[35] A. Ruhe, *An algorithm for numerical determination of the structure of a general matrix*, BIT, 10 (1970), pp. 196–216.

[36] M. Wicks and R. DeCarlo, *Computing the distance to an uncontrollable system*, IEEE Trans. Automat. Control, 36 (1991), pp. 39–49.