Continuation of Invariant Subspaces for Large Bifurcation Problems

David Samuel Bindel
James Demmel
Mark Friedman

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2006-13
http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-13.html

February 13, 2006

Acknowledgement

Mark Friedman was partly supported under NSF DMS-0209536.

CONTINUATION OF INVARIANT SUBSPACES FOR LARGE BIFURCATION PROBLEMS

DAVID BINDEL *, JAMES DEMMEL †, AND MARK FRIEDMAN ‡

Abstract. We summarize an algorithm developed in [17] for computing a smooth orthonormal basis for an invariant subspace of a parameter-dependent matrix, and describe how to extend it for numerical bifurcation analysis. We adapt the continued subspace to track behavior relevant to bifurcations, and use projection methods to deal with large problems. To test our ideas, we have integrated our code into Matcont, a program for numerical continuation and bifurcation analysis.

1. Introduction. Parameter-dependent Jacobian matrices provide important information about dynamical systems

$$\frac{du}{dt} = f(u, \alpha), \text{ where } u \in \mathbb{R}^n, \ \alpha \in \mathbb{R}, \ f(u, \alpha) \in \mathbb{R}^n. \tag{1.1}$$

For example, to analyze stability at branches $(u(s), \alpha(s))$ of steady states

$$f(u, \alpha) = 0, \tag{1.2}$$

we look at the linearization $A(s) = D_u f(u(s), \alpha(s))$. If the system comes from a spatial discretization of a partial differential equation, then $A(s)$ will typically be large and sparse. In this case, an invariant subspace $\mathcal{R}(s)$ corresponding to a few eigenvalues near the imaginary axis provides information about stability and bifurcations.

Recently, we developed with collaborators the *CIS algorithm* for the continuation of invariant subspaces of a parameter-dependent matrix [17, 20, 25, 26, 7]. In this report, we extend the CIS algorithm to make it more suitable to numerical bifurcation analysis. Our goal is to extend numerical bifurcation techniques developed for small systems to larger systems. We also wish to ensure that bifurcations are detected reliably; this goal becomes especially relevant for non-normal matrices, where a small perturbation of a matrix may result in a large change to its eigenvalues [40, 41]. To this end, we make the following contributions to the development of the method: we derive new sufficient conditions for the existence of a continuously-defined invariant subspace; we introduce logic to adapt or reinitialize the subspace during continuation so that it is always well-defined and always includes information relevant to bifurcation analysis; we extend the algorithm to use Galerkin projection methods when n is large and direct methods are expensive; and we integrate our method into the MATCONT bifurcation analysis tool [18].

The CIS algorithm consists of a predictor based on first derivative information, and a corrector based on iterative refinement of an approximate invariant subspace (see [38], [16] and references therein). The algorithm evaluates a smoothly varying orthonormal basis for $\mathcal{R}(s)$ at sample points $s_0 < s_1 < \ldots < s_{N-1} < s_N$. This basis approximately minimizes arclength over all orthonormal bases for $\mathcal{R}(s)$, in a sense we will make precise in Section 2.1. The step size is adapted so that $h_i = s_i - s_{i-1}$ decreases when $\mathcal{R}(s)$ changes fast and increases when $\mathcal{R}(s)$ changes slowly. When the eigenvalues corresponding to $\mathcal{R}(s)$ come too near the rest of the spectrum, the

*Computer Science Division, University of California, Berkeley

†Computer Science Division and Department of Mathematics, University of California, Berkeley

‡Mathematical Sciences Department, University of Alabama, Huntsville. The author was supported under NSF DMS-0209536.

Script capitals ($\mathcal{Z}$)	Subspaces of $\mathcal{R}^m$
San-serif capitals (S)	Operators on matrix spaces (e.g. Sylvester operators)
Standard roman capitals (Z)	Matrices and bases
$\text{Grass}(n,m)$	The Grassmann manifold of m-dimensional subspaces of $\mathbb{R}^n$
$\text{Stief}(n,m)$	The Stiefel manifold of orthogonal bases of elements of $\text{Grass}(n,m)$
$O(n)$	Orthogonal matrices in $\mathbb{R}^{n\times n}$
$f(u,\alpha)$	Right-hand side in a dynamical system $\frac{du}{dt} = f(u,\alpha)$
$(u(s),\alpha(s))$	A branch of equilibria of $\frac{du}{dt} = f(u,\alpha)$
$A(s)$	A parameter-dependent matrix ($A : \mathbb{R} \to \mathbb{R}^{n\times n}$). Typically, $A(s) = D_u f(u(s),\alpha(s))$.
$A[s_0,s_1] = \frac{A(s_0)-A(s_1)}{s_0-s_1}$	Newton divided difference of A
$\mathcal{R}(s)$	A continuous maximal invariant subspace of $A(s)$
$Q(s) = \begin{bmatrix} Q_1(s) & Q_2(s) \end{bmatrix}$	A continuous basis for $\mathbb{R}$ such that $\mathcal{R}(s) = \text{span}(Q_1(s))$
$T(s) = \begin{bmatrix} T_{11}(s) & T_{12}(s) \\ 0 & T_{21}(s) \end{bmatrix}$	A continuous block Schur factor: $A(s)Q(s) = Q(s)T(s)$
$Y(s) = Q_2(s_0)^T Q_1(s)$	Riccati equation unknown
$\bar{Q}_1(s) = Q(s_0) \begin{bmatrix} I \\ Y(s) \end{bmatrix}$	Alternately normalized basis for $\mathcal{R}(s)$
$\hat{T}(s) = \begin{bmatrix} \hat{T}_{11} & \hat{T}_{12} \\ \hat{E}_{11} & \hat{T}_{21} \end{bmatrix}$	Approximate Schur form at s near s_0: $\hat{T}(s) = Q(s_0)^T A(s) Q(s_0)$
$\Lambda(s) = \{\lambda_i\}_{i=1}^n$	The spectrum of $A(s)$
$\Lambda_1(s) = \{\lambda_i\}_{i=1}^m$	The spectrum of $A(s)\|_{\mathcal{R}(s)}$
$\Lambda_2(s) = \{\lambda_i\}_{i=m+1}^n$	The spectrum of $A(s)\|_{\mathcal{R}(s)^\perp}$
$P(s)$	A (skew) eigenprojector associated with $\mathcal{R}(s)$
$\mathsf{S}Y = YA_{11} - A_{22}Y$	Sylvester operator associated with A ($A_{ij} = Q_i^T A Q_j$)
$\text{sep}(B,C)$	The smallest singular value of the Sylvester map $X \to BX - XC$
$\mathcal{V}$	A projection space for Galerkin approximation
V	An orthonormal basis for $\mathcal{V}$

FIG. 0.1. *Table of notation*

continuation procedure breaks down. In this case, the size of the continued subspace is adapted, and continuation proceeds with a larger or smaller subspace.

The rest of the paper is organized as follows. After discussing related work in the remainder of this section, we turn to the theory of existence and uniqueness of continuously-defined invariant subspaces and prove our new result on sufficient conditions for existence in Section 2. In Section 3, we describe the CIS algorithm and our new algorithms for initializing and updating the invariant subspace during the continuation process. In Section 4, we describe how to modify the CIS algorithm to use projection methods; and in Section 5, we illustrate the usefulness of the modified algorithm in bifurcation analysis through the solution of a model problem in MATCONT. We conclude and present our plans for future work in Section 6.

1.1. Related work. The local behavior of eigendecompositions and other matrix factorizations when viewed as matrix functions is of long-standing interest, and is treated in detail in the book by Stewart and Sun [39], as well as in the authoritative tome of Kato [32]. The local behavior of invariant subspaces can be analyzed by representing the subspaces near some reference subspace in terms of an orthogonal departure from that reference space; such analysis leads directly to an algebraic Ric-

cati equation. In [16], this Riccati equation was used as the basis for a unified analyis of several algorithms for refining approximate invariant subspaces; and in more recent work [9], new algorithms for invariant subspace approximation are proposed which combine a Galerkin approximate solution to an algebraic Riccati equation with the subspace construction ideas of the Jacobi-Davidson algorithm. In [24], Edelman and his colleagues proposed a more global approach to the analysis of linear algebra algorithms based on Grassmann manifolds and Stiefel manifolds (manifolds of subspaces and of orthonormal subspace bases, respectively); this approach has inspired several new methods for invariant subspace refinement, four of which are summarized and analyzed in [1].

No algorithm can produce globally continuous eigendecompositions, even for the set of diagonalizable matrices. However, one can smoothly define an invariant subspace basis along a path through matrix space, assuming the path crosses no singularities that would render the subspace discontinuous. In [19], a variety of continuous eigendecompositions for one-parameter matrix functions are described, including continuous Schur and block Schur decompositions. In a paper by Govarets, Guckenheimer, and Khibnik [31] which motivated our work on invariant subspace continuation, a low-dimensional invariant subspace of the Jacobian matrix, corresponding to the eigenvalues with largest real parts, was computed at each point along a continuation path and used to detect Hopf bifurcations via the bialternate matrix product. The authors concluded that subspace reduction can be combined with complicated bifurcation computations and should be tried for large problems.

The CIS algorithm was presented and analyzed in [17] and further studied in [20], [25], with additional practical developments in [26] and [7]. The algorithm of [20] constructs a smooth block 2-by-2 Schur decomposition; in [22], the approach is extended to the case of more blocks, and a new method is proposed to compute a smooth similarity reduction to block bidiagonal form. In [21], the approach described in [17] for using subspace continuation to compute connecting orbits between equilibria was extended to compute connecting orbits between periodic orbits. To continue low-dimensional invariant subspaces of sparse matrices, the authors of [6] use a bordered Bartels-Stewart algorithm to solve each corrector iteration; in [8], this approach is combined with ideas from [20, 25]. Though [6] and [8] deal with methods for sparse matrices, they differ from our current work in that they use different predictors and correctors, and they do not analyze and update the subspace during continuation to ensure it retains all information relevant to bifurcations.

Numerical continuation for large nonlinear systems arising from ODEs and discretized PDEs is an active area of research, and the idea of subspace projection is common in many methods being developed. The continuation algorithms are typically based on Krylov subspaces, or on recursive projection methods which use a time integrator instead of a Jacobian multiplication as a black box to identify the low-dimensional invariant subspace where interesting dynamics take place; see e.g. [3, 37, 29, 11, 30, 23, 27, 10, 13], and references there.

2. Continuous invariant subspaces. Let $A \in C^k([0,1],\ \mathbb{R}^{n\times n})$ be a k-times continuously differentiable parameter-dependent matrix. We can write the spectrum $\Lambda(s)$ of $A(s)$ as n continuous functions $\lambda_1(s), \ldots, \lambda_n(s)$ [32]. At parameter values where $\lambda_i(s)$ is a multiple eigenvalue, $\lambda_i(s)$ may not be differentiable, and it may be impossible to define a continuous right eigenvector. However, $\lambda_i(s)$ is a C^k function with a C^k right eigenvector as long as $\lambda_i(s)$ has algebraic multiplicity 1. More

generally, define

$$\begin{array}{l}\Lambda_1(s) := \{\lambda_i(s)\}_{i=1}^{m} \\ \Lambda_2(s) := \{\lambda_i(s)\}_{i=m+1}^{n} \\ \Lambda(s) := \Lambda_1(s) \cup \Lambda_2(s)\end{array} . \tag{2.1}$$

While $\Lambda_1(s)$ and $\Lambda_2(s)$ remain disjoint, there is a well-defined maximal right invariant subspace $\mathcal{R}(s)$ corresponding to $\Lambda_1(s)$, and $\mathcal{R}(s)$ is C^k. There are several ways to prove this fact, which we review in sections 2.2, 2.3, and 2.4.

In what follows, we will primarily use the Frobenius matrix norm: $\|A\|_F = \sqrt{\mathrm{tr}(A^T A)}$. We also assume that complex conjugate pairs are not split between Λ_1 and Λ_2.

2.1. The geometry of subspaces. We begin with a brief review of the geometry of subspaces and orthonormal bases (see [24] for a more complete treatment). The *Stiefel manifold* $\mathrm{Stief}(n,m)$ is the set of matrices with orthonormal columns:

$$\mathrm{Stief}(n,m) := \{Z \in \mathbb{R}^{n\times m} : Z^T Z = I\} \tag{2.2}$$

where $m \leq n$. We can also write

$$\mathrm{Stief}(n,m) = \{QI_{n,m} : Q \in O(n), I_{n,m} = \text{leading } m \text{ columns of } I_n\} \tag{2.3}$$

Well-known examples of Stiefel manifolds are the unit sphere (for $m = 1$) and the orthogonal group $O(n)$ (for $m = n$).

The *Grassmann manifold* $\mathrm{Grass}(n,m)$ is the set of all m-dimensional subspaces of $\mathbb{R}^n$. We represent elements of $\mathrm{Grass}(n,m)$ by equivalence classes of members of $\mathrm{Stief}(n,m)$ spanning the same space. That is,

$$\mathrm{Grass}(n,m) = \mathrm{Stief}(n,m)/\left[Z \sim ZU,\, U \in O(m)\right]. \tag{2.4}$$

The tangent directions to the orthogonal group $O(n)$ at Q_0 are translations of the skew symmetric matrices. For any $Q \in O(n)$ near Q_0,

$$Q = Q_0 + Q_0 H + \text{higher order terms, where } H = -H^T. \tag{2.5}$$

The tangents to $\mathrm{Stief}(n,m)$ have a related structure. If $Z_0 = Q_0 I_{n,m} \in \mathrm{Stief}(n,m)$ for $Q_0 \in O(n)$, then for any nearby $Z \in \mathrm{Stief}(n,m)$,

$$Z = Z_0 + Q_0 H I_{n,m} + \text{higher order terms, where } H = -H^T \tag{2.6}$$

In block form, these tangent directions look like $Q_0 \begin{bmatrix} H_{11} \\ H_{21} \end{bmatrix}$, where $H_{11} \in \mathbb{R}^{m\times m}$ is skew-symmetric and $H_{21} \in \mathbb{R}^{(n-m)\times m}$ is arbitrary.

