ON EIGEN AND SINGULAR VALUE CLUSTERS [1]

E. E. Tyrtyshnikov, N. L. Zamarashkin [2]

Abstract - Usually when singular values are clustered, the eigenvalues behave similarly. However, it is not the case if we make no assumptions. Here we present examples when the singular values are clustered whereas the eigenvalues are not, and vice versa. Besides, the necessary and sufficient assumptions are discussed under which the former implies the latter. We also present a new algebraic approach to one-point clusters.

1. **What are clusters?**

Sometimes the most of (though not all) eigen or singular values are amassed near some set on the complex plane (one or several points, as a rule). Such a set is said to be a cluster. However, we need to bring in something more into the picture. Below we put the definitions proposed in [6] (see also [7, 8]).

Consider a sequence of matrices $A_n \in \mathbb{C}^{n \times n}$ with the eigenvalues $\lambda_i(A_n)$ and a subset M of complex numbers. For any $\varepsilon > 0$, denote by $\gamma_n(\varepsilon) = \gamma_n(\varepsilon, M)$ the number of those eigenvalues of A_n that fall outside the ε distance from M. Then M is called a *(general) cluster* if

$$\lim_{n\to\infty} \frac{\gamma_n(\varepsilon)}{n} = 0 \quad \forall\, \varepsilon > 0,$$

and a *proper cluster* if

$$\gamma_n(\varepsilon) \le c(\varepsilon) \quad \forall\, n, \quad \forall\, \varepsilon > 0.$$

We chiefly consider the clusters consisting of one or several (finitely many) points ($M = \mathbb{C}$ is never of interest).

— Received: 15/11/1996

(1) The work was supported in part by the Russian Fund of Basic Research and also by the Volkswagen-Stiftung.

(2) Institute of Numerical Mathematics, Russian Academy of Sciences, Gubkina 8, Moscow 117333, Russia. E-mail: tee@inm.ras.ru

One might be interested as well in the singular value clusters. To distinguish between the eigen and the singular value cases let us write $\gamma_n(\varepsilon; \lambda) = \gamma_n(\varepsilon, M; \lambda)$ and $\gamma_n(\varepsilon; \sigma) = \gamma_n(\varepsilon, M; \sigma)$, respectively. For brevity, let us mark that M is a cluster by $\lambda(A_n) \sim M$ or $\sigma(A_n) \sim M$.

Of course, we tacitly assume that A_n are the elements of some common process (for instance, they could arise from a digitization of some operator equation on a sequence of meshes).

Clusters are very important in the convergence analysis of iterational methods [1, 3, 5, 7]. For the minimal residual methods, the eigenvalue clusters are particularly important. Still, when devising preconditioners it is easier to fight for the singular value clusters. Fortunately if the singular values are clustered then under rather mild assumptions the eigenvalues are also clustered (it is one of our results).

2. Singular value clusters

To prove that a sequence has a cluster, we can try to find a "close" sequence for which this is already established. That "closeness" can be treated in a rather broad sense.

Theorem 2.1. [6, 7] Suppose A_n and B_n are such that

(2.1) $||A_n - B_n||_F^2 = o(n)$

or, alternatively,

(2.2) $\operatorname{rank}(A_n - B_n) = o(n)$.

In either case, any singular value cluster for A_n is also a singular value cluster for B_n, and vice versa.

3. Eigenvalue clusters

In order to state that the eigenvalue clusters for A_n and B_n coincide, we ought to add something to the premises of Theorem 2.1. For example, we can formulate the following

Theorem 3.1. Suppose $A_n, B_n \in \mathbb{C}^{n \times n}$ are diagonalizable for every n and denote by P_n, Q_n the corresponding eigenvector matrices. If

$$\operatorname{cond}_2^2 P_n \operatorname{cond}_2^2 Q_n \, ||A_n - B_n||_F^2 \;=\; o(n)$$

then any eigenvalue cluster for A_n is also an eigenvalue cluster for B_n, and vice versa.

PROOF. Using the extension of the Hoffman–Wielandt theorem obtained by Sun and Zhang (see [4]), we are to follow the same lines as when proving Theorem 2.1. ■

Some stronger statements can be made if one of the sequences is a constant matrix, for example, the zero one.

THEOREM 3.2. Suppose $\sigma(A_n) \sim 0$ (A_n has the singular value cluster at zero) and, additionally, uniformly in all sufficiently small $\varepsilon > 0$,

$$\log ||A_n||_2 = O\left(\frac{n}{\gamma_n(\varepsilon;\sigma)}\right).$$

Then $\lambda(A_n) \sim 0$.

