SIAM J. NUMER. ANAL.
Vol. 25, No. 6, December 1988

012

ISOSPECTRAL FLOWS AND ABSTRACT MATRIX FACTORIZATIONS*

MOODY T. CHU† AND LARRY K. NORRIS†

Abstract. A general framework for constructing isospectral flows in the space gl(n) of n by n matrices is proposed. Depending upon how gl(n) is split, this framework gives rise to different types of abstract matrix factorizations. When sampled at integer times, these flows naturally define special iterative processes, and each flow is associated with the sequence generated by the corresponding abstract factorization. The proposed theory unifies as special cases the well-known matrix decomposition techniques used in numerical linear algebra and is likely to offer a broader approach to the general matrix factorization problem.

Key words. isospectral flows, matrix factorization, projections, abstract decompositions

AMS(MOS) subject classification. 65F15

1. Introduction.

Let $\mathrm{Gl}(n) = \mathrm{Gl}\,(n, K)$ denote the *Lie group* of all n by n nonsingular matrices over the field K. K may be either the real or complex numbers. The tangent space to the n^2-dimensional manifold $\mathrm{Gl}\,(n)$ at the identity is the *Lie algebra* $\mathrm{gl}\,(n)$ of all n by n matrices. Given a matrix $X_0 \in \mathrm{gl}\,(n)$, we are interested in studying the dynamical behavior of curves on the surface $\mathcal{M}(X_0)$ in $\mathrm{gl}\,(n)$ defined by

$$\mathcal{M}(X_0) := \{g^{-1}X_0 g \mid g \in \mathrm{Gl}\,(n)\}. \tag{1.1}$$

Let $g(t)$ be a one-parameter family of matrices in $\mathrm{Gl}\,(n)$ with $g(0) = I$. Then

$$X(t) := g^{-1}(t)X_0 g(t) \tag{1.2}$$

defines a differentiable one-parameter family of matrices, with $X(0) = X_0$, on the surface $\mathcal{M}(X_0)$. The derivative $dX(t)/dt$ can be expressed as

$$\frac{dX(t)}{dt} = [X(t), k(t)], \tag{1.3}$$

where the n by n matrices $k(t)$ are defined by

$$k(t) := g^{-1}(t)\frac{dg(t)}{dt}, \tag{1.4}$$

and the bilinear operator $[\cdot, \cdot]$ is the *Lie bracket* defined by

$$[A, B] := AB - BA. \tag{1.5}$$

Therefore, the initial value problem (IVP)

$$\frac{dX(t)}{dt} = [X(t), k(t)], \qquad X(0) = X_0 \tag{1.6}$$

with $k(t)$ a continuous one-parameter family of matrices in $\mathrm{gl}\,(n)$ characterizes a differentiable curve on $\mathcal{M}(X_0)$. We note that the above argument can be reversed. That is, given an arbitrary one-parameter family of matrices $k(t)$ in $\mathrm{gl}\,(n)$, then the solution of (1.6) will be of the form (1.2) for some differentiable $g(t)$ satisfying

$$\frac{dg(t)}{dt} = g(t)k(t), \qquad g(0) = I. \tag{1.7}$$

---

* Received by the editors March 2, 1987; accepted for publication (in revised form) November 18, 1987.
† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695-8205.

Solutions of IVPs (1.6) with different choices of $k(t)$ are closely related to many important numerical techniques used in linear algebra. Some recent results and further references can be found in the review papers by Chu [2] and Watkins [12]. For example, one of the most efficient ways of solving the linear algebraic eigenvalue problem

$$Ax = \lambda x \tag{1.8}$$

uses the so-called *QR algorithm* [5], [9]. Suppose $A_0 = A \in \mathrm{gl}\,(n, \mathbb{R})$. The unshifted QR algorithm generates a sequence of orthogonally similar matrices $\{A_k\}$ from the scheme

$$A_k = Q_k R_k, \quad A_{k+1} = R_k Q_k, \quad k = 0, 1, \cdots \tag{1.9}$$

where each Q_k represents an orthogonal matrix and R_k an upper triangular matrix. Consider the homogeneous quadratic differential system

