Gohberg-Kaashoek Numbers and Forward Stability of the Schur Canonical Form

Anastasiia Minenkova[a], Evelyn Nitch-Griffin[b], Vadim Olshevsky[c]

[a] *University of Hartford, West Hartford, CT, USA. Email: minenkova@hartford.edu*
[b] *Landmark College, Putney, VT, USA. Email: evelynnitchgriffin@landmark.edu*
[c] *University of Connecticut, Storrs, CT, USA. Email: olshevsky@uconn.edu*

Abstract

In the present paper we complete the description of classes of matrices and their perturbation with the stable Schur decomposition depending on the Jordan structure of the perturbation, started in [9].

Keywords: structure-preserving perturbations, Schur decomposition, the Gohberg-Kaashoek numbers, invariant subspaces, gaps.

1. Introduction

For every quadratic matrix A there is a decomposition $A = UTU^*$ where U is unitary and T is upper triangular. This triangular matrix T is called a *Schur Triangular form* and the factorization is called the *Schur Decomposition.* The norm throughout this paper is unitarily invariant. In [9] it was shown that in general there is no forward stability for the Schur form.

Thus, the natural question to ask is when the forward stability is possible. We give the answer in this paper.

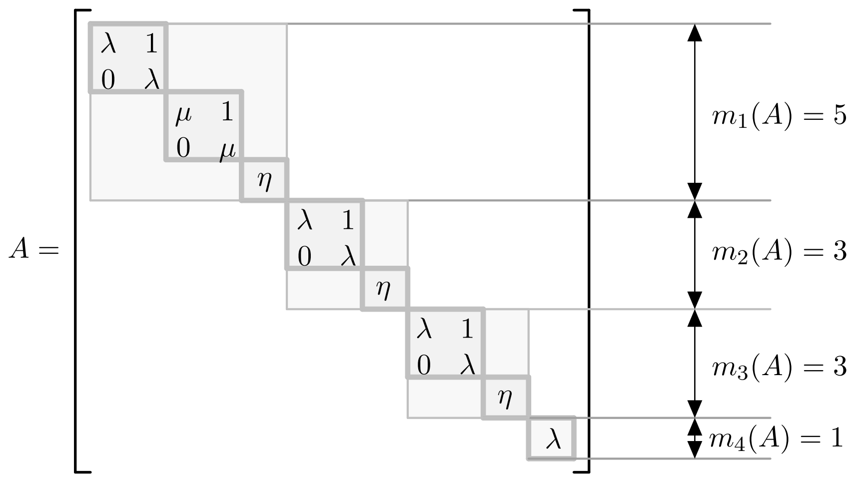
We start with the structure of the original matrix. It is of interest to find the classes of matrices A_0 for which the Schur form is always stable. It turns out, one such case is when A_0 is non-derogatory, i.e. having only one Jordan block per eigenvalue. We discovered that the following result holds.

Theorem 1.1 (Non-derogatory case. Hölder forward stability)**.** *Let A_0 be a non-derogatory matrix. Then its Schur canonical form is forward Hölder stable, i.e. if $A_0 = U_0T_0U_0^*$ then there exist constants $K, \varepsilon > 0$ (depending on A_0 only) such that for all A with $\|A - A_0\| < \varepsilon$ there exists a Schur factorization UTU^* of A such that we have*

$$\|U - U_0\| + \|T - T_0\| \leq K\|A - A_0\|^{1/n}. \tag{1.1}$$

What happens if we consider the structure preserving perturbation? Before answering this question let us discuss how to quantify the structure of a matrix and introduce a concept that will serve this purpose.

Consider $A \in \mathbb{C}^{n\times n}$. Denote by $\sigma(A)$ the set of all its eigenvalues. Let $m = \left[\begin{smallmatrix} m_1 & m_2 & \dots & m_n \end{smallmatrix}\right]^\top$ be a vector with integer entries such that $m_i \geq m_{i+1}$ for $i = 1, \dots, n-1$. The vector $k = \left[\begin{smallmatrix} k_1 & k_2 & \dots & k_n \end{smallmatrix}\right]^\top$ with $k_i = \max\limits_{1\leq l\leq n} \{l : m_l \geq i\}$ is called *dual* to m. In terms of the Gohberg-Kaashoek numbers m_j's it means that if we have



then we can also put the Jordan chains for λ, μ, and η in the following order:

				$m_j(A,\lambda)$				$m_j(A,\mu)$			$m_j(A,\eta)$
	$\mathbf{e_2} \to$	$\mathbf{e_1} \to$	$\mathbf{0}$	2	$\mathbf{e_4} \to$	$\mathbf{e_3} \to$	$\mathbf{0}$	2	$\mathbf{e_5} \to$	$\mathbf{0}$	1
	$\mathbf{e_7} \to$	$\mathbf{e_6} \to$	$\mathbf{0}$	2				0	$\mathbf{e_8} \to$	$\mathbf{0}$	1
	$\mathbf{e_{10}} \to$	$\mathbf{e_9} \to$	$\mathbf{0}$	2				0	$\mathbf{e_{11}} \to$	$\mathbf{0}$	1
	$\mathbf{e_{12}} \to$	$\mathbf{0}$		1				0			0
k_i	4	3		k_i	1	1		k_i	3		

we look at the **size** of the Jordan blocks for each eigenvalue and group them together in decreasing order to get the bigger blocks. Hence, $m(A) = [5,3,3,1,0,0,0,0,0,0,0,0]^\top$ whereas the entries of the dual vector are going to be the **number** of bigger blocks with the size greater or equal to $j = 1, 2, \dots$, i.e. $k(A) = [4,3,3,1,1,0,0,0,0,0,0,0]^\top$.