The tangent space at $Z_0 \in \mathrm{Stief}(n,m)$ is a direct sum of two orthogonal spaces: the vertical space and the horizontal space (Figure 2.1). The *vertical space* is

$$\{\Delta Z \in \mathbb{R}^{n\times m} : \Delta Z = Z_0 H_{11} \text{ and } H_{11} \in \mathbb{R}^{m\times m} \text{ is skew}\}, \tag{2.7}$$

and the *horizontal space* is

$$\{\Delta Z \in \mathbb{R}^{n\times m} : Z_0^T \Delta Z = 0\}. \tag{2.8}$$

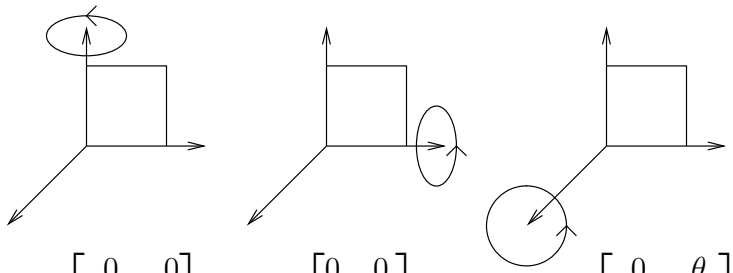

FIG. 2.1. *Two horizontal tangents (left) and one vertical tangent (right) at* $[e_1, e_2] \in \mathrm{Stief}(3,2)$

The set of matrices in $\mathrm{Stief}(n,m)$ spanning the same space as Z_0 is $\{Z \in \mathrm{Stief}(n,m) : Z = Z_0 U,\ U \in O(m)\}$. The vertical directions are exactly the tangents to this set. So vertical motion "spins" vectors without changing the subspace, while horizontal motion changes the subspace spanned.

We define the differentiable structure of $\mathrm{Grass}(n,m)$ in terms of the structure of $\mathrm{Stief}(n,m)$: a path $\mathcal{Z}(s)$ in $\mathrm{Grass}(n,m)$ is C^k if there is a C^k basis $Z : [0,1] \to \mathrm{Stief}(n,m)$ such that $\mathcal{Z}(s) = \mathrm{span}(Z(s))$. This basis is not unique; however, given a basis $Z_0 \in \mathrm{Stief}(n,m)$ for $\mathcal{Z}(0)$, there is a unique C^k basis starting from Z_0 which moves only horizontally. We describe the basis in the following lemma.

LEMMA 2.1. *Let* $\mathcal{Z} : [0,1] \to \mathrm{Grass}(n,m)$ *be a* C^k *parameter-dependent space* ($k > 0$). *Then for any* $Z_0 \in \mathrm{Stief}(n,m)$ *such that* $\mathcal{Z}(0) = \mathrm{span}(Z_0)$, *there is a unique* C^k *basis* $Z : [0,1] \to \mathrm{Stief}(n,m)$ *for* $\mathcal{Z}(s)$ *such that* $Z(0) = Z_0$ *and*

$$Z(s)^T Z'(s) = 0. \tag{2.9}$$

This basis minimizes the Euclidean arclength

$$l(Z) = \int_0^1 \|Z'(s)\|_F \, ds \tag{2.10}$$

over all C^k *orthonormal bases for* $\mathcal{Z}(s)$.

Proof.

Let $\hat{Z} : [0,1] \to \mathrm{Stief}(n,m)$ be one C^k orthonormal basis for $\mathcal{Z}$. Any other C^k orthonormal basis for $\mathcal{Z}$ can be written $Z = \hat{Z}U$ for some C^k function $U : [0,1] \to O(m)$. By the Pythagorean theorem,

$$\|(\hat{Z}U)'\|_F^2 = \|(I - \hat{Z}\hat{Z}^T)(\hat{Z}U)'\|_F^2 + \|\hat{Z}\hat{Z}^T(\hat{Z}U)'\|_F^2 \tag{2.11}$$

where the first term corresponds to horizontal motion, and the second term to vertical motion. Since the Frobenius norm is invariant under unitary transformations, we can show the first term depends only on $\mathcal{Z}$, and not on the particular choice of basis:

$$\|(I - \hat{Z}\hat{Z}^T)(\hat{Z}U)'\|_F = \|(I - \hat{Z}\hat{Z}^T)(\hat{Z}'U + \hat{Z}U')\|_F \tag{2.12}$$

$$= \|(I - \hat{Z}\hat{Z}^T)\hat{Z}'U\|_F \tag{2.13}$$

$$= \|(I - \hat{Z}\hat{Z}^T)\hat{Z}'\|_F. \tag{2.14}$$

By again using unitary invariance of the norm, we rewrite the second term as

$$\|\hat{Z}\hat{Z}^T(\hat{Z}U)'\|_F = \|\hat{Z}^T(\hat{Z}U)'\|_F = \|(\hat{Z}U)^T(\hat{Z}U)'\|_F. \tag{2.15}$$

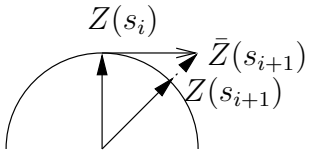

FIG. 2.2. *Discrete approximation to* $Z(s)^T Z'(s) = 0$ *for* $m = 1$

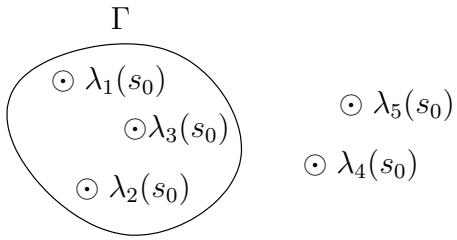

FIG. 2.3. *Contour* Γ *in* $\mathbb{C}$ *enclosing* $\Lambda_1 \subset \Lambda$

Therefore, the minimum attainable arclength should occur when

$$0 = (\hat{Z}U)^T(\hat{Z}U)' = U^T(\hat{Z}^T\hat{Z}'U + U') \tag{2.16}$$

or equivalently,

$$U' = -\hat{Z}\hat{Z}'U. \tag{2.17}$$

By the standard theory for linear ODEs, there is a unique U which satisfies (2.17) together with the initial condition $\hat{Z}(0)U(0) = Z_0$. Therefore, there is a unique orthonormal basis $Z = \hat{Z}U$ which satisfies (2.9) and $Z(0) = Z_0$. Furthermore, Z has minimal arclength.
□

For computation, we can approximate the equation $Z(s)^T Z'(s) = 0$ by the condition

$$Z(s_i)^T(\bar{Z}(s_{i+1}) - Z(s_i)) = 0. \tag{2.18}$$

As long as no vectors in $\mathcal{Z}(s_i)$ are normal to $\mathcal{Z}(s_{i+1})$, such a $\bar{Z}(s_{i+1}) \in \mathbb{R}^{n\times m}$ exists. Then we let the computed $Z(s_{i+1})$ be the element of $\mathrm{Stief}(n,m)$ nearest $\bar{Z}(s_{i+1})$ that spans the same space (Figure 2.2). The problem of finding the nearest element of $\mathrm{Stief}(n,m)$ to a given full-rank matrix in $\mathbb{R}^{n\times m}$ is called the *orthogonal Procrustes problem* [28, p. 582], and we will return to it later.

2.2. Complex-analytic characterization. In [32], Kato characterizes continuity of invariant subspaces in terms of the associated eigenprojections. If Γ is a union of disjoint positively-oriented simple closed contours in $\mathbb{C}$ with $\Lambda_1(s_0)$ inside Γ and $\Lambda_2(s_0)$ outside Γ (see Figure 2.3), then

$$P(s) := -\frac{1}{2\pi i}\int_\Gamma (A(s) - \xi I)^{-1}\, d\xi \tag{2.19}$$

is well-defined for any s near s_0. The matrix $P(s)$ is a projection with range $\mathcal{R}(s)$.

Suppose $X_0 \in \mathbb{R}^{n\times m}$ is a basis for $\mathcal{R}(s_0)$. Then we use $P(s)$ to locally produce a continuous basis $X(s)$ for $\mathcal{R}(s)$:

$$X(s) := P(s)X_0. \tag{2.20}$$

Because $X(s_0) = X_0$ is full rank, and the full rank matrices form an open subset of $\mathbb{R}^{n\times m}$, by continuity $X(s)$ will have full rank for all s sufficiently near s_0.

2.3. Differential equation characterization. We can also prove the existence of a C^k invariant subspace by writing a differential equation for a Schur factorization. This is the approach used in [19], [17], and [20]; we summarize their result in the following theorem.

THEOREM 2.2. *([19, 17, 20]) Suppose $\Lambda_1(s)$ and $\Lambda_2(s)$ are disjoint for all $s \in [0,1]$. Then there is an orthogonal matrix Q and block upper triangular matrix T, each with C^k dependence on s, so that*

$$A(s) = Q(s)T(s)Q(s)^T \tag{2.21}$$

$$= \begin{bmatrix} Q_1(s) & Q_2(s) \end{bmatrix} \begin{bmatrix} T_{11}(s) & T_{12}(s) \\ 0 & T_{22}(s) \end{bmatrix} \begin{bmatrix} Q_1(s) & Q_2(s) \end{bmatrix}^T. \tag{2.22}$$

where $Q_1(s) \in \mathbb{R}^{n\times m}$ is a basis for the subspace $\mathcal{R}(s)$ corresponding to $\Lambda_1(s)$, and $Q_2(s) \in \mathbb{R}^{n\times(n-m)}$ is a basis for $\mathcal{R}(s)^\perp$.

Proof. The proof is written in detail in the cited references, so we only sketch the main ideas here. We differentiate the relation $A = QTQ^T$ to get

$$A' = Q'TQ^T + QTQ'^T + QT'Q^T. \tag{2.23}$$

Because Q is orthogonal, $H = Q^TQ'$ must be skew; by multiplying by Q^T and Q on the left and right, respectively, we have

$$Q^TA'Q = HT - TH + T'. \tag{2.24}$$

Since $T_{21} = T'_{21} = 0$, H_{21} satisfies

$$Q_2^TA'Q_1 = H_{21}T_{11} - T_{22}H_{21} \tag{2.25}$$

The spectra of T_{11} and T_{22} (Λ_1 and Λ_2 respectively) remain disjoint by hypothesis, so there is a unique solution H_{21} for equation (2.25). We specify that $H_{11}(s) = Q_1(s)^TQ'_1(s) = 0$ and $H_{22}(s) = Q_2(s)^TQ'_2(s) = 0$ to get a unique solution for equation (2.23) given an initial factorization $A(0) = Q(0)T(0)Q(0)^T$. Because we have constrained Q_1 and Q_2 to move only horizontally, Q_1 is a minimal arclength basis for $\mathcal{R}$ and Q_2 is a minimal arclength basis for $\mathcal{R}^\perp$. ∎

2.4. Algebraic characterization.

2.4.1. A Riccati equation. Suppose $A \in C^k([0,1], \mathbb{R}^{n\times n})$ and at some $s_0 \in [0,1]$, $\Lambda_1(s_0)$ and $\Lambda_2(s_0)$ are disjoint. Then by the results in previous sections, there is a (non-unique) continuous block Schur decomposition for s near s_0, which at s_0 is

$$A(s_0) = \begin{bmatrix} Q_1(s_0) & Q_2(s_0) \end{bmatrix} \begin{bmatrix} T_{11}(s_0) & T_{12}(s_0) \\ 0 & T_{22}(s_0) \end{bmatrix} \begin{bmatrix} Q_1(s_0) & Q_2(s_0) \end{bmatrix}^T \tag{2.26}$$

where the the spectrum of $T_{ii}(s_0)$ is $\Lambda_i(s_0)$. Sufficiently near s_0, continuity demands that no nonzero vector in $\mathcal{R}(s)$ be orthogonal to $\mathcal{R}(s_0)$, so we may write

$$\mathcal{R}(s) = \operatorname{span}\left(Q(s_0) \begin{bmatrix} I \\ Y(s) \end{bmatrix} \right) \tag{2.27}$$

for some continuous Y with $Y(s_0) = 0$. The function $Y(s)$ must satisfy an algebraic Riccati equation, which we describe in the following lemma.

LEMMA 2.3. *([17, 20]) Let $A \in C^k([0,1], \mathbb{R}^{n\times n})$ have a block Schur decomposition at s_0 as in (2.26), where the diagonal blocks of $T(s_0)$ have disjoint spectra. Define*

$$\widehat{T}(s) = \begin{bmatrix} \widehat{T}_{11}(s) & \widehat{T}_{12}(s) \\ E_{21}(s) & \widehat{T}_{22}(s) \end{bmatrix} := Q(s_0)^T A(s) Q(s_0). \tag{2.28}$$

Then for s near s_0, there is a unique, continuous, minimum-norm solution $Y(s) \in \mathbb{R}^{(n-m)\times m}$ to the Riccati equation

$$F(Y) := \widehat{T}_{22}(s)Y - Y\widehat{T}_{11}(s) + E_{21}(s) - Y\widehat{T}_{12}(s)Y = 0 \tag{2.29}$$

and there is a continuous block Schur decomposition

$$A(s) = Q(s)T(s)Q(s)^T \tag{2.30}$$

where

$$Q(s) = \bar{Q}(s)\left(\bar{Q}(s)^T\bar{Q}(s)\right)^{-1/2} \tag{2.31}$$

$$\bar{Q}(s) = Q(s_0)\begin{bmatrix} I & -Y(s)^T \\ Y(s) & I \end{bmatrix}. \tag{2.32}$$

This theorem is stated in [17] and [20], and extends results proved by Demmel [16], Stewart [38], and Stewart and Sun [39, section V.2]. For completeness, we repeat the proof here.