Instead of proving this, we propose and prove another, to some extent more general, theorem.

THEOREM 3.3. Assume that matrices A_n are nonsingular,

$$\sigma(A_n) \sim M \equiv \{x \in \mathbb{R} : \ s \ \le \ x \ \le \ r\},$$

and, additionally, uniformly in all sufficiently small $\varepsilon > 0$,

$$\log ||A_n^{\pm 1}||_2 = o\left(\frac{n}{\gamma_n(\varepsilon;\sigma)}\right),$$

where $\gamma_n(\varepsilon;\sigma) = \gamma_n(\varepsilon, M;\sigma)$. Then

$$\lambda(A_n) \sim R \ \equiv \ \{z \in \mathbb{C} : \ s \ \le \ |z| \ \le \ r\}.$$

PROOF. Assume that the eigenvalues and singular values are indexed as follows:

$$|\lambda_1(A_n)| \ge \ldots \ge |\lambda_n(A_n)| \quad \text{and} \quad \sigma_1(A_n) \ge \ldots \ge \sigma_n(A_n).$$

We make use of the following Weyl inequalities:

$$\prod_{k=1}^{m} |\lambda_k(A_n)| \ \le \ \prod_{k=1}^{m} \sigma_k(A_n), \qquad m = 1, \ldots, n. \tag{3.1}$$

First of all, we prove that

$$\lambda(A_n) \sim B(r) \equiv \{z \in \mathbb{C} : \ |z| \ \le \ r\}.$$

By contradiction, suppose it is not the case. Then there exist ε_0, $c_0 > 0$ and some subset of increasing indices $\mathcal{N} = \{n_1, n_2, \ldots\}$ such that

$$\gamma_n(\varepsilon_0;\lambda) \ge c_0 n \quad \forall\, n \in \mathcal{N},$$

where $\gamma_n(\varepsilon_0; \lambda) = \gamma_n(\varepsilon_0, B; \lambda)$. Without loss of generality, assume that $n_k = k \quad \forall\, k$. Choose any $\varepsilon > 0$. Using (3.1) we obtain

$$(r+\varepsilon_0)^{\gamma_n(\varepsilon_0;\lambda)} \le \prod_{k=1}^{\gamma_n(\varepsilon_0;\lambda)} |\lambda_k(A_n)| \;\le\; \prod_{k=1}^{\gamma_n(\varepsilon_0;\lambda)} \sigma_k(A_n)$$

$$\le ||A_n||_2^{\gamma_n(\varepsilon;\sigma)} \, (r+\varepsilon)^{\gamma_n(\varepsilon_0;\lambda)-\gamma_n(\varepsilon;\sigma)}$$

$$\Rightarrow \quad \left(\frac{r+\varepsilon_0}{r+\varepsilon}\right)^{\frac{\gamma_n(\varepsilon_0;\lambda)}{n}} \le \left(\frac{||A_n||_2}{r+\varepsilon}\right)^{\frac{\gamma_n(\varepsilon;\sigma)}{n}} .$$

By the contradictory assumption, if $\varepsilon < \varepsilon_0$ then the left-hand side is lower-bounded by a positive constant. For sufficiently small ε, this constant can be made greater than 1 whereas the right-hand side tends to 1 as n grows to infinity.

By similar arguments we can prove that $\lambda(A_n^{-1}) \sim B(s^{-1})$. In case $s = 0$ this is trivial (if we agree that $B(\infty) = \mathbb{C}$). It remains only to note that R is an intersection of $B(r)$ and $B(s^{-1})$. ∎

Remark 3.1. We can relax the hypotheses of the above theorem. It remains valid if the upper estimate on $\|A_n^{\pm 1}\|_2$ is replaced by the following requirements:

$$\log ||A_n||_2 = o\left(\frac{n}{\gamma_n(\varepsilon, B(r); \sigma)}\right)$$

and

$$\log ||A_n^{-1}||_2 = o\left(\frac{n}{\gamma_n(\varepsilon, B(s^{-1}); \sigma)}\right).$$

As is clear from the above proof, we could easily have even a more precise assertion of which the requirements on the norms are changed onto some relationships between the norms and the radii.