If A has the same Jordan structure as A_0 then we can get the Lipschitz stability of the Schur form. Recall if A has q distinct eigenvalues then $\Omega(A) = \{(m_i(A,\lambda_j))_{i=1}^n, j = 1,\dots,q\}$ is called the *Jordan structure* of A and the set of all matrices of the Jordan structure Ω we denote by $\mathcal{J}(\Omega)$.

Theorem 1.2 (Same Jordan structure. Lipschitz forward stability)**.** *Let $A_0 = U_0 T_0 U_0^*$. Then, there exist constants $K, \varepsilon > 0$ (depending on A_0 only) such that for all $A \in \mathcal{J}(\Omega(A_0))$ with $\|A - A_0\| < \varepsilon$ there exists a Schur factorization UTU^* of A such that $\|U - U_0\| + \|T - T_0\| \leq K\|A - A_0\|$.*

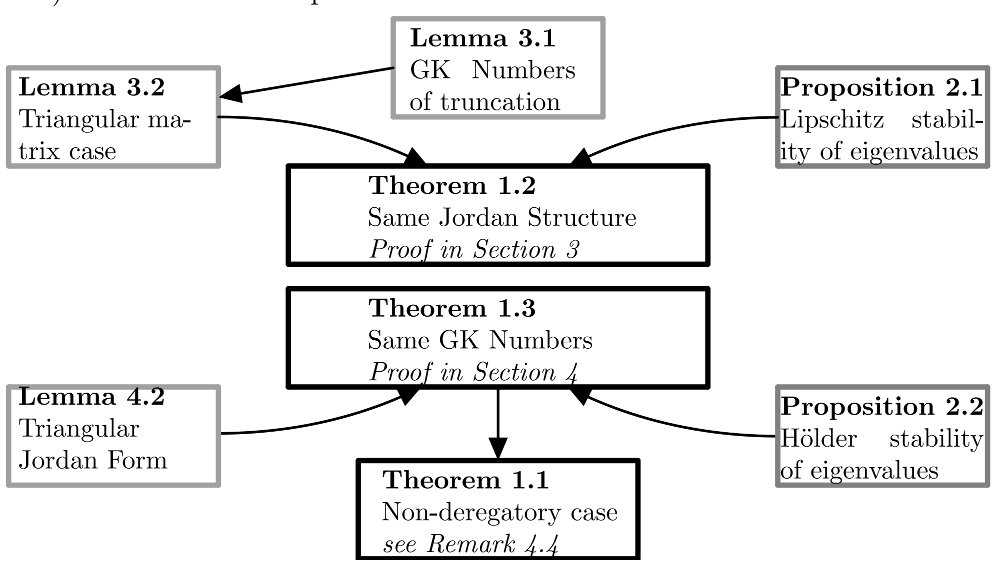
If we extend the class of perturbation to those having the same GK numbers as the original matrix, we get a different forward stability result.

Theorem 1.3 (Same GK numbers. Hölder forward stability)**.** *Let A_0 be given. Any Schur canonical form is forward Hölder stable in the class of matrices A having the same Gohberg-Kaashoek numbers as A_0. This means that there exist constants $K, \varepsilon > 0$ (depending on A_0 only) such that for all A with the same GK numbers as A_0 and $\|A - A_0\| < \varepsilon$ there exists a Schur factorization $A = UTU^*$ of A such that* (1.1)

Combining this fact with results from [9], we conclude the investigation of the stability of the Schur canonical form.

Structure of the paper.

Section 2 is devoted to some facts from the perturbation theory and theory of gaps/semigaps that are the backbone of our discourse. You can find proof of Theorem 1.2 in Section 3 and proof of Theorem 1.3 in Section 4. We introduce some useful technical lemmas in both Section 3 and Section 4 to aid us in proving main results. Note that Theorem 1.1 (the non-derogatory case) is the direct consequence of Theorem 1.3 as discussed in Remark 4.4.



2. Auxiliary Results

2.1. Stability of Eigenvalues

Note that entries on the main diagonal of T in the Schur decomposition are the eigenvalues of A. That is why the eigenvalues stability results give us the confidence to consider stability of the Schur decomposition.

The following result can be found in [1].

Proposition 2.1. *Let A_0 be an $n \times n$ matrix and $\{\lambda_1, \ldots, \lambda_n\}$ be its eigenvalues, and A being its perturbation with $\|A - A_0\| < \varepsilon$ for sufficiently small ε depending on A_0 and the eigenvalues μ_j. If the number of distinct eigenvalues of A_0 is the same as of A, then there is a certain ordering of them such that for some positive $K = K(A_0)$*

$$|\mu_i - \lambda_i| \leq K\|A - A_0\|, \quad i = 1, 2, \ldots, |\sigma(A_0)|.$$

For the general case of the eigenvalues stability we have the following result (see [12, Appendix K]).

Proposition 2.2. *Let A_0 be an $n \times n$ matrix and $\{\lambda_1, \ldots, \lambda_n\}$ be its eigenvalues. Then, there is an ordering of λ_j's that for every A with $\|A - A_0\| < \varepsilon$ for sufficiently small ε depending on A_0 there is an ordering of its eigenvalues μ_j's and a positive constant $K = K(A_0)$ such that*

$$|\mu_j - \lambda_j| \leq K\|A - A_0\|^{1/n}. \tag{2.2}$$

2.2. Gap and Semi-gap

We discuss stability of vectors related to the invariant subspaces and hence need some topological properties of the set of subspaces in $\mathbb{C}^n$.