Proof. We want the matrix $\bar{Q}_1(s)$, which is exactly the matrix used in (2.27), to be a basis for $\mathcal{R}(s)$. To span an invariant subspace, $\bar{Q}_1(s)$ must satisfy the equation

$$A(s)\bar{Q}_1(s) = \bar{Q}_1(s)\bar{T}_{11}(s) \tag{2.33}$$

for some matrix $\bar{T}_{11}(s)$. As we saw in Section 2.1, (2.33) has a continuous set of solutions. To specify a unique solution, we add a normalizing equation:

$$Q_1(s_0)^T\bar{Q}_1(s) = I, \tag{2.34}$$

which implies

$$\bar{Q}_1(s) = Q(s_0)\begin{bmatrix} I \\ Y(s) \end{bmatrix}. \tag{2.35}$$

In order to have $\bar{Q}_1(s_0) = Q_1(s_0)$, we require $Y(s_0) = 0$.

If we multiply (2.33) on the left by $Q(s_0)^T$ and substitute (2.35) for $\bar{Q}_1(s)$ we have

$$Q(s_0)^T A(s) Q(s_0)\begin{bmatrix} I \\ Y(s) \end{bmatrix} = \begin{bmatrix} I \\ Y(s) \end{bmatrix}\bar{T}_{11}(s). \tag{2.36}$$

Now we rewrite $Q(s_0)^T A(s) Q(s_0)$ using (2.28):

$$\begin{bmatrix} \widehat{T}_{11}(s) & \widehat{T}_{12}(s) \\ E_{21}(s) & \widehat{T}_{22}(s) \end{bmatrix}\begin{bmatrix} I \\ Y(s) \end{bmatrix} = \begin{bmatrix} I \\ Y(s) \end{bmatrix}\bar{T}_{11}(s). \tag{2.37}$$

The first row of (2.37) gives us an expression for $\bar{T}_{11}(s)$:

$$\bar{T}_{11}(s) = \widehat{T}_{11}(s) + \widehat{T}_{12}(s)Y(s). \tag{2.38}$$

We substitute into the second row to get

$$E_{21}(s) + \widehat{T}_{22}(s)Y(s) = Y(s)\left(\widehat{T}_{11}(s) + \widehat{T}_{12}(s)Y(s)\right) \tag{2.39}$$

and rearrange terms to get (2.29).

Note that the computed $Q(s)$ is an orthogonal matrix, and is designed so that the leading columns span $\mathcal{R}(s)$. □

2.4.2. A constructive existence proof. Lemma 2.3 says that near s_0 we can write $\mathcal{R}(s)$ in terms of a continuous solution to an algebraic Riccati equation, but it says nothing about the size of the neighborhood or the magnitude of the Riccati solution. To get more detailed information about $Y(s)$, we extend a theorem due to Stewart [38], [39, section V.2].

THEOREM 2.4.

Define $\Omega := C([a,b], \mathbb{R}^{(n-m)\times m})$. For $Y \in \Omega$, we will suppress the argument s to write $\|Y\|$ for the function $s \mapsto \|Y(s)\|$. The norm $\|\cdot\|$ may be any consistent norm. We use $|||Y||| = \max_{s\in[a,b]} \|Y(s)\|$ to denote the norm on Ω.

Let $Y_0 \in \Omega$ be given. Define a Sylvester operator $\mathsf{S} : \Omega \to \Omega$ and a bilinear function $\phi : \Omega \times \Omega \to \Omega$ by

$$\mathsf{S}Z := Z(\hat{T}_{11} + \hat{T}_{12}Y_0) - (\hat{T}_{22} - Y_0\hat{T}_{12})Z \tag{2.40}$$

$$\phi(X,Y) := \mathsf{S}^{-1}(X\hat{T}_{12}Y). \tag{2.41}$$

Suppose S is invertible on $[a,b]$. Then we can define continuous functions $\alpha, \beta : [a,b] \to \mathbb{R}$ by

$$\alpha := \|\mathsf{S}^{-1}(F(Y_0))\| \tag{2.42}$$

$$\beta := \max_{\|X\|=\|Y\|=1} \|\phi(X,Y)\|. \tag{2.43}$$

Suppose also that $4\alpha\beta < 1$ on $[a,b]$, and define

$$\xi_* := \frac{2\alpha}{1+\sqrt{1-4\alpha\beta}}. \tag{2.44}$$

Then there is a unique continuous solution Y_* to the Riccati equation (2.29) such that $\|Y_* - Y_0\| \le \xi_*$.

Proof.

Let $Z := Y - Y_0$. Then we rewrite (2.29) as

$$0 = F(Y_0 + Z) = F(Y_0) - \mathsf{S}(Z + \phi(Z,Z)), \tag{2.45}$$

which we can rearrange to get

$$Z = \mathsf{S}^{-1}\left(F(Y_0)\right) - \phi(Z,Z). \tag{2.46}$$

So the map $\psi : \Omega \to \Omega$ given by

$$\psi(Z) := \mathsf{S}^{-1}(F(Y_0)) - \phi(Z,Z) \tag{2.47}$$

has fixed points where (2.29) has solutions. Now define $\Omega_0 = \{Z \in \Omega : \|Z\| \le \xi_*\}$. We will show that $\psi(\Omega_0) \subseteq \Omega_0$ and ψ is contractive on Ω_0, so by the contraction mapping theorem the iteration $Z_{i+1} = \psi(Z_i)$ will converge to a unique fixed point $Z_* \in \Omega_0$ starting from any $Z_1 \in \Omega_0$.

1. $\psi(\Omega_0) \subseteq \Omega_0$:
   By the definition of α and β,

$$\|\psi(Z)\| \leq \alpha + \beta\|Z\|^2$$

   Define $\tau(\xi) = \alpha + \beta\xi^2$. The quadratic equation $\xi = \tau(\xi)$ has two real solutions when $4\alpha\beta < 1$; the smaller solution is

$$\xi_* = \frac{1 - \sqrt{1 - 4\alpha\beta}}{2\beta} = \frac{2\alpha}{1 + \sqrt{1 - 4\alpha\beta}}$$

   Because $\beta \geq 0$, τ is monotonically nondecreasing for positive arguments. So for $0 \leq \|Z\| \leq \xi_*$,

$$0 \leq \|\psi(Z)\| \leq \tau(\|Z\|) \leq \tau(\xi_*) = \xi_*$$

   So $\psi(\Omega_0) \subset \Omega_0$. Therefore all the iterates Z_i remain in Ω_0.
2. ψ is contractive on Ω_0:
   For any $X, Y \in \Omega_0$,

$$\begin{aligned}
\|\psi(X) - \psi(Y)\| &= \|\phi(X, X) - \phi(Y, Y)\| \\
&= \|\phi(X, X - Y) + \phi(X - Y, Y)\| \\
&\leq \beta\left(\|X\|\|X - Y\| + \|X - Y\|\|Y\|\right) \\
&\leq 2\beta\xi_*\|X - Y\| \\
&= \frac{4\alpha\beta}{1 + \sqrt{1 - 4\alpha\beta}}\|X - Y\| \\
&< 4\alpha\beta\|X - Y\|
\end{aligned}$$

   Let $\gamma := \max_{s\in[a,b]} 4\alpha\beta$; by hypothesis, $\gamma < 1$. Then we have

$$|||\psi(X) - \psi(Y)||| < \gamma\,|||X - Y|||\,.$$

Therefore, ψ has a unique fixed point Z_* in Ω_0; and there is a unique continuous solution $Y_* = Y_0 + Z_*$ to the Riccati equation (2.29) such that $\|Y_* - Y_0\| \leq \xi_*$.
□

The *separation* of matrices B and C is the smallest singular value of the Sylvester operator $\mathsf{A}(X) = BX - XC$:

$$\operatorname{sep}(B, C) := \sigma_{\min}(\mathsf{A}) = \min_{\|X\|_F=1} \|BX - XC\|_F = \frac{1}{\|\mathsf{A}^{-1}\|_2}. \tag{2.48}$$

$\operatorname{sep}(B, C)$ is zero when B and C have a common eigenvalue, and it is small if a small perturbation makes them share an eigenvalue. If B and C are normal, $\operatorname{sep}(B, C)$ is the distance between their spectra, but in general $\operatorname{sep}(B, C)$ may be much smaller, since a small change to a non-normal matrix can cause a relatively large change to the spectrum [40, 41]. By manipulating norm inequalities and using the notion of matrix separation, we can bound the quantities α and β defined in Theorem 2.4.

LEMMA 2.5. *Let α and β be defined as in Theorem 2.4. Then the following inequalities hold pointwise for $s \in [a, b]$:*

$$\|\mathsf{S}\|_2^{-1}\|F(Y_0)\|_F \;\leq\; \alpha \;\leq\; \|\mathsf{S}^{-1}\|_2\|F(Y_0)\|_F \tag{2.49}$$

and

$$\frac{1}{\sqrt{(n-m)m}}\|\mathsf{S}^{-1}\|_2\|\widehat{T}_{12}\|_2 \;\le\; \beta \;\le\; \|\mathsf{S}^{-1}\|_2\|\widehat{T}_{12}\|_2, \tag{2.50}$$

where

$$\|\mathsf{S}^{-1}\|_2 = \frac{1}{\operatorname{sep}(\widehat{T}_{11}+\widehat{T}_{12}Y_0, \widehat{T}_{22}-Y_0\widehat{T}_{12})}. \tag{2.51}$$

Proof. Equation (2.51) is simply the definition of sep, while (2.49) follows from the basic properties of an operator two-norm. To see the upper bound in (2.50), observe that

$$\|\phi(X,Y)\|_F = \|X\widehat{T}_{12}Y\|_F \le \|\widehat{T}_{12}\|_2\|X\|_F\|Y\|_F.$$

To see the lower bound in (2.50), let $X = e_iu^T$ and $Y = ve_j^T$, where u and v are left and right singular vectors for $\sigma_{\max}(\widehat{T}_{12})$ and i and j are chosen to maximize $\|\mathsf{S}^{-1}(e_ie_j^T)\|$. Then $\|X\|_F = \|Y\|_F = 1$, and

$$\begin{aligned}
\beta &\ge \|\mathsf{S}^{-1}\left((e_iu^T)\widehat{T}_{12}(ve_j)\right)\|_F && (2.52)\\
&= \|\widehat{T}_{12}\|_2\|\mathsf{S}^{-1}(e_ie_j^T)\|_F && (2.53)\\
&\ge \frac{1}{\sqrt{(n-m)m}}\|\mathsf{S}^{-1}\|_2\|\widehat{T}_{12}\|_2 && (2.54)
\end{aligned}$$

We can see the last inequality by viewing $\mathsf{S}^{-1}(e_ie_j)^T$ as the column of greatest norm from S^{-1} when S^{-1} is viewed in Kronecker product form as an $(n-m)m$-by-$(n-m)m$ matrix. ∎

The bounds (2.49) and (2.50) together with Theorem 2.4 yield the following theorem.

THEOREM 2.6. *([17, 20]) Let $Y_0 : [a,b] \to \mathbb{R}^{(n-m)\times m}$ be continuous, and define*

$$\kappa(\widehat{T}) := \frac{\|\widehat{T}_{12}\|_2\,\|F(Y_0)\|_F}{\operatorname{sep}^2(\widehat{T}_{11}+\widehat{T}_{12}Y_0, \widehat{T}_{22}-Y_0\widehat{T}_{12})} \tag{2.55}$$

In any neighborhood in which $\kappa(\widehat{T}) < 1/4$, the Riccati equation (2.29) has a unique continuous solution $Y_*(s)$ such that

$$\|Y_*\|_F < \frac{2\|F(Y_0)\|_F}{\operatorname{sep}(\widehat{T}_{11}+\widehat{T}_{12}Y_0, \widehat{T}_{22}-Y_0\widehat{T}_{12})}. \tag{2.56}$$

In any neighborhood where $\kappa(\widehat{T}) < 1/12$, Newton's method on the Riccati equation (2.29) will converge quadratically pointwise to Y_* starting from Y_0.

Proof.

To prove the existence statement, substitute (2.49) and (2.50) into Theorem 2.4.

In [16], Demmel proved the quadratic convergence of Newton's iteration for $\kappa(\widehat{T}) < 1/12$. To extend to the case of a parameter-dependent matrix, we simply apply Demmel's theorem pointwise.

∎

2.5. Connecting subspaces. Suppose we are given bases for invariant subspaces of $A(s)$ at $s = 0$ and $s = h$. How can we check that the two end points are connected by a continuously defined invariant subspace basis on $[0, h]$? This question has practical significance for our continuation algorithm, since we would like to avoid branch-jumping behavior when two subspaces come close to each other, and we would like to detect when a continued invariant subspace ceases to be continuously defined.

Theorem 2.6 partially answers the question of how to check for a continuous connecting invariant subspace. But to apply the theorem, we need to bound $\kappa(\widehat{T})$ on the interval $[0, h]$. In the remainder of this section, we describe how to construct bounds which incorporate information from both $s = 0$ and $s = h$ using interpolation. Our ultimate goal is Theorem 2.12, but first we need some technical lemmas.

We first turn to the problem of bounding $\|B^{-1}\|_2$, where $B \in C^1([0, h], \mathbb{R}^{p\times p})$ is some parameterized operator on a Euclidean space. Since S is also a linear operator on a Euclidean space ($\mathbb{R}^{(n-m)\times m}$ with the Frobenius inner product), all our results apply directly to S as well. We begin by reviewing a simple result about matrix interpolation.