$$\frac{dX(t)}{dt} = [X(t), \Pi_0(X(t))] \tag{1.10}$$

where $\Pi_0(X) := (X^-) - (X^-)^T$ and X^- is the strictly lower triangular part of X. This system, known as the *Toda lattice*, is a special case of (1.3) with $k(t) := \Pi_0(X(t))$, a special one-parameter family of skew-symmetric matrices. Recently it has been found [1], [4], [8], [10], [13] that the Toda flow, when sampled at integer times, gives exactly the same sequence as does the QR algorithm applied to the matrix $A_0 = \exp(X(0))$. The convergence of the QR algorithm, therefore, can be understood from ordinary differential equations theory [1], [4].

In this paper we want to continue the study of the relationship between the solutions of the ODEs (1.3) for general $k(t)$ and the corresponding matrix factorizations. We think that the isospectral flow phenomenon could be better understood by using differential geometry and Lie group concepts and that a unified theory of the various matrix factorization techniques could be established which, in turn, might shed light on the design of new algorithms. We are especially interested in characterizing the matrices $k(t)$ so that the resulting solution $X(t)$ of (1.3) would have nice and useful asymptotic behavior. This paper contains some of the results we have found in these aspects.

The paper is organized as follows. We begin in the next section with some preliminary results. These results are easy to prove but serve as the foundation for various approaches to the isospectral flow problems. In § 3 we demonstrate how the proposed abstract framework can be practically applied to unify three well-known matrix factorization techniques used in numerical linear algebra. Motivated by the success in interpreting old results, we continue to discuss other types of matrix factorizations in § 4. Some of the suggested factorizations are new, and we believe that there are many more that deserved to be explored.

2. General framework. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two subspaces of $\mathrm{gl}\,(n)$ such that each element in $\mathrm{gl}\,(n)$ can be represented *uniquely* as the sum of elements from $\mathcal{T}_1$ and $\mathcal{T}_2$. Note that the decomposition $\mathrm{gl}\,(n) = \mathcal{T}_1 + \mathcal{T}_2$ is not assumed to be a direct sum. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be the natural *projection mappings* from $\mathrm{gl}\,(n)$ into the subspaces $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively. Given an arbitrary polynomial $p(x)$, consider the IVP:

$$\frac{dX(t)}{dt} = [X(t), \mathcal{P}_1(p(X(t)))], \qquad X(0) = X_0. \tag{2.1}$$

Since $[X(t), p(X(t))] = 0$, the solution of (2.1) also satisfies the IVP

$$\frac{dX(t)}{dt} = [\mathcal{P}_2(p(X(t))), X(t)], \qquad X(0) = X_0. \tag{2.2}$$

Let $g_1(t)$ and $g_2(t)$ be solutions of the implicitly defined IVPs

$$\frac{dg_1(t)}{dt} = g_1(t)\mathscr{P}_1(p(X(t))), \qquad g_1(0) = I, \tag{2.3}$$

$$\frac{dg_2(t)}{dt} = \mathscr{P}_2(p(X(t)))g_2(t), \qquad g_2(0) = I. \tag{2.4}$$

Then $g_1(t)$ and $g_2(t)$ are defined and necessarily nonsingular for all t for which $X(t)$ exists. Furthermore, it is easy to see the following theorem.

THEOREM 2.1. *Suppose $X(t)$ solves the* IVP (2.1). *Then*

$$X(t) = g_1(t)^{-1}X_0 g_1(t) = g_2(t)X_0 g_2(t)^{-1}. \tag{2.5}$$

Regardless of how the Lie algebra gl (n) is split, we claim the following theorem.