Let $\mathcal{M}, \mathcal{N}$ be subspaces of $\mathbb{C}^n$. A matrix $P_{\mathcal{M}}$ is called an *orthogonal projector* onto a subspace $\mathcal{M} \subset \mathbb{C}^n$ if

- it is surjective: $\mathrm{Im} P_{\mathcal{M}} = \mathcal{M}$;
- applying it twice yields the same result: $P_{\mathcal{M}}^2 = P_{\mathcal{M}}$;
- it is Hermitian: $P_{\mathcal{M}}^* = P_{\mathcal{M}}$.

The following concept is the key definition.

The **gap** $\theta(\mathcal{M}, \mathcal{N})$ between $\mathcal{M}$ and $\mathcal{N}$ is defined as follows

$$\theta(\mathcal{M}, \mathcal{N}) = \|P_{\mathcal{M}} - P_{\mathcal{N}}\|$$

or, equivalently,

$$\theta(\mathcal{M},\mathcal{N}) = \max\left\{ \sup_{\substack{x\in\mathcal{M}\\ \|x\|=1}} \inf_{y\in\mathcal{N}} \|x-y\|,\ \sup_{\substack{y\in\mathcal{N}\\ \|y\|=1}} \inf_{x\in\mathcal{M}} \|x-y\| \right\}.$$

Note that $\theta(\mathcal{M},\mathcal{N})$ is a metric on the set of all subspaces in $\mathbb{C}^n$. Moreover, $\theta(\mathcal{M},\mathcal{N}) \leq 1$. The Hausdorff distance between sets Inv A and Inv B of all invariant subspaces matrices A and B can be defined as follows

$$\text{dist } (\text{Inv}\, A, \text{Inv}\, B) = \max\{ \sup_{\mathcal{M}\in \text{Inv }A} \theta(\mathcal{M}, \text{Inv }B),\ \sup_{\mathcal{N}\in \text{Inv }B} \theta(\mathcal{N}, \text{Inv }A)\}.$$

This distance is a metric as well.

Let $\mathcal{M},\mathcal{N}$ be subspaces of $\mathbb{C}^n$. The quantity

$$\theta_0(\mathcal{M},\mathcal{N}) = \sup_{\substack{x\in\mathcal{M}\\ \|x\|=1}} \inf_{y\in\mathcal{N}} \|x-y\|$$

is called the *semigap* (or one-sided gap) from $\mathcal{M}$ to $\mathcal{N}$.

The next result can be found in [9].

Proposition 2.3. *Let A_0 be fixed. Then, there exist $\epsilon, K > 0$ such that for all A with $\|A - A_0\| < \epsilon$, we have*

$$\theta_0(\ker(A), \ker(A_0)) \leq K\|A - A_0\|. \tag{2.3}$$

3. The Same Jordan Structure and Forward Stability

Before proving Theorem 1.2 we need a couple of technical lemmas.

The next result describes the recursion we will use. In particular, we want to figure out what happens to the GK numbers during each step of recursion. Here is the idea behind it:

$$m_1(A,\lambda_t) \geq m_2(A,\lambda_t) \geq m_3(A,\lambda_t) \geq \ldots \geq m_{l-1}(A,\lambda_t) \geq m_l(A,\lambda_t)$$

$\Longrightarrow$ The corresponding Jordan chains stay the same.

$m_j(A,\lambda_t) < m_l(A,\lambda_t)$ for $j > l$

So what happens when $m_j(A, \lambda_t) = m_l(A, \lambda_t)$ for some j's greater than l? Let j^* be the maximal such index.

Recursion decreases this chain by one vector.

$$\boxed{m_l(A, \lambda_t)} = m_{l+1}(A, \lambda_t) = \ldots = m_{j^*}(A, \lambda_t) > m_l(A, \lambda_t) - 1$$

Now let us formalize it.

Lemma 3.1. *Consider matrix B with the eigenvalues $\{\lambda_j\}$'s, having the GK numbers $\{m_j(B, \lambda_i)\}$ and e_1 as its eigenvector corresponding to the Jordan chain for λ_t and $m_l(B, \lambda_t)$, i.e.*

$$B = \left[\begin{array}{c|ccc} \lambda_t & \star & \cdots & \star \\ \hline 0 & & & \\ \vdots & & C & \\ 0 & & & \end{array}\right].$$

Then

- $m_j(C, \lambda_i) = m_j(B, \lambda_i)$ *for all* $i \neq t$ *or* $i = t$ *and* $j > l + 1$*;*
- $m_l(C, \lambda_t) = m_l(B, \lambda_t) - 1$*,* $m_{l+1}(C, \lambda_t) = m_{l+1}(B, \lambda_t)$ *if* $m_l(A, \lambda_t) > m_{l+1}(B, \lambda_t)$*;*
- $m_{j^*}(C, \lambda_t) = m_l(B, \lambda_t) - 1$*,* $m_j(C, \lambda_t) = m_{j+1}(B, \lambda_t)$ *for* $j = l, \ldots, j^* - 1$ *if* $m_l(B, \lambda_t) = m_{l+1}(B, \lambda_t) = \ldots = m_{j^*}(B, \lambda_t)$ *and* j^* *is the maximal such index;*
- $m_j(C, \lambda_t) = m_j(B, \lambda_t)$ *if* $m_j(B, \lambda_1) < m_l(B, \lambda_t)$*.*

The proof of this lemma can be found in [9].

