Isospectral flows for displacement structured matrices

Dario Fasino
Dipartimento di Matematica e Informatica, Università di Udine
Via delle Scienze 208, 33100 Udine (Italy)

Summary

This paper concerns eigenvalue computations with displacement structured matrices, for example, Toeplitz or Toeplitz-plus-Hankel. A technique using isospectral flows is introduced. The flow is enforced to preserve the displacement structure of the original matrix by means of a suitable constraint added in its formulation. In order to fulfill the constraint, the numerical integration of the flow requires, to evaluate the gradient, the solution of a matrix equation.

1 Introduction

Isospectral flows can be regarded as continuous counterparts of more familiar discrete methods, as the QR algorithm, for the computation of matrix eigenvalues. Let \( X(t) \) be a matrix-valued function satisfying the differential equation

\[
\frac{dX}{dt} = XA - AX, \quad X(0) = X_0, \quad t \geq 0,
\]

for a suitable matrix coefficient \( A(t) \equiv A(X,t) \). For all \( t > 0 \) for which \( X(t) \) is defined, \( X(t) \) is similar to \( X_0 \), hence the above differential equation is said to generate an isospectral flow \([4, 8, 18]\). Indeed, consider the solution \( S(t) \) of the initial value problem

\[
\frac{dS}{dt} = AS, \quad S(0) = I, \quad t \geq 0,
\]

the so-called Flaschka formulation of (1.1). By direct substitution, one obtains

\[
X(t) = S(t)X_0S^{-1}(t).
\]

The two equations (1.1) and (1.2) are equivalent in the sense that, if one of these two ODEs can be integrated for \( t \) in some interval \([0,\tilde{t}]\), then also the other can be solved in the same interval. If \( A(t) \) is chosen appropriately, one can make \( X(t) \) well defined for all real \( t \), moreover, its limiting matrices for \( t \) approaching infinity are diagonal or triangular, hence showing explicitly the eigenvalues of \( X_0 \).

The goal of this paper is to describe how one can generate isospectral flows evolving on spaces of matrices having a displacement structure. These are spaces defined in terms of a displacement operator \( \mathcal{L} \).
\[ L(X) = C_l X - X C_r, \]  
\( (1.4) \)

mapping all matrices in that space into matrices whose rank is a small fixed integer, independent on their order. Mostly known examples of displacement structured matrices are the Toeplitz matrices and the Toeplitz-plus-Hankel matrices [14].

Within displacement structured spaces there are special matrices whose eigenvalues can be computed explicitly. These matrices belong to simultaneously diagonalizable sets, usually matrix algebras, related to the matrices \( C_l, C_r \) in (1.4). For illustration, considered here are the algebra of circulant matrices within the Toeplitz matrix space, and the so called \( \tau \)-class [3] within the Toeplitz-plus-Hankel matrix space, but the approach outlined here deserves a greater generality.

In this paper a suitable construction of the coefficient \( A(t) \) is proposed, leading to isospectral flows such that:

- for all \( t \geq 0 \), the matrix \( X(t) \) has the assigned displacement structure, and
- the matrix \( X(1) \) belongs to a specified algebra of simultaneously diagonalizable matrices.

These structured isospectral flows can be integrated numerically by explicit methods, for example of Runge-Kutta type. Obviously, the isospectrality is not retained in general by a numerical scheme, but the departure from it can be estimated during the integration. The main reason motivating this study is the possibility of a structured backward error estimate (see [13] for the use of backward error analysis in ODEs): If the numerical scheme used to integrate (1.1) satisfies some mild conditions, the computed approximation to \( X(1) \) can be regarded as the exact outcome of a flow starting from a perturbed initial point, belonging to the same displacement space as \( X_0 \).

Preliminary results, both theoretical and numerical, show that isospectral flows do exist in the structured spaces quoted above. The structure of the matrices involved can be exploited in order to devise efficient algorithms to compute the derivative of \( X(t) \) and to advance in time its numerical integration.

In Section 2, the problem under consideration is put in the context of the relevant literature, and the novelty of the present approach is clarified. In Section 3 the construction of structured isospectral flows is presented and analyzed. Some numerical examples are shown in Section 4.