THEOREM 2.2. *Let $X(t)$, $g_1(t)$ and $g_2(t)$ be solutions of the* IVP*s* (2.1), (2.3), *and* (2.4), *respectively, on the interval* $[0, T]$, $T > 0$. *Then*

$$\exp(p(X_0)\cdot t) = g_1(t)g_2(t). \tag{2.6}$$

Proof. It is trivial that $\exp(p(X_0)\cdot t)$ satisfies the IVP

$$\frac{dY(t)}{dt} = p(X_0)Y(t), \qquad Y(0) = I.$$

Let $Z(t) := g_1(t)g_2(t)$. Then $Z(0) = I$. Also

$$\begin{aligned}\frac{dZ(t)}{dt} &= (dg_1(t)/dt)g_2(t) + g_1(t)(dg_2(t)/dt)\\ &= g_1(t)\mathscr{P}_1(p(X(t)))g_2(t) + g_1(t)\mathscr{P}_2(p(X(t)))g_2(t)\\ &= g_1(t)p(X(t))g_2(t)\\ &= p(X_0)Z(t) \quad (\text{by } (2.5)).\end{aligned}$$

By the uniqueness theorem for initial value problems, it follows that $Z(t) = \exp(p(X_0)\cdot t)$. □

THEOREM 2.3. *Let $X(t)$, $g_1(t)$, and $g_2(t)$ be solutions of the* IVP*s* (2.1), (2.3), *and* (2.4), *respectively, on the interval* $[0, T]$, $T > 0$. *Then*

$$\exp(p(X(t))\cdot t) = g_2(t)g_1(t). \tag{2.7}$$

Proof. By (2.6), we have

$$\begin{aligned}g_2(t)g_1(t) &= g_1(t)^{-1}\exp(p(X_0)\cdot t)g_1(t)\\ &= \exp(g_1(t)^{-1}p(X_0)g_1(t)\cdot t)\\ &= \exp(p(X(t))\cdot t) \quad (\text{by } (2.5)).\end{aligned}$$ □

Remark 2.1. If we set $t = 1$, then Theorems 2.2 and 2.3 imply that

$$\exp(p(X(0))) = g_1(1)g_2(1) \tag{2.8a}$$

and

$$\exp(p(X(1))) = g_2(1)g_1(1). \tag{2.8b}$$

Since the differential system (2.1) is *autonomous*, we know that the phenomenon (2.8) will occur at every feasible integer time. In accordance with conventional numerical linear algebra, the iterative process (2.8) for all integers will be called the *abstract g_1g_2-algorithm.*

Remark 2.2. It is understood in Lie theory [3], [7], [11] that corresponding to a *Lie algebra decomposition* of gl (n), there is a *Lie group decomposition* of Gl (n) in a neighborhood of I. What we have shown above is a generalization of this concept, that is, the two subspaces $\mathcal{T}_1$ and $\mathcal{T}_2$ are not necessarily Lie subalgebras. However, every matrix (i.e., the matrix $\exp(p(X_0)\cdot t)$) in a neighborhood of I can still be written as the product of two elements (i.e., $g_1(t)$ and $g_2(t)$) of Gl (n). The locality of this neighborhood depends upon the maximal interval of existence for the problem (2.1). The right-hand side of (2.6) will be called the *abstract* g_1g_2*-decomposition* of the matrix $\exp(p(X_0)\cdot t)$, where $g_1(t)$ and $g_2(t)$ are solutions of the associated IVPs (2.3) and (2.4), respectively. In the next section we will show how the QR, LU and other classical matrix factorizations are special cases of this abstract g_1g_2-decomposition.

Remark 2.3. If f is any analytic function defined on a domain containing the spectrum of X_0, then $[f(X(t)), X(t)]=0$ (see [6]). Thus the above theorems remain true if $p(x)$ is replaced by $f(x)$.

3. Applications. We now interpret the above abstract results in the terminology of standard numerical linear algebra. The following three examples are extracted from the review article by Watkins [12]. Our point here is to demonstrate how the arguments given in the previous section unify various methods for constructing iterative processes based on matrix factorizations.