The following fact allows us to use recursion in showing the Lipshitz stability of the Schur form under small perturbations preserving the Jordan structure. Before proceeding let us warn the reader. We need to be cautious and not to fall in the pitfall that is right in front of us. We start with close matrices with the same Jordan structure: $m_1(T_0, 0) = m_1(B, 0) = 2$ and $m_2(T_0, 0) = m_2(B, 0) = 1$.

$$T_0 = \left[\begin{array}{c|cc} 0 & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \quad \text{and} \quad B = \left[\begin{array}{c|cc} 0 & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & \varepsilon & 0 \end{array}\right].$$

We might think that after using recursion we might get different Jordan structures for submatrices of T_0 and B. However, it is not true since e_1 is the eigenvector of T_0 corresponding to the Jordan chain of length 2. We need to find another eigenvector of B and not just e_1 that corresponds to the Jordan chain of length 2 as well. In this case we could take $e_1 + \varepsilon e_3$ and build our recursion from there. So the "close" Jordan chains actually are:

$$e_2 \to e_1 \to 0 \quad \text{vs.} \quad e_2 \to e_1 + \varepsilon e_3 \to 0.$$

Then we construct a unitary matrix where the first column is $\frac{e_1+\varepsilon e_3}{\sqrt{1+\varepsilon^2}}$ that will be close to identity

$$V_1 = \begin{bmatrix} \frac{1}{\sqrt{1+\varepsilon^2}} & 0 & -\frac{\varepsilon}{\sqrt{1+\varepsilon^2}} \\ 0 & 1 & 0 \\ \frac{\varepsilon}{\sqrt{1+\varepsilon^2}} & 0 & \frac{1}{\sqrt{1+\varepsilon^2}} \end{bmatrix}.$$

This gives us the following Schur factorizations.

$$T_0 = I \left[\begin{array}{c|cc} 0 & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] I^* \quad \text{and} \quad B = V_1 \left[\begin{array}{c|cc} 0 & \sqrt{1+\varepsilon^2} & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] V_1^*.$$

The general result holds as well.

Lemma 3.2. *Let T_0 be an upper triangular matrix with the eigenvalues $\{\lambda_i\}$'s. Then, for all $B \in \mathcal{J}(\Omega(T_0))$ with the eigenvalues $\{\mu_j\}$'s there exist a suitable order of $\{\lambda_i\}$'s and $\{\mu_j\}$'s and unitary matrix V_1 such that matrices T_1 and B_1 in*

$$T_0 = \left[\begin{array}{c|ccc} \lambda_1 & \star & \cdots & \star \\ \hline 0 & & & \\ \vdots & & T_1 & \\ 0 & & & \end{array}\right] \quad \textit{and} \quad V_1^* B V_1 = \left[\begin{array}{c|ccc} \mu_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & B_1 & \\ 0 & & & \end{array}\right]$$

have the same Jordan structure, i.e. $m_k(T_1, \lambda_l) = m_k(B_1, \mu_l)$. Moreover,

$$\|I - V_1\| \le K \|T_0 - B\|. \tag{3.4}$$

Proof. Recall that by Proposition 2.1 we have the Lipshitz stability of the eigenvalues. That is for some order of $\{\lambda_i\}$'s and $\{\mu_j\}$'s we get

$$|\mu_i - \lambda_i| \le K \|B - T_0\|, \quad i = 1, 2, \ldots, |\sigma(T_0)|.$$

So we fix this exact order of the eigenvalues, without loss of generality assume that the first entry of T_0 is λ_1, and construct $\{f_j\}_{j=1}^m$, the Jordan chain corresponding to λ_1 for T_0 and having the eigenvector $f_1 = e_1 := [1\ 0\ 0\ \ldots\ 0]^\top$ in it, where $m = m_l(B, \mu_1) = m_l(T_0, \lambda_1)$ for some l.

Our next step is constructing the corresponding "close" chain. According to [1, Propostition 1.5], there is $\{g_k\}$ corresponding to μ_1 of B that obey the desired Lipshitz bound, that is

$$\|g_j - f_j\| \le K_j \|B - T_0\|,$$

where K_j for $j = 1, \ldots, m$ depends on T_0 only. Now, we construct V_1 which is close to the identity and unitary whose first column is parallel to g_1, i.e. $v_1 = \frac{g_1}{\|g_1\|}$. Since g_1 is close to e_1, the same is true about v_1 and e_1.

$$\|v_1 - e_1\| = \|v_1 - \frac{e_1}{\|g_1\|} + \frac{e_1}{\|g_1\|} - e_1\| \le \frac{\|g_1 - e_1\|}{\|g_1\|} + \left| \frac{1}{\|g_1\|} - 1 \right| \le$$

$$\le \frac{2K_1\|B - T_0\|}{1 - K_1\|B - T_0\|} \le \widehat{K}\|B - T_0\|.$$

Moreover, the new orthonormal basis $\{v_1, \ldots, v_n\}$ which form V_1 satisfies $\|e_i - v_i\| \le \widehat{K}\|T_0 - B\|$.

Therefore, we have $\|V_1 - I\| \le \sqrt{n}\widehat{K}\|B - T_0\|$.

The last question left is whether the Jordan structures of T_1 and B_1 are still the same. Recall, we applied our argument to Jordan chains of T_0 and B of the same length. By Lemma 3.1 applied to both resulting matrices, we see that T_1 and B_1 must have the same Jordan structure. $\square$

Now we have enough machinery to prove Theorem 1.2.