LEMMA 2.7. *Suppose $B \in C^1([0, h], \mathbb{R}^{p\times p})$ and B' is Lipschitz with constant M. Then*

$$B(s) = B(0) + B[0, h]s + B[0, h, s]s(s - h) \tag{2.57}$$

where $B[0, h]$ and $B[0, h, s]$ are first and second Newton divided differences and

$$\begin{aligned} \|B[0, h]\|_2 &\le \max_{\xi\in[0,h]} \|B'(\xi)\|_2 \\ \|B[0, h, s]\|_2 &\le M. \end{aligned}$$

Proof.

For any $u, v \in \mathbb{R}^p$ and any distinct $a, b \in [0, h]$, $a < b$, the mean value theorem applied to the scalar function $u^T B(s)v$ implies

$$u^T B[a, b]v = u^T B(\xi)v \tag{2.58}$$

for some $\xi \in [a, b]$. Therefore, $\|B[a, b]\|_2 \le \max_{\xi\in[a,b]} \|B(\xi)\|_2$.

Now we compute

$$\begin{aligned} u^T B[0, h, s]v &= (u^T B[0, s]v - u^T B[h, s]v)/h && (2.59)\\ &= (u^T B'(\xi_1)v - u^T B'(\xi_2)v)/h && (2.60)\\ &\le \frac{\|B'(\xi_1) - B'(\xi_2)\|_2}{h}\|u\|_2\|v\|_2 && (2.61)\\ &\le M\|u\|_2\|v\|_2. && (2.62) \end{aligned}$$

So $\|B[0, h, s]\|_2 \le M$.

□

We can now show a very simple bound on the minimal singular value of B.

LEMMA 2.8. *Suppose $B \in C^1([0, h], \mathbb{R}^{p\times p})$ and B' is Lipschitz with constant M. Then*

$$\sigma_{\min}(B(s)) \ge \sigma_{\min}(B(0)) - \|B[0, h]\|_2 s - Ms(h - s) \tag{2.63}$$

Proof. By the previous lemma,

$$\|B(s) - B(0)\|_2 = \|B[0,h]s + B[0,h,s]s(s-h)\| \leq \|B[0,h]\|_2 s + Ms(s-h).$$

To complete the proof, recall (e.g. from [28]) that

$$|\sigma_{\min}(B(s)) - \sigma_{\min}(B(0))| \leq \|B(s) - B(0)\|_2.$$

□

Lemma 2.8 uses only the norm of $B(s) - B(0)$; we can refine the bound by using the direction as well as the magnitude.

LEMMA 2.9. *Suppose $B \in C^1([0,h], \mathbb{R}^{p\times p})$ and B' is Lipschitz with constant M. Then*

$$\sigma_{\min}(B(s)) \geq \sigma_{\min}(B(0))(1 - \|B(0)^{-1}B[0,h]\|_2 s) - Ms(h-s) \tag{2.64}$$

Proof. Let $E(s) = B[0,h]s$. If $\|B(0)^{-1}E(s)\|_2 \geq 1$, then the lemma is trivial. Otherwise, $I + B(0)^{-1}E(s)$ is invertible, and

$$(B(0) + E(s))^{-1} = \left(I + B(0)^{-1}E(s)\right)^{-1} B(0)^{-1} \tag{2.65}$$

$$= \sum_{k=0}^{\infty} \left(-B(0)^{-1}E(s)\right)^k B(0)^{-1} \tag{2.66}$$

so

$$\left\|(B(0) + E(s))^{-1}\right\|_2 \leq \frac{\|B(0)^{-1}\|_2}{1 - \|B(0)^{-1}E(s)\|_2}. \tag{2.67}$$

Taking inverses on both sides, we have

$$\sigma_{\min}(B(0) + E(s)) \geq \sigma_{\min}(B(0))(1 - \|B(0)^{-1}E(s)\|). \tag{2.68}$$

Therefore

$$\sigma_{\min}(B(s)) = \sigma_{\min}(B(0) + B[0,s]s + B[0,h,s]s(s-h)) \tag{2.69}$$

$$\geq \sigma_{\min}(B(0) + B[0,s]s) - Ms(h-s) \tag{2.70}$$

$$\geq \sigma_{\min}(B(0))(1 - \|B(0)^{-1}B[0,h]\|_2 s) - Ms(h-s) \tag{2.71}$$

□

We now turn to the problem of bounding $\|F(Y_0)\|_F$ in (2.29) for a specific choice of Y_0. Suppose $\begin{bmatrix} I \\ hZ \end{bmatrix}$ is a basis for a given invariant subspace of $\widehat{T}(h)$ (see(2.28)); then we linearly interpolate $Y_0(s) = sZ$, so that the residual $F(Y_0)$ is zero at both $s = 0$ and $s = h$.

LEMMA 2.10. *Suppose $\widehat{T} \in C^1$ and $\widehat{T}'$ has Lipschitz constant M. Also suppose $\begin{bmatrix} I \\ hZ \end{bmatrix}$ spans an invariant subspace of $\widehat{T}(h)$, and define*

$$G(s) := \widehat{T}_{22}[0,h]Z - Z\widehat{T}_{11}[0,h] - Z\left(\widehat{T}_{12}(0) + (s+h)\widehat{T}_{12}[0,h]\right)Z. \tag{2.72}$$

Then for $Y_0(s) = sZ$, and for any $s \in [0, h]$,

$$\|F(Y_0)\|_F \leq \frac{h^2}{2}\left\{\max(\|G(0)\|_F, \|G(h)\|_F) + \sqrt{m}M(1 + h\|Z\|_2)^2\right\} \tag{2.73}$$

Proof.
We write $F(Y_0(s))$ as the product

$$F(Y_0(s)) = \begin{bmatrix}-Y_0(s) & I\end{bmatrix}\widehat{T}(s)\begin{bmatrix}I \\ Y_0(s)\end{bmatrix}. = \begin{bmatrix}-sZ & I\end{bmatrix}\widehat{T}(s)\begin{bmatrix}I \\ sZ\end{bmatrix}. \tag{2.74}$$

Using the Newton form of the interpolant,

$$\widehat{T}(s) = \widehat{T}(0) + \widehat{T}[0, h]s + \widehat{T}[0, h, s]s(s - h); \tag{2.75}$$

we can therefore write $F(Y_0(s))$ as

$$F(Y_0(s)) = F_1(Y_0(s)) + F_2(Y_0(s)) \tag{2.76}$$

$$F_1(Y_0(s)) = \begin{bmatrix}-sZ & I\end{bmatrix}\left(\widehat{T}(0) + \widehat{T}[0, h]s\right)\begin{bmatrix}I \\ sZ\end{bmatrix} \tag{2.77}$$

$$F_2(Y_0(s)) = \begin{bmatrix}-sZ & I\end{bmatrix}\left(\widehat{T}[0, h, s]s(s - h)\right)\begin{bmatrix}I \\ sZ\end{bmatrix}. \tag{2.78}$$

We now bound the norms of $F_1(Y_0(s))$ and $F_2(Y_0(s))$ independently.

To bound $F_1(Y_0(s))$, we expand and collect terms at each order in s:

$$\begin{aligned} F_1(Y_0(s)) = {} & E_{21}(0) \\ & + s\left(\widehat{T}_{22}(0)Z - Z\widehat{T}_{11}(0) + E_{21}[0, h]\right) \\ & + s^2\left(\widehat{T}_{22}[0, h]Z - Z\widehat{T}_{11}[0, h] - Z\widehat{T}_{12}(0)Z\right) \\ & + s^3\left(-Z\widehat{T}_{12}[0, h]Z\right) \end{aligned} \tag{2.79, 2.80}$$

Since $F(Y_0(s))|_{s=0} = 0$, we know $E_{21}(0) = 0$. Similarly, since $F(Y_0(s))|_{s=h} = 0$, we know

$$\begin{aligned} & \widehat{T}_{22}(0)Z - Z\widehat{T}_{11}(0) + E_{21}[0, h] \\ & = -h\left(\widehat{T}_{22}[0, h]Z - Z\widehat{T}_{11}[0, h] - Z\widehat{T}_{12}(0)Z\right) \\ & \quad -h^2\left(-Z\widehat{T}_{12}[0, h]Z\right). \end{aligned} \tag{2.81}$$

Substituting (2.81) into (2.80), we have

$$\begin{aligned} F_1(Y_0(s)) = {} & (s^2 - sh)\left(\widehat{T}_{22}[0, h]Z - Z\widehat{T}_{11}[0, h] - Z\widehat{T}_{12}(0)Z\right) + \\ & (s^3 - sh^2)\left(-Z\widehat{T}_{12}[0, h]Z\right). \end{aligned} \tag{2.82}$$

Factoring out $s(s - h)$ from both terms, we have

$$F_1(Y_0(s)) = s(s - h)G(s). \tag{2.83}$$

Note that $G(s)$ is linear, so by convexity of norms,

$$\|G(s)\|_F \leq \max\left(\|G(0)\|_F, \|G(h)\|_F\right) \text{ for } s \in [0,h]. \tag{2.84}$$

Therefore

$$\|F_1(Y_0(s))\|_F \leq \frac{h^2}{2}\max\left(\|G(0)\|_F, \|G(h)\|_F\right) \text{ for } s \in [0,h]. \tag{2.85}$$

We use a cruder bound for $F_2(Y_0(s))$. Since $F_2(Y_0(s)) \in \mathbb{R}^{(n-m)\times m}$, $\|F_2(Y_0(s))\|_F \leq \sqrt{m}\|F_2(Y_0(s))\|_2$. Both $\begin{bmatrix}-sZ & I\end{bmatrix}$ and $\begin{bmatrix}I\\ hZ\end{bmatrix}$ are bounded in 2-norm by $1+h\|Z\|_2$; and by 2.7, $\|\widehat{T}[0,h,s]\| \leq M$. Therefore

$$\|F_2(Y_0(s))\|_2 \leq \left\|\begin{bmatrix}-sZ & I\end{bmatrix}\right\|_2 \left\|\widehat{T}[0,h,s]\right\|_2 \left\|\begin{bmatrix}I\\ sZ\end{bmatrix}\right\|_2 s(s-h) \tag{2.86}$$

$$\leq \frac{h^2}{2}M(1+h\|Z\|_2)^2. \tag{2.87}$$

Substituting the above bounds into $\|F(Y_0(s))\|_F \leq \|F_1(Y_0(s))\|_F + \|F_2(Y_0(s))\|_F$ concludes the proof.
□

Now we bound $\|\widehat{T}_{12}(s)\|_2$ on $[0,h]$.

LEMMA 2.11. *Suppose $\widehat{T} \in C^1$ and $\widehat{T}'$ has Lipschitz constant M. Then for $s \in [0,h]$,*

$$\|\widehat{T}_{12}(s)\|_2 \leq \max\left(\|\widehat{T}_{12}(0)\|_2, \|\widehat{T}_{12}(h)\|_2\right) + \frac{1}{2}Ms(h-s) \tag{2.88}$$

Proof. By Lemma 2.7,

$$\|T_{12}(s)\|_2 = \|T_{12}(0) + T_{12}[0,h]s + T_{12}[0,h,s]s(s-h)\|_2 \tag{2.89}$$

$$\leq \|T_{12}(0) + T_{12}[0,h]s\|_2 + Ms(h-s), \tag{2.90}$$

and because norms are convex functions,

$$\|T_{12}(0) + T_{12}[0,h]s\|_2 \leq \max\left(\|T_{12}(0)\|_2, \|T_{12}(h)\|_2\right). \tag{2.91}$$

□

Putting together the preceding bounds, we have the following theorem.

THEOREM 2.12. *Suppose $\widehat{T}(s)$ is C^2 and $\widehat{T}'$ is Lipschitz with constant M. Suppose $\begin{bmatrix}I\\ 0\end{bmatrix}$ and $\begin{bmatrix}I\\ hZ\end{bmatrix}$ span invariant subspaces at 0 and h respectively. Let S be defined as in (2.40). Then if*

$$\sigma_{\min}(\mathsf{S}(0))(1 - h\|\mathsf{S}(0)^{-1}\mathsf{S}[0,h]\|_2) - \frac{1}{2}Mh^2 > 0 \tag{2.92}$$

the operator S is invertible for all $s \in [0,h]$. Further, the constants α and β defined

in (2.42) and (2.43) are bounded for all $s \in [0,h]$ by

$$\alpha \leq \frac{h^2}{2} \frac{\max\left(\|G(0)\|_F, \|G(h)\|_F\right) + \sqrt{m}M(1+h\|Z\|_2)^2}{\sigma_{\min}(\mathsf{S}(0))(1-h\|\mathsf{S}(0)^{-1}\mathsf{S}[0,h]\|_2) - \frac{1}{2}Mh^2} \tag{2.93}$$

$$= \frac{h^2}{2} \frac{\max\left(\|G(0)\|_F, \|G(h)\|_F\right) + \sqrt{m}M}{\sigma_{\min}(\mathsf{S}(0))} + O(h^3) \tag{2.94}$$

$$\beta \leq \frac{\max\left(\|\widehat{T}_{12}(0)\|_2, \|\widehat{T}_{12}(h)\|_2\right) + \frac{1}{2}Mh^2}{\sigma_{\min}(\mathsf{S}(0))(1-h\|\mathsf{S}(0)^{-1}\mathsf{S}[0,h]\|_2) - \frac{1}{2}Mh^2} \tag{2.95}$$

$$= \frac{\max\left(\|\widehat{T}_{12}(0)\|_2, \|\widehat{T}_{12}(h)\|_2\right)}{\sigma_{\min}(\mathsf{S}(0))} + O(h) \tag{2.96}$$

where

$$G(s) = \widehat{T}_{22}[0,h]Z - Z\widehat{T}_{11}[0,h] - Z\left(\widehat{T}_{12}(0) + (s+h)\widehat{T}_{12}[0,h]\right)Z.$$

Therefore, by Theorem 2.4, if the resulting upper bound on $4\alpha\beta$ is bounded below one, there is a continuous connecting invariant subspace between $\begin{bmatrix} I \\ 0 \end{bmatrix}$ at $s = 0$ and $\begin{bmatrix} I \\ hZ \end{bmatrix}$ at $s = h$.