For simplicity we will consider only the case $K=\mathbb{R}$ and $p(x)=x$, and we adopt the following notation:

$o(n)$:= The set of all skew-symmetric matrices in gl (n);
$O(n)$:= The set of all orthogonal matrices in Gl (n);
$r(n)$:= The set of all strictly upper triangular matrices in gl (n);
$R(n)$:= The set of all upper triangular matrices in Gl (n);
$l(n)$:= The set of all strictly lower triangular matrices in gl (n);
$L(n)$:= The set of all lower triangular matrices in Gl (n);
$d(n)$:= The set of all diagonal matrices in Gl (n);
X^+ := The strictly upper triangular part of the matrix X;
X^0 := The diagonal part of the matrix X;
X^- := The strictly lower triangular part of the matrix X.

The three cases considered in [12] and their relationships to our notions mentioned in the previous section are summarized in Table 1. For example, if we choose to decompose the space gl $(n,\mathbb{R})$ as the direct sum of the subspace $\mathcal{T}_1$ of all skew-symmetric matrices and the subspace $\mathcal{T}_2$ of all upper triangular matrices, then the corresponding projections $\mathcal{P}_1(X)$ and $\mathcal{P}_2(X)$ are necessarily of the form $\mathcal{P}_1(X)=X^- -(X^-)^T$ and $\mathcal{P}_2(X)=X^+ + X^0 + (X^-)^T$, respectively. In this case the solutions $g_1(t)$ and $g_2(t)$ of (2.3) and (2.4) are necessarily orthogonal and upper triangular, respectively. According to Remark 2.1, we understand that with $A_0=\exp(X_0)$, the QR algorithm produces a sequence of isospectral matrices $\{A_k\}$ which is exactly the sequence $\{\exp(X(k))\}$.

Observe that the decomposition gl $(n)=\mathcal{T}_1+\mathcal{T}_2$ in Cases 1 and 2 of Table 1 are direct sum decompositions, whereas the decomposition in Case 3 is not a direct sum.

Even though the motivation for splitting gl (n) in the specific ways shown above may not be that straightforward, in each of the above cases both g_1 and g_2 manifest fairly obvious structure. That is, the products g_1g_2 for the first two cases are understood by numerical analysts to be the QR *decomposition* and the LU *decomposition* of a matrix, respectively, whereas the product g_1g_2 for the third case is precisely the *Cholesky decomposition* of a symmetric matrix if the initial value X_0 is symmetric to begin with.

TABLE 1

	Case 1	Case 2	Case 3
$\mathcal{T}_1$	$o(n)$	$l(n)$	$l(n)+d(n)/2$
$\mathcal{T}_2$	$r(n)+d(n)$	$r(n)+d(n)$	$r(n)+d(n)/2$
$k(t)=\mathcal{P}_1(X(t))$	$X^- -(X^-)^T$	X^-	$X^- + X^0/2$
$\mathcal{P}_2(X(t))$	$X^+ + X^0 + (X^-)^T$	$X^+ + X^0$	$X^+ + X^0/2$
$g_1(t)$	$Q(t) \in O(n)$	$L(t) \in L(n)$	$G(t) \in L(n)$
$g_2(t)$	$R(t) \in R(n)$	$U(t) \in R(n)$	$H(t) \in R(n)$
Numerical Algorithm	QR Algorithm	LU Algorithm	Cholesky Algorithm

The theory concerning the asymptotic behavior of either the dynamical flow or the discrete iteration for any of the above three cases has been well developed [1], [4], [8], [10], [12]. It is natural to ask whether other kinds of splittings of gl (n) would induce (or relate to) other kinds of useful dynamical systems or iterative schemes for solving the eigenvalue problem. In the next section we will demonstrate some of the interesting results we have found.

4. More examples. For illustration purposes, we will assume $K = \mathbb{R}$ and $p(x) = x$.

Example 1. Suppose X_0 is symmetric. We want to choose $\mathcal{T}_1$ to be a proper subspace of $o(n)$, and then let $\mathcal{T}_2$ be a complement of $\mathcal{T}_1$ in gl (n). In doing so, by (2.3), $g_1(t)$ must be orthogonal and, by Theorem 2.1, $X(t)$ remains symmetric and $\|X(t)\|_F = \|X_0\|_F$ where $\|\cdot\|_F$ is the Frobenius matrix norm. It follows that $X(t)$ is defined for all $t \in \mathbb{R}$.