Proof of Theorem 1.2. Let us note that instead of matrices A_0 and A we consider the upper triangular matrix T_0 and the matrix $B = U_0^* A U_0$. Here we are using the fact that

$$\|A - A_0\| = \|A - U_0 T_0 U_0^*\| = \|U_0(B - T_0)U_0^*\| = \|B - T_0\|,$$

since U_0 is unitary.

Let $\lambda_1, \ldots, \lambda_p$ and $\mu_1, \ldots, \mu_p$ be the distinct eigenvalues of T_0 and B respectively. According to Lemma 3.2 there is a unitary matrix V_1 such that T_1 and B_1 in

$$T_0 = \left[\begin{array}{c|ccc} \lambda_1 & \star & \cdots & \star \\ \hline 0 & & & \\ \vdots & & T_1 & \\ 0 & & & \end{array}\right] \quad \text{and} \quad V_1^* B V_1 = \left[\begin{array}{c|ccc} \mu_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & B_1 & \\ 0 & & & \end{array}\right]$$

have the same Jordan structure, i.e. $m_k(T_1, \lambda_l) = m_k(B_1, \mu_l)$ with

$$\|I - V_1\| \leq \widehat{K}_1 \|T_0 - B\|.$$

Moreover,

$$\|T_1 - B_1\| \leq \|T_0 - V_1^* B V_1\| \leq \|T_0 - T_0 V_1 + T_0 V_1 - B V_1 + B V_1 - V_1^* B V_1\| \leq$$

$$\leq \|T_0\| \|I - V_1\| + \|V_1\| \|T_0 - B\| + \|V_1\| \|I - V_1^*\| \|B\| \leq$$

$$\leq (\|B\| + \|T_0\|) \|I - V_1\| + \|T_0 - B\| \leq (\varepsilon + 2\|T_0\|) \|I - V_1\| + \|T_0 - B\| \leq \widetilde{K}_1 \|T_0 - B\|.$$

Next, we apply Lemma 3.2 to T_1 and B_1, i.e. there is a unitary matrix V_2 such that

$$T_1 = \left[\begin{array}{c|ccc} \lambda_? & \star & \cdots & \star \\ \hline 0 & & & \\ \vdots & & T_2 & \\ 0 & & & \end{array}\right] \quad \text{and} \quad V_2^* B_1 V_2 = \left[\begin{array}{c|ccc} \mu_? & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & B_2 & \\ 0 & & & \end{array}\right]$$

have the same Jordan structure, i.e. $m_k(T_2, \lambda_l) = m_k(B_2, \mu_l)$, with

$$\|I - V_2\| \leq \widehat{K}_2 \|T_0 - B\|$$

and $\|T_2 - B_2\| \leq \|T_1 - V_2^* B_1 V_2\| \leq \widetilde{K}_2 \|T_0 - B\|$. We continue to recursively apply Lemma 3.2.

The last step is combining all the matrices. Define $\widehat{V}_j = I_j \otimes V_j$ and $\widehat{V} = \widehat{V}_1 \cdot \ldots \cdot \widehat{V}_n$, and $T := \widehat{V}^* B \widehat{V}$. We know that $\|I - \widehat{V}_j\| \leq \widehat{K} \|T_0 - B\|$ where $\widehat{K} = \max_j \widehat{K}_j$. Thus,

$$\|I - \widehat{V}\| = \|I - \widehat{V}_1 \cdot \ldots \cdot \widehat{V}_n\| =$$

$$\|I - \widehat{V}_2 \cdot \ldots \cdot \widehat{V}_n + \widehat{V}_2 \cdot \ldots \cdot \widehat{V}_n - \widehat{V}_1 \cdot \ldots \cdot \widehat{V}_n\| \leq$$

$$\|I-\widehat{V}_2\cdot\ldots\cdot\widehat{V}_n\|+\|I-\widehat{V}_1\|\leq\ldots\leq\sum_1^n\|I-\widehat{V}_j\|\leq n\widehat{K}\|T_0-B\|$$

and, similarly,

$$\|T_0-T\|\leq\widetilde{K}\|T_0-B\|.$$

By taking $K=n\widehat{K}+\widetilde{K}$ we obtain the statement in question and hence conclude the proof of this theorem. $\square$

4. Gohberg-Kaashoek Numbers and Forward Stability

What happens when the eigenvalues split in general? What do Gohberg-Kaashoek numbers of the original matrix and its perturbation tell us about the forward stability of the Schur form? These are the questions that we are going to address in this section.

First of all, notice that Lemma 3.1 does not work here. Here is why.

Example 4.1. *Consider*

$$T_0=\left[\begin{array}{cccc|ccc} 0 & 1 & & & & & \\ & 0 & 1 & & & & \\ & & 0 & 1 & & & \\ & & & 0 & & & \\ \hline & & & & 0 & 1 & \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{array}\right] \quad \textit{and } B=\left[\begin{array}{cccc|ccc} 0 & 1 & & & & & \\ & -\varepsilon & 1 & & & & \\ & & \varepsilon & 1 & & & \\ & & & 0 & & & \\ \hline & & & & 0 & 1 & \\ & & & & & 0 & 1 \\ & & & & & & \varepsilon \end{array}\right]$$