Dropping higher-order terms, we have

$$\alpha \leq \frac{h^2}{2} \frac{\max\left(\|G(0)\|_F, \|G(h)\|_F\right) + \sqrt{m}M}{\sigma_{\min}(\mathsf{S}(0))} + O(h^3) \tag{2.97}$$

$$\beta \leq \frac{\max\left(\|\widehat{T}_{12}(0)\|_2, \|\widehat{T}_{12}(h)\|_2\right)}{\sigma_{\min}(\mathsf{S}(0))} + O(h) \tag{2.98}$$

Besides $\mathrm{sep}(\widehat{T}_{11}(0), \widehat{T}_{22}(0)) = \sigma_{\min}(\mathsf{S}(0))$ and $\|\mathsf{S}(0)^{-1}\mathsf{S}[0,h]\|_2$, the quantities in the bounds of the above theorem are cheap and simple to compute.

3. The CIS algorithm: direct methods. We now describe the CIS algorithm in the case when we can use direct solvers. Much of this work is described in [17], [20], [25], and [26]. Here, we emphasize parts of the computation that we perform differently, or which are particularly relevant to the sparse case.

At the highest level, our algorithm is as follows:

1. Choose an initial invariant subspace.
2. Compute a continuation step.
3. Normalize the solution.
4. Adapt the space and step size to improve convergence and resolve features of interest.

We can continue either $Q_1(s)$ and $T_{11}(s)$ or the full $Q(s)$ and $T(s)$ matrices. Currently, our dense code computes the full Schur factors at each step. When we continue only the first part of the decomposition, as we do in the sparse case, we also compute a few extra eigenvalues from $\Lambda_2(s)$. We use these eigenvalues to decide whether the algorithm should be reinitialized with a different partitioning of the spectrum.

3.1. Initialization. To initialize the algorithm at s_0, we compute a Schur decomposition of $A(s_0)$ and use standard LAPACK routines [2] to sort the decomposition so selected eigenvalues appear in $T_{11}(s_0)$. For bifurcation problems, we assume

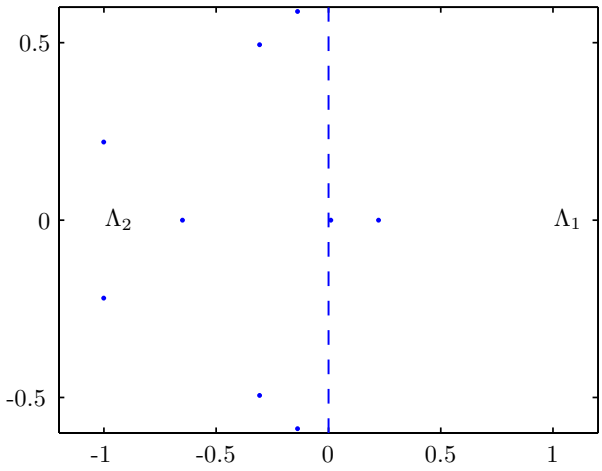

FIG. 3.1. *Selected eigenvalues during initialization*

that only a small part of the spectrum is unstable; therefore, we include in our m-dimensional subspace vectors corresponding to all the unstable eigenvalues as well as a few stable eigenvalues nearest the imaginary axis (see Figure 3.1).

We require that $\Lambda_1(s_0)$ contains any unstable eigenvalues and some specified number of stable eigenvalues; but we may include additional eigenvalues in order to simplify the subsequent continuation process. For example, we include an extra eigenvalue in order to avoid splitting a complex conjugate pair of eigenvalues between $\Lambda_1(s_0)$ and $\Lambda_2(s_0)$. More generally, we would like to choose $\Lambda_1(s_0)$ so that the gap between the real parts of the leftmost eigenvalue in $\Lambda_1(s_0)$ and the rightmost eigenvalue in $\Lambda_2(s_0)$ are greater than some threshold. In this way, we hope to keep track of all eigenvalues that might cross the imaginary axis.

In the dense case, the same LAPACK routine used to sort the Schur form also estimates the sensitivity of the selected subspace, and so we may choose a larger subspace if the smallest feasible subspace is very sensitive. Though the cost of the computations at a single point increases as we increase the size of our subspace, continuing a less sensitive subspace will allow us to take larger steps.

We summarize the initialization procedure in Algorithm 1.

3.2. Choosing a subspace. We have considered three strategies for computing $\mathcal{R}(s_1)$ starting from $\mathcal{R}(s_0)$:

- As in the construction of Theorem 2.6, apply a predictor and then use a Newton corrector.
- Choose a subspace which minimizes the distance between eigenvalues in the computed $\Lambda(s_1)$ and eigenvalues in $\Lambda(s_0)$.
- Choose a subspace by finding the m eigenvectors of $A(s_1)$ which most nearly lie in $\mathcal{R}(s_0)$, or which most nearly lie in a predicted subspace.

We currently use an approximate Euler predictor and a Newton corrector. We use the convergence of the corrector to govern our step size: if it converges slowly or fails to converge, we reduce the step size, or reinitialize the continuation process with a larger or smaller subspace. If the corrector converges quickly, we increase the step size.

Algorithm 1 Choose an initial subspace

Input: $A(s_0)$,
$n_{\min}$, $n_{\max}$, {bounds on subspace size}
$n_{\text{stableref}}$, {number of stable reference eigenvalues}
ϵ_{gap}, {minimum gap between $\Lambda_1(s_0)$ and $\Lambda_2(s_0)$}

Output: $Q_1(s_0)$ and $T_{11}(s_0)$

Compute a Schur decomposition $A(s_0) = QTQ^T$
$t :=$ real parts of converged eigenvalues sorted in descending order
Find smallest m so that $\begin{cases} n_{\min} \le m \le n_{\max} \\ m \ge (\#\text{ unstable eigenvalues}) + n_{\text{stableref}} \\ t(m) - t(m+1) > \epsilon_{\text{gap}} \end{cases}$

if no such m exists **then**
  error "Spectrum too tightly clustered"
else
  Sort subspace for rightmost m eigenvalues to the front of Q, T
  Return $Q_1 = Q(:, 1:m)$, $T_{11} = T(1:m, 1:m)$
end if

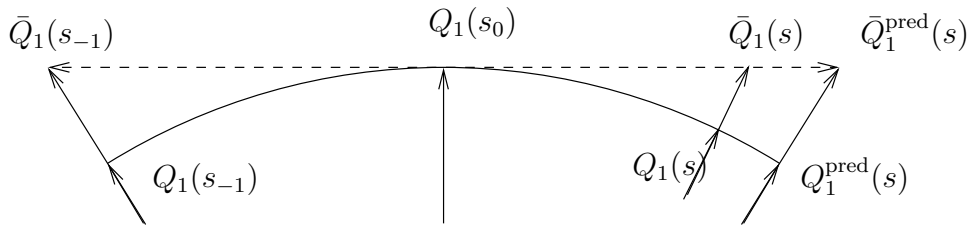

FIG. 3.2. *Choosing a consistent normalization for secant prediction*

3.2.1. Subspace predictors. We build an Euler predictor for $\mathcal{R}(s_1)$ by differentiating the Schur factorization as in (2.23) and substituting finite difference approximations for Q' and T'. Alternatively, we could differentiate the Riccati equation (2.29) and substitute a finite difference approximation for Y'. Either way, this gives us the equation

$$T_{22}(s_0)Y_0(s_1) - Y_0(s_1)T_{11}(s_0) = -(s_1 - s_0)E'_{21}(s_1) \tag{3.1}$$

If derivatives of A are unavailable, we can substitute a finite difference approximation for $E'_{21}(s)$ to get the approximate Euler predictor equation

$$T_{22}(s_0)Y_0(s_1) - Y_0(s_1)T_{11}(s_0) = -E_{21}(s_1). \tag{3.2}$$

We can also build a secant predictor; but to do so, we must consider how consecutive steps are normalized. In a single predictor-corrector step, we normalize the basis for a space $\mathcal{X}$ by requiring that $Q(s_0)^T X = I$; however, this normalization changes with each step. If $\mathcal{R}(s_{-1})$ is the invariant subspace from a previous continuation step, we must choose a basis $\bar{Q}_1(s_{-1})$ for $\mathcal{R}(s_{-1})$ which is consistent with the current normalization (see Figure 3.2). Because $\bar{Q}_1(s_{-1})$ spans the same space as $Q_1(s_{-1})$, there must be some invertible $B(s_{-1}) \in \mathbb{R}^{m \times m}$ such that

$$\bar{Q}_1(s_{-1}) = Q_1(s_{-1})B(s_{-1}), \tag{3.3}$$

and the normalizing condition is

$$I = Q_1(s_0)^T \bar{Q}_1(s_{-1}) = Q_1(s_0)^T Q_1(s_{-1}) B(s_{-1}). \tag{3.4}$$

Therefore

$$B(s_{-1}) = \left(Q_1(s_0)^T Q_1(s_{-1})\right)^{-1} \tag{3.5}$$

$$\bar{Q}_1(s_{-1}) = Q_1(s_{-1}) \left(Q_1(s_0)^T Q_1(s_{-1})\right)^{-1}. \tag{3.6}$$

By linear extrapolation, the secant predictor for $\bar{Q}_1(s_1)$ is

$$\bar{Q}_1^{\text{pred}}(s_1) = Q_1(s_0) + \frac{s_1 - s_0}{s_0 - s_{-1}} \left(Q_1(s_0) - \bar{Q}_1(s_{-1})\right) \tag{3.7}$$

The Riccati unknown has the form $Y(s) = Q_2(s_0)^T \bar{Q}_1(s)$ with $Y(s_0) = 0$, so we can rewrite the predictor (3.7) as

$$Y_0(s_1) = -\frac{s_1 - s_0}{s_0 - s_{-1}} Y(s_{-1}), \tag{3.8}$$

where

$$Y(s_{-1}) = Q_2(s_0)^T \bar{Q}_1(s_{-1}). \tag{3.9}$$

We similarly write higher-order polynomial predictors by choosing a consistent normalization for several steps and using polynomial extrapolation.

3.2.2. Direct Newton corrector iterations. One way to find $\bar{Q}_1(s)$ is to simultaneously solve residual equations for the the eigensystem and the normalization:

$$R = \begin{bmatrix} A(s)\bar{Q}_1(s_1) - \bar{Q}_1(s_1)\bar{T}_{11}(s_1) \\ Q_1(s_0)^T \bar{Q}_1(s_1) - I \end{bmatrix} = 0 \tag{3.10}$$

We can compute a Newton step for (3.10) using a bordered Bartels-Stewart algorithm [6]. Alternately, we can eliminate $\bar{T}_{11}(s_1)$ and perform Newton iteration on the Riccati equation (2.29). A Newton step for the Riccati equation can be solved using an ordinary Bartels-Stewart algorithm [28, p. 367].

Newton iterations on the reduced and unreduced systems are equivalent in exact arithmetic, assuming that the initial iterate in the unreduced case satisfies the normalization condition $Q_1(s_0)^T \bar{Q}_1^{\text{pred}}(s_1) = I$. However, while reducing (3.10) to a Riccati equation reduces the problem size by a modest amount, the reduced system will usually be dense, even if (3.10) is sparse. For small problems, we use dense methods, and the loss of sparsity matters little; for large problems, we sidestep the issue by using projection methods, as described in Section 4. For medium-sized problems, it may be better to use sparse direct solvers to take Newton steps on the unreduced system of equations.

3.3. Normalizing the solution. After we compute a basis $\bar{Q}_1(s_1)$ for $\mathcal{R}(s_1)$, we normalize to find another basis $Q_1(s_1)$ which is as near as possible to $Q_1(s_0)$. This normalization approximates the minimal arclength condition described in Section 2.1. We describe several ways to write the normalization in the following lemma.

LEMMA 3.1. *Let $\bar{Q}_1(s_1)$ be a basis for $\mathcal{R}(s_1)$ with $Q_1(s_0)^T \bar{Q}_1(s_1) = I$. Let $\bar{Q}_1(s_1) = U\Sigma V^T$ be a singular value decomposition with $U \in \mathbb{R}^{n\times m}$ and $\Sigma, V \in \mathbb{R}^{m\times m}$,*

and let $Y(s_1) = Q_2(s_0)\bar{Q}_1(s_1)$. *Then the orthonormal basis* $Q_1(s_1) \in \mathrm{Stief}(n,m)$ *for* $\mathcal{R}(s_1)$ *which minimizes* $\|Q_1(s_1) - Q_1(s_0)\|_F$ *can be written in the following ways:*

$$Q_1(s_1) = UV^T \tag{3.11}$$

$$Q_1(s_1) = \bar{Q}_1(s_1)\left(\bar{Q}_1(s_1)^T\bar{Q}_1(s_1)\right)^{-1/2} \tag{3.12}$$

$$Q_1(s_1) = Q_1(s_0)\begin{bmatrix} I \\ Y(s_1)\end{bmatrix}\left(I + Y(s_1)^TY(s_1)\right)^{-1/2}. \tag{3.13}$$

Proof. If $\bar{Q}_1(s_1) = U\Sigma V^T$, then one orthonormal basis for $\mathcal{R}(s_1)$ is UV^T. We can write any other orthonormal basis for $\mathcal{R}(s_1)$ as UV^TW for some orthogonal matrix $W \in O(m)$.