Evidently there are many ways to choose $\mathcal{T}_1$ and its counterpart $\mathcal{T}_2$. The first case presented in § 3 amounts to the extreme case that $\mathcal{T}_1$ is precisely the $(n(n-1)/2)$-dimensional subspace $o(n)$ and that $\mathcal{T}_2$ is the $(n(n+1)/2)$-dimensional subspace of all upper triangular matrices in gl (n). We note that corresponding to the same $\mathcal{T}_1 = o(n)$, we could choose $\mathcal{T}_2$ to be the subspace $s(n) :=$ The set of all symmetric matrices in gl (n). But then the vector field of (2.1) would be identically zero and this kind of splitting of gl (n) leads to a trivial case.

It is more convenient to work with the projection $\mathcal{P}_1$ than the subspace $\mathcal{T}_1$. Given X_0, let the index subset $\Delta \subset \{(i,j) \mid 1 \leqq j < i \leqq n\}$ represent the portion of the lower triangular part of X_0 which we want to *zero out* by orthogonal similarity transformations. For each $X(t)$, let $\hat{X}(t)$ be the strictly lower triangular matrix that is made up of the portion of $X(t)$ corresponding to Δ. We claim the following theorem.

THEOREM 4.1. *Suppose the projection maps* $\mathcal{P}_1 : \mathrm{gl}\,(n) \to o(n)$ *and* $\mathcal{P}_2 : \mathrm{gl}\,(n) \to \mathrm{gl}\,(n)$ *are defined by*

$$\mathcal{P}_1(X(t)) := \hat{X}(t) - \hat{X}(t))^T \tag{4.1}$$

and

$$\mathcal{P}_2(X(t)) := X(t) - \mathcal{P}_1(X(t)), \tag{4.2}$$

respectively. Then, for all $(i,j) \in \Delta$, $x_{ij}(t) \to 0$ *as* $t \to \infty$.

Proof. Due to the symmetry of $X(t)$ and the fact that $\hat{X}(t)$ is strictly lower triangular, it is easy to check that for each $i=1,\cdots,n$,

$$\begin{aligned}\frac{dx_{ii}(t)}{dt}&=\sum_k x_{ik}(t)(\hat{x}_{ki}(t)-\hat{x}_{ik}(t))\\&=2\sum_{i<\alpha} x_{\alpha i}(t)\hat{x}_{\alpha i}(t)-2\sum_{i>\beta} x_{i\beta}(t)\hat{x}_{i\beta}(t).\end{aligned}\tag{4.3}$$

Note that $\hat{x}_{ij}(t)=0$ unless $(i,j)\in\Delta$. Note also that if $\hat{x}_{ij}(t)\neq 0$, then $\hat{x}_{ij}(t)=x_{ij}(t)$. So both terms on the right-hand side of (4.3) are sums of perfect squares and involve only indices in Δ.

Let J_1 denote the smallest column number in Δ. Then for all $i<J_1$, $dx_{ii}(t)/dt=0$ and

$$\frac{dx_{J_1J_1}(t)}{dt}=2\sum_{J_1<\alpha} x_{\alpha J_1}(t)\hat{x}_{\alpha J_1}(t)\geqq 0.\tag{4.4}$$

Since $x_{J_1J_1}(t)$ is monotone and is bounded for all t, both limits of $x_{J_1J_1}(t)$ as $t\to\infty$ and $t\to-\infty$ exist. From (4.4),

$$x_{J_1J_1}(T)-x_{J_1J_1}(-T)=2\int_{-T}^{T}\left(\sum_{J_1<\alpha} x_{\alpha J_1}(t)\hat{x}_{\alpha J_1}(t)\right)dt$$

is bounded as $T\to\infty$. It follows that for all $(\alpha,J_1)\in\Delta$, $x_{\alpha J_1}(t)\in\mathscr{L}^2(-\infty,\infty)$. On the other hand, it is clear that for all $(\alpha,J_1)\in\Delta$, $|(d/dt)x^2_{\alpha J_1}(t)|\leqq M$ for some constant M. We claim that

$$\lim_{t\to\infty} x^2_{\alpha J_1}(t)=0\quad\text{for each }(\alpha,J_1)\in\Delta.\tag{4.5}$$