### 2 Related work

In a paper dating back to 1983 [8], Deift, Nanda and Tomei suggested that competitive numerical eigenvalue algorithms for eigenvalue computation might be based on the integration of certain differential equations. Further studies showed that familiar algorithms as the QR algorithm can be better understood with the help of certain isospectral flows, and their connection with differential geometry and matrix factorizations has become an active research topic, see e.g., [1, 5, 7, 18].

On the other hand, it was proved in [4] that the flow (1.1) cannot be integrated numerically, while retaining isospectrality in the discrete solution, even in exact arithmetics. In order to retain isospectrality, the Flaschka formulation (1.2) was explored deeply in recent papers, e.g., [4, 10]. In particular, in [10] the ODE (1.2) is translated by means
of a Cayley transform into a corresponding differential equation with a skew-symmetric matrix-valued solution. Since any reasonable numerical method retains skew-symmetry, the transformed equation can be integrated accurately, and the underlying structure is preserved. By means of an inverse Cayley transform one computes the discrete steps $S(t_k)$ first, and the sequence $X(t_k)$ is recovered using (1.3). With the help of techniques analogous to the above, isospectral flows were discovered, that evolve on matrix spaces having some nice structure, notably, the symmetric matrices and the Hamiltonian matrices.

However, displacement spaces seem to be much less tractable, under this aspect. In the recent literature there are many attempts of describing sets of matrices with a particular displacement structure and prescribed spectra, see for example, [1, 6, 11]. The approach followed is to characterize all linear preservers of such sets [11], and is based on the fact that, the function $S(t)$ in (1.2) evolves in a Lie group iff $A(t)$ belongs to the associated Lie algebra [18]. This approach eventually led to negative results, because no nontrivial displacement structure is preserved under the action of a Lie group. In the case of Toeplitz matrices, this conclusion is made clear in a paper of Chu [6]: There, the author shows that the set of all real orthogonal $n \times n$ matrices $Q$ such that $Q T Q^T$ is Toeplitz whenever $T$ is a symmetric Toeplitz matrix, has exactly eight elements. In fact, no algorithms based on continuation or isospectral methods appear in the recent survey [9] on numerical algorithms for eigenvalue and singular value computations with Toeplitz matrices. Probably, this situation does not changes moving to more general displacement structures. Indeed, in the words of [11]:

We originally hoped that we could move around somewhat freely [...] by means of the linear preservers of such sets. This hope motivated our study of linear preservers. Unfortunately, our hope was too optimistic. Our results show that there are not enough such linear preservers.

In summary, the failure of previous attempts to “move around” isospectral sets of displacement structured matrices is due to the attitude to characterize the coefficient $A(t)$ suitably, so that $dX/dt$ possesses the desired structure. In the next section, a technique enforcing the structure directly on $dX/dt$ will be shown.

3 The basic algorithm

From now on, the following notations are used: All matrices are $n \times n$. The bilinear operator $[\cdot ,\cdot ]$ is the commutator product of two matrices, $[A,B] = AB - BA$. The set of all matrices commuting with $X$ is denoted by $Z(X)$. Moreover, suppose that $C = C_l = C_r$ in (1.4) is normal, the spectral decomposition $C = U \Lambda U^*$ is explicitly known, and its eigenvalues are all distinct. The latter hypotheses are motivated by the relevant cases

$$C_1 = \begin{pmatrix} 0 & 1 & \cdots & \cdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \ddots & \cdots \\ 1 & \cdots & \cdots & 0 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 0 & 1 & \cdots & \cdots \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \cdots \\ \cdots & \cdots & \cdots & 1 \end{pmatrix}.$$

In fact, Toeplitz matrices can be defined in terms of a displacement operator like (1.4) having $C_l = C_r = C_1$, while a displacement operator where $C_l = C_r = C_2$ is associated to
the space of Toeplitz-plus-Hankel matrices. Observe that $Z(C_1)$ is the algebra of circulant matrices, and $Z(C_2)$ is the $\tau$-class [3].