4. The Cauchy-Toeplitz example

Consider the Cauchy-Toeplitz matrices

$$A_n = \left[\frac{1}{i-j+\frac{1}{2}}\right]_{n\times n} .$$

We know (see [2, 9]) that

$$\sigma(A_n) \sim \pi \quad \text{and} \quad \|A_n\|_2 \le \pi .$$

Moreover, it was proved in [9] that

$$\gamma_n(\varepsilon;\sigma) = O(\log^2 n)$$

and

$$\|A_n^{-1}\|_2 \le c \log n.$$

Since trivially

$$\log\log n = o(\frac{n}{\log^2 n}),$$

all the hypotheses of Theorem 3.3 are fulfilled. Hence, we can state that

$$\lambda(A_n) \sim \{z \in \mathbb{C} :\ |z| = \pi\}.$$

5. **How neat are the assumptions?**

We are aware now that next to always the property $\sigma(A_n) \sim 0$ implies $\lambda(A_n) \sim 0$. The only assumption we need is the one on the behavior of the 2-norms of A_n. How neat is it?

We now produce an example when $\sigma(A_n) \sim 0$ whereas $\lambda(A_n) \sim 0$ does not hold. Sure the norms of A_n must grow, and it is interesting to realise how fast they have to grow.

LEMMA 5.1. Suppose $D \in \mathbb{C}^{n\times n}$ is a diagonal matrix. Then for any $0 < \varepsilon < 1$ there exists a rank-one matrix $A = D + L + U$, where L is strictly lower triangular with all nonzero entries in modulus less than or equal to ε, and U is strictly upper triangular.

PROOF. Write $A = uv^T$ and try to satisfy the following demands:

(5.1) $u_i v_i = d_i,\ i = 1, \ldots, n; \quad |u_i v_j| \le \varepsilon,\ i > j.$

Choose

$$\zeta = \min\{1,\ \frac{\varepsilon}{\max\limits_i |d_i|}\}$$

and set

$$u_i = \zeta^i, \quad v_j = d_j \zeta^{-j}.$$

Obviously, the demands (5.1) are met. ■

Now, take $d_i = i$. For each n let us take $\varepsilon_n = \frac{\delta_n}{n^2}$, where $\delta_n > 0$, $\delta_n \to 0$, and construct a rank-one matrix $A_n = D_n + L_n + U_n$ using the above lemma. Then we set $B_n = A_n - L_n$. Since B_n is upper triangular, its eigenvalues are easy to find. They are $d_1, \ \dots \ , d_n$. At the same time, all the singular values of A_n save for one are equal to zero, and hence $\sigma(A_n) \sim 0$. Since $\|A_n - B_n\|_F^2 \le \delta_n$, from Theorem 2.1 we conclude that $\sigma(B_n) \sim 0$. Finally,

$$\sigma(B_n) \sim 0 \quad \text{but} \quad \lambda(B_n) \not\sim 0.$$

Since

$$\log \|B_n\|_2 \;\ge\; c\frac{n}{\delta_n},$$

the assumption imposed on the norms is quite accurate and can not be weakened, at least for the whole of matrices.

Note that it might be as well so that

$$\lambda(A_n) \sim 0 \quad \text{but} \quad \sigma(A_n) \not\sim 0.$$

To produce an example, for instance, we can take up

$$A_n = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}_{n \times n}$$

6. A pseudo-ideal approach to clusters

In this section we present a train of ideas on how we might study one-point clusters. The approach proposed seems to be absolutely different from those that were used previously. Below we do not use any more the Hoffman-Wielandt theorem, neither the interlacing properties. What made it possible is probably that the one-point clusters are *more algebraic* in nature than more complicated clusters.

Let $\mathcal{A}$ be a sequence of matrices $A_n \in \mathbb{C}^{n \times n}$ $(n = 1, 2, \ \dots \ )$. If $\mathcal{N} = \{n_k\}$ is a sequence of increasing indices $n_1 < n_2 < \dots$, then denote by $\mathcal{A}(\mathcal{N})$ a sequence $A_{n_k}, n_k \in \mathcal{N}$. Let M comprise all sequences. If $K = \{\mathcal{A}\} \subset M$ then $K(\mathcal{N}) = \{\mathcal{A}(\mathcal{N})\}$ consists of the corresponding subsequences.

If $\mathcal{A}, \mathcal{B} \in M$ then we define $\mathcal{A}\mathcal{B}$ and $\mathcal{A} + \mathcal{B}$ as pairwise product and sum of two sequences.

Denote by $L \subset M$ all sequences of matrices A_n with the 2-norms uniformly bounded in n (the bound itself may depend upon the sequence).

We call $K \subset M$ a *pseudo-ideal* if the following 2 properties are fulfilled:

(1) if $\mathcal{A},\mathcal{B} \in K$ then $\mathcal{A}+\mathcal{B} \in K$;
(2) if $\mathcal{A} \in K$ and $\mathcal{B} \in L$ then $\mathcal{AB},\mathcal{BA} \in K$.