Indeed, otherwise there would exist a sequence of numbers $\{t_k\}$ such that $x^2_{\alpha J_1}(t_k)\geqq\delta>0$ for some δ and such that $t_{k+1}-t_k\geqq\delta/M$. Then it would follow that $x^2_{\alpha J_1}(t)\geqq\delta/2$ whenever $|t-t_k|\leqq\delta/2M$ and that $x_{\alpha J_1}(t_k)$ would not be $\mathscr{L}^2$-integrable. This is a contradiction. So (4.5) is proved.

Now let I denote the largest row number in Δ. Then for all $i>I$, $dx_{ii}(t)/dt=0$ and

$$\frac{dx_{II}(t)}{dt}=-2\sum_{\beta<I} x_{I\beta}(t)\hat{x}_{I\beta}(t)\leqq 0.\tag{4.6}$$

Again we see that $x_{II}(t)$ is monotone and bounded for all t. By an argument similar to the above, we conclude that

$$\lim_{t\to\infty} x_{I\beta}(t)=0\quad\text{for each }(I,\beta)\in\Delta.\tag{4.7}$$

We now consider the case when $J_1<i<I$. Let J_2 be the next smallest column number in Δ. Then

$$\frac{dx_{J_2J_2}(t)}{dt}=2\sum_{J_2<\alpha} x_{\alpha J_2}(t)\hat{x}_{\alpha J}(t)-2x_{J_2J_1}(t)\hat{x}_{J_2J_1}(t).\tag{4.8}$$

If (J_2,J_1) is not in Δ, then $x_{J_2J_2}(t)$ is monotone and bounded for all t. For the case where $(J_2,J_1)\in\Delta$, we consider $y_{J_2J_2}(t)$ defined by

$$y_{J_2J_2}(t)=x_{J_2J_2}(t)+2\int_{-\infty}^{t} x_{J_2J_1}(s)\hat{x}_{J_2J_1}(s)\,ds.$$

Then $y_{J_2J_2}(t)$ is bounded for all t since $x_{J_2J_1}(s)\in\mathscr{L}^2(-\infty,\infty)$. Also, $y_{J_2J_2}(t)$ is monotone since $dy_{J_2J_2}(t)/dt=2\sum_{J_2<\alpha}x_{\alpha J_2}(t)\hat{x}_{\alpha J}(t)\geqq 0$. So, regardless of whether $(J_2,J_1)\in\Delta$, we may now apply the same argument as above to conclude that

$$\lim_{t\to\infty}x_{\alpha J_2}(t)=0 \quad \text{for each } (\alpha,J_2)\in\Delta. \tag{4.9}$$

By induction, we continue the above process from the leftmost column to the rightmost column involved in Δ. Every term involved in the second summation of (4.3) has been proved previously to be $\mathscr{L}^2$-integrable and converge to zero. So $\sum_{i<\alpha}x_{\alpha i}(t)\hat{x}_{\alpha i}(t)\to 0$ as $t\to 0$ for every $(\alpha,i)\in\Delta$. The assertion is proved. □

Remark 4.1. Theorem 4.1 suggests a constructive way to knock out *any* off-diagonal portion of X_0 by orthogonal similarity transformations. This can be interpreted as a generalization of the well-known *Schur theorem* (which guarantees the knock-out of the strictly lower triangular part only). There are many interesting applications of this theorem. For example, if we choose $\Delta=\{(i,j)\,|\,1\leqq j<i-1\leqq n-1\}$, then the corresponding dynamical system (2.1) represents a *continuous tridiagonalization process* on the symmetric matrix X_0.