Now, let $\Delta_0 = CX_0 - X_0C$ and consider the initial value problem (1.1) where $A(t)$ is imposed to fulfill the following matrix equation:

$$[A,X] = Y, \quad [C,Y] = -\Delta_0, \quad t \geq 0.$$  

(3.5)

The above equation is to be solved for $A(t)$ and $Y(t)$, given $X(t)$. Moreover, suppose that $A(t)$, and hence $X(t)$, is well defined for $0 \leq t \leq 1$. Then the following properties are easily verified:

**Theorem 1** If $X(t)$ is well defined for $t \in [0,1]$ from (1.1) and (3.5), then:

a) $X(t)$ has the same eigenvalues and Jordan structure of $X_0$;

b) $X(t) = (1-t)X_0 + \Gamma(t)$, for some $\Gamma(t) \in Z(C)$;

c) $X(1) \in Z(C)$.

From the property (b) above, we see that $X(t)$ has the same displacement structure of $X_0$. In particular, if $X_0$ and $C$ are both real and symmetric, then $X(t)$, if exists, is real and symmetric. Furthermore, from property (c), we see that $X(1) = U\Lambda_0 U^*$, where $\Lambda_0$ is a diagonal matrix with the eigenvalues of $X_0$.

For the solution of (3.5), we can follow a classical approach and translate it into a matrix-vector form, by means of the Vec operator (see, e.g., [17], Chapter 12): For any $n \times n$ matrix $A$, let $\text{Vec}(A)$ be the $n^2\times 1$-order vector obtained stacking downwards its columns, into one long vector. Hence, let $a = \text{Vec}(A)$, $y = \text{Vec}(Y)$ and $\delta = \text{Vec}(\Delta_0)$.

Using well known properties of Vec, one finds the following equivalent form of equation (3.5):

$$\begin{pmatrix} I \otimes X - X^t \otimes I \\ O \otimes O \end{pmatrix} \begin{pmatrix} a \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ -\delta \end{pmatrix}. \quad (3.6)$$

Denote $M$ the matrix on the left hand side. One sees that $M$ is singular: Indeed, for any $B_1 \in Z(X)$ and for any $B_2 \in Z(C^t)$ it holds

$$M \begin{pmatrix} b_1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad M^t \begin{pmatrix} 0 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (3.7)$$

where $b_i = \text{Vec}(B_i)$. So, we need a criterion to establish whether (3.6) is consistent. Introducing the matrices $U$ and $V$ and their vectorizations $u = \text{Vec}(U)$ and $v = \text{Vec}(V)$, we observe that the homogeneous equation $(u^t,v^t)M = 0$ reduces to the relations

$$[X^t,U] = O, \quad [C^t,V] = U. \quad (3.8)$$

Since the system (3.6) is consistent if and only if the right hand side is orthogonal to any vector $(u^t,v^t)$ such that $(u^t,v^t)M = 0$, using (3.8) we discover a necessary and sufficient condition:

$$0 = u^t \delta = \text{Trace}(V^tCX_0 - V^tX_0C) = \text{Trace}(V^tCX_0 - CV^tX_0) = \text{Trace}(U^tX_0).$$
Theorem 2 The system (3.6) is consistent if and only if the following implication is true:
\[ X^t U = O \quad \text{and} \quad C^t V = U \implies \text{Trace}(U^t X_0) = 0. \]

The condition in the above theorem is trivially fulfilled when all vectors in the kernel of $M^t$ are those described in the second equation of (3.7), as $U = O$ in this case. In fact, the following result holds:

Theorem 3 If $\text{Null}(M) = \text{Dim } Z(C)$ then equation (3.6) is consistent.