We start with $m_1(T_0)=m_1(T_0,0)=m_1(B,0)+m_1(B,\varepsilon)+m_1(B,-\varepsilon)=m_1(B)=4$ and $m_2(T_0)=m_2(T_0,0)=m_2(B,0)+m_2(B,\varepsilon)=m_2(B)=3$. If we consider the truncation as before we will get

$$T_1=\left[\begin{array}{ccc|ccc} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & 0 & & & \\ \hline & & & 0 & 1 & \\ & & & & 0 & 1 \\ & & & & & 0 \end{array}\right] \text{ and } B_1=\left[\begin{array}{ccc|ccc} -\varepsilon & 1 & & & & \\ & \varepsilon & 1 & & & \\ & & 0 & & & \\ \hline & & & 0 & 1 & \\ & & & & 0 & 1 \\ & & & & & \varepsilon \end{array}\right].$$

That is $m_1(T_1)=3$ and $m_1(B_1)=4$, $m_2(T_1)=3$ and $m_2(B_1)=2$. However, it does not mean the result of Theorem 1.3 is not true. We just need to find a different approach.

We need an alternative to Jordan bases. Recall that every matrix A is similar to its Jordan form J. We could group together all Jordan blocks corresponding to the longest chains for each distinct eigenvalue. Denote by A_1 the direct sum of these blocks.Its size is $m_1(A) \times m_1(A)$. Remove these blocks from the consideration and take a look at the Jordan blocks that are left. Now group all Jordan blocks corresponding to the longest (the second longest for the original set of Jordan blocks) chains for each distinct eigenvalue again. Call A_2 their direct sum and its size is $m_2(A) \times m_2(A)$. We continue the process until there is no more Jordan blocks to consider.

Note that $A_1 \oplus A_2 \oplus \ldots \oplus A_{k_1(A)}$ is permutation similar to J. This construction has several peculiar properties. The characteristic polynomial of A_1 is the minimal polynomial of A. The characteristic polynomials of A_k are called the invariant factors of A and their degrees are non-increasing. Let us denote them by $p_k(A) := |A_k - \lambda I|$. Two matrices are similar if and only if their invariant factors are identical.

What we are interested in is the following. Let P be an invertible matrix such that $A = P(A_1 \oplus A_2 \oplus \ldots \oplus A_{k_1(A)})P^{-1}$. In this case P consists of a Jordan basis for A. Moreover, there is an invertible S such that

$$\widehat{A}_t = \begin{bmatrix} J_t(\lambda_1) & U & & & \\ & J_t(\lambda_2) & U & & \\ & & \ddots & \ddots & \\ & & & \ddots & U \\ & & & & J_t(\lambda_{s_t}) \end{bmatrix},$$

where all the entries of U are zero except the lower left one which is 1, and $A = S(\widehat{A}_1 \oplus \widehat{A}_2 \oplus \ldots \oplus \widehat{A}_{k_1(A)})S^{-1}$. We call the basis that forms S *the Frobenius basis of A.* Note such S exists for any matrix. Since $\lambda_1 \neq \lambda_2$,

$$\widehat{A}_t = \left[\begin{array}{c|c} J_t(\lambda_1) & 1 \\ \hline & J_t(\lambda_2) \end{array}\right] = R_t \left[\begin{array}{c|c} J_t(\lambda_1) & 0 \\ \hline & J_t(\lambda_2) \end{array}\right] R_t^{-1} = R_t A_t R_t^{-1}.$$

Let us rewrite it a bit for our purposes.

Lemma 4.2. *Let T_0 be an upper triangular matrix.*

$$T_0 = \begin{bmatrix} \lambda_1 & * & \ldots & * \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix},$$

where λ_j's might have repetitions. Then $T_0 = S_0 \widehat{J}_0 S_0^{-1}$, where S_0 is an upper triangular invertible matrix and J_0 is also upper triangular and a *permutation of the invariant factor form such that $\widehat{J}_0$ has the same main diagonal as T_0.*

Proof. Let $\{\lambda^{(i)}\}_{i=1}^d$ be the set of unique eigenvalues and $\{f_j^{(l)}(\lambda^{(i)})\}$ be a Jordan basis of T_0 such that:

- $f_j^{(l)}(\lambda^{(i)}) = (T_0 - \lambda^{(i)})f_{j+1}^{(l)}(\lambda^{(i)})$ for $l = 1, \dots, k_1(T_0, \lambda^{(i)})$ and $j = 1, \dots, m_l(T_0, \lambda^{(i)}) - 1$,
- $f_{m_l(T_0,\lambda^{(i)})}^{(l)}(\lambda^{(i)}) \in \ker(T_0 - \lambda^{(i)})^{m_l(T_0,\lambda^{(i)})} \setminus (\ker(T_0 - \lambda^{(i)})^{m_l(T_0,\lambda^{(i)})-1} \cup \operatorname{span}\{f_j^{(s)}(\lambda^{(i)})\}_{\substack{j = 1, \dots, m_s(T_0, \lambda^{(i)}) \\ s = 1, \dots, l-1.}})$.

We require additional restrictions on $\{\lambda^{(i)}\}_{i=1}^d$ that we are about to discuss.

Recall that T_0 is upper triangular which means that if

$$f_j^{(l)}(\lambda^{(i)}) = \sum_{t=1}^{s} \alpha_t e_t \tag{4.5}$$

with $\alpha_s \neq 0$ then $\lambda^{(i)} = \lambda_s$ since that corresponds to non-pivot columns of $T_0 - \lambda^{(i)} I$. Moreover, we can choose S_0 to be upper triangular. This will guarantee the minimal possible s for each $f_j^{(l)}(\lambda^{(i)})$ in (4.5). That is for each $s = 1, \dots, n$ there is exactly one such a vector $f_j^{(l)}(\lambda^{(i)})$ for some i, j, and l among our Jordan basis such that decomposition (4.5) with $\alpha_s \neq 0$ holds true.