Remark 4.2. The corresponding iterative scheme of Theorem 4.1, as is suggested by Remark 2.1, can now be understood only in the abstract sense because there is no plain way to describe the structure of $g_2(t)$ in general. This is in contrast to all of the examples in § 2 where $g_2(t)$ is always upper triangular. Nevertheless, we do know here that $g_1(t)$ is always orthogonal.

Example 2. The proof of Theorem 4.1 depends essentially upon the symmetry of $X(t)$. Given a general $X_0\in\mathrm{gl}\,(n)$, however, Theorem 2.2 always relates $g_1(t)$ and $g_2(t)$ to powers of the matrix $\exp(X_0)$. We now demonstrate how this property can be applied to study the asymptotic behavior for some nonsymmetric cases.

For simplicity in the discussion we will assume X_0 has n distinct eigenvalues. For general cases without this assumption, the following arguments can be modified easily. Let $\Delta:=\{(i,j)\,|\,1\leqq j\leqq k-1,\ k\leqq i\leqq n\}$ be a *rectangular* index set for some fixed $2\leqq k\leqq n$. For each $X(t)$, let $\hat{X}(t)$ be the strictly lower triangular matrix which is made of the portion of $X(t)$ corresponding to Δ. We are again interested in orthogonal similarity transformations, so we define $\mathscr{P}_1(X(T))$ and $\mathscr{P}_2(X(t))$ as in (4.1) and (4.2), respectively. Then the corresponding solution $X(t)$ of (2.1) exists for all $t\in\mathbb{R}$. We claim the following theorem.

THEOREM 4.2. *For all* $(i,j)\in\Delta$, $x_{ij}(t)\to 0$ *as* $t\to\infty$.

Proof. (The idea in the following proof is similar to that in [12].) It is clear that the matrix $\mathscr{P}_2(X(t))$ is of upper triangular 2 by 2 block form with the (2,1)-block identically zero. Since $g_2(0)=I$, it is easy to argue by induction that $g_2(t)$ must also be of the same form. In particular, the (2,1)-block of $g_2(t)$ is constantly zero for all t. Therefore, by Theorem 2.2, the ith columns of $\exp(X_0\cdot t)$ for $1\leqq i\leqq k$ is a linear combination of the first k columns of $g_1(t)$. On the other hand, by the theory of the simultaneous iteration method [5], [9], the subspace spanned by the first k columns of $\exp(X_0\cdot t)$ converges as $t\to\infty$ to the invariant subspace spanned by the first k eigenvectors (corresponding to the first k eigenvalues arranged in the descending order) of X_0. Together with the fact $X_0g_1=g_1X(t)$ (Theorem 2.1), it follows that the (2,1)-block of $X(t)$ must converge to zero as $t\to\infty$. □

The idea in the above proof can easily be extended to include the following general case which is the continuous realization of the so-called *treppeniteration* [5].

COROLLARY 4.1. *So long as* Δ *is an index subset of* $\{(i,j)\,|\,1\leqq i\leqq j\leqq n\}$ *such that its complement represents a block upper triangular matrix, if we define* $\mathscr{P}_1(X(t))$ *and*

$\mathcal{P}_2(X(t))$ *according to* (4.1) *and* (4.2), *respectively, then for all* $(i, j) \in \Delta$, $x_{ij}(t) \to 0$ *as* $t \to \infty$.

Example 3. For $0 < k < n$, let Δ be an index subset of $\{(i, j) \mid 1 \leqq i \leqq j \leqq k\}$ such that its complement represents a k by k block upper triangular matrix. Let $X_{11}(t)$ denote the k by k leading principal submatrix of the n by n matrix $X(t)$, and let $\hat{}$, $\mathcal{P}_1$, and $\mathcal{P}_2$ be defined in the same way as in Example 1. Then

$$\mathcal{P}_1(X(t)) = \begin{bmatrix} \mathcal{P}_1(X_{11}(t)), & 0 \\ 0 & 0 \end{bmatrix}.$$

It follows that for each $(i, j) \in \Delta$, $x_{ij}(t)$ of $X(t)$ converges to zero as $t \to \infty$ since, by Corollary 4.1, each element of $X_{11}(t) \to 0$.