The analysis of equation (3.6) conducted so far, allows us to assess existence of the gradient of the flow (1.1) in just one point $X$. Now let consider the question of existence of a solution of (1.1) in the interval $[0, 1]$. Let
\[ W = \begin{pmatrix} U^t \otimes U & U^t \otimes U \end{pmatrix}. \] (3.9)

By observing that if $X \in Z(C)$ then $U^* X U = D$, a diagonal matrix, we can recognize by a close look at the decomposition
\[ W^* M W = \begin{pmatrix} I \otimes D - D \otimes I & -I \otimes I \\ O \otimes O & I \otimes \Lambda - \Lambda \otimes I \end{pmatrix} \]
that equation (3.6) is consistent for all $X \in Z(C)$ and for all $\delta \in \text{Range}(I \otimes C - C^t \otimes I)$. Moreover, under these hypotheses, $y$ is uniquely determined. Thus, using a perturbation argument, we can prove the local existence of structured isospectral flows:

Theorem 4 Let $X_0 = X_{00} + X_{01}$, with $X_{00} \in Z(C)$. If $\|X_{01}\|$ is sufficiently small, then the solution $X(t)$ of (1.1) and (3.5) is well defined for all $0 \leq t \leq 1$.

At present, there is no clear analysis of singularities of (1.1). But the conjecture arises naturally, that if the displacement space used is not too narrow, then the set of initial conditions $X_0$ for which the flow cannot be integrated along $[0, 1]$ has zero measure in it.

Before stating the algorithm, we need an analysis of the stiffness of (1.1), subject to the constraints (3.5), in order to make an appropriate choice of the numerical scheme. The procedure to follow is made clear in [15]: We look for an upper bound for the Jacobian of the map $X \mapsto Y$. That map is well defined if $\text{Null}(M) = \text{Dim } Z(X)$. Indeed, in this case, all solutions of (3.6) have the form
\[ \begin{pmatrix} a \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ 0 \end{pmatrix} \]
where $b_1$ is as in the first relation of (3.7). In this way, and so $Y$, is uniquely determined by any solution of (3.6). However, observe that uniqueness of $Y$ is of less concern than its existence, in order to define a working algorithm based on (1.1).

Theorem 5 Assume $X(t)$ are normal matrices, having distinct eigenvalues. Let $\Phi$ be the map implicitly defined via (3.6) by $\Phi(X) = Y$. Then, in the spectral operator norm,
\[ \|\partial \Phi / \partial X\| \leq \frac{\|X_0\|_F}{c \min_{i \neq j} |\lambda_i - \lambda_j|}, \] (3.10)
where $\lambda_i$ are the eigenvalues of $X_0$ and $c$ is the cosine of the angle between $Z(C)$ and $Z(X)$, under the Frobenius inner product of matrices.
The proof of the above theorem can be carried out by means of arguments similar to the ones found in [2], using the hypothesis that $C$ has distinct eigenvalues. However, numerical experiments suggest that stiffness estimates obtained from (3.10) are in practice too stringent. Detailed proofs of all preceding theorems will appear in a forthcoming paper.

We concentrate now on the numerical solution of (3.6). The matrix $M$ has order $2n^2$, so it may become large, but it possesses a displacement structure, related to the one of $X$: If we let $\Delta = CX - XC, \text{Rank}(\Delta) = r$, and

$K_1 = \begin{pmatrix} C^t \otimes I & C^t \otimes I \end{pmatrix}$

$K_2 = \begin{pmatrix} I \otimes C & I \otimes C \end{pmatrix}$,

then it holds

$[K_1, M] = \begin{pmatrix} \Delta^t \otimes I & O \otimes O \end{pmatrix}$

$[K_2, M] = \begin{pmatrix} I \otimes \Delta & O \otimes O \end{pmatrix}$,

so we have $\text{Rank}([K_i, M]) = nr, i = 1, 2$ (moreover, $[K_1, [K_2, M]] = O$). Hence, for the solution of (3.6), we can adopt the fast implementation of Gaussian elimination with partial pivoting proposed in [12, 16] for the solution of displacement structured linear systems. The paper [12] showed how Toeplitz-like and other related matrices can be efficiently transformed into Cauchy-like matrices using fast trigonometric transforms, and described a stable and efficient implementation of Gaussian elimination with partial pivoting for such matrices. If the matrices involved are symmetric, as in the Toeplitz-plus-Hankel case, their symmetry can be preserved, using the modifications proposed in the paper [16]. In our case, the matrix $M$ is a mosaic of block-structured matrices with structured blocks. Nevertheless, since the matrices $K_{1,2}$ above are diagonalized by the matrix $W$ in (3.9), it can be shown that the matrix $W^*MW$ has a Cauchy-like displacement structure, hence algorithms as those in [12, 16] can be used for the solution of system (3.6). The resulting computational complexity is $O(nr^3)$.