Next consider $v_j = \sum_{t=1}^d f_{m_j(T_0,\lambda_t)}^{(j)}(\lambda_t)$ for $j = 1, \dots, k_1(T_0)$. For each j construct a chain

$$v_j \to (T_0 - \lambda_{s_1})v_j \to (T_0 - \lambda_{s_2})(T_0 - \lambda_{s_1})v_j \to \dots \to \left(\prod_{t=1}^{m_j(T_0)} (T_0 - \lambda_{s_{m_j(T_0)-t}}) \right) v_j$$

where each s_j corresponds to the largest index in the decomposition in the standard basis of the previous vector in the chain. That is $s_1 > s_2 > \dots > s_{m_j(T_0)}$. Note $\left(\prod_{t=1}^{m_j(T_0)} (T_0 - \lambda_{s_{m_j(T_0)-t}})\right) v_j = p_j(T_0)v_j = 0$. All the indexes s_t's are coming from the Jordan chains involving $\{f_{m_j(T_0,\lambda_t)}^{(j)}(\lambda_t)\}_t$.

We are ready to construct S_0 and $\widehat{J}_0$. For each j we put vector v_j as s_1th column of S and all the vectors in the chain we described above at their corresponding s_tth columns of S. In $\widehat{J}_0$ we have the only non-zero entry as $\lambda_{s_{m_j(T_0)}}$ in the $s_{m_j(T_0)}$th column which is the main diagonal entry. Each s_tth column for $t = 1, \dots, m_j(T_0) - 1$ consists of λ_{s_t} on the main diagonal entry and 1 as (s_t, s_{t+1}) entry, the rest entries are 0.

Hence, we got S_0, an upper triangular matrix with no zeros on the main diagonal (i.e. invertible), and $\widehat{J}_0$, an upper triangular matrix which is a row-column permutation of the invariant form of T_0. By our construction $T_0 S_0 = S_0 \widehat{J}_0$ or $T_0 = S_0 \widehat{J}_0 S_0^{-1}$ which is exactly what we wanted to show. $\square$

Proof of Theorem 1.3. Let us start with the given Schur decomposition of A_0, that is $A_0 = U_0 T_0 U_0^*$ where U_0 is unitary and T_0 is upper triangular. As before we consider matrices T_0 and $B = U_0^* A U_0$. From the discussion in this chapter we know that $T_0 = S_0 \widehat{J}_0 S_0^{-1}$ (according to Lemma 4.2, $\widehat{J}_0$ and S_0 could be chosen to be upper triangular, since T_0 is upper triangular) and $B = S \widehat{J} S^{-1}$, where $\widehat{J}_0$ and $\widehat{J}$ are the direct sums of invariant factors and S_0 and S consist of the invariant factors bases of T_0 and B respectively.

We can make these factors close to each other. We can choose ε small enough so we can guarantee the separation of eigenvalues, i.e. it is not only that the GK numbers for matrices are the same, but also each eigenvalue λ_j of T_0 has exactly the same number of the eigenvalues $\mu_{j_1}, \dots \mu_{j_?}$ of B counting multiplicities as the multiplicity of λ_j and so $m_k(T_0, \lambda_j) = m_k(B, \mu_{j_1}) + \dots + m_k(B, \mu_{j_?})$. That is why there is an order of λ's and μ's such that

$$\|\widehat{J}_0 - \widehat{J}\| \le \sum_{i=1}^{n} |\lambda_? - \mu_?| \le K_0 \|A_0 - A\|^{1/n}.$$

Let us decompose S_0 into the basis of T_0 described in Lemma 4.2 and pick $\{f_j\}_{j=1}^m$, a chain from this basis, where $m = m_l(B) = m_l(T_0)$. Our next step is constructing the corresponding "close" chain for $\{\mu_j\}$'s. For

$$f_m \xrightarrow{T_0 - \lambda_m I} f_{m-1} \xrightarrow{T_0 - \lambda_{m-1} I} \dots \xrightarrow{T_0 - \lambda_2 I} f_1 \xrightarrow{T_0 - \lambda_1 I} 0$$

we find a close to f_m vector g_m, using Proposition 2.3,

$$\|f_m - g_m\| \le \theta \left(\prod_{j=1}^{m} (T_0 - \lambda_j I), \prod_{j=1}^{m} (B - \mu_j I) \right) \le K_m \|T_0 - B\|^{1/n}$$

and construct a chain $\{g_j\}_{j=1}^m$ consisting of vectors $g_k = (B - \mu_{k+1}) g_{k+1}$ with $\|\mu_{k+1} - \lambda_{k+1}\| < K \|B - T_0\|^{1/n}$. They satisfy the following relations

$$\|g_k - f_k\| = \|(B - \mu_{k+1} I) g_{k+1} - (T_0 - \lambda_{k+1} I) f_{k+1}\| =$$

$$\|(B - T_0) g_{k+1} + T_0(g_{k+1} - f_{k+1}) - (\mu_{k+1} - \lambda_{k+1}) g_{k+1} - \lambda_{k+1}(g_{k+1} - f_{k+1})\| \le$$

$$\|B - T_0\|\|g_{k+1}\| + \|T_0\|\|g_{k+1} - f_{k+1}\| + |\mu_{k+1} - \lambda_{k+1}|\|g_{k+1}\|$$
$$+|\lambda|\|g_{k+1} - f_{k+1}\| \le K_k\|B - T_0\|^{1/n},$$

where K_j's $(j = 1, \ldots, m)$ still depend on T_0 only and the power $1/n$ appears due to Proposition 2.2. Define $\widetilde{K} = \max\{K_j\}$. That is

$$\|g_k - f_k\| \le \widetilde{K}\|B - T_0\|^{1/n}$$

for all $k = 1, \ldots, m$. We use g's to construct S close to S_0, i.e.