Now we solve an orthogonal Procrustes problem ([28, p. 582]) to find W corresponding to the orthonormal basis nearest Q_0. Choose W to minimize

$$\|Q_1(s_0) - UV^TW\|_F^2. \tag{3.14}$$

Because the Frobenius norm is invariant under unitary transformations, we have

$$\begin{aligned}
&\left\|Q_1(s_0) - UV^TW\right\|_F^2 \\
&= \left\|Q(s_0)\left(\begin{bmatrix} I_m \\ 0\end{bmatrix} - \begin{bmatrix} Q_1(s_0)^TUV^TW \\ Q_2(s_0)^TUV^TW\end{bmatrix}\right)\right\|_F^2 \\
&= \left\|\begin{bmatrix} I_m - Q_1(s_0)^TUV^TW \\ -Q_2(s_0)^TUV^TW\end{bmatrix}\right\|_F^2
\end{aligned}$$

and by the Pythagorean theorem,

$$\begin{aligned}
\left\|Q_1(s_0) - UV^TW\right\|_F^2 = \left\|I_m - Q_1(s_0)^TUV^TW\right\|_F^2 + \\
\left\|-Q_2(s_0)^TUV^TW\right\|_F^2.
\end{aligned}$$

The second term of the sum does not depend on W, since W is orthogonal. Therefore, we minimize $\left\|Q_1(s_0) - UV^TW\right\|_F^2$ by minimizing

$$\left\|I_m - Q_1(s_0)^TUV^TW\right\|_F^2 \tag{3.15}$$

By hypothesis,

$$I = Q_1(s_0)^T\bar{Q}_1(s_0) = Q_1(s_0)^TU\Sigma V^T. \tag{3.16}$$

If we substitute (3.16) into (3.15) and use the unitary invariance of the Frobenius norm yet again, we have

$$\begin{aligned}
\left\|I_m - Q_1(s_0)^TUV^TW\right\|_F^2 &= \left\|Q_1(s_0)^TU\left(\Sigma - V^TWV\right)V^T\right\|_F^2 \\
&= \left\|\Sigma - V^TWV\right\|_F^2
\end{aligned}$$

The matrix V^TWV is orthogonal, and the closest orthogonal matrix to the positive diagonal matrix Σ is the identity. Therefore, (3.15) is minimized when $V^TWV = I$. Thus $\|Q_1(s_0) - UV^TW\|_F^2$ is minimized for $W = I$, and so $Q_1(s_1) = UV^T$. This proves (3.11).

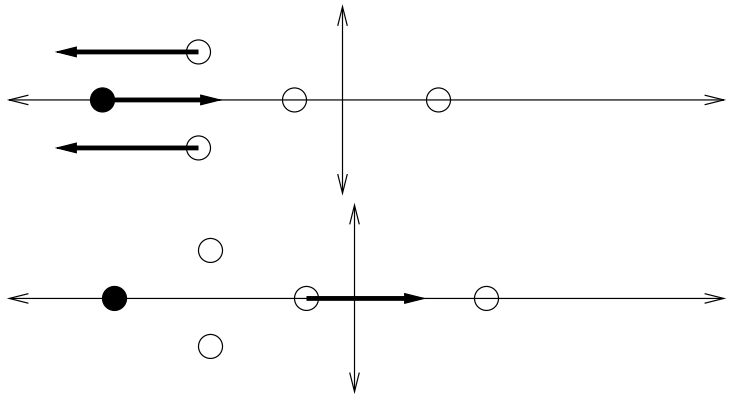

FIG. 3.3. *Examples of overlap and bifurcation. In the top example (overlap), one of the eigenvalues from $\Lambda_1(s)$ (open circles) changes position with one of the eigenvalues of $\Lambda_2(s)$. In the bottom example, an eigenvalue crosses over the imaginary axis (a bifurcation), so that $\Lambda_1(s)$ contains fewer stable eigenvalues.*

To show (3.12), we write

$$\begin{aligned}
\bar{Q}_1(s_1)\left(\bar{Q}_1(s_1)^T\bar{Q}_1(s_1)\right)^{-1/2} &= U\Sigma V^T\left(V\Sigma U^T U\Sigma V^T\right)^{-1/2} \\
&= U\Sigma V^T\left(V\Sigma^2 V^T\right)^{-1/2} \\
&= U\Sigma V^T V\left(\Sigma^2\right)^{-1/2} V^T \\
&= UV^T \\
&= Q_1(s_1)
\end{aligned}$$

If we write

$$\bar{Q}_1(s_1) = Q(s_0)\begin{bmatrix} I \\ Y(s_1) \end{bmatrix}, \tag{3.17}$$

then

$$\begin{aligned}
\bar{Q}_1(s_1)^T\bar{Q}_1(s) &= \begin{bmatrix} I \\ Y(s_1) \end{bmatrix}^T Q(s_0)^T Q(s_0)\begin{bmatrix} I \\ Y(s_1) \end{bmatrix} \\
&= \left(I + Y(s_1)^T Y(s_1)\right)
\end{aligned} \tag{3.18}$$

Now substitute (3.17) and (3.18) into (3.12) to get (3.13).
□

3.4. Subspace analysis and adaptation.

3.4.1. Bifurcations and overlaps. When the CIS algorithm is initialized, the set $\Lambda_1(s_0)$ contains all the unstable eigenvalues of $A(s_0)$ and a few of the stable eigenvalues nearest the imaginary axis. The set $\Lambda_2(s_0)$ lies strictly left of $\Lambda_1(s_0)$ in the complex plane. During continuation, eigenvalues from $\Lambda_1(s)$ may cross the imaginary axis (a bifurcation), or $\Lambda_2(s)$ may cease to lie strictly to the left of $\Lambda_1(s)$ (an overlap). These situations are illustrated in Figure 3.3. When bifurcation or overlap occurs, we reinitialize the continuation procedure.

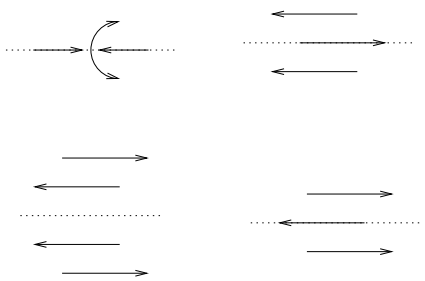

FIG. 3.4. *Generic overlap situations. On the left, two real eigenvalues collide and produce a complex pair (top), and the real parts of two complex conjugate eigenvalue pairs change order (bottom). On the right, a complex conjugate pair and a real eigenvalue change places in two ways.*

A *generic* overlap or bifurcation is one which persists when the function $A(s)$ is perturbed. For steady-state continuation problems, the only generic bifurcations are fold bifurcations, in which an isolated real eigenvalue crosses the imaginary axis; and Hopf bifurcations, in which an isolated complex conjugate pair of eigenvalues crosses the imaginary axis. There are four generic types of overlap (see Figure 3.4). In three cases, a single real eigenvalue or complex conjugate pair from $\Lambda_2(s)$ moves right of some element of $\Lambda_1(s)$. In the fourth case, a single eigenvalue from $\Lambda_2(s)$ collides with an eigenvalue from $\Lambda_1(s)$ to form a complex conjugate pair. $Q_1(s)$ corresponding to $\Lambda_1(s)$ will cease to be continuously defined, and we expect that the Newton iteration will not converge. Complex conjugate eigenvalues in the spectrum may also generically collide and become real eigenvalues, but because we do not allow complex conjugate pairs to be split between $\Lambda_1(s)$ and $\Lambda_2(s)$, this behavior does not result in an overlap.

3.4.2. Step size and subspace adaptation. Standard bifurcation analysis algorithms [30] involve computing functions of $A(s)$. We adapt these methods to large problems by computing the same functions of the much smaller $T_{11}(s)$. Therefore, we try to ensure that only eigenvalues from $\Lambda_1(s)$ can cross the imaginary axis, so that $T_{11}(s)$ will provide all the relevant information about bifurcations. To prevent eigenvalues from $\Lambda_2(s)$ from crossing the imaginary axis, we adapt the step size and the size of the $\Lambda_1(s)$ so that overlaps and bifurcations are not allowed in the same step. We summarize the step size and subspace adaptation logic in Algorithm 2.

When an overlap occurs because two real eigenvalues collide to form a conjugate pair, the Newton iteration will fail to converge. To detect other types of overlap at s, we compute the overlap set:

$$\{(\lambda_i(s), \lambda_j(s)) \in \Lambda_1(s) \times \Lambda_2(s) : \mathrm{Re}(\lambda_i(s)) < \mathrm{Re}(\lambda_j(s))\}.$$

If this set is non-empty, then an overlap has occurred. To decide whether multiple overlaps have occurred, we count the number of $(\lambda_i(s), \lambda_j(s))$ pairs in the overlap set. To avoid double-counting overlaps involving complex conjugate pairs, we only count the pairs such that $\mathrm{Im}(\lambda_i(s)) \leq 0$ and $\mathrm{Im}(\lambda_j(s)) \leq 0$.

Only one overlap is allowed in a step. If we detect multiple overlaps, we retry with a smaller step size until only one overlap is left. If we reach the minimum step size and still have multiple overlaps when continuing from s_i, we reinitialize the continuation process at s_i so that the overlap set from the failed step belongs entirely to $\Lambda_1(s_i)$ or entirely to $\Lambda_2(s_i)$.

We detect bifurcations by counting the unstable eigenvalues. If the total number of unstable eigenvalues at s_{i+1} differs from the total number of unstable eigenvalues at

s_i, then a bifurcation occurred during the step. If this total number changed by more than one real eigenvalue or one complex conjugate eigenpair, we assume that multiple bifurcations have occurred, and we try to resolve them by decreasing the step size. If we cannot resolve the behavior with the minimum step size, then the algorithm fails with a diagnostic message. Unless we fail or a bifurcation and an overlap both occur during the step, we assume that $\Lambda_1(s)$ contains all information about bifurcations.

If an overlap or bifurcation occurs in an accepted step from s_i to s_{i+1}, we will reinitialize the computation at s_{i+1} before attempting another step. This way, the new spectral sets will not overlap, and the new $\Lambda_1(s_{i+1})$ will include no more or fewer eigenvalues than necessary after a bifurcation.

Algorithm 2 Continue and adapt invariant subspace of $A(s)$

Input:	$A(s)$	{matrix-valued function}
	s_0,	{starting parameter}
	h_{initial},	{starting step size}
	$h_{\min}$, $h_{\max}$	{bounds on the step size}
Output:	$Q(s)$ and $T(s)$	

```
Compute initial point Q(s_0), T(s_0) using Algorithm 1.
s := s_0, h := h_initial
while not done do
  Compute a candidate step and candidate step size ĥ
  Test for bifurcation and overlap
  if subspace did not converge then
    Reinitialize at s using Algorithm 1
    Reset step size to h_initial
  else if multiple overlap, multiple bifurcation, or overlap and bifurcation then
    if h > h_min then
      Decrease h
    else if multiple bifurcation then
      error "Could not resolve nongeneric bifurcation"
    else
      Reinitialize at s using Algorithm 1
    end if
  else
    Record the decomposition and diagnostic information
    s := s + h, h := min(h_max, ĥ)
    if bifurcation or overlap occurred in accepted step then
      Reinitialize at s using Algorithm 1
      h := h_initial
    end if
  end if
end while
```

4. The CIS algorithm: projection methods. We now turn to the case when the dimension n of $A(s)$ is large and we are interested in a space $\mathcal{R}(s)$ of dimension $m \ll n$. In this case, direct methods are expensive; however, if we can multiply by $A(s)$ quickly, we can use projection methods.

4.1. Choosing a projection space. In the direct case, we consider two spectral sets: $\Lambda_1(s)$, which contains the unstable eigenvalues and a few of the rightmost stable eigenvalues; and $\Lambda_2(s)$, which contains the remaining eigenvalues. In the projection case, we consider three spectral sets: $\Lambda_1(s)$, a set of m elements which contains the unstable eigenvalues and a few of the rightmost stable eigenvalues; $\Lambda_2(s)$, a set of $p-m$ elements which contains a few of the rightmost eigenvalues not in $\Lambda_1(s)$; and $\Lambda_3(s)$, a set of $n-p$ elements which contains the remainder of the spectrum. Our basic strategy in the projected CIS algorithm is to build a projection space $\mathcal{V}$ of dimension p, where $m < p \ll n$, so that the restriction of $A(s)$ to $\mathcal{V}$ provides good approximations to $\Lambda_1(s)$ and $\Lambda_2(s)$.

4.2. Initialization. During initialization, we may not know how large $\mathcal{V}$ must be to find all the unstable eigenvalues plus a few stable eigenvalues. Therefore, the projected version of the initialization routine calls Algorithm 1 in a loop. While not enough stable eigenvalues converge or there are no sufficiently large gaps between stable eigenvalues in the converged part spectrum, more eigenvalues are requested. If a suitable subspace cannot be found when a specified maximum number of eigenvalues are requested, the code exits with a diagnostic message.

4.3. Projected normalization and residual equations. Suppose $V \in \mathbb{R}^{p\times n}$ is an orthonormal basis for a projection space $\mathcal{V}$. Recall the n-by-m residual equation (2.33)

$$A(s)\bar{Q}_1(s) - \bar{Q}_1(s)\bar{T}_{11}(s) = 0.$$

We approximate the equation by assuming that $\bar{Q}_1(s) \approx \bar{Q}_1^h(s) := V\hat{Q}_1(s)$ and choosing $\bar{Q}_1^h(s)$ to satisfy the Galerkin condition

$$\begin{aligned} 0 &= V^T\left(A(s)\bar{Q}_1^h(s) - \bar{Q}_1^h(s)\bar{T}_{11}^h(s)\right) && (4.1)\\ &= V^T A(s) V\hat{Q}_1(s) - \hat{Q}_1(s)\bar{T}_{11}^h(s) && (4.2)\end{aligned}$$

We assume the same normalizing condition we used before:

$$Q_1(s_0)^T\bar{Q}_1^h(s) = \left(V^TQ_1(s_0)\right)^T\hat{Q}_1(s) = I \qquad (4.3)$$

Once $\bar{Q}_1^h(s)$ has been computed, we can use Lemma 3.1 to compute the orthonormal basis $Q_1^h(s)$ for the same space which is closest to $Q_1(s_0)$ in the Frobenius norm. We will let $Q_2^h(s) \in \mathbb{R}^{n\times(p-m)}$ be an orthonormal basis for the orthogonal complement of $\mathrm{span}(Q_1^h(s))$ in $\mathcal{V}$. Though we require continuity of $Q_1^h(s)$, it will not be important for our purposes to continuously define $Q_2^h(s)$.