Remark 4.3. It is easy to see that all of the above results remain true if the underlying index subset Δ is such that it can be rearranged by permutations to be any one of the forms we have mentioned. Thus, we have considerable flexibility in choosing Δ to knock out portions of X_0.

Example 4. We now consider the Hamiltonian eigenvalue problem. A matrix $X \in \mathrm{gl}\,(2n, \mathbb{R})$ is called *Hamiltonian* if and only if it is of the form

$$X = \begin{bmatrix} A, & N \\ K, & -A^T \end{bmatrix} \tag{4.10}$$

where K, $N \in \mathrm{gl}\,(n, \mathbb{R})$ are symmetric. Suppose we define

$$\mathcal{P}_1(X) = \begin{bmatrix} 0, & -K \\ K, & 0 \end{bmatrix}. \tag{4.11}$$

Then it is easy to see that $[X, \mathcal{P}_1(X)]$ is also Hamiltonian. With this in mind, given a Hamiltonian matrix X_0, we may define the dynamical system (2.1) with $\mathcal{P}_1$ given by (4.11). Then, the corresponding solution $g_1(t)$ of (2.3) is both orthogonal and *symplectic*, and $X(t) = g_1(t)^{-1} X_0 g_1(t)$ remains Hamiltonian for all t. Furthermore, from Theorem 4.2, it follows that $K(t) \to 0$ as $t \to \infty$. Practically speaking, we then only need to consider the eigenvalues of $\lim_{t \to \infty} A(t)$.

Remark 4.4. Thus far, we have considered only the case where $\mathcal{P}_1(X(t))$ is required to be skew-symmetric. We could have chosen, for example, $\mathcal{P}_1(X(t)) = \hat{X}(t)$ and $\mathcal{P}_2(X(t)) = X(t) - \hat{X}(t)$ in Example 2. Then the resulting dynamics would be analogous to that of the LU algorithm.

Remark 4.5. In our numerical experimentation with Example 2, we have *never* failed in zeroing out *any* prescribed index subset Δ. We conjecture that *Theorem* 4.1 *will remain true even for nonsymmetric cases.* Based on the center manifold theory, a proof similar to that developed in [1] can be established for the local convergence; however, we have not yet been able to provide a proof of the global convergence.

REFERENCES

[1] M. T. CHU, *The generalized Toda flow, the* QR *algorithm and the center manifold theory*, SIAM J. Algebraic Discrete Methods, 5 (1984), pp. 187–201.

[2] ———, *On the continuous realization of iterative processes*, SIAM Rev. 30 (1988), pp. 375–386.

[3] M. L. CURTIS, *Matrix Groups*, Springer-Verlag, Berlin, New York, 1979.

[4] P. DEIFT, T. NANDA, AND C. TOMEI, *Differential equations for the symmetric eigenvalue problem*, SIAM J. Numer. Anal., 20 (1983), pp. 1–22.

[5] G. H. GOLUB AND C. E. VAN LOAN, *Matrix Computation*, Johns Hopkins University Press, Baltimore, MD, 1983.

[6] F. R. GANTMACHER, *The Theory of Matrices*, Vols. 1-2, Chelsea, New York, 1959.
[7] S. HELGASON, *Differential Geometry, Lie Groups and Symmetric Spaces*, Academic Press, New York, 1978.
[8] T. NANDA, *Isospectral flows on band matrices*, Ph.D. disseration, New York University, New York, 1982.
[9] B. N. PARLETT, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
[10] W. W. SYMES, *The* QR-*algorithm and scattering for the finite nonperiodic Toda Lattice*, Phys. D, 4 (1982), pp. 275-280.
[11] F. W. WARNER, *Foundations of Differentiable Manifolds and Lie Groups*, Scott, Foresman and Company, Glenview, 1970.
[12] D. S. WATKINS, *Isospectral flows*, SIAM Rev., 26 (1984), pp. 379-391.
[13] D. S. WATKINS AND L. ELSNER, *Selfsimilar flows*, Washington State University, Pullman, WA, 1987.