$$\|S - S_0\| \le n\widetilde{K}\|B - T_0\|^{1/n}.$$

Remark 4.3. *The QR decompositions are Lipschitz stable in the class of invertible matrices, i.e. if $S_0 = Q_0R_0$ is invertible then for any $S = QR$ sufficiently close to A we have*

$$\|Q - Q_0\| + \|R - R_0\| \le K\|S - S_0\|.$$

This fact together with its proof could be found in [2] for example.

Let us fix the following QR decomposition of S_0: $S_0 = IS_0$, where I is the identity matrix. Then according to Remark 4.3, since S_0 is invertible, we can find a QR decomposition of $S = QR$ such that

$$\|Q - I\| + \|R - S_0\| \le K\|S - S_0\| \le nK\widetilde{K}\|B - T_0\|^{1/n}. \tag{4.6}$$

If we define matrix $U = U_0Q$ then we get

$$\|U_0 - U\| = \|I - Q\| \le nK\widetilde{K}\|A_0 - A\|^{1/n}$$

for some $K_1 > 0$. Moreover, $T := U^*AU = R\widehat{J}R^{-1}$ is upper triangular as a product of upper triangular matrices with

$$\|T_0 - T\| = \|U_0^*A_0U_0 - U^*AU\| \le \|U_0^*A_0U_0 - U_0^*A_0U\| +$$
$$+\|U_0^*A_0U - U_0^*AU\| + \|U_0^*AU - U^*AU\| \le \|A_0\|\|U_0 - U\| +$$
$$+\|A_0 - A\| + \|U_0^* - U^*\|\|A\| \le \widehat{K}K\|A - A_0\|^{1/n}.$$

Hence, (1.1) holds true.

□

Remark 4.4. *Note that instead of asking the question:* "For which class of matrices can we guarantee the forward stability of the Schur form?" *we could ask ourselves:* "Which class of matrices has the same GK numbers as any of its perturbations in a small enough neighborhood?" *If an $n \times n$ matrix A_0 is non-derogatory then $m(A_0) = [n, 0, 0, \ldots, 0]^\top$. We will have the same picture for any of its permutations A with $\|A - A_0\| < \epsilon$ for small enough ϵ depending on A_0 only (see [10]).*

Hence, Theorem 1.1 is the direct consequence of Theorem 1.3.

To summarize, we have characterized the class of perturbations of matrices for which Schur forms are forward stable. Thus, completing the study of the Schur form stability under a small perturbation.

References

[1] T. Bella, V. Olshevsky, U. Prasad, *Lipschitz stability of canonical Jordan bases of H-selfadjoint matrices under structure-preserving perturbations*, Linear Algebra and its Applications, **428**, 8–9, 2008, 2130–2176.

[2] R.Bhatia, *Matrix Analysis*, Springer, 1996, 347 pp.

[3] H. Den Boer, G.Ph. Thijsse, *Semi-stability of sums of partial multiplicities under additive perturbations*, Integral Equations and Operator Theory, **3**, 1980, 23–42.

[4] I. Gohberg, M. A. Kaashoek, *Unsolved problems in matrix and operator theory*, Integral Equations Operator Theory, **1**, 1978, 278–283.

[5] I. Gohberg, P. Lancaster, L. Rodman, *Invariant Subspaces of Matrices with Applications*, Canadian Mathematical Society Series of Monographs and Advanced Texts. A Wiley-Interscience Publication. John Wiley& Sons, Inc., New York, 1986, xviii+692 pp.

[6] R. Horn, C. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1985. xiii+561 pp.

[7] T. Kato, *Perturbation theory for linear operators*, Die Grundlehren der mathematischen Wissenschaften, Band 132 Springer-Verlag New York, Inc., New York, 1966, xix+592 pp.

[8] A. Markus, E. Parilis, *The change of the Jordan structure of a matrix under small perturbations*, Mat. Issled., **54** (1980), 98–109 (*in Russian*), *English translation*: Linear Algebra Appl., **54**, 1983, 139–152.

[9] A. Minenkova, V. Olshevsky, E. Nitch-Griffin, *Backward stability of the Schur decomposition under small perturbation*, accepred to Linear Algebra and Applications, arXiv.2108.02312

[10] V. Olshevsky, *A condition for the nearness of sets of invariant subspaces of near matrices in terms of their Jordan structures.*, (in Russian), Siberian Math. Journal, **30**, 4, 1989, 102–110, English translation: Siberian Math. Journal, Plenum publishing corp., **30**, 4, 1989, 580–586.

[11] V. Matsaev, V. Olshevsky, *Cyclic dimensions, kernel multiplicities, and Gohberg-Kaashoek numbers*, Linear Algebra Appl., **239**, 1996, 161–174.

[12] A.M. Ostrowski, *Solution of equations in Euclidean and Banach spaces. Third edition of Solution of equations and systems of equations. Pure and Applied Mathematics*, Vol. 9. Academic Press, New York-London, 1973. xx+412 pp.