We typically will use a projection space $\mathcal{V}$ which is itself an approximate maximal invariant subspace computed by an Arnoldi method. Suppose that $A(s_1)\mathcal{V} \subset \mathcal{V}$, and let $V^\perp \in \mathbb{R}^{n\times(n-p)}$ be an orthonormal basis for $\mathcal{V}^\perp$. Then at s_1, solutions to the Galerkin equation (4.2) span invariant subspaces of $A(s_1)$.

If $\mathcal{V}$ is a p-dimensional maximal invariant subspace corresponding to the rightmost part of the spectrum of $A(s_1)$, then we compute the leading two-by-two part of a three-by-three block Schur form

$$\begin{aligned} A(s_1) = &\begin{bmatrix} Q_1^h(s_1) & Q_2^h(s_1) & V^\perp\end{bmatrix}\\ &\begin{bmatrix} T_{11}^h(s_1) & T_{12}^h(s_1) & T_{13}^h(s_1)\\ 0 & T_{22}^h(s_1) & T_{23}^h(s_1)\\ 0 & 0 & T_{33}^h(s_1)\end{bmatrix}\\ &\begin{bmatrix} Q_1^h(s_1) & Q_2^h(s_1) & V^\perp\end{bmatrix}^T. \end{aligned}$$

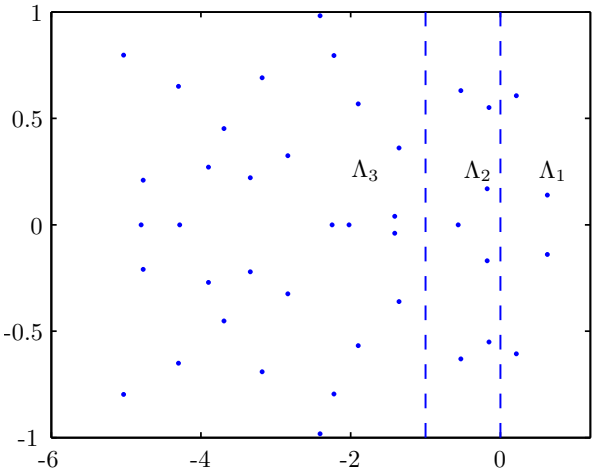

FIG. 4.1. *Eigenvalue sets in the projected CIS algorithm. In practice, Λ_3 will contain many more eigenvalues than Λ_1 and Λ_2.*

The spectrum of the $T_{11}^h(s)$ block is the continued set of eigenvalues $\Lambda_1(s)$. The $T_{22}^h(s)$ block has a few of the rightmost remaining eigenvalues, which we use to diagnose overlap. The eigenvalues of the uncomputed block $T_{33}^h(s)$ are part of the spectrum which lies further from the imaginary axis. Figure 4.1 illustrates the three spectral sets corresponding to $T_{11}^h(s)$, $T_{22}^h(s)$, and $T_{33}^h(s)$ in the case when no overlap has occurred.

As in the dense case, we can eliminate $\bar{T}_{11}^h(s)$ from equation (4.2); we summarize this calculation in the following lemma.

LEMMA 4.1. *Let $V^T Q_1(s_0)$ have the singular value decomposition*

$$V^T Q_1(s_0) = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} R^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} R^T \tag{4.4}$$

where $U \in \mathbb{R}^{p\times p}$, $\Sigma \in \mathbb{R}^{m\times m}$, and $R \in \mathbb{R}^{m\times m}$. Let

$$\widehat{T}^h(s) = \begin{bmatrix} \widehat{T}_{11}^h(s) & \widehat{T}_{12}^h(s) \\ E_{21}^h(s) & \widehat{T}_{22}^h(s) \end{bmatrix} := \begin{bmatrix} \Sigma & 0 \\ 0 & I \end{bmatrix} U^T V^T A(s) V U \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I \end{bmatrix}. \tag{4.5}$$

Then any solution to the Galerkin equation (4.2) and normalizing condition (4.3) can be written as

$$\hat{Q}_1(s) = U \begin{bmatrix} \Sigma^{-1} \\ \hat{Y}^h(s) \end{bmatrix} R^T \tag{4.6}$$

where $\hat{Y}^h(s) \in \mathbb{R}^{(p-m)\times m}$ is a solution to the Riccati equation

$$F^h(Y^h(s)) := \widehat{T}_{22}^h(s) Y^h(s) - Y^h(s) \widehat{T}_{11}^h(s) + E_{21}^h(s) - Y^h(s) \widehat{T}_{12}^h(s) Y^h(s) = 0. \tag{4.7}$$

Proof. Let $B(s) = U^T \hat{Q}_1^h(s) R$. Substituting the SVD (4.4) into (4.3), we have

$$I = R \begin{bmatrix} \Sigma & 0 \end{bmatrix} U^T \hat{Q}_1(s) \tag{4.8}$$

$$= R \begin{bmatrix} \Sigma & 0 \end{bmatrix} B(s) R^T \tag{4.9}$$

If we multiply on the left by R^T and on the right by R, we have

$$I = \begin{bmatrix} \Sigma & 0 \end{bmatrix} B(s). \tag{4.10}$$

Therefore, for some $Y^h(s) \in \mathbb{R}^{(p-m)\times m}$, $B(s)$ can be written as

$$B(s) = \begin{bmatrix} \Sigma^{-1} \\ Y^h(s) \end{bmatrix} = \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I \\ Y^h(s) \end{bmatrix}. \tag{4.11}$$

Now we substitute $\hat{Q}_1^h(s) = UB(s)R^T$ into the projected residual equation (4.2):

$$V^T A(s) V U \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I \\ Y^h(s) \end{bmatrix} R^T - U \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I \\ Y^h(s) \end{bmatrix} R^T \bar{T}_{11}^h(s) = 0. \tag{4.12}$$

If we multiply by $\begin{bmatrix} \Sigma & 0 \\ 0 & I \end{bmatrix} U^T$ on the left and by R on the right, we have

$$\widehat{T}^h(s) \begin{bmatrix} I \\ Y^h(s) \end{bmatrix} = \begin{bmatrix} I \\ Y^h(s) \end{bmatrix} \left(R^T \bar{T}_{11}^h(s) R \right). \tag{4.13}$$

The first row of (4.13) gives an expression for $R^T \bar{T}_{11}^h(s) R$, which we can substitute into the second row to get the Riccati equation (4.7):

$$\begin{aligned} R^T \bar{T}_{11}^h(s) R &= \widehat{T}_{11}^h(s) + \widehat{T}_{12}^h(s) Y^h(s) \\ E_{21}^h(s) + \widehat{T}_{22}^h(s) Y^h(s) &= Y^h(s) \left(R^T \bar{T}_{11}^h(s) R \right) \\ &= Y^h(s) \widehat{T}_{11}^h(s) + Y^h(s) \widehat{T}_{12}^h(s) Y^h(s). \end{aligned}$$

□

In Theorem 2.3, we saw that for s sufficiently near s_0, the normalized basis for $\mathcal{R}(s)$ corresponded to the minimum norm solution for the Riccati equation (2.29). The norm of the Riccati unknown $Y(s)$ is equal to the distance $\|\bar{Q}_1(s) - Q_1(s_0)\|_F$. We now show that $\|Y^h(s)\|_F$ is similarly related to $\|\bar{Q}_1^h(s) - Q_1(s_0)\|_F$.

LEMMA 4.2. *In the previous lemma, the distance from $\bar{Q}_1^h$ to $Q_1(s_0)$ is*

$$\|\bar{Q}_1^h(s) - Q_1(s_0)\|_F^2 = \|Y^h\|_F^2 + \|\Sigma^{-1}\|_F^2 - m. \tag{4.14}$$

Proof.

We decompose $Q_1(s_0)$ and $\bar{Q}_1^h(s)$ into components in three orthogonal spaces spanned by $V^\perp$, VU_1, and VU_2:

$$Q_1(s_0) = V^\perp (V^\perp)^T Q_1(s_0) + VU_1 \Sigma R^T \tag{4.15}$$

$$Q_1^h(s) = VU_1 \Sigma^{-1} R^T + VU_2 Y^h(s) R^T \tag{4.16}$$

where the first equation is a consequence of (4.4) and the second equation follows from (4.6). The difference is

$$Q_1(s_0) - Q_1^h(s) = \left\{ \begin{array}{l} V^\perp (V^\perp)^T Q_1(s_0) + \\ VU_1(\Sigma - \Sigma^{-1}) R^T + \\ VU_2 Y_h(s) R^T \end{array} \right\}. \tag{4.17}$$

Because the three components are orthogonal, the squared Frobenius norm is the sum of the squares of the Frobenius norms; that is

$$\|Q_1(s_0) - Q_1^h(s)\|_F^2 = \left\{ \begin{array}{l} \|V^\perp (V^\perp)^T Q_1(s_0)\|_F^2 + \\ \|VU_1(\Sigma - \Sigma^{-1})R^T\|_F^2 + \\ \|VU_2 Y_h(s) R^T\|_F^2 \end{array} \right\}. \tag{4.18}$$

Because multiplication by an orthonormal matrix does not change the Frobenius norm, we can write

$$\|Q_1(s_0) - Q_1^h(s)\|_F^2 = \left\{ \begin{array}{l} \|(V^\perp)^T Q_1(s_0)\|_F^2 + \\ \|\Sigma - \Sigma^{-1}\|_F^2 + \\ \|Y_h(s)\|_F^2 \end{array} \right\} \tag{4.19}$$

$$= \left\{ \begin{array}{l} \|(V^\perp)^T Q_1(s_0)\|_F^2 + \\ (\|\Sigma\|_F^2 + \|\Sigma^{-1}|_F^2 - 2m) + \\ \|Y_h(s)\|_F^2 \end{array} \right\}. \tag{4.20}$$

Note that

$$m = \|Q_1(s_0)\|_F^2 = \|(V^\perp)^T Q_1(s_0)\|_F^2 + \|V^T Q_1(s_0)\|_F^2 \tag{4.21}$$

$$= \|(V^\perp)^T Q_1(s_0)\|_F^2 + \|\Sigma\|_F^2. \tag{4.22}$$

Now substitute

$$\|(V^\perp)^T Q_1(s_0)\|_F^2 = m - \|\Sigma\|_F^2 \tag{4.23}$$

into (4.20) to obtain the desired result.
□

Therefore, if s_1 is sufficiently near s_0 and $\mathcal{V}$ is itself an invariant subspace of $A(s_1)$ such that $\mathcal{R}(s_1) \subset \mathcal{V}$, the minimal norm solution to the projected Riccati equation (4.7) corresponds exactly to the minimal norm solution to the Riccati equation (2.29).

4.4. Projected predictors and correctors. The Euler predictor (3.1) and the finite difference version of the Euler predictor (3.2) are subtly different in the projected case. A projection subspace $\mathcal{V}$ which is an invariant subspace for $A(s_1)$ will generally not contain $\mathcal{R}(s_0)$; consequently, $Q_1(s_0)$ will not correspond to a solution to the projected Riccati equation (4.7) at $s = s_0$. Worse, $E_{21}^h(s_0)$ will usually be nonzero. If we naively differentiate the relation $F^h(Y^h(s)) = 0$ and use the resulting differential equation to form an Euler-like approximation $Y_0^h(s_1)$ starting from a value of 0 for $Y^h(s_0)$, then to first order $F^h(Y_0^h(s_1))$ will be $E_{21}(s_0)$.

We can remedy this problem by requiring $\mathcal{R}(s_0) \subset \mathcal{V}$. However, a more straightforward alternative is to compute a secant prediction $\bar{Q}_1^{\text{pred}}(s_1)$ using (3.7), and then project

$$\bar{Q}_1^{h,\text{pred}}(s_1) = VV^T \bar{Q}_1^{\text{pred}}(s_1). \tag{4.24}$$

The corresponding projected Riccati predictor is then

$$Y_0^h(s_1) = U_2^T V^T \bar{Q}_1^{\text{pred}}(s_1) R \tag{4.25}$$

In the current code, we use the trivial predictor $Y_0^h(s_1) = 0$.

Once we have a predicted value $Y_0^h(s_1)$, we solve the projected Riccati equation with a Newton iteration, just as we did in the direct methods. We note that the projected matrix $V^T A(s) V$ will usually be dense, and so there seems to be little benefit to solving the unreduced equations. Just as in the direct case, alternate subspace selection methods based on eigenvalues and eigenvectors are possible.

5. Integrating the CIS algorithm into *MATCONT*. In the introduction, we described how invariant subspace continuation can be used to adapt bifurcation analysis methods for small problems in order to analyze much larger systems. In this section, we discuss one example of our work to use the CIS algorithm in this way to extend the bifurcation analysis code MATCONT [18]: using projected test functions to detect and locate Hopf bifurcations.

5.1. Detecting and locating Hopf bifurcations. Let $x(s) = (u(s), \alpha(s)) \in \mathbb{R}^n \times \mathbb{R}$ be a smooth local parameterization of a solution branch of the stationary problem (1.2):

$$f(x(s)) = f(u(s), \alpha(s)) = 0.$$

We write the Jacobian matrix along this path as $A(s) := f_u(x(s))$. A solution point $x(s_0)$ is a *bifurcation point* if $\operatorname{Re}\lambda_i(s_0) = 0$ for at least one eigenvalue $\lambda_i(s_0)$ of $A(s_0)$. The point $x(s_0)$ is a *simple Hopf bifurcation* if the simple eigenvalue $\lambda_i(s_0)$ is a pure imaginary number and $\operatorname{Re}\left(\frac{d\lambda_i}{ds}(s_0)\right) \neq 0$.