If a pseudo-ideal K is such that $L(\mathcal{N}) \not\subset K(\mathcal{N})$ for any $\mathcal{N}$, then it will be termed a *super-ideal.*

An evident property of any super-ideal K is that any its sequence is not allowed to contain a subsequence of the identity matrices.

THEOREM 6.1. A pseudo-ideal $K \subset M$ is such that any its sequence has the singular value cluster at zero if and only if K is a super-ideal.

PROOF. Suppose any sequence in K has the singular value cluster at zero. Then, as is readily seen, if $\sigma(A_n) \sim 0$ and $\sigma(B_n) \sim 0$ then $\sigma(A_n + B_n) \sim 0$. If $\sigma(A_n) \sim 0$ and $\|B_n\| \le c$ uniformly in n then $\sigma(A_nB_n) \sim 0$ and $\sigma(B_nA_n) \sim 0$. It is clear also that $\sigma(A_{n_k}) \sim 0$ can not hold for arbitrary matrices A_{n_k}. Therefore, K must be a super-ideal.

Now, assume that K is a super-ideal but still contains a sequence A_n such that $\sigma(A_n) \not\sim 0$. Consequently, there exist $\varepsilon, c_0 > 0$ and $\mathcal{N} = \{n_k\}$ such that

$$\gamma_m(\varepsilon;\sigma) \ge c_0 m \;\forall\; m \in \mathcal{N}.$$

Consider the singular value decompositions

$$A_m = V_m \Sigma_m U_m^*, \quad m \in \mathcal{N}.$$

Since U_m and V_m are unitary, they belong to $L(\mathcal{N})$ and hence $\{\Sigma_m\}$ belongs to $K(\mathcal{N})$. What is more, by multiplications by diagonal matrices with the entries less than or equal to ε_0^{-1} we can obtain from Σ_m a diagonal matrix $D_m = \operatorname{diag}\{d_i\}$ with units for all i such that $\sigma_i > \varepsilon_0$, and zeroes elsewhere. Using permutations we can obtain from D_m a diagonal matrix with arbitrarily prescribed positions for those units. A sum of properly chosen such matrices will yield the unity matrix of order m, and obviously we are to add not more than $1 + c_0^{-1}$ matrices. Thus, we are led to infer that $K(\mathcal{N}) \supset L(\mathcal{N})$, which contradicts the assumptions we started with. ■

As an example, consider K consisting of sequences $\{A_n\}$ such that

$$\exists\; \Delta_n \in \mathbb{C}^{n\times n}:\; \|A_n + \Delta_n\| = o(\sqrt{n}) \quad \text{and} \quad \operatorname{rank}\Delta_n = o(n).$$

It is not difficult to verify that K is a pseudo-ideal. By the above theorem, $\sigma(A_n) \sim 0$ for any sequence $\{A_n\} \in K$.

We'd like to thank Paolo Tilli for his remarks on this note.

REFERENCES

[1] O. AXELSSON, G. LINDSKOG, *On the rate of convergence of the preconditioned CG method,* Numer. Math. **48**, (1986) 499–523.

[2] S. V. PARTER, *On the distribution of the singular values of Toeplitz matrices,* Linear Algebra Appl. **80**, (1986) 115–130.

[3] R. CHAN, G. STRANG, *Toeplitz equations by conjugate gradients with circulant preconditioners,* SIAM J. Sci. Stat. Comput. **10**, (1989) 104–119.

[4] G. W. STEWART, J. SUN, Matrix Perturbation Theory, Academic Press, Inc., 1990.

[5] V. V. STRELA, E. E. TYRTYSHNIKOV, *Which circulant preconditioner is better?* Math. Comput. **65 (213)**, (1996) 137–150.

[6] E. E. TYRTYSHNIKOV, *Distributions and clusters,* Matrix methods and algorithms, INM RAS, 1993, 124–167. (In Russian.)

[7] E. E. TYRTYSHNIKOV, *Influence of matrix operations on the distribution of eigenvalues and singular values of Toeplitz matrices,* Linear Algebra Appl. **216**, (1994) 1–23.

[8] E. E. TYRTYSHNIKOV, *Circulant preconditioners with unbounded inverses,* Linear Algebra Appl. **216**, (1995) 1–23.

[9] E. E. TYRTYSHNIKOV, *Singular values of Cauchy-Toeplitz matrices,* Linear Algebra Appl. **161**, (1992) 99–116.