Finally, the proposed algorithm is the following: Integrate the differential equation (1.1) with an explicit method, for $0 \leq t \leq 1$. Choose the stepsize according to the estimate (3.10) and the stability region of the integrator. The gradient $Y = dX/dt$ is obtained from the solution of system (3.6), exploiting the displacement structure of $M$. The eigenvalues of $X_0$ are then recovered from the diagonal of $U^*X(1)U$.

4 Numerical experiments

In this section the results of runs of the above algorithm on a Toeplitz matrix are shown. The experiments were performed using Matlab. The matrix $X_0$ chosen is a $10 \times 10$ symmetric matrix with pseudorandom entries, $\|X_0\| \approx 5.7, C = C_2$ and $\|\Delta_0\| \approx 3.13$, in the Frobenius norm. The flow was integrated using the explicit Euler method, the Heun method, and a third order explicit Runge-Kutta method, with stepsize 0.1: $X_i \approx X(i/10)$. During these experiments, all matrices $M$ encountered had rank 190, and the equalities $\text{Null}(M) = \text{Dim} \mathcal{Z}(X) = \text{Dim} \mathcal{Z}(C) = 10$ were true. The systems (3.6) were always found numerically consistent, the residual norm being in the range $6-12 \cdot 10^{-16}$.

The results are shown in the following table, where each row corresponds to a step along the interval $[0, 1]$. The first three columns are for the Euler method, the next two for Heun, the last two for the third order method. The column labeled $\|\Delta_i\|$ report the Frobenius...
norm of the displacement $\Delta_i = CX_i - X_iC$; this is shown only for the Euler method, as the two others agree to all decimals shown. Its linearly decreasing behaviour is in agreement with part (b) of Theorem 1. The columns labeled $\|Y_i\|$ report the Frobenius norm of the step $Y_i = X_i - X_{i-1}$; and $E_i$ stands for the error in the spectrum of $X_i$ with respect to $X_0$, $E_i = \text{Dist}(\sigma(X_i), \sigma(X_0))$. The latter is a measure of departure from isospectrality due to the numerical scheme, and illustrates the improvement in accuracy obtained by means of higher order methods.

\[
\begin{array}{cccccc}
 i & \|\Delta_i\| & \|Y_i\| & E_i & \|Y_i\| & E_i \\
\hline
1 & 2.818 & 0.366 & 0.022 & 0.344 & 0.0003 & 0.344 & 1.1 \times 10^{-4} \\
2 & 2.505 & 0.324 & 0.038 & 0.313 & 0.0004 & 0.313 & 1.4 \times 10^{-4} \\
3 & 2.191 & 0.299 & 0.052 & 0.293 & 0.0005 & 0.293 & 1.6 \times 10^{-4} \\
4 & 1.878 & 0.283 & 0.064 & 0.279 & 0.0005 & 0.279 & 1.6 \times 10^{-4} \\
5 & 1.565 & 0.272 & 0.074 & 0.269 & 0.0005 & 0.269 & 1.7 \times 10^{-4} \\
6 & 1.252 & 0.264 & 0.083 & 0.262 & 0.0006 & 0.262 & 1.7 \times 10^{-4} \\
7 & 0.939 & 0.259 & 0.092 & 0.257 & 0.0007 & 0.257 & 1.7 \times 10^{-4} \\
8 & 0.626 & 0.255 & 0.100 & 0.254 & 0.0008 & 0.254 & 1.7 \times 10^{-4} \\
9 & 0.313 & 0.252 & 0.107 & 0.252 & 0.0009 & 0.252 & 1.7 \times 10^{-4} \\
10 & 0.000 & 0.251 & 0.114 & 0.251 & 0.0010 & 0.251 & 1.7 \times 10^{-4} \\
\end{array}
\]

**Bibliography**