A *test function* $\psi(x(s))$ is a (typically) smooth scalar function that has a regular zero at a bifurcation point. A bifurcation point between consecutive continuation points $x(s_k)$ and $x(s_{k+1})$ is *detected* when

$$\psi(x(s_k))\psi(x(s_{k+1})) < 0. \tag{5.1}$$

Once a bifurcation point has been detected, it can be *located* by solving a system of the form

$$\begin{cases} f(x) = 0, \\ g(x) = 0 \end{cases} \tag{5.2}$$

where g may be ψ or may be some other function which has a regular zero at the bifurcation point.

To detect Hopf points, MATCONT uses the test function

$$\psi(x(s)) := \det\left[2A(s) \odot I_n\right] = \prod_{i<j} \left(\lambda_i(s) + \lambda_j(s)\right), \tag{5.3}$$

where $\odot$ is the bialternate product [30]. Using the projection computed from the CIS algorithm, we introduce the analogous test function

$$\hat{\psi}(x(s)) := \det\left[2T_{11}(s) \odot I_m\right] = \prod_{i<j\leq m} \left(\lambda_i(s) + \lambda_j(s)\right). \tag{5.4}$$

Clearly, $\psi(x(s))$ and $\hat{\psi}(x(s))$ are zero if $A(s)$ has a pure imaginary pair of eigenvalues $(\pm i\kappa)$, and so ψ and $\hat{\psi}$ can be used to test for Hopf bifurcations. However, these functions may also be zero because of a pair of real eigenvalues which sum to zero. Therefore, we also introduce a parity function which counts the number of unstable complex conjugate pairs:

$$\chi(x(s)) = (-1)^{\#\left\{\lambda_i(s):\operatorname{Re}\lambda_i(s)\geq 0 \text{ and } \operatorname{Im}\lambda_i(s)>0\right\}}. \tag{5.5}$$

We detect a Hopf bifurcation when

$$\hat{\psi}(x(s_k))\hat{\psi}(x(s_{k+1})) < 0 \text{ and } \chi(x(s_k))\chi(x(s_{k+1})) < 0. \tag{5.6}$$

A well-known method to locate a Hopf point (see e.g. e.g. [33, 30, 5]) is to solve the system

$$\begin{cases} f(x) = 0, \\ f_u(x)r - i\omega r = 0, \\ r^* r_0 - 1 = 0 \end{cases} \tag{5.7}$$

where $x \in \mathbb{R}^{n+1}$, $r \in \mathbb{C}^n$, and $\omega \in \mathbb{R}$. The reference vector $r_0 \in \mathbb{C}^n$ is given. Usually, the system (5.7) is converted to a system of $3n + 2$ real unknowns. Based on the CIS algorithm, we replace (5.7) with the system

$$\begin{cases} f(x) = 0, \\ T_{11}(x)r - i\omega r = 0, \\ r^* r_0 - 1 = 0 \end{cases} \tag{5.8}$$

where r and r_0 are now vectors in $\mathbb{C}^m$. In contrast to (5.7), the system (5.8) involves $n + 2m + 2$ real unknowns.

5.2. The one-dimensional Brusselator. The *1D Brusselator* [34] is a well known model system for autocatalytic chemical reactions with diffusion. The problem is defined on $\Omega = (0, 1)$ by coupled differential equations for unknowns u and v

$$\frac{d_1}{l^2}u'' - (b+1)u + u^2 v + a = 0$$

$$\frac{d_2}{l^2}v'' + bu - u^2 v = 0$$

with boundary conditions

$$u(0) = u(1) = a \text{ and } v(0) = v(1) = \frac{b}{a}. \tag{5.9}$$

This problem exhibits a rich bifurcation scenario and has been used in the literature as a standard model for bifurcation analysis [36, 14, 15, 4, 12, 35]. Utilizing a second-order finite difference discretization

$$f'' \approx \frac{1}{h^2}(f_{i-1} - 2f_i + f_{i+1})$$

with $h = (N + 1)^{-1}$, the resulting discrete problem can be written in the form (1.2). This discretization of the Brusselator is used in a MATCONT example [18].

In order to verify the accuracy of locating a Hopf point, we continue a constant solution branch: $u(x) = a$, $v(x) = \frac{b}{a}$, with respect to b. In this case the values of b where Hopf bifurcation occurs are known analytically as a function of N, see e.g. [12, Eq. (24)]. Using MATLAB 7.0 on a 1.67 GHz G4, we located this bifurcation to at least eight correct digits for problems with $N = 1024$ to $N = 8192$ grid points; since there are two unknowns per grid point, the total size is $n = 2N$. Because these problems have only one-dimensional connectivity, the Jacobian may be reordered into a very narrowly banded form, and so the size to solve a linear system involving the Jacobian scales linearly with N. For each problem size, about 65% of the time was spent on spectrally transformed Arnoldi iterations using ARPACK; 12% of the time was spent on solving bordered systems for the corrector during continuation and for the Newton steps; and 7% of the time was spent on forming the Jacobian matrix. The cost of one Newton step for locating the bifurcation was approximately the same as the cost of one Newton step during the continuation, and to locate each bifurcation took three Newton steps. At $N = 8192$, the total time for fifteen steps of continuation and for locating one Hopf bifurcation was 158 seconds.

6. Conclusions and Future Work. In this paper, we have discussed the CIS algorithm for computing a smooth orthonormal basis for an invariant subspace of a parameter-dependent matrix, and we have extended it to make it more suitable for numerical bifurcation analysis. In particular, we have made the following contributions:

1. We have derived new sufficient conditions for the existence of a continuous invariant subspace connecting invariant subspaces of matrices at the end of a parameterized matrix curve.
2. We have extended the original CIS algorithm for dense problems with logic for adapting the continued subspace in order to ensure that it always includes information relevant to bifurcation analysis. Such adaptation is necessary when an bifurcation occurs or when there is an *overlap*: that is, when the real parts of eigenvalues change order.
3. We have extended our algorithm to work efficiently on large sparse matrices by exploiting Galerkin projection methods. The original CIS algorithm used direct methods for dense matrices, and so cost $O(n^3)$ work at each step.
4. We have incorporated the projection-based CIS algorithm into the MATCONT bifurcation analysis package, and we have applied the combined code to the Brusselator model problem.

Future work includes the following topics. We are still actively investigating how the information can most effectively be used for finding bifurcations from non-static equilibria, and how to best use the CIS algorithm in detecting and computing codimension-2 bifurcations along branches of Hopf and limit points. We are also involved in using of the CIS algorithm in order to study the dependence of resonant frequencies of mechanical devices as design parameters are varied.

REFERENCES

[1] P.-A. Absil, R. Sepulchre, P. V. Dooren, and R. Mahony, *Cubically convergent iterations for invariant subspace computation*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 70–96.

[2] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, third ed., 1999.

[3] P. Ashwin, K. Böhmer, and Z. Mei, *A numerical Liapunov-Schmidt method with applications to Hopf bifurcation on a square*, Math. Comp., 64 (1995), pp. 649–670.

[4] P. Ashwin and Z. Mei, *A Hopf bifurcation with Robin boundary conditions*, J. Dynamics and Differential Equations, 6 (1994), pp. 487–503.

[5] W.-J. Beyn, A. Champneys, E. J. Doedel, Y. A. Kuznetsov, B. Sandstede, and W. Govaerts, *Numerical continuation and computation of normal forms*, in Handbook of Dynamical Systems III: Towards Applications, B. Fiedler, ed., Elsevier, 2001, ch. 4.

[6] W.-J. Beyn, W. Kless, and V. Thümmler, *Continuation of low-dimensional invariant subspaces in dynamical systems of large dimension*, in Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems, B. Fiedler, ed., Springer, 2001, pp. 47–72.

[7] D. Bindel, J. Demmel, and M. Friedman, *Continuation of invariant subspaces for large bifurcation problems*, in Proceedings of the SIAM Conference on Linear Algebra, Williamsburg, VA, 2003.

[8] J. Bosec, *Continuation of Invariant Subspaces in Bifurcation Problems*, PhD thesis, University of Marburg, 2002.

[9] J. H. Brandts, *The Riccati method for eigenvalues and invariant subspaces of matrices with inexpensive action*, Linear Algebra and its Applications, 358 (2003), pp. 335–365.

[10] E. A. Burroughs, R. B. Lehoucq, L. A. Romero, and A. J. Salinger, *Linear stability of flow in a differentially heated cavity via large-scale eigenvalue calculations*, Tech. Rep. SAND2002-3036J, Sandia National Laboratories, 2002.

[11] C. S. Chien and M. H. Chen, *Multiple bifurcations in a reaction-diffusion problem*, Computers Math. Applic., 35 (1998), pp. 15–39.

[12] C. S. Chien, Z. Mei, and C. L. Shen, *Numerical continuation at double bifurcation points of a reaction-diffusion problem*, Int. J. Bifur. and Chaos, 8 (1997), pp. 117–139.
[13] K. A. Cliffe, A. Spence, and S. J. Tavener, *The numerical analysis of bifurcations problems with application to fluid mechanics*, Acta Numerica, (2000), pp. 1–93.
[14] M. Colubitsky and D. G. Schaeffer, *Singularities and groups in bifurcation theory, Vol 1*, Springer-Verlag, 1985.
[15] G. Dangelmayr, *Degenerate bifurcations near a double eigenvalue in the Brusselator*, J. Austral. Math. Soc. Ser. B, 28 (1987), pp. 486–535.
[16] J. W. Demmel, *Three methods for refining estimates of invariant subspaces*, Computing, 38 (1987), pp. 43–57.
[17] J. W. Demmel, L. Dieci, and M. J. Freidman, *Computing connecting orbits via an improved algorithm for continuing invariant subspaces*, SIAM J. Sci. Comput., 22 (2001), pp. 81–94.
[18] A. Dhooge, W. Govaerts, Y. Kuznetsov, W. Mestrom, and A. M. Riet, *MATLAB continuation software package CL_MATCONT*, Jan. 2003. `http://www.math.uu.nl/people/kuznet/cm/`.
[19] L. Dieci and T. Eirola, *On smooth decompositions of matrices*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 800–819.
[20] L. Dieci and M. J. Friedman, *Continuation of invariant subspaces*, Numerical Linear Algebra Applications, 8 (2001), pp. 317–327.
[21] L. Dieci and A. Papini, *Point-to-periodic and periodic-to-periodic connections*. Preprint, 2003.
[22] L. Dieci and A. J. Rebaza, *Continuation of eigendecompositions*, To appear in FGCS: Future Generation Computer Systems – Computational Science, (2002).
[23] E. J. Doedel and H. Sharifi, *Collocation methods for continuation problems in nonlinear elliptic PDEs, issue on continuation*, 74 (2000), pp. 105–118.
[24] A. Edelman, T. Arias, and S. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353.
[25] M. J. Friedman, *Improved detection of bifurcations in large nonlinear systems via the Continuation of Invariant Subspaces algorithm*, Int. J. Bif. and Chaos, 11 (2001), pp. 2277–2285.
[26] M. J. Friedman and M. E. Jackson, *An improved RLV stability analysis via a continuation approach*, tech. rep., NASA Marshall Space Flight Center, 2002.
[27] K. Georg, *Matrix-free numerical continuation and bifurcation*, Numerical Functional Analysis and Optimization, 22 (2001), pp. 303–320.
[28] G. H. Golub and C. F. V. Loan, *Matrix Computations*, The John Hopkins University Press, 1989.
[29] W. Govaerts, *Computation of singularities in large nonlinear systems*, SIAM J. Numer. Anal., 34 (1997), pp. 867–880.
[30] ———, *Numerical methods for bifurcations of dynamical equilibria*, SIAM Publications, Philadelphia, 2000.
[31] W. Govaerts, J. Guckenheimer, and A. Khibnik, *Defining functions for multiple Hopf bifurcations*, SIAM J. Numer. Anal, 34 (1997), pp. 1269–1288.
[32] T. Kato:1995:PTL, *Perturbation Theory for Linear Operators*, Springer-Verlag, corrected printing of the second edition ed., 1995.
[33] Y. A. Kuznetsov, *Elements of Applied Bifurcation Theory, Second edition*, Springer-Verlag, New York, 1998.
[34] R. Lefever and I. Prigogine, *Symmetry-breaking instabilities in dissipative systems II*, J. Chem. Phys., 48 (1968), pp. 1695–1700.
[35] Z. Mei, *Numerical bifurcation analysis for reaction-diffusion equations*, PhD thesis, University of Marburg, 1997.
[36] D. Schaeffer and M. Golubitsky, *Bifurcation analysis near a double eigenvalue of a model chemical reaction*, Arch. Rational Mech. Anal., 75 (1981), pp. 315–347.
[37] G. M. Shroff and H. B. Keller, *Stabilization of unstable procedures: The recursive projection method*, SIAM J. Numer. Anal, 30 (1993), pp. 1099–1120.
[38] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Review, 4 (1973), pp. 727–764.
[39] G. W. Stewart and J. guang Sun, *Matrix Perturbation Theory*, Academic Press, San Diego, CA, 1990.
[40] L. N. Trefethen, *Pseudospectra of matrices*, in Numerical Analysis 1991, D. F. Griffiths and G. A. Watson, eds., Longman Scientific and Technical, Harlow, Essex, UK, 1991, pp. 234–262.
[41] J. M. Varah, *On the separation of two matrices*, SIAM J. Num. Anal., (1979), pp. 212